Estadística Aplicada III ITAM
16 de febrero de 2024 2024-I

Tarea 3. Fecha de entrega: 27 de febrero de 2024.

Problemas

1. Encuentra los estimados máximo-verosímiles del vector promedio $\boldsymbol{\mu}$ y la matriz de covarianzas $\boldsymbol{\Sigma}$ de la muestra aleatoria:

$$\mathbf{X} = \begin{bmatrix} 3 & 6 \\ 4 & 4 \\ 5 & 7 \\ 4 & 7 \end{bmatrix}$$

de una población normal bivariada.

2. Sean $\mathbf{x}_1, \ldots, \mathbf{x}_{20}$ una muestra de una población $\mathcal{N}_6(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Especificar cada una de las distribuciones completamente de las siguientes variables:

   a. $(\mathbf{x}_1 - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}_1 - \boldsymbol{\mu})$
   b. $\bar{\mathbf{x}}$ y $\sqrt{n}(\bar{\mathbf{x}} - \boldsymbol{\mu})$
   c. $(n-1)\mathbf{S}$
   d. $n(\bar{\mathbf{x}} - \boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ (distribución aproximada)

3. Harry Roberts, un naturalista del Departamento de Pesca de Alaska, estudia a los osos grizzly con el objetivo de mantener una población sana. Medidas de $n = 61$ osos dan los siguientes estimados de $\bar{\mathbf{x}}$ y $\mathbf{S}$. Las variables son: peso (kg), longitud del cuerpo (cm), cuello (cm), circunferencia (cm), longitud de la cabeza (cm), ancho de la cabeza (cm):

```r
xbar <- c(95.52, 164.38, 55.69, 93.39, 17.98, 31.13)
S <- matrix(c(3266.46, 1343.97, 731.54, 1175.50, 162.68, 238.37,
              1343.97, 721.91, 324.25, 537.35, 80.17, 117.73,
              731.54, 324.25, 179.28, 281.17, 39.15, 56.80,
              1175.50, 537.35, 281.17, 474.98, 63.73, 94.85,
              162.68, 80.17, 39.15, 63.73, 9.95, 13.88,
              238.37, 117.73, 56.80, 94.85, 13.88, 21.26), nrow = 6)
```

   a. Obtener los intervalos de confianza simultáneos de 95 % para muestras grandes para las seis medias poblacionales de medidas corporales.
   b. Obtener la elipse de confianza simultánea del 95 % de grandes muestras para el peso medio y la circunferencia media.
   c. Obtener los intervalos de confianza Bonferronizados del 95 % para las seis medias en la parte (a).
   d. Referirse a pa parte (b). Construir el rectángulo de confianza Bonferroni de 95 % para el peso medio y la circunferencia media usando $m = 6$. Comparar este rectángulo con la elipse de confianza en la parte (b).
   e. Obtener el intevalo de confianza de 95 % de Bonferroni para la diferencia del ancho medio de la cabeza y el largo medio de la cabeza usando $m = 6 + 1 = 7$ para considerar esta afirmación así como para cada media individual.

4. Dados los datos $\mathbf{X} = \begin{bmatrix} 3 & 6 & 0 \\ 4 & 4 & 3 \\ NA & 8 & 3 \\ 5 & NA & NA \end{bmatrix}$ con componentes faltantes, usar el algoritmo EM para estimar $\boldsymbol{\mu}$ y $\boldsymbol{\Sigma}$. Determinar los estimados iniciales, e iterar para encontrar los primeros estimados revisados.

5. Ejercicio 5.20. Este ejercicio se refiere al ave conocida como milano picogarfio:

Los datos de este ejercicio son los siguientes:

```
X <- read.delim("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J&W/T5-12.DAT",
                sep = "", header = F)
str(X)

'data.frame': 45 obs. of  2 variables:
 $ V1: int  191 197 208 180 180 188 210 196 191 179 ...
 $ V2: int  284 285 288 273 275 280 283 288 271 257 ...
```

a. Encuentra y bosqueja la elipse de 95 % de confianza para las medias poblacionales μ_1 y μ_2. Supongan que se sabe que $\mu_1 = 190$mm y $\mu_2 = 275$mm para machos milanos picogarfios. ¿Son valores plausibles para la longitud media de la cola y longitud media del ala para los pájaros hembras? Explicar.
b. Construir los intervalos simultáneos T^2 de 95 % para μ_1 y μ_2 y los intervalos de Bonferroni de 95 % de confianza para las mismas cantidades. Comparar los dos conjuntos de intervalos. ¿Qué ventaja, si la hay tienen los intervalos T^2 sobre los intervalos Bonferroni?
c. Es la distribución normal bivariada un modelo poblacional viable? Explicar con referencia a qq-plots y un diagrama de dispersión.

6. Consideren las 30 observaciones de cráneos de hombres egipcios para el primer periodo de tiempo:

```
X <- read.delim("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J&W/T6-13.DAT",
                sep = "", header = F)
names(X) <- c("MaxBreath", "BasHeight", "BasLength", "NasHeight", "TimePeriod")
Y <- subset(X,TimePeriod==1)
Y$TimePeriod <- NULL
str(Y)

'data.frame': 30 obs. of  4 variables:
 $ MaxBreath: int  131 125 131 119 136 138 139 125 131 134 ...
 $ BasHeight: int  138 131 132 132 143 137 130 136 134 134 ...
 $ BasLength: int  89 92 99 96 100 89 108 93 102 99 ...
 $ NasHeight: int  49 48 50 44 54 56 48 48 51 51 ...
```

a. Construir qq-plots de las distribuciones marginales para las cuatro variables. También construir la gráfica ji-cuadrada de las distribución multivariada. ¿Los datos se ven normales multivariados? Explicar.

b. Construir los intervalos de Bonferroni del 95 % para cada una de las variables. También los simultáneos T^2 correspondientes y comparar.

7. Como parte de un estudio del proceso de ensamble de sus láminas, un fabricante de autos utiliza sensores que registran la desviación del grosor nominal (milímetros) en seis ubicaciones sobre un auto. Los primeros cuatro son medidos cuando el cuerpo del auto está completo y los últimos dos son medidos en el cuerpo inferior en una etapa temprana del ensamblado. Los datos de 50 autos están en la siguiente tabla (tabla 5.14)

```
X <- read.delim("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J&W/T5-14.dat", sep = "", header = F)
str(X)
```

```
'data.frame':	50 obs. of  6 variables:
 $ V1: num  -0.12 -0.6 -0.13 -0.46 -0.46 -0.46 -0.46 -0.13 -0.31 -0.37 ...
 $ V2: num  0.36 -0.35 0.05 -0.37 -0.24 -0.16 -0.24 0.05 -0.16 -0.24 ...
 $ V3: num  0.4 0.04 0.84 0.3 0.37 0.07 0.13 -0.01 -0.2 0.37 ...
 $ V4: num  0.25 -0.28 0.61 0 0.13 0.1 0.02 0.09 0.23 0.21 ...
 $ V5: num  1.37 -0.25 1.45 -0.12 0.78 1.15 0.26 -0.15 0.65 1.15 ...
 $ V6: num  -0.13 -0.15 0.25 -0.25 -0.15 -0.18 -0.2 -0.18 0.15 0.05 ...
```

a. El proceso parece estable para los primeros 30 casos. Usar estos casos para estimar $\bar{\mathbf{x}}$ y $\mathbf{S}$. Luego construir una gráfica T^2 usando todas las variables. Incluyendo los 50 autos.

b. ¿Qué ubicaciones indican una preocupación?

8. Con respecto a los datos de tormentas de nieve que se muestran a continuación:

```
X <- read.delim("https://raw.githubusercontent.com/jvega68/EA3/master/datos/J&W/T3-2.DAT",
                sep = "", header = F)
str(X)
```

```
'data.frame':	25 obs. of  2 variables:
 $ V1: num  12.5 14.5 8 9 19.5 8 9 7 7 9 ...
 $ V2: num  13.7 16.5 17.4 11 23.6 13.2 32.1 12.3 11.8 24.4 ...
```

a. Encontrar la región de confianza del 95 % para el vector media, después de realizar una transformación apropiada.

b. En la misma escala,encontrar los intervalos bonferronizados del 95 % para las dos componentes de las medias