University of California, Davis
Department of Statistics

Name ..Solution..........
P R I N T P L E A S E

STA 135

Sample
Midterm II

Instructions: 1. **WORK ALL PROBLEMS**. Please, give details and explanations and **SHOW ALL YOUR WORK** so that partial credits can be given.
2. You may use **three** sheets of **notes** and a **calculator** but **no** other reference materials.

Points

1. Let $\underline{X}_1$, $\underline{X}_2$, $\underline{X}_3$, $\underline{X}_4$, be random samples from a p-dimensional multivariate normal distribution with mean vector $\underline{\mu}$ and covariance matrix Σ.

(a) Find the distribution of $(\underline{X}_1 + \underline{X}_2 + \underline{X}_3 + \underline{X}_4) / 4$.
(b) Find the distribution of $(\underline{X}_1 - \underline{X}_2 + \underline{X}_3 - \underline{X}_4) / 2$.
(c) Suppose the original population is not multivariate normal and we have increased the sample size from 4 to 400. What is the approximate distribution of the sample mean vector.
(d) Let S denote the sample covariance matrix. With the information given in part (c), what is the approximate distribution of $n(\bar{\underline{X}} - \underline{\mu})' S^{-1} (\bar{\underline{X}} - \underline{\mu})$.

(25)

(a) $\underline{X}_1, \dots, \underline{X}_4 \overset{iid}{\sim} N(\underline{\mu}, \Sigma)$

$E[\frac{1}{4}(\underline{X}_1 + \underline{X}_2 + \underline{X}_3 + \underline{X}_4)] = \frac{1}{4}[E(\underline{X}_1) + \dots + E(\underline{X}_4)] = \frac{1}{4}(4\underline{\mu}) = \underline{\mu}$

$Cov[\frac{1}{4}(\underline{X}_1 + \underline{X}_2 + \underline{X}_3 + \underline{X}_4)] = \frac{1}{16}[Cov(\underline{X}_1) + \dots + Cov(\underline{X}_4)] = \frac{1}{16}(4\Sigma) = \frac{1}{4}\Sigma$

$(\underline{X}_1 + \underline{X}_2 + \underline{X}_3 + \underline{X}_4)/4 \sim N(\underline{\mu}, \frac{1}{4}\Sigma)$

(b) $E[(\underline{X}_1 - \underline{X}_2 + \underline{X}_3 - \underline{X}_4)/2] = \frac{1}{2}[E(\underline{X}_1) - E(\underline{X}_2) + E(\underline{X}_3) - E(\underline{X}_4)]$
$= \frac{1}{2}[\underline{\mu} - \underline{\mu} + \underline{\mu} - \underline{\mu}] = \underline{0}$

$Cov[(\underline{X} - \underline{X}_2 + \underline{X}_3 - \underline{X}_4)/2] = \frac{1}{4}[Cov(\underline{X}_1) + Cov(\underline{X}_2) + Cov(\underline{X}_3) + Cov(\underline{X}_4)]$
$= \frac{1}{4}(4\Sigma) = \Sigma$

$(\underline{X}_1 - \underline{X}_2 + \underline{X}_3 - \underline{X}_4)/2 \sim N(\underline{0}, \Sigma)$

(c) Let $\bar{\underline{X}} = \frac{1}{n}\sum_{i=1}^{n} \underline{X}_i$. Then

$E(\bar{\underline{X}}) = \frac{1}{n}\sum_{i=1}^{n} E(\underline{X}_i) = \frac{1}{n}(n\underline{\mu}) = \underline{\mu}$

$Cov(\bar{\underline{X}}) = \frac{1}{n^2}\sum_{i=1}^{n} Cov(\underline{X}_i) = \frac{1}{n^2}(n\Sigma) = \frac{1}{n}\Sigma$

Based on the Central Limit theorem $\bar{\underline{X}}$ is approx. $N(\underline{\mu}, \frac{1}{400}\Sigma)$

(d) The large-sample distribution of $n(\bar{\underline{X}} - \underline{\mu})' S^{-1} (\bar{\underline{X}} - \underline{\mu})$ is approx. χ^2 with $p=4$ df.

2. In a study of grizzly bears the following summary statistics on head length (cm) and head width (cm) were obtained for n= 61 bears.

$$\bar{\underline{x}} = [17.98 \quad 31.13]', \qquad S = \begin{bmatrix} 9.95 & 13.88 \\ 13.88 & 21.26 \end{bmatrix}$$

(a) Obtain the large-sample 95% simultaneous confidence intervals for the means of each one of these two measurements. ($\chi^2_2(0.05) = 5.99$)

(b) Obtain the large-sample 95% confidence region for mean head length and head width. ($\chi^2_2(0.05) = 5.99$)

(c) Obtained the 95% large-sample Bonferroni confidence interval for means of these two measurements. ($Z(.0125) = 2.24$)

(25)

(a) The $100(1-\alpha)\%$ large-sample simult. Conf. int. for μ_{k} is

$$\bar{x}_{k} \pm \sqrt{\chi^2_p(\alpha)}\sqrt{\frac{S_{kk}}{n}}$$

The 95% Confidence intervals are

Head Length : $17.98 \pm \sqrt{5.99}\sqrt{\frac{9.95}{61}} \Rightarrow 17.98 \pm .99 \Rightarrow [16.99, 18.97]$

Head Width : $31.13 \pm \sqrt{5.99}\sqrt{\frac{21.26}{61}} \Rightarrow 31.13 \pm 1.44 \quad [29.69, 32.57]$

(b) The $100(1-\alpha)\%$ large sample Confidence region for $\underline{\mu}$ is

$$n(\bar{\underline{x}}-\underline{\mu})' S^{-1} (\bar{\underline{x}}-\underline{\mu}) \leq \chi^2_p(\alpha)$$

Since $S^{-1} = \begin{bmatrix} 1.13 & -.74 \\ -.74 & .53 \end{bmatrix}$, the 95% Confidence region is

$$\underline{\mu}: \; 61 \begin{bmatrix} (17.98-\mu_1) & (31.13-\mu_2) \end{bmatrix} \begin{bmatrix} 1.13 & -.74 \\ -.74 & .53 \end{bmatrix} \begin{bmatrix} 17.98-\mu_1 \\ 31.13-\mu_2 \end{bmatrix} \leq 5.99$$

(c) The $100(1-\alpha)\%$ large-Sample Bonferroni interval for μ_{k} is

$$\bar{x}_{k} \pm z(\alpha/2p)\sqrt{\frac{S_{kk}}{n}}$$

The 95% Confidence intervals are:

μ_1 : $17.98 \pm 2.24\sqrt{\frac{9.95}{61}} \Rightarrow 17.98 \pm .90 \Rightarrow [17.08, 18.80]$

μ_2 : $31.13 \pm 2.24\sqrt{\frac{21.26}{61}} \Rightarrow 31.13 \pm 1.32 \Rightarrow [29.81, 32.45]$

3. The following data matrix is observed for a two-dimensional random vector $\underline{X}$.

$$\mathbf{X} = \begin{bmatrix} 3 & 4 \\ 6 & 2 \\ 3 & 3 \end{bmatrix}$$

Assume that the population is multivariate normal with unknown mean vector $\underline{\mu}$ and unknown covariance matrix Σ.

(a) Use the Hotelling T^2 to test $H_0 : \underline{\mu} = [\,3 \quad 2\,]'$ against $H_0 : \underline{\mu} \neq [\,3 \quad 2\,]'$ at 0.05 level of significance. ($F_{2,1}(.05) = 200$).

(b) Construct a 95% confidence region for mean vector $\underline{\mu}$ and use that to test the hypothesis stated in (a)

(25)

(a) Computing the sampl mean vector and covariance matrix

We have:

$$\bar{\underline{X}} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}, \quad S = \begin{bmatrix} 3 & -3/2 \\ -3/2 & 1 \end{bmatrix}. \quad \text{Then} \quad S^{-1} = \begin{bmatrix} 4/3 & 2 \\ 2 & 4 \end{bmatrix}.$$

We have

$$T^2 = n(\bar{\underline{X}} - \underline{\mu}_0)' S^{-1} (\bar{\underline{X}} - \underline{\mu}_0)$$

$$= 3\,[(4-3) \quad (3-2)] \begin{bmatrix} 4/3 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 4-3 \\ 3-2 \end{bmatrix}$$

$$= 3\,[\,11/4 \quad 6\,] \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \frac{3(35)}{4} = 26.25$$

$$\frac{(n-1)p}{n-p} F_{p,n-p}(\alpha) = \frac{(3-1)(2)}{3-2} F_{2,1}(.05) = 2(200) = 400.$$

Since $26.25 < 400$, we cannot reject $H_0 : \underline{\mu} = \begin{bmatrix} 3 \\ 2 \end{bmatrix}$ at .05 level.

(b) The 95% confidence region for $\underline{\mu}$ is

$$\underline{\mu} : \; n(\bar{\underline{X}} - \underline{\mu})' S^{-1} (\bar{\underline{X}} - \underline{\mu}) \leq \frac{p(n-p)}{np} F_{p,n-p}(\alpha)$$

$$3[(4-\mu_1) \quad (3-\mu_2)] \begin{pmatrix} 4/3 & 2 \\ 2 & 4 \end{pmatrix} \begin{bmatrix} 4-\mu_1 \\ 3-\mu_2 \end{bmatrix} \leq \frac{2(3-1)}{3-2} F_{2,1}(.05) = 400$$

For $\underline{\mu} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$, we we have $26.25 < 400$, so that $\underline{\mu} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}$ is inside the confidence region and we cannot reject H_0 at .05 level.

4. A researcher considered three indices measuring the severity of heart attacks. All three indices were evaluated for each patient. The values of these indices for n= 40 heart attack patients arriving at a hospital emergency room produced the summary statistics:

$$\bar{\underline{x}} = [46.1 \;\; 57.3 \;\; 50.4]'$$

$$S = \begin{bmatrix} 101.3 & 63.0 & 71.0 \\ 63.0 & 80.2 & 55.6 \\ 71.0 & 55.6 & 97.4 \end{bmatrix}$$

Assume that the population is multivariate normal with unknown mean vector $\underline{\mu}$ and unknown covariance matrix Σ, and test the equality of mean indices at .05 level. ($F_{2,38}(.05) = 3.25$)

(25)

We would like to test $H_0: \mu_1 = \mu_2 = \mu_3$ vs $H_1: H_0$ is not true.

We can write Ho as $H_0: C\underline{\mu} = 0$, where

$$C = \begin{bmatrix} 1 & -1 & 0 \\ 0 & 1 & -1 \end{bmatrix} \text{ and } \underline{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix}.$$

Then,

$$C\bar{\underline{X}} = \begin{bmatrix} -11.2 \\ 6.9 \end{bmatrix}, \quad CSC' = \begin{bmatrix} 55.5 & -32.6 \\ -32.6 & 66.4 \end{bmatrix}$$

and

$$T^2 = n(C\bar{\underline{x}})'(CSC')^{-1}(C\bar{\underline{x}}) = 90.4.$$

With $n = 40$ and $q = 3$

$$\frac{(n-1)(q-1)}{(n-q+1)} F_{q-1,\,n-q+1}(.05) = \frac{(40-1)(3-1)}{40-3+1} F_{2,38}(.05)$$

$$= \frac{(39)(2)}{38}(3.25)$$

$$= 6.67$$

Since $T^2 = 90.4 > 6.67$, we can reject Ho at .05 level.