Mathematical Sciences 807
Applied Multivariate Statistical Analysis
Midterm

I. Suppose the weight(lbs.), X_1, height(in.), X_2, and age(yrs.) X_3 of a randomly chosen patient in a psychiatric study is multivariate normally distributed with means 170, 68, and 40 and variances 400, 16 and 256, respectively. Furthermore, suppose we have the covariances between each pair of variables, cov(X_1, X_2) = 64, cov(X_1, X_3) = 128, cov(X_2, X_3) = 0.

i) Determine the conditional distribution of height given weight.

ii) Determine the conditional distribution of weight given height and age.

iii) How important is age in the determining an individual's weight, if the individual is 72 inches tall ? Give your solution as a percentage reduction or increase in the conditional mean and variance.

II. Given the previous exercise, suppose a sample of 20 patients yielded the following results: $\overline{x}_1$ = 165, $\overline{x}_2$= 64, and $\overline{x}_3$ = 43 with variances 441, 25 and 256, respectively cov(X_1, X_2) = 60, cov(X_1, X_3) = 128, cov(X_2, X_3)=0. Test the hypothesis:

$$H_0 : \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix} = \begin{bmatrix} 170 \\ 45 \end{bmatrix}$$

vs

$$H_A : \begin{bmatrix} \mu_1 \\ \mu_3 \end{bmatrix} \neq \begin{bmatrix} 170 \\ 45 \end{bmatrix}$$

at α = 0.01. Sketch the rejection region for the univariate tests (should be rectangular) and for the multivariate test.

III. The correlation matrix of data published by Rollet (1889) on the stature (S) of 50 men, dissected at Lyons, France in order to get measurements of the length of femur (F), humerus (H), tibia (T), and radius (R) is given below:

$$R = \begin{array}{c} \\ F \\ H \\ T \\ R \\ S \end{array} \begin{array}{c} \begin{array}{ccccc} F & H & T & R & S \end{array} \\ \begin{bmatrix} 1 & 0.8421 & 0.8058 & 0.7439 & 0.8105 \\ 0.8421 & 1 & 0.8601 & 0.8451 & 0.8091 \\ 0.8058 & 0.8601 & 1 & 0.7804 & 0.7769 \\ 0.7439 & 0.8451 & 0.7804 & 1 & 0.6956 \\ 0.8105 & 0.8091 & 0.7769 & 0.6956 & 1 \end{bmatrix} \end{array}$$

with eigenvalues

Variable	F	H	T	R	S
Eigenvalue	4.1905	0.3242	0.1931	0.1812	0.1111

i) What does this say about the variation in body measurements ?

ii) Eliminating F gives the partial correlations:

$$R = \begin{array}{c} \\ H \\ T \\ R \\ S \end{array} \begin{array}{c} \begin{array}{cccc} H & T & R & S \end{array} \\ \begin{bmatrix} 1 & 0.5682 & 0.6068 & 0.4007 \\ 0.5682 & 1 & 0.4574 & 0.3569 \\ 0.6068 & 0.4574 & 1 & 0.2376 \\ 0.4007 & 0.3569 & 0.2376 & 1 \end{bmatrix} \end{array}$$

Do these correlations offer anymore information about the relationship be-

Variable	H	T	R	S
Eigenvalue	2.3397	0.7864	0.5226	0.3512

tween F and the remaining variables ?

iii) Next, consider the elimination of H:

$$R = \begin{array}{c} \\ T \\ R \\ S \end{array} \begin{array}{c} \begin{array}{ccc} T & R & S \end{array} \\ \begin{bmatrix} 1 & 0.1772 & 0.1714 \\ 0.1722 & 1 & -0.0076 \\ 0.1714 & -0.0076 & 1 \end{bmatrix} \end{array}$$

with eigenvalues: Do these correlations offer anymore information about the

Variable	T	R	S
Eigenvalue	1.2410	1.0076	0.7514

relationship between F and H removed and the remaining variables ?

IV. Table 1 gives statistics of bivariate observations extracted by Seber (1984) from Lubischew (1962) on specimens of three species of male flea beetles. The two variables measured were maximal width of aedeagus in the forepart (x_1) and front angle of aedeagus (x_2); there were n_1 = 21 specimens of Chaetocnema concinna (G_1), n_2 = 31 specimens of Chaetocnema heikertingeri (G_2) and n_3 = 22 specimens of Chaetocnema heptapotamica (G_3).

Table 1: Mean, Covariance, Between, Within, and Total varation matrices

$$\overline{\mathbf{G}}_1 = \begin{bmatrix} 146.1905 \\ 14.0952 \end{bmatrix} \qquad \mathbf{S}_1 = \begin{bmatrix} 31.6619 & -0.9691 \\ -0.9691 & 0.7905 \end{bmatrix} \qquad \mathbf{B} = \begin{bmatrix} 6205.9053 & -351.3439 \\ -351.3439 & 274.8343 \end{bmatrix}$$

$$\overline{\mathbf{G}}_2 = \begin{bmatrix} 124.6452 \\ 14.2903 \end{bmatrix} \qquad \mathbf{S}_2 = \begin{bmatrix} 21.3699 & -0.3269 \\ -0.3269 & 1.2129 \end{bmatrix} \qquad \mathbf{W} = \begin{bmatrix} 1634.6985 & -39.7329 \\ -39.7329 & 72.01489 \end{bmatrix}$$

$$\overline{\mathbf{G}}_3 = \begin{bmatrix} 138.2727 \\ 10.0909 \end{bmatrix} \qquad \mathbf{S}_3 = \begin{bmatrix} 17.1602 & -0.5022 \\ -0.5022 & 0.9437 \end{bmatrix} \qquad \mathbf{T} = \begin{bmatrix} 7840.6038 & -391.0768 \\ -391.0768 & 346.8491 \end{bmatrix}$$

i) Is it reasonable to assume that the variance-covariance matrices are not significantly different, ie. equal ? Explain.

ii) Box's criteria for common variance-covariance matrices is given by:

$$\begin{aligned} M &= k\,\{(n_1+n_2+n_3-3)\,ln \mid \mathbf{W} \mid \\ &- (n_1-1)\,ln \mid \mathbf{S}_1 \mid - (n_2-1)\,ln \mid \mathbf{S}_2 \mid - (n_3-1)\,ln \mid \mathbf{S}_3 \mid\} \end{aligned}$$

where k = 0.9156, $\mid \mathbf{W} \mid$ = 116143.7898, $\mid \mathbf{S}_1 \mid$ = 24.0889, $\mid \mathbf{S}_2 \mid$ = 25.8126, and $\mid \mathbf{S}_3 \mid$ = 15.9423. M$\sim \chi^2$ with $\frac{1}{2}p(p+1)$ degrees of freedom, where we would reject H_0: common variance-covariance for large values of *M*. Calculate *M*. Does this confirm your assessment from part (i) ?

iii) Assuming that your answer to parts i) and ii) where affirmative, determine Wilk's Lambda in order to test if there are differences in the three type of beetles.

iv) Test if there are differences in the three types of beetles. Use α = 0.01.

V. The table below gives observations of four measurements, chest, mid-upper arm, height, in centimeters, and age in months for 9 young children in a growth control study.

Table 2: Four measurements taken on each of nine young children

Child	Chest circumference (cm)	Mid upper arm circumference (cm)	Height (cm)	Age (Months)
1	58.4	14.0	80	21
2	59.2	15.0	75	27
3	60.3	15.0	78	27
4	57.4	13.0	75	22
5	59.5	14.0	79	26
6	58.1	14.5	78	26
7	58.0	12.5	75	23
8	55.5	11.0	64	22
9	59.2	12.5	80	22

(a) Assuming that these four variables are multivariate normally distributed, ie. $\begin{bmatrix} y_1 \\ y_2 \\ x_1 \\ x_2 \end{bmatrix} \sim \mathbf{N}_4 \left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$ find the estimated mean and variance of the conditional expectation of chest and mid-upper arm circumference given height and age based on this data.