Stat524 Midterm (Oct. 4 2000)

1). Suppose $\vec{X}' = (X_1, X_2, X_3, X_4)$ is multivariate normal with mean $\mu = (1, 1, 1, 1)$ and covariance matrix

$$\Sigma = \begin{pmatrix} 6.5 & -2 & 0 & 1 \\ -2 & 6.5 & 0 & -1 \\ 0 & 0 & 3 & 0 \\ 1 & -1 & 0 & 5 \end{pmatrix}.$$

(a)(2) What is the joint distribution of (X_1, X_4) given $X_3 = x_3$?

(b)(3) Two of the four eigenvalues of Σ are known to be 4.5 and 9, determine the other two eigenvalues.

(c)(3) Determine the shortest axis of the constant density ellipsoid that contains 95% of the probability.

(d)(2) Suppose $\vec{X}_1, \ldots, \vec{X}_{10}$ is a random sample from the population. What is the distribution of $\bar{X} = (\vec{X}_1 + \cdots + \vec{X}_{10})/10$?

2). Three psychological tests were given to 20 men, and part of the data is show in the following table. The variables are: y_1 =pictorial inconsistencies, y_2=tool recognition and y_3= vocabulary.

case	y_1	y_2	y_3
1	15	24	14
2	17	32	26
3	15	29	23
⋮	⋮	⋮	⋮
15	18	34	26
16	16	16	16
17	16	28	27
18	18	32	23
19	17	21	21
20	11	25	23

The sample mean is $\bar{x}'_m = (15.7, 26.3, 22.5)$, and the sample covariance matrix is

$$S_m = \begin{pmatrix} 6.536842 & 8.884211 & 7.157895 \\ 8.884211 & 38.957895 & 19.578947 \\ 7.157895 & 19.578947 & 20.052632 \end{pmatrix}$$

and the inverse of S_m is

$$S_m^{-1} = \begin{pmatrix} 0.26310731 & -0.02513394 & -0.06937735 \\ -0.02513394 & 0.05280054 & -0.04258159 \\ -0.06937735 & -0.04258159 & 0.11620911 \end{pmatrix}$$

(a)(3) Test whether the average scores of men in the three exams are 15, 29 and 24. Give the explicit expression to calculate the test statistic T^2. In fact, $T^2 = 13.42$. What is your conclusion at $\alpha = 5\%$?

(b)(3) Derive the 95% simultaneous confidence interval for μ_1, the average score of men in the first test.

(c)(3) The simultaneous confidence intervals for μ_2 and μ_3 are (21.7, 30.9) and (19.2, 25.8). Observe that the simultaneous confidence intervals contain 15, 29 and 24 respectively. Is this contradictory to your conclusion in (a)? Explain.

(d)(5) Suppose the researchers were interested in whether the average scores of men in the three tests are the same. Which procedure you can apply to answer the question. Show the steps explicitly. You are not required to get the numeric results.

In the study, 20 women were also selected to take the same three tests. Their scores are summarized as $\bar{x}_f = (13.15, 16.75, 22.65)$ and

$$S_f = \begin{pmatrix} 4.134211 & 5.355263 & 2.739474 \\ 5.355263 & 31.881579 & 12.171053 \\ 2.739474 & 12.171053 & 32.239474 \end{pmatrix}$$

(e)(6) Is there any significant difference between the scores of men and women in those tests? Show the steps that lead to the answer. What assumptions are needed to make the procedure valid?

3). In a classical experiment, apple trees of different rootstocks were compared. The data for eight trees from each of three rootstocks were given in the following table. And the variables are y_1=trunk girth at 4 years (mm×100) and y_2=extension growth at 4 years (m).

root	trunk	exten
1	1.11	2.569
1	1.19	2.928
1	1.09	2.865
1	1.25	3.844
1	1.11	3.027
1	1.08	2.336
1	1.11	3.211
1	1.16	3.037
2	1.25	2.074
2	1.47	2.885
2	1.21	3.378
2	1.65	3.906
2	1.27	2.782
2	1.35	3.018
2	1.47	3.383
2	1.39	3.447
3	1.07	2.505
3	0.99	2.315
3	1.06	2.667
3	1.02	2.390
3	1.15	3.021
3	1.20	3.085
3	1.20	3.308
3	1.17	3.231

The summary statistics are as follows. For all the apple trees, the sample mean vector is $\bar{x}' = (1.21, 2.97)$ and the total sample covariance matrix is

$$S = \begin{pmatrix} 0.0254 & 0.0447 \\ 0.0447 & 0.2159 \end{pmatrix}$$

For the apple trees of rootstock 1, the sample mean vector is $\bar{x}_1' = (1.14, 2.98)$, and the sample covariance matrix is

$$S_1 = \begin{pmatrix} 0.0034 & 0.0203 \\ 0.0203 & 0.2007 \end{pmatrix}.$$

For the apple trees of rootstock 2, the sample mean vector is $\bar{x}_2' = (1.38, 3.11)$ and the sample covariance matrix is

$$S_2 = \begin{pmatrix} 0.0211 & 0.0501 \\ 0.0501 & 0.3048 \end{pmatrix}.$$

For the apple trees of rootstock 3, the sample mean vector is $\bar{x}_3' = (1.11, 2.82)$ and the sample covariance matrix is

$$S_3 = \begin{pmatrix} 0.0069 & 0.0314 \\ 0.0314 & 0.1543 \end{pmatrix}.$$

(a)(6) Construct the MANOVA table for testing that there no difference among apple trees of different rootstocks.

(b)(4) Let B be the treatment matrix of SSP and W the residual matrix of SSP in the MANOVA table. Suppose the eigenvalues of $W^{-1}B$ are $\lambda_1 = 2.56$, $\lambda_2 = 0.02$. Calculate the Wilks' test statistic.

(c)(5)The null hypothesis was indeed rejected. Bonferroni confidence intervals can be used to detect where the differences are. Derive the 95% confidence intervals for the comparison between rootstock 1 and rootstock 2.

4). Six hematology variables were measured on 20 workers. The variables are y_1=hemoglobin concentration, y_2= packed cell volume, y_3 =white block cell count, y_4=lymphocte count, y_5=neutrophil count and y_6=serum lead concentration. This data is shown in the following table.

Observation	y_1	y_2	y_3	y_4	y_5	y_6
1	13.4	39	410	14	25	17
2	14.6	46	500	15	30	20
3	13.5	42	450	19	21	18
4	15.0	46	460	23	16	18
5	14.6	44	510	17	31	19
6	14.0	44	490	20	24	19
7	16.4	49	430	21	17	18
8	14.8	44	440	16	26	29
9	15.2	46	410	27	13	27
10	15.5	48	840	34	42	36
11	15.2	47	560	26	27	22
12	16.9	50	510	28	17	23
13	14.8	44	470	24	20	23
14	16.2	45	560	26	25	19
15	14.7	43	400	23	13	17
16	14.7	42	340	9	22	13
17	16.5	45	540	18	32	17
18	15.4	45	690	28	36	24
19	15.1	45	460	17	29	17
20	14.2	46	420	14	25	28

The sample mean vector is $\bar{x}' = (15, 45, 494.5, 20.1, 24.6, 21.2)$. The sample covariance matrix S is

	y_1	y_2	y_3	y_4	y_5	y_6
y_1	0.8823947	1.7578947	35.67632	2.849211	0.1902632	0.5715789
y_2	1.7578947	6.3157895	113.68421	8.736842	0.5789474	5.9473684
y_3	35.6763158	113.6842105	12289.21053	474.973684	632.6578947	343.7894737
y_4	2.8492105	8.7368421	474.97368	37.839474	3.7657895	17.8000000
y_5	0.1902632	0.5789474	632.65789	3.765789	57.1026316	14.6210526
y_6	0.5715789	5.9473684	343.78947	17.800000	14.6210526	29.4315789

(a)(5) Derive the first principal component based on S. What is the contribution of the first component to the total sample variance?

The result of (a) suggests that a more informative approach should be based on the sample correlation matrix R, which is

	y_1	y_2	y_3	y_4	y_5	y_6
y_1	0.99999988	0.74464202	0.3425994	0.49308354	0.02680381	0.1121600
y_2	0.74464202	0.99999994	0.4080603	0.56515563	0.03048577	0.4362186
y_3	0.34259939	0.40806028	1.0000001	0.69652230	0.75522941	0.5716416
y_4	0.49308354	0.56515563	0.6965223	0.99999994	0.08101314	0.5333849
y_5	0.02680381	0.03048577	0.7552294	0.08101314	0.99999988	0.3566513
y_6	0.11215997	0.43621865	0.5716416	0.53338486	0.35665128	1.0000001

The eigenvalues and eigenvectors of R are:

values:

λ_1	λ_2	λ_3	λ_4	λ_5	λ_6
3.13830828	1.46844737	0.74461661	0.47456551	0.16391780	0.01014431

vectors:

e_1	e_2	e_3	e_4	e_5	e_6
0.3589955	−0.4872426	−0.49558993	−0.05094799	−0.61887954	−0.04999924
0.4268959	−0.4177236	−0.05382989	−0.45903750	0.65178284	0.06951970
0.4917640	0.3404033	−0.15317549	0.30288498	0.17197182	−0.70534913
0.4652894	−0.1526545	0.32666195	0.65633971	0.03645569	0.47051406
0.2683168	0.6334820	−0.47289802	−0.18071950	0.00846533	0.51994513
0.3981987	0.2182302	0.63062641	−0.48115301	−0.40148756	−0.05850865

(b)(5) Derive the major principal components. Justify your choices.

(c)(5) Can you detect any collinearity in the original variables? If yes, derive a possible approximate linear relationship explicitly.