Question 1:

a. Given **X** is a p-variate random vector with mean vector μ and covariance matrix Σ. Write down the necessary and sufficient condition (without proof) for **X** to be a multivariate normal random vector.

b. Let $U_i(i = 0, 1, 2, 3)$ be independent standard normal random variables and $X^T = (X_1, X_2, X_3)$ where

$$X_1 = \frac{1}{\sqrt{2}}U_0 + \frac{1}{\sqrt{2}}U_1,$$

$$X_2 = \frac{1}{\sqrt{2}}U_2 + \frac{1}{\sqrt{2}}U_3 \text{ and}$$

$$X_3 = \frac{1}{\sqrt{2}}U_0 + \frac{1}{\sqrt{2}}U_3.$$

(i). Obtain the mean, $\mu = E(X)$ and $\Sigma = Cov(X)$.
(ii) Which of the following random variables are independent? Explain.
(1) X_1 and X_2
(2) X_2 and X_3
(3) (X_1, X_3) and X_2

Question 2:
Answer this question using the SAS program and its output given in Appendix 1. The data set given below contains 9 independent measurements of a bivariate random vector $X^T = (X_1, X_2)$ with unknown mean μ and covariance matrix Σ.

x_1	6.5	7.1	6.1	5.2	7.2	6.1	4.6	5.1	5.4
x_2	2.3	1.0	2.2	1.4	2.5	1.6	2.6	2.5	2.0

(a) Estimate the unknown parameters μ and Σ.
(b) Test the hypothesis $H_0 : \mu^T = (6.5, 3.0)$ against $H_1 : \mu^T \ 6\neq (6.5, 3.0)$ at level 0.05.
(c) Obtain 95% confidence region for the mean vector, μ and draw the confidence region.
(d) Compare the results in part (b) and (c).
(e) List any assumptions used for the above parts.

Question 3:

Given $X^T = (X_1, X_2, X_3, X_4)$ is a normal random vector with mean vector $\mu^T = (\mu_1, \mu_2, \mu_3, \mu_4)$ and covariance matrix Σ. The sample mean vector and the sample covariance matrix computed using 44 random observations of X are given below.

$$\bar{X}=\begin{pmatrix}2.0\\3.5\\4.5\\2.8\end{pmatrix};\qquad S=\begin{pmatrix}3.6 & 0.4 & 2.0 & 2.8\\0.4 & 5.6 & 0.8 & 3.0\\2.0 & 0.8 & 3.6 & 1.8\\2.8 & 3.0 & 1.8 & 6.4\end{pmatrix}$$

Answer the following questions using the SAS program and its output given in Appendix 2.

(a) Write down the first two principal components of X and their variances.
(b) Compute a 95% confidence interval for λ_2, the second largest eigenvalue of Σ.

Question 4:

(a) Let X be a random vector with mean μ and covariance matrix Σ given below:

$$\mu=\begin{pmatrix}3\\2\\2\end{pmatrix};\qquad \begin{pmatrix}9 & 3.78 & 6.75\\3.78 & 4 & 3.5\\6.75 & 3.5 & 25\end{pmatrix}$$

Obtain a factor model with one common factor. List any assumptions used to obtain the model.

(b) The distance between pairs of five items, A, B, C, D and E are as follows:

	A	B	C	D	E
A	0				
B	2	0			
C	2	8	0		
D	1	3	2	0	
E	1	2	3	1	0

Performed the Hierarchical clustering procedure on these five items using the complete linkage method. Draw a dendrogram.

Question 5:

The manager of the commercial loan department of a bank wants to develop a rule to use in determining whether or not to approve various request for a loan. The manager believes that three key characteristics of a company's performance are important in making this decision: liquidity, profitability and activity. The manager measures liquidity as the ratio of current assets to current liabilities. Profitability is measured as the ratio of net profit to sales. Activity is measured as the ratio of sales to fixed assets.
The manager has obtained the following data from a random sample of 18 loans processed in the past five years. These loans have been classified into two groups: 1 for the loans that were accepted and 2 for those loans that were rejected.

Data	Group	Liquidity	Profitability	Activity
1	1	0.90	0.34	1.53
2	1	0.88	0.23	1.67

3	1	0.92	0.28	1.43
4	1	0.89	0.14	1.24
5	1	0.78	0.35	1.80
6	1	0.81	0.26	2.01
7	1	0.72	0.18	1.75
8	1	0.93	0.22	0.99
9	1	0.82	0.26	1.40
10	2	0.78	0.26	1.34
11	2	0.78	0.27	1.67
12	2	0.72	0.18	1.53
13	2	0.69	0.16	1.20
14	2	0.63	0.15	0.88
15	2	0.58	0.22	1.42
16	2	0.81	0.18	1.59
17	2	0.67	0.21	1.21
18	2	0.65	0.16	1.37

Use the SAS Statistical package output given in Appendix 3 to answer the following questions.

(a) List the assumptions used in the discriminant analysis given in SAS output.

(b) Construct Fisher's linear discriminant function.

(c) Evaluate the performance of the above discriminant function by computing the apparent error rate (APER) and the estimated expected actual error rate $\hat{E}(AER)$.

(d) Suppose that the manager receives loan applications from two companies with the following financial information. Which of these loans will the manager approve?

Data	Liquidity	Profitability	Activity
A	0.78	0.27	1.58
B	0.67	0.26	1.23

Appendix 1

```
Data temp;
  input x1 x2;
datalines;
 6.5 2.3
 7.1 1.0
 6.1 2.2
 5.2 1.4
 7.2 2.5
 6.1 1.6
 4.6 2.6
 5.1 2.5
 5.4 2.0
 ;
  proc IML;
   use temp;
   vname={x1 x2};
   read all var vname into x;
;
  n=nrow(x);
  p=ncol(x);
  reset print;
  ax=round(T(x[+,]/n),0.01);
  xpx=round(t(x)*x-n*ax*t(ax),0.01);
  cx=round(xpx/(n-1),0.01);
  ix=round(inv(cx),0.0001);
  mu={6.5,3.0};
  xx=ax-mu;
  tt=t(xx)*ix*xx;
  f1=finv(0.95,1,7);
  f2=finv(0.95,2,7);
  f3=finv(0.95,2,8);
  t1=tinv(0.95,7);
  t2=tinv(0.95,8);
  t3=tinv(0.95,9);
  t5=tinv(0.975,7);
  t6=tinv(0.975,8);
  t7=tinv(0.975,9);
  v1=round(eigval(cx),0.01);
  v2=round(eigvec(cx),0.01);
  v3=round(eigval(ix),0.01);
  v4=round(eigvec(ix),0.01);
quit;
```

```
 AX       2 rows       1 col      (numeric)
                       5.92
                       2.01

XPX       2 rows       2 cols     (numeric)
                6.87      -1.07
               -1.07       2.55

 CX       2 rows       2 cols     (numeric)
                0.86      -0.13
               -0.13       0.32

 IX       2 rows       2 cols     (numeric)
              1.2389    0.5033
              0.5033    3.3295

 MU       2 rows       1 col      (numeric)
                        6.5
                          3

 XX       2 rows       1 col      (numeric)
                      -0.58
                      -0.99

 TT       1 row        1 col      (numeric)
                4.2579986

 F1       1 row        1 col      (numeric)
                5.5914479
 F2       1 row        1 col      (numeric)
                4.7374141

 F3       1 row        1 col      (numeric)
                4.4589701


 T1       1 row        1 col      (numeric)
                1.8945786

 T2       1 row        1 col      (numeric)
                 1.859548

 T3       1 row        1 col      (numeric)
                1.8331129
```

```
T5      1 row       1 col     (numeric)
             2.3646243

T6      1 row       1 col     (numeric)
             2.3060041

T7      1 row       1 col     (numeric)
             2.2621572

V1      2 rows      1 col     (numeric)
                  0.89
                  0.29

V2      2 rows      2 cols    (numeric)
             0.97       0.22
            -0.22       0.97

V3      2 rows      1 col     (numeric)
                  3.44
                  1.12

V4      2 rows      2 cols    (numeric)
             0.22       0.97
             0.97      -0.22
```

Appendix 2

```
 Proc IML;
  S={3.6 0.4 2.0 2.8,
     0.4 5.6 0.8 3.0,
     2.0 0.8 3.6 1.8,
     2.8 3.0 1.8 6.4};
;
  reset print;
  InS=round(inv(S),0.01);
;
  u1=round(eigval(InS),0.01);
  u2=round(eigvec(InS),0.01);
  v1=round(eigval(S),0.01);
  v2=round(eigvec(S),0.01);
quit;
```

```
INS       4 rows      4 cols     (numeric)
      0.59        0.12       -0.23       -0.25
      0.12        0.27       -0.05       -0.17
     -0.23       -0.05        0.41           0
     -0.25       -0.17           0        0.34

U1       4 rows      1 col     (numeric)
                      0.9
                     0.41
                     0.22
                     0.08

U2       4 rows      4 cols     (numeric)
      0.76       -0.08        0.51        0.39
      0.29        0.35       -0.75        0.48
     -0.39        0.74        0.42        0.35
     -0.43       -0.57        0.01         0.7

V1       4 rows      1 col     (numeric)
                    10.86
                     4.77
                     2.45
                     1.12

V2       4 rows      4 cols     (numeric)
      0.39        0.52       -0.07        0.76
      0.48       -0.75        0.35        0.29
      0.34        0.42        0.75       -0.39
      0.71        0.03       -0.56       -0.44
```

Appendix 3

```
Data ExamQ4;
  input id Group Liquidity  Profitability Activity;
  datalines;
   1 1 0.90 0.34 1.53
   2 1 0.88 0.23 1.67
   3 1 0.92 0.28 1.43
   4 1 0.89 0.14 1.24
   5 1 0.78 0.35 1.80
   6 1 0.81 0.26 2.01
   7 1 0.72 0.18 1.75
   8 1 0.93 0.22 0.99
   9 1 0.82 0.26 1.40
  10 2 0.78 0.26 1.34
  11 2 0.78 0.27 1.67
  12 2 0.72 0.18 1.53
  13 2 0.69 0.16 1.20
  14 2 0.63 0.15 0.88
  15 2 0.58 0.22 1.42
  16 2 0.81 0.18 1.59
  17 2 0.67 0.21 1.21
  18 2 0.65 0.16 1.37
;
 proc discrim data=ExamQ4 method=NORMAL pool=yes
        list listerr crosslisterr;
   priors '1'=0.75 '2'=0.25;
   class Group;
   var Liquidity  Profitability Activity;
   run;
```

The DISCRIM Procedure

Observations	18	DF Total	17
Variables	3	DF Within Classes	16
Classes	2	DF Between Classes	1

Class Level Information

Group	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	_1	9	9.0000	0.500000	0.750000
2	_2	9	9.0000	0.500000	0.250000

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	-13.69973

The DISCRIM Procedure

Pairwise Generalized Squared Distances Between Groups

$$D^2(i|j) = (\bar{X}_i - \bar{X}_j)' \, COV^{-1} (\bar{X}_i - \bar{X}_j) - 2 \ln PRIOR_j$$

Generalized Squared Distance to Group

From Group	1	2
1	0.57536	7.58439
2	5.38716	2.77259

Linear Discriminant Function

$$\text{Constant} = -.5 \, \bar{X}_j' \, COV^{-1} \bar{X}_j + \ln PRIOR_j \qquad \text{Coefficient Vector} = COV^{-1} \bar{X}_j$$

Linear Discriminant Function for Group

Variable	1	2
Constant	-88.86548	-63.69619
Liquidity	164.85727	137.46255
Profitability	-8.06202	-12.64690
Activity	25.43143	22.67218

The DISCRIM Procedure

Classification Results for Calibration Data: WORK.EXAMQ4

Resubstitution Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in Each Group

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, \sum_k \exp(-.5 \, D_k^2(X))$$

Posterior Probability of Membership in Group

Obs	From Group	Classified into Group	1	2
1	1	1	0.9949	0.0051
2	1	1	0.9901	0.0099
3	1	1	0.9948	0.0052
4	1	1	0.9636	0.0364
5	1	1	0.9410	0.0590
6	1	1	0.9772	0.0228
7	1	1	0.5520	0.4480
8	1	1	0.9828	0.0172
9	1	1	0.9129	0.0871
10	2	1 *	0.7480	0.2520
11	2	1 *	0.8854	0.1146
12	2	2	0.4018	0.5982
13	2	2	0.0978	0.9022
14	2	2	0.0082	0.9918
15	2	2	0.0127	0.9873
16	2	1 *	0.9032	0.0968
17	2	2	0.0749	0.9251
18	2	2	0.0547	0.9453

* Misclassified observation

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.EXAMQ4

Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV^{-1} \, (X-\bar{X}_j)$$

Posterior Probability of Membership in Each Group

$$Pr(j|X) = \exp(-.5 \, D_j^2(X)) \, / \, SUM_k \exp(-.5 \, D_k^2(X))$$

Number of Observations and Percent Classified into Group

From Group	1	2	Total
1	9	0	9
	100.00	0.00	100.00
2	3	6	9
	33.33	66.67	100.00
Total	12	6	18
	66.67	33.33	100.00
Priors	0.75	0.25	

Error Count Estimates for Group

	1	2	Total
Rate	0.0000	0.3333	0.0833
Priors	0.7500	0.2500	

Error Count Estimates for Group

	1	2	Total
Rate	0.0000	0.3333	0.0833
Priors	0.7500	0.2500	

The DISCRIM Procedure

Classification Results for Calibration Data: WORK.EXAMQ4

Cross-validation Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \, COV_{(X)}^{-1} \, (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each Group

$$Pr(j|X) = \exp(-.5\ D_j^2(X)) \ / \ SUM_k \exp(-.5\ D_k^2(X))$$

Posterior Probability of Membership in Group

Obs	From Group	Classified into Group	1	2
7	1	2 *	0.1288	0.8712
10	2	1 *	0.8238	0.1762
11	2	1 *	0.9458	0.0542
16	2	1 *	0.9787	0.0213

* Misclassified observation

The DISCRIM Procedure

Classification Summary for Calibration Data: WORK.EXAMQ4

Cross-validation Summary using Linear Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_{(X)j})' \ COV_{(X)}^{-1} \ (X-\bar{X}_{(X)j})$$

Posterior Probability of Membership in Each Group

$$Pr(j|X) = \exp(-.5\ D_j^2(X)) \ / \ SUM_k \exp(-.5\ D_k^2(X))$$

Number of Observations and Percent Classified into Group

From Group	1	2	Total
1	8	1	9
	88.89	11.11	100.00
2	3	6	9
	33.33	66.67	100.00
Total	11	7	18
	61.11	38.89	100.00
Priors	0.75	0.25	

Error Count Estimates for Group

	1	2	Total
Rate	0.1111	0.3333	0.1667
Priors	0.7500	0.2500	