For the data shown in Table 8.1, divide the chemical companies into two groups: group I consists of those companies with a P/E less than 9, and group II consists of those companies with a P/E greater than or equal to 9. Group I should be considered mature or troubled firms, and group II should be considered growth firms. Perform a discriminant function analysis, using ROR5, D/E, SALESGR5, EPS5, NPM1, and PAYOUTR1. Assume equal prior probabilities and costs of misclassification. Test the hypothesis that the population D2 = 0. Produce a graph of the posterior probability of belonging to group I versus the value of the discriminant function. Estimate the probabilities of misclassification by several methods.

CODE:

```
data firms;
set chem;
if pe<9 then pe= 0;
else if pe>=9 then pe= 1;
proc discrim distance data = firms ;
class pe;
var ROR5 DE SALESGR5 EPS5 NPM1 PAYOUTR1;
run;
proc discrim data = firms crossvalidate;
class pe;
var ROR5 DE SALESGR5 EPS5 NPM1 PAYOUTR1;
run;
```

The SAS System

The DISCRIM Procedure

Total Sample Size	30	DF Total	29
Variables	6	DF Within Classes	28
Classes	2	DF Between Classes	1

Number of Observations Read	42
Number of Observations Used	30

Class Level Information

pe	Variable Name	Frequency	Weight	Proportion	Prior Probability
0	_0	12	12.0000	0.400000	0.500000
1	_1	18	18.0000	0.600000	0.500000

Pooled Covariance Matrix Information

Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
6	4.18306

The SAS System

The DISCRIM Procedure

Squared Distance to pe		
From pe	0	1
0	0	1.82693
1	1.82693	0

F Statistics, NDF=6, DDF=23 for Squared Distance to pe		
From pe	0	1
0	0	1.80083
1	1.80083	0

Prob > Mahalanobis Distance for Squared Distance to pe		
From pe	0	1
0	1.0000	0.1434
1	0.1434	1.0000

Prob > Mahalanobis distance is within 0.15

⇨ Fail To reject null hypothesis $D^2 = 0$ (Bonferroni inequality)

Generalized Squared Distance to pe		
From pe	0	1
0	0	1.82693
1	1.82693	0

Linear Discriminant Function for pe			
Variable	Label	0	1
Constant		-31.90052	-38.28292
ror5	ror5	1.24581	1.28715
de	de	27.30944	25.74322
salesgr5	salesgr5	1.37095	1.58963
eps5	eps5	-0.19533	-0.25480
npm1	npm1	0.64526	0.84870
payoutr1	payoutr1	47.65038	55.07132

Discrimination Function Analysis Resubstitution Summary=>

The SAS System

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.FIRMS
Resubstitution Summary using Linear Discriminant Function

Number of Observations and Percent Classified into pe			
From pe	0	1	Total
0	10 83.33	2 16.67	12 100.00
1	4 22.22	14 77.78	18 100.00
Total	14 46.67	16 53.33	30 100.00
Priors	0.5	0.5	

Error Count Estimates for pe			
	0	1	Total
Rate	0.1667	0.2222	0.1944
Priors	0.5000	0.5000	

Discrimination Function Analysis Cross-Validation Summary=>

The SAS System

The DISCRIM Procedure
Classification Summary for Calibration Data: WORK.FIRMS
Cross-validation Summary using Linear Discriminant Function

Number of Observations and Percent Classified into pe			
From pe	0	1	Total
0	9 75.00	3 25.00	12 100.00
1	6 33.33	12 66.67	18 100.00
Total	15 50.00	15 50.00	30 100.00
Priors	0.5	0.5	

Error Count Estimates for pe			
	0	1	Total
Rate	0.2500	0.3333	0.2917
Priors	0.5000	0.5000	

Is it possible to distinguish between men and women in the depression data set on the basis of income and level of depression? What is the classification function? What are your prior probabilities? Test whether the following variables help discriminate: EDUCAT, EMPLOY, HEALTH.

CODE:

```
proc discrim data = dep1;
class SEX;
var INCOME CESD;
run;
proc discrim data = dep1;
class SEX;
var EDUCAT EMPLOY HEALTH;
run;
proc discrim data = dep1;
class SEX;
var INCOME CESD EDUCAT EMPLOY HEALTH;
run;
```

Taking variables INCOME and Level of Depression:

Total Sample Size	294	DF Total	293
Variables	2	DF Within Classes	292
Classes	2	DF Between Classes	1

Number of Observations Read	294
Number of Observations Used	294

Class Level Information

SEX	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	_1	111	111.0000	0.377551	0.500000
2	_2	183	183.0000	0.622449	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
2	9.74816

Generalized Squared Distance to SEX		
From SEX	1	2
1	0	0.18425
2	0.18425	0

Linear Discriminant Function for SEX		
Variable	1	2
Constant	-1.87083	-1.58347
INCOME	0.11635	0.09326
CESD	0.12511	0.14879

Equations=>

Male = -1.87 +.116* INCOME + .125 * CESD
Female = -1.58 + .09*INCOME + .14 * CESD

WORK.DEP1 Resubstitution Summary using using LDA

Number of Observations and Percent Classified into SEX			
From SEX	1	2	Total
1	57 51.35	54 48.65	111 100.00
2	60 32.79	123 67.21	183 100.00
Total	117 39.80	177 60.20	294 100.00
Priors	0.5	0.5	

Error Count Estimates for SEX			
	1	2	Total
Rate	0.4865	0.3279	0.4072
Priors	0.5000	0.5000	

=>Variables INCOME and Level of Depression for classification of sex gives error rate as 40.72 %

Taking variables EDUCAT, EMPLOY, HEALTH:

Total Sample Size	294	DF Total	293
Variables	3	DF Within Classes	292
Classes	2	DF Between Classes	1

Number of Observations Read	294
Number of Observations Used	294

Class Level Information					
SEX	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	_1	111	111.0000	0.377551	0.500000
2	_2	183	183.0000	0.622449	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
3	0.72043

Generalized Squared Distance to SEX		
From SEX	1	2
1	0	0.22465
2	0.22465	0

Linear Discriminant Function for SEX		
Variable	1	2
Constant	-9.18416	-9.73107
EDUCAT	3.13499	3.09751
EMPLOY	1.16444	1.45247
HEALTH	2.97652	3.03149

Equations=>

Male = -9.18 + 3.13*EDUCAT + 1.16 * EMPLOY+2.97*HEALTH
Female = -9.73 + 3.09*EDUCAT + 1.45 * EMPLOY+3.03*HEALTH

WORK.DEP1 Resubstitution Summary using using LDA:

Number of Observations and Percent Classified into SEX			
From SEX	1	2	Total
1	86 77.48	25 22.52	111 100.00
2	101 55.19	82 44.81	183 100.00
Total	187 63.61	107 36.39	294 100.00
Priors	0.5	0.5	

Error Count Estimates for SEX			
	1	2	Total
Rate	0.2252	0.5519	0.3886
Priors	0.5000	0.5000	

Variables EDUCAT, EMPLOY, HEALTH for classification of sex gives error rate as 38.86 %.

Taking variables INCOME, CESD, EDUCAT, EMPLOY and HEALTH:

Total Sample Size	294	DF Total	293
Variables	5	DF Within Classes	292
Classes	2	DF Between Classes	1

Number of Observations Read	294
Number of Observations Used	294

Class Level Information					
SEX	Variable Name	Frequency	Weight	Proportion	Prior Probability
1	_1	111	111.0000	0.377551	0.500000
2	_2	183	183.0000	0.622449	0.500000

Pooled Covariance Matrix Information	
Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
5	10.23090

Generalized Squared Distance to SEX		
From SEX	1	2
1	0	0.33114
2	0.33114	0

Linear Discriminant Function for SEX		
Variable	1	2
Constant	-9.65008	-10.13496
INCOME	0.06030	0.04092
CESD	0.06939	0.08893
EDUCAT	2.87860	2.92996
EMPLOY	1.17998	1.44978
HEALTH	2.89879	2.90040

Eqautions=>

MALE=-9.65+.06*INCOME+.06*CESD+2.87*EDUCAT+1.17*EMPLOY+2.89*HEALTH
FEMALE=-10.13+.04*INCOME+.088*CESD+2.92*EDUCAT+1.44*EMPLOY +2.90*HEALTH

Number of Observations and Percent Classified into SEX			
From SEX	1	2	Total
1	86 77.48	25 22.52	111 100.00
2	84 45.90	99 54.10	183 100.00
Total	170 57.82	124 42.18	294 100.00
Priors	0.5	0.5	

Error Count Estimates for SEX			
	1	2	Total
Rate	0.2252	0.4590	0.3421
Priors	0.5000	0.5000	

Variables INCOME, CESD, EDUCAT, EMPLOY and HEALTH for classification of sex gives error rate as 34.21%.