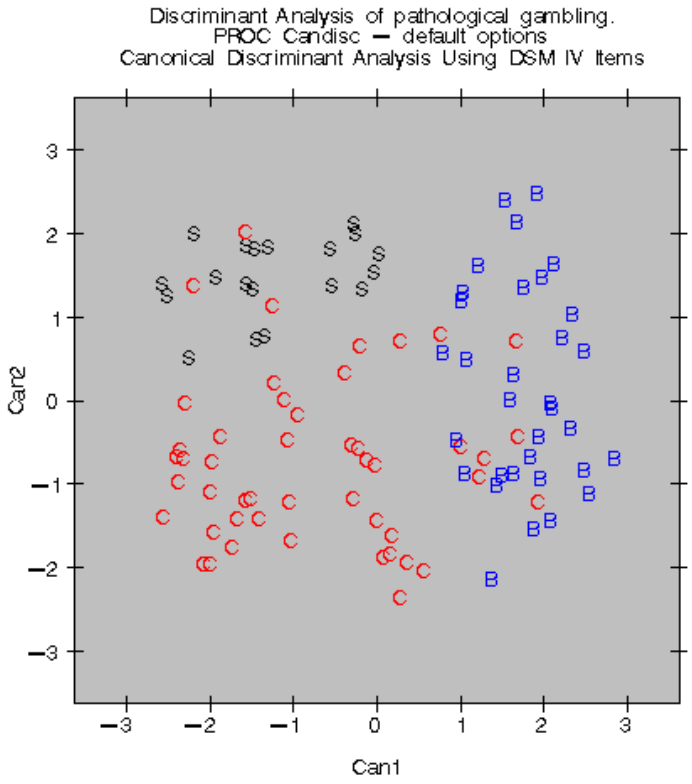
```
1          *** CH04S#D# ***;
2          ********************************************************************************;
3          *** The following example uses the gamble data set, which is composed of  ***;
4          *** data from two instruments that address pathological gambling:         ***;
5          *** The first instrument is a 12-item measure developed by Edward Johnson ***;
6          *** that is based on the ten DSM-IV diagnostic criteria for pathological  ***;
7          *** gambling. These items are named dsm1-dsm12.                           ***;
8          *** The second instrument is a 20-item questionnaire developed and used   ***;
9          *** by Gamblers Anonymous (GA) to help prospective members decide         ***;
10         *** whether they need help. These items are named ga1-ga20.               ***;
11         ********************************************************************************;
12         dm "output;clear;log;clear";
13         options ps=256 ls=99 nocenter nodate nonumber nolabel;
14
15         ods html style=minimal File='C:\EXST7037\Discrim\Gambling example\ch4_All01.html';
NOTE: Writing HTML Body file: C:\EXST7037\Discrim\Gambling example\ch4_All01.html
16         Title1 "Discriminant Analysis of pathological gambling.";
17         Libname amul "C:\EXST7037\Discrim\Gambling example\";
NOTE: Libref AMUL was successfully assigned as follows:
      Engine:        V9
      Physical Name: C:\EXST7037\Discrim\Gambling example
18
19         data gamble; set amul.gamble;
20           label  dsm1 = 'Wished stop thkg re gambling'
21                  dsm2 = 'Wished stop thkg re get money'
22                  dsm3 = 'Felt need to bet more and more'
23                  dsm4 = 'Rely on others for funds'
24                  dsm5 = 'Gamble to escape'
25                  dsm6 = 'Lie about how much I gamble'
26                  dsm7 = 'Relaxing difficult if not gambling'
27                  dsm8 = 'Win back money next day'
28                  dsm9 = 'Felt I should cut back on gambling'
29                  dsm10 = 'Illegal acts to pay for gambling'
30                  dsm11 = 'Danger of losing relationship'
31                  dsm12 = 'Danger of losing job'
32         run;
33
34         *** ch4s1d1.sas ***;
35         Title2 "PROC Candisc - default options";
NOTE: There were 100 observations read from the data set AMUL.GAMBLE.
NOTE: The data set WORK.GAMBLE has 100 observations and 33 variables.
NOTE: DATA statement used (Total process time):
      real time           0.01 seconds
      cpu time            0.01 seconds
36         proc candisc data = amul.gamblegrp out=candout;
37           class type;
38           var dsm1-dsm12;
39           title3 'Canonical Discriminant Analysis Using DSM IV Items';
40         run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: The data set WORK.CANDOUT has 100 observations and 46 variables.
NOTE: The PROCEDURE CANDISC printed pages 1-5.
NOTE: PROCEDURE CANDISC used (Total process time):
      real time           0.12 seconds
      cpu time            0.06 seconds
41
42         *symlen=1 sets symbol length to 1 - only the first letter
43         of the Type variable will be used as a symbol on the plot.;
44         %let plotitop =  cback   = white, cframe  = ligr, color   = black,
45                     colors  = black red blue ;
46
47         %plotit (data=candout, plotvars=can2 can1, labelvar=_blank_,
48         symvar=type, typevar=type, symsize = 1, symlen=1);
Overridden Parameters:
cback=white
cframe=ligr
color=black
colors=black red blue
```

```
                    Types Legend  | Steady    Control   Binge
                    --------------+-----------------------------
                    Symbol Types  | symbol    symbol    symbol
                    Symbols       |
                    Symbol Colors | black     red       blue
                    Label  Colors | black     red       blue
                    Symbol Sizes  | 1         1         1
                    Label  Sizes  | 1         1         1
                    Symbol Fonts  | swiss     swiss     swiss
                    Label  Fonts  | swiss     swiss     swiss
                    ---------------------------------------
                    Iterative Scatter Plot of Labeled Points Macro
                Iteration    Place    Line Size    Page Size    Penalty
                ---------------------------------------------------------
                    1          2         65           45           0
The following code will create the (empty) printer plot on which the graphical plot is based:
options nonumber ls=65 ps=45;
proc plot nolegend formchar='|----|+|---' data=preproc vtoh=2;
   plot Can2 * Can1 $ _blank_ = _symbol_ / haxis=by 1 vaxis=by 1 box list=1
        placement=((h=2 -2 : s=right left) (v=1 -1 * h=0 -1 to -5 by alt));
   label Can2 = '#' Can1 = '#';
   run; quit;
The plot was created with the following goptions:
goptions reset=goptions erase hpos=129 vpos=40 hsize=15.00in vsize=9.34in device=WIN;
The OUT=anno Annotate data set has 186 observations.
The PLOTIT macro used 2.2 seconds to create OUT=anno.
49         title;

Discriminant Analysis of pathological gambling.
PROC Candisc - default options
Canonical Discriminant Analysis Using DSM IV Items

The CANDISC Procedure
Observations      100         DF Total                   99
Variables          12         DF Within Classes          97
Classes             3         DF Between Classes          2

Class Level Information
             Variable
type         Name         Frequency      Weight    Proportion
Binge        Binge               33     33.0000      0.330000
Control      Control             48     48.0000      0.480000
Steady       Steady              19     19.0000      0.190000

                Multivariate Statistics and F Approximations
                                S=2    M=4.5    N=42
Statistic                      Value    F Value    Num DF    Den DF    Pr > F
Wilks' Lambda             0.22283947       8.02        24       172    <.0001
Pillai's Trace            1.03412267       7.76        24       174    <.0001
Hotelling-Lawley Trace    2.33440867       8.28        24    144.92    <.0001
Roy's Greatest Root       1.62463115      11.78        12        87    <.0001
         NOTE: F Statistic for Roy's Greatest Root is an upper bound.
                NOTE: F Statistic for Wilks' Lambda is exact.

                                 Adjusted    Approximate      Squared
                      Canonical  Canonical      Standard    Canonical
                    Correlation  Correlation       Error  Correlation
                1      0.786762     0.755131    0.038293     0.618994
                2      0.644305     0.607476    0.058782     0.415129

Test of H0: The canonical correlations in the current row and all that follow are zero
            Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq)
                                                Likelihood Approximate
  Eigenvalue Difference Proportion Cumulative      Ratio F Value Num DF Den DF Pr > F
1     1.6246     0.9149     0.6959     0.6959 0.22283947    8.02     24    172 <.0001
2     0.7098                0.3041     1.0000 0.58487142    5.61     11     87 <.0001
```

S=2 M=4.5 N=42

Given the two matrices H and E Where p is the rank of (H+E), which is less than or equal to the number of columns of M. Let q be the rank of $L(X'WX)^{-}L'$ and v be the error d.f., then **S** = min(p,q), **M** = ($|p-q|$) and **N** = (v-p-1)/2.



Discriminant Analysis of pathological gambling.
PROC Candisc - default options
Canonical Discriminant Analysis Using DSM IV Items

The CANDISC Procedure

Total Canonical Structure

Variable	Can1	Can2
dsm1	-0.104959	0.883139
dsm2	-0.130310	0.771425
dsm3	0.012647	0.706136
dsm4	0.966578	0.155140
dsm5	0.114538	0.011325
dsm6	0.218150	-0.047926
dsm7	0.067765	-0.082651
dsm8	0.786913	-0.120752
dsm9	-0.173155	0.806900
dsm10	0.726576	0.061913
dsm11	0.149955	-0.124164
dsm12	0.751927	0.173628

Between Canonical Structure

Variable	Can1	Can2
dsm1	-0.143621	0.989633
dsm2	-0.202018	0.979382
dsm3	0.021866	0.999761
dsm4	0.991472	0.130321
dsm5	0.996738	0.080709
dsm6	0.984198	-0.177070
dsm7	0.707520	-0.706693
dsm8	0.992196	-0.124685
dsm9	-0.253481	0.967340
dsm10	0.997574	0.069614
dsm11	0.827663	-0.561226
dsm12	0.982586	0.185807

Pooled Within Canonical Structure

Variable	Can1	Can2
dsm1	-0.079185	0.825494
dsm2	-0.093349	0.684685
dsm3	0.008767	0.606468
dsm4	0.929851	0.184911
dsm5	0.070990	0.008697
dsm6	0.136750	-0.037223
dsm7	0.041948	-0.063389
dsm8	0.621581	-0.118176
dsm9	-0.126741	0.731758
dsm10	0.547243	0.057776
dsm11	0.093516	-0.095937
dsm12	0.581297	0.166305

Total-Sample Standardized Canonical Coefficients

Variable	Can1	Can2
dsm1	-0.588914684	0.738446884
dsm2	0.289005550	0.132345075
dsm3	0.124911015	-0.077526194
dsm4	1.553115936	0.480518256
dsm5	0.215763165	0.254302063
dsm6	0.043031944	-0.076191207
dsm7	-0.205764414	-0.151074644
dsm8	0.128692189	-0.912529408
dsm9	-0.026719098	0.403795951
dsm10	0.057461067	0.174112395
dsm11	-0.032670157	0.079537705
dsm12	-0.114665563	0.370479761

Pooled Within-Class Standardized Canonical Coefficients

Variable	Can1	Can2
dsm1	-0.486776415	0.610374537
dsm2	0.251576944	0.115205294
dsm3	0.112368187	-0.069741471
dsm4	1.006758447	0.311480812
dsm5	0.217083499	0.255858231
dsm6	0.042807167	-0.075793224
dsm7	-0.207283842	-0.152190226
dsm8	0.101596478	-0.720399386
dsm9	-0.022763336	0.344013962
dsm10	0.047574238	0.144154380
dsm11	-0.032668207	0.079532958
dsm12	-0.092492864	0.298840674

Raw Canonical Coefficients

Variable	Can1	Can2
dsm1	-0.414338207	0.519543436
dsm2	0.203333428	0.093113014
dsm3	0.087882689	-0.054544512
dsm4	1.092713919	0.338074560
dsm5	0.151802842	0.178917361
dsm6	0.030275656	-0.053605265
dsm7	-0.144768097	-0.106290433
dsm8	0.090542981	-0.642021348
dsm9	-0.018798552	0.284095634
dsm10	0.040427444	0.122498928
dsm11	-0.022985493	0.055959736
dsm12	-0.080674374	0.260655616

Class Means on Canonical Variables

type	Can1	Can2
Binge	1.770181344	0.169802865
Control	-0.733300659	-0.714794789
Steady	-1.221976459	1.510876596

```
52         options ps=256 ls=99 nocenter nodate nonumber nolabel;
53         *** ch4s2d1.sas ***;
54         Title1 "Discriminant Analysis of pathological gambling.";
55         Title2 "PROC Discrim - proportional priors";
56         proc discrim data = amul.gamblegrp list anova;
57           class type;
58           priors prop;
59           var dsm1-dsm12;
60         run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: The PROCEDURE DISCRIM printed pages 6-10.
NOTE: PROCEDURE DISCRIM used (Total process time):
      real time           0.23 seconds
      cpu time            0.18 seconds

Discriminant Analysis of pathological gambling.
PROC Discrim - proportional priors

The DISCRIM Procedure
Observations     100          DF Total                 99
Variables         12          DF Within Classes        97
Classes            3          DF Between Classes        2
```

Class Level Information

type	Variable Name	Frequency	Weight	Proportion	Prior Probability
Binge	Binge	33	33.0000	0.330000	0.330000
Control	Control	48	48.0000	0.480000	0.480000
Steady	Steady	19	19.0000	0.190000	0.190000

```
Pooled Covariance Matrix Information
                Natural Log of the
 Covariance     Determinant of the
Matrix Rank      Covariance Matrix
         12               -1.27952

Pairwise Generalized Squared Distances Between Groups
      2                          -1
     D (i|j) = (X  - X )' COV    (X  - X ) - 2 ln PRIOR
                 i    j          i    j             j

      Generalized Squared Distance to type
From
type             Binge       Control        Steady
Binge          2.21733       8.51787      14.07295
Control        9.26726       1.46794       8.51388
Steady        12.96881       6.66036       3.32146
```

Univariate Test Statistics
F Statistics, Num DF=2, Den DF=97

Variable	Total Standard Deviation	Pooled Standard Deviation	Between Standard Deviation	R-Square	R-Square / (1-RSq)	F Value	Pr > F
dsm1	1.4213	1.1748	0.9959	0.3306	0.4939	23.95	<.0001
dsm2	1.4213	1.2373	0.8790	0.2576	0.3469	16.82	<.0001
dsm3	1.4213	1.2786	0.7882	0.2071	0.2612	12.67	<.0001
dsm4	1.4213	0.9213	1.3285	0.5883	1.4290	69.30	<.0001
dsm5	1.4213	1.4300	0.1566	0.0082	0.0082	0.40	0.6716
dsm6	1.4213	1.4139	0.3020	0.0304	0.0314	1.52	0.2236
dsm7	1.4213	1.4318	0.1305	0.0057	0.0057	0.28	0.7587
dsm8	1.4213	1.1221	1.0808	0.3894	0.6376	30.92	<.0001
dsm9	1.4213	1.2109	0.9309	0.2888	0.4062	19.70	<.0001
dsm10	1.4213	1.1768	0.9925	0.3284	0.4889	23.71	<.0001
dsm11	1.4213	1.4213	0.2469	0.0203	0.0207	1.01	0.3695
dsm12	1.4213	1.1465	1.0428	0.3625	0.5686	27.58	<.0001

Average R-Square	
Unweighted	0.2347647
Weighted by Variance	0.2347647

Linear Discriminant Function

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

Variable	Linear Discriminant Function for type Binge	Control	Steady
Constant	-13.19550	-4.91699	-10.65029
dsm1	-0.72609	-0.14839	1.21042
dsm2	0.87634	0.28493	0.39280
dsm3	0.43956	0.26780	0.10346
dsm4	3.89561	0.86096	1.07942
dsm5	1.06698	0.52868	0.85271
dsm6	0.38691	0.35853	0.22443
dsm7	0.07963	0.53608	0.37026
dsm8	-0.26089	0.08036	-1.39281
dsm9	1.21308	1.00883	1.65032
dsm10	0.42288	0.21331	0.46620
dsm11	0.68002	0.68806	0.82384
dsm12	0.46495	0.43634	1.05590

Classification Results for Calibration Data: AMUL.GAMBLEGRP
Resubstitution Results using Linear Discriminant Function

Generalized Squared Distance Function

$$D^2_j(X) = (X-\bar{X}_j)'\ \text{COV}^{-1}\ (X-\bar{X}_j) - 2 \ln \text{PRIOR}_j$$

Posterior Probability of Membership in Each type

$$\Pr(j|X) = \exp(-.5\ D^2_j(X)) \ /\ \text{SUM}_k \exp(-.5\ D^2_k(X))$$

Posterior Probability of Membership in type

Obs	From type	Classified into type	Binge	Control	Steady
1	Steady	Steady	0.0061	0.4764	0.5175
2	Control	Control	0.0014	0.9061	0.0925
3	Binge	Binge	0.9945	0.0049	0.0006
4	Binge	Binge	0.5289	0.4708	0.0002
5	Control	Control	0.0003	0.9961	0.0036
6	Control	Control	0.1052	0.8781	0.0167
7	Control	Control	0.0020	0.9822	0.0158

8	Binge	Binge		0.9938	0.0046	0.0016
9	Control	Control		0.1495	0.5672	0.2833
10	Control	Control		0.0135	0.8317	0.1549
11	Control	Control		0.0006	0.7676	0.2318
12	Control	Steady	*	0.0088	0.2913	0.6998
13	Control	Binge	*	0.9610	0.0316	0.0074
14	Control	Control		0.0686	0.9299	0.0015
15	Steady	Steady		0.0002	0.0370	0.9627
16	Binge	Binge		0.9817	0.0183	0.0000
17	Binge	Binge		0.8265	0.0815	0.0920
18	Binge	Binge		0.8200	0.1513	0.0287
19	Control	Control		0.0856	0.8944	0.0200
20	Control	Control		0.0009	0.9474	0.0517
21	Control	Control		0.0924	0.7128	0.1948
22	Binge	Binge		0.9249	0.0750	0.0001
23	Control	Control		0.0010	0.9887	0.0102
24	Control	Control		0.0562	0.9423	0.0015
25	Steady	Steady		0.0024	0.0746	0.9230
26	Binge	Binge		0.9560	0.0433	0.0007
27	Control	Control		0.0001	0.9839	0.0160
28	Steady	Steady		0.0080	0.4701	0.5219
29	Steady	Steady		0.0029	0.1643	0.8328
30	Steady	Steady		0.0639	0.0851	0.8510
31	Control	Control		0.0003	0.9270	0.0727
32	Binge	Binge		0.8912	0.0094	0.0994
33	Steady	Steady		0.0540	0.0668	0.8792
34	Control	Control		0.0645	0.9023	0.0332
35	Control	Binge	*	0.7046	0.2932	0.0022
36	Control	Steady	*	0.0004	0.1315	0.8681
37	Control	Binge	*	0.7005	0.2088	0.0907
38	Steady	Steady		0.0037	0.0751	0.9212
39	Steady	Steady		0.0002	0.1441	0.8557
40	Steady	Steady		0.0350	0.1109	0.8541
41	Control	Binge	*	0.6535	0.3401	0.0063
42	Control	Control		0.0002	0.9961	0.0037
43	Control	Control		0.1006	0.8984	0.0011
44	Steady	Steady		0.1348	0.2608	0.6045
45	Control	Control		0.1462	0.8531	0.0007
46	Control	Control		0.0003	0.9621	0.0377
47	Steady	Steady		0.0001	0.1067	0.8932
48	Control	Binge	*	0.9159	0.0838	0.0002
49	Control	Control		0.0005	0.9915	0.0080
50	Steady	Steady		0.1651	0.1348	0.7001
51	Binge	Binge		0.9938	0.0062	0.0000
52	Control	Control		0.0007	0.9944	0.0049
53	Control	Control		0.0020	0.9891	0.0089
54	Control	Control		0.0102	0.9299	0.0599
55	Control	Control		0.0415	0.9502	0.0084
56	Control	Control		0.0016	0.9830	0.0154
57	Steady	Steady		0.0037	0.1869	0.8095
58	Steady	Steady		0.0017	0.0668	0.9315
59	Control	Control		0.0042	0.9917	0.0041
60	Control	Control		0.0761	0.8946	0.0294
61	Control	Control		0.0107	0.7528	0.2366
62	Control	Control		0.0004	0.9336	0.0661
63	Control	Control		0.3927	0.4165	0.1908
64	Control	Control		0.0004	0.9167	0.0829
65	Binge	Binge		0.8724	0.0405	0.0871
66	Binge	Binge		0.9298	0.0696	0.0007
67	Binge	Binge		0.9827	0.0173	0.0001
68	Steady	Control	*	0.0007	0.5066	0.4927
69	Binge	Binge		0.8256	0.0922	0.0823
70	Binge	Binge		0.9284	0.0683	0.0033
71	Binge	Binge		0.9715	0.0148	0.0137
72	Control	Binge	*	0.7705	0.2268	0.0028
73	Binge	Binge		0.8702	0.1290	0.0009
74	Steady	Steady		0.0526	0.2387	0.7087
75	Steady	Steady		0.0009	0.1251	0.8740

76	Binge	Binge		0.9463	0.0490	0.0047
77	Binge	Binge		0.9844	0.0076	0.0081
78	Binge	Binge		0.9884	0.0047	0.0069
79	Binge	Binge		0.9411	0.0090	0.0498
80	Control	Control		0.0006	0.9754	0.0240
81	Binge	Binge		0.6237	0.3676	0.0087
82	Binge	Binge		0.8751	0.1247	0.0002
83	Control	Control		0.0059	0.9824	0.0117
84	Binge	Binge		0.6979	0.2418	0.0604
85	Control	Binge	*	0.9179	0.0807	0.0014
86	Binge	Binge		0.9902	0.0083	0.0016
87	Control	Control		0.0589	0.9407	0.0004
88	Binge	Binge		0.9838	0.0159	0.0003
89	Steady	Steady		0.1750	0.1896	0.6354
90	Binge	Binge		0.7807	0.2182	0.0012
91	Binge	Binge		0.8259	0.1729	0.0012
92	Binge	Binge		0.6046	0.3918	0.0036
93	Control	Steady	*	0.0013	0.0456	0.9531
94	Control	Control		0.0648	0.9313	0.0040
95	Binge	Binge		0.9769	0.0224	0.0007
96	Control	Control		0.0179	0.8801	0.1020
97	Binge	Binge		0.9767	0.0225	0.0008
98	Binge	Binge		0.9337	0.0659	0.0003
99	Binge	Binge		0.9604	0.0036	0.0360
100	Control	Control		0.0879	0.9097	0.0024

* Misclassified observation

Classification Summary for Calibration Data: AMUL.GAMBLEGRP
Resubstitution Summary using Linear Discriminant Function

Generalized Squared Distance Function
Generalized Squared Distance Function

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

Posterior Probability of Membership in Each type

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

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	33	0	0	33
	100.00	0.00	0.00	100.00
Control	7	38	3	48
	14.58	79.17	6.25	100.00
Steady	0	1	18	19
	0.00	5.26	94.74	100.00
Total	40	39	21	100
	40.00	39.00	21.00	100.00
Priors	0.33	0.48	0.19	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.0000	0.2083	0.0526	0.1100
Priors	0.3300	0.4800	0.1900	

```
61
62          *** ch4s2d2.sas ***;
63          Title2 "PROC Discrim - equal priors";
64          proc discrim data = amul.gamblegrp list;
65            class type;
66            priors equal;
67            var dsm1-dsm12;
68          run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: The PROCEDURE DISCRIM printed pages 11-14.
NOTE: PROCEDURE DISCRIM used (Total process time):
      real time           0.12 seconds
      cpu time            0.07 seconds


Discriminant Analysis of pathological gambling.
PROC Discrim - equal priors

Class Level Information

                Variable                                                Prior
type            Name          Frequency       Weight    Proportion  Probability
Binge           Binge                33      33.0000      0.330000     0.333333
Control         Control              48      48.0000      0.480000     0.333333
Steady          Steady               19      19.0000      0.190000     0.333333


  Number of Observations and Percent Classified into type
From
type              Binge       Control        Steady         Total
Binge                33             0             0            33
                100.00          0.00          0.00       100.00

Control               8            36             4            48
                  16.67         75.00          8.33       100.00

Steady                0             0            19            19
                   0.00          0.00        100.00       100.00

Total                41            36            23           100
                  41.00         36.00         23.00       100.00

Priors         0.33333       0.33333       0.33333


                Error Count Estimates for type
                   Binge       Control        Steady         Total

Rate              0.0000        0.2500        0.0000        0.0833
Priors            0.3333        0.3333        0.3333


69
70          *** ch4s2d3.sas ***;
71          Title2 "PROC Discrim - priors set 0.4, 0.4, 0.2";
72          proc discrim data = amul.gamblegrp list;
73            class type;
74            priors 'Binge'=.4 'Steady'=.4 'Control'=.2;
75            var dsm1-dsm12;
76          run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: The PROCEDURE DISCRIM printed pages 15-18.
NOTE: PROCEDURE DISCRIM used (Total process time):
      real time           0.14 seconds
      cpu time            0.09 seconds
```

Discriminant Analysis of pathological gambling.
PROC Discrim - priors set 0.4, 0.4, 0.2

Class Level Information

type	Variable Name	Frequency	Weight	Proportion	Prior Probability
Binge	Binge	33	33.0000	0.330000	0.400000
Control	Control	48	48.0000	0.480000	0.200000
Steady	Steady	19	19.0000	0.190000	0.400000

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	33	0	0	33
	100.00	0.00	0.00	100.00
Control	8	33	7	48
	16.67	68.75	14.58	100.00
Steady	0	0	19	19
	0.00	0.00	100.00	100.00
Total	41	33	26	100
	41.00	33.00	26.00	100.00
Priors	0.4	0.2	0.4	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.0000	0.3125	0.0000	0.0625
Priors	0.4000	0.2000	0.4000	

```
77
78         *** ch4s3d1.sas ***;
79         Title2 "PROC Discrim - proportional priors - variance test";
80         proc discrim data = amul.gamblegrp pool = test slpool = .05;
81           title3 'Test for equality of covariance matrices';
82           title4 'and quadratic discriminant analysis';
83            class type;
84           priors prop;
85           var dsm1-dsm12;
86         run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: The PROCEDURE DISCRIM printed pages 19-22.
NOTE: PROCEDURE DISCRIM used (Total process time):
      real time           0.23 seconds
      cpu time            0.04 seconds
86       !      title;
```

Discriminant Analysis of pathological gambling.
PROC Discrim - proportional priors - variance test
Test for equality of covariance matrices
and quadratic discriminant analysis

Class Level Information

type	Variable Name	Frequency	Weight	Proportion	Prior Probability
Binge	Binge	33	33.0000	0.330000	0.330000
Control	Control	48	48.0000	0.480000	0.480000
Steady	Steady	19	19.0000	0.190000	0.190000

Discriminant Analysis of pathological gambling.
PROC Discrim - proportional priors - variance test
Test for equality of covariance matrices
and quadratic discriminant analysis

Within Covariance Matrix Information

type	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
Binge	12	-5.61076
Control	12	-1.26247
Steady	12	-10.13772
Pooled	12	-1.27952

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Notation: K = Number of Groups

P = Number of Variables

N = Total Number of Observations - Number of Groups

N(i) = Number of Observations in the i'th Group - 1

$$V = \frac{\prod_{i} |\text{Within SS Matrix}(i)|^{N(i)/2}}{|\text{Pooled SS Matrix}|^{N/2}}$$

$$RHO = 1.0 - \left(\sum_{i} \frac{1}{N(i)} - \frac{1}{N} \right) \frac{2P^2 + 3P - 1}{6(P+1)(K-1)}$$

$$DF = .5(K-1)P(P+1)$$

Under the null hypothesis:

$$-2\ RHO\ \ln \frac{N^{PN/2}\ V}{\prod_{i} N(i)^{PN(i)/2}}$$

is distributed approximately as Chi-Square(DF).

Chi-Square	DF	Pr > ChiSq
237.071443	156	<.0001

Since the Chi-Square value is significant at the 0.05 level, the within covariance matrices will be used in the discriminant function.
Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p252.

Pairwise Generalized Squared Distances Between Groups

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

Generalized Squared Distance to type

From type	Binge	Control	Steady
Binge	-3.39344	5.15437	137.68427
Control	27.58445	0.20546	74.79533
Steady	40.85911	6.07826	-6.81626

```
Discriminant Analysis of pathological gambling.
PROC Discrim - proportional priors - variance test
Test for equality of covariance matrices
and quadratic discriminant analysis

The DISCRIM Procedure
Classification Summary for Calibration Data: AMUL.GAMBLEGRP
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

 2           _         -1    _
D (X) = (X-X )' COV   (X-X ) + ln |COV | - 2 ln PRIOR
 j          j      j     j           j              j

Posterior Probability of Membership in Each type

                     2                     2
Pr(j|X) = exp(-.5 D (X)) / SUM exp(-.5 D (X))
                     j        k            k
```

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	32	1	0	33
	96.97	3.03	0.00	100.00
Control	1	45	2	48
	2.08	93.75	4.17	100.00
Steady	0	0	19	19
	0.00	0.00	100.00	100.00
Total	33	46	21	100
	33.00	46.00	21.00	100.00
Priors	0.33	0.48	0.19	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.0303	0.0625	0.0000	0.0400
Priors	0.3300	0.4800	0.1900	

```
88         *** ch4s4d1.sas ***;
89         Title2 "PROC Discrim - proportional priors - no pooling";
90         proc discrim data = amul.gamblegrp pool=no testdata = amul.gamblegrp2 testlist;
91           class type;
92           priors prop;
93           var dsm1-dsm12;
94         run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP2.
NOTE: The PROCEDURE DISCRIM printed pages 23-27.
NOTE: PROCEDURE DISCRIM used (Total process time):
      real time           0.21 seconds
      cpu time            0.17 seconds
95

PROC Discrim - proportional priors - no pooling

Class Level Information
           Variable                                                  Prior
type       Name         Frequency       Weight    Proportion    Probability
Binge      Binge               33      33.0000      0.330000       0.330000
Control    Control             48      48.0000      0.480000       0.480000
Steady     Steady              19      19.0000      0.190000       0.190000
```

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	32	1	0	33
	96.97	3.03	0.00	100.00
Control	1	45	2	48
	2.08	93.75	4.17	100.00
Steady	0	0	19	19
	0.00	0.00	100.00	100.00
Total	33	46	21	100
	33.00	46.00	21.00	100.00
Priors	0.33	0.48	0.19	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.0303	0.0625	0.0000	0.0400
Priors	0.3300	0.4800	0.1900	

PROC Discrim - proportional priors - no pooling

The DISCRIM Procedure
Classification Summary for Test Data: AMUL.GAMBLEGRP2
Classification Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D^2_j(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j| - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each type

$$Pr(j|X) = \exp(-.5 \, D^2_j(X)) \, / \, SUM_k \exp(-.5 \, D^2_k(X))$$

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	26	9	0	35
	74.29	25.71	0.00	100.00
Control	2	39	1	42
	4.76	92.86	2.38	100.00
Steady	1	8	14	23
	4.35	34.78	60.87	100.00
Total	29	56	15	100
	29.00	56.00	15.00	100.00
Priors	0.33	0.48	0.19	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.2571	0.0714	0.3913	0.1935
Priors	0.3300	0.4800	0.1900	

```
96         *** ch4s5d1.sas ***;
97         Title2 "PROC StepDisc";
98         proc stepdisc data = amul.gamblegrp method=stepwise;
99           class type;
100          var dsm1-dsm12;
101        run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: The PROCEDURE STEPDISC printed pages 28-34.
NOTE: PROCEDURE STEPDISC used (Total process time):
      real time           0.09 seconds
      cpu time            0.07 seconds
102
```

PROC StepDisc

The STEPDISC Procedure

The Method for Selecting Variables is STEPWISE

Observations	100	Variable(s) in the Analysis	12
Class Levels	3	Variable(s) will be Included	0
		Significance Level to Enter	0.15
		Significance Level to Stay	0.15

Class Level Information

type	Variable Name	Frequency	Weight	Proportion
Binge	Binge	33	33.0000	0.330000
Control	Control	48	48.0000	0.480000
Steady	Steady	19	19.0000	0.190000

Stepwise Selection: Step 1

Statistics for Entry, DF = 2, 97

Variable	R-Square	F Value	Pr > F	Tolerance
dsm1	0.3306	23.95	<.0001	1.0000
dsm2	0.2576	16.82	<.0001	1.0000
dsm3	0.2071	12.67	<.0001	1.0000
dsm4	0.5883	69.30	<.0001	1.0000
dsm5	0.0082	0.40	0.6716	1.0000
dsm6	0.0304	1.52	0.2236	1.0000
dsm7	0.0057	0.28	0.7587	1.0000
dsm8	0.3894	30.92	<.0001	1.0000
dsm9	0.2888	19.70	<.0001	1.0000
dsm10	0.3284	23.71	<.0001	1.0000
dsm11	0.0203	1.01	0.3695	1.0000
dsm12	0.3625	27.58	<.0001	1.0000

Variable dsm4 will be entered.

Variable(s) that have been Entered
dsm4

MultivariateStatistics

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.411699	69.30	2	97	<.0001
Pillai's Trace	0.588301	69.30	2	97	<.0001
Average Squared Canonical Correlation	0.294151				

Stepwise Selection: Step 2

Statistics for Removal, DF = 2, 97

Variable	R-Square	F Value	Pr > F
dsm4	0.5883	69.30	<.0001

No variables can be removed.

Statistics for Entry, DF = 2, 96

Variable	Partial R-Square	F Value	Pr > F	Tolerance
dsm1	0.3443	25.21	<.0001	0.9936
dsm2	0.2581	16.70	<.0001	0.9951
dsm3	0.2041	12.31	<.0001	0.9958
dsm5	0.0066	0.32	0.7280	0.9975
dsm6	0.0031	0.15	0.8602	0.9580
dsm7	0.0036	0.17	0.8401	0.9958
dsm8	0.0638	3.27	0.0423	0.3916
dsm9	0.2867	19.29	<.0001	0.9970
dsm10	0.0053	0.25	0.7765	0.4888
dsm11	0.0103	0.50	0.6088	0.9879
dsm12	0.0030	0.14	0.8669	0.3916

Variable dsm1 will be entered.

Variable(s) that have been Entered
dsm1 dsm4

MultivariateStatistics

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.269939	44.39	4	192	<.0001
Pillai's Trace	0.925766	41.80	4	194	<.0001
Average Squared Canonical Correlation	0.462883				

Stepwise Selection: Step 3

Statistics for Removal, DF = 2, 96

Variable	Partial R-Square	F Value	Pr > F
dsm1	0.3443	25.21	<.0001
dsm4	0.5967	71.03	<.0001

No variables can be removed.

Statistics for Entry, DF = 2, 95

Variable	Partial R-Square	F Value	Pr > F	Tolerance
dsm2	0.0420	2.08	0.1301	0.3575
dsm3	0.0269	1.31	0.2738	0.4947
dsm5	0.0077	0.37	0.6941	0.9865
dsm6	0.0015	0.07	0.9322	0.9507
dsm7	0.0004	0.02	0.9823	0.9814
dsm8	0.0720	3.68	0.0288	0.3879
dsm9	0.0210	1.02	0.3656	0.3036
dsm10	0.0022	0.10	0.9024	0.4785
dsm11	0.0010	0.05	0.9557	0.9699
dsm12	0.0040	0.19	0.8265	0.3907

Variable dsm8 will be entered.

Variable(s) that have been Entered
dsm1 dsm4 dsm8

MultivariateStatistics

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.250508	31.60	6	190	<.0001
Pillai's Trace	0.973315	30.34	6	192	<.0001
Average Squared Canonical Correlation	0.486657				

PROC StepDisc

The STEPDISC Procedure
Stepwise Selection: Step 4

Statistics for Removal, DF = 2, 95

Variable	Partial R-Square	F Value	Pr > F
dsm1	0.3501	25.58	<.0001
dsm4	0.3650	27.31	<.0001
dsm8	0.0720	3.68	0.0288

No variables can be removed.

Statistics for Entry, DF = 2, 94

Variable	Partial R-Square	F Value	Pr > F	Tolerance
dsm2	0.0405	1.98	0.1436	0.3570
dsm3	0.0256	1.24	0.2949	0.3870
dsm5	0.0126	0.60	0.5501	0.3841
dsm6	0.0006	0.03	0.9707	0.3848
dsm7	0.0012	0.06	0.9443	0.3876
dsm9	0.0200	0.96	0.3871	0.3036
dsm10	0.0215	1.03	0.3598	0.3343
dsm11	0.0007	0.03	0.9658	0.3811
dsm12	0.0256	1.23	0.2960	0.2951

Variable dsm2 will be entered.

Variable(s) that have been Entered
dsm1 dsm2 dsm4 dsm8

Multivariate Statistics

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.240375	24.43	8	188	<.0001
Pillai's Trace	0.992137	23.38	8	190	<.0001
Average Squared Canonical Correlation	0.496068				

Stepwise Selection: Step 5

Statistics for Removal, DF = 2, 94

Variable	Partial R-Square	F Value	Pr > F
dsm1	0.1622	9.10	0.0002
dsm2	0.0405	1.98	0.1436
dsm4	0.3839	29.29	<.0001
dsm8	0.0705	3.56	0.0323

No variables can be removed.

Statistics for Entry, DF = 2, 93

Variable	Partial R-Square	F Value	Pr > F	Tolerance
dsm3	0.0067	0.31	0.7309	0.2885
dsm5	0.0087	0.41	0.6661	0.3497
dsm6	0.0008	0.04	0.9613	0.3549
dsm7	0.0013	0.06	0.9396	0.3549
dsm9	0.0111	0.52	0.5938	0.2335
dsm10	0.0195	0.93	0.3998	0.3322
dsm11	0.0006	0.03	0.9736	0.3345
dsm12	0.0255	1.21	0.3015	0.2927

No variables can be entered.
No further steps are possible.

Stepwise Selection Summary

Step	Number In	Entered	Removed	Partial R-Square	F Value	Pr > F	Wilks' Lambda	Pr < Lambda	Average Squared Canonical Correlation	Pr > ASCC
1	1	dsm4		0.5883	69.30	<.0001	0.41169856	<.0001	0.29415072	<.0001
2	2	dsm1		0.3443	25.21	<.0001	0.26993945	<.0001	0.46288291	<.0001
3	3	dsm8		0.0720	3.68	0.0288	0.25050786	<.0001	0.48665727	<.0001
4	4	dsm2		0.0405	1.98	0.1436	0.24037460	<.0001	0.49606842	<.0001

```
103         *** ch4s5d2.sas ***;
104         Title2 "PROC Discrim - proportional priors - validation of stepwise discriminant
104       ! functions";
105         proc discrim data = amul.gamblegrp pool = no testdata = amul.gamblegrp2;
106           class type;
107           priors prop;
108           var dsm1 dsm2 dsm4 dsm8;
109         run;
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP.
NOTE: There were 100 observations read from the data set AMUL.GAMBLEGRP2.
NOTE: The PROCEDURE DISCRIM printed pages 35-38.
NOTE: PROCEDURE DISCRIM used (Total process time):
      real time           0.06 seconds
      cpu time            0.03 seconds
```

PROC Discrim - proportional priors - validation of stepwise discriminant functions

The DISCRIM Procedure

Observations	100	DF Total	99
Variables	4	DF Within Classes	97
Classes	3	DF Between Classes	2

Class Level Information

type	Variable Name	Frequency	Weight	Proportion	Prior Probability
Binge	Binge	33	33.0000	0.330000	0.330000
Control	Control	48	48.0000	0.480000	0.480000
Steady	Steady	19	19.0000	0.190000	0.190000

Within Covariance Matrix Information

type	Covariance Matrix Rank	Natural Log of the Determinant of the Covariance Matrix
Binge	4	-1.79957
Control	4	0.20860
Steady	4	-3.52194

Pairwise Generalized Squared Distances Between Groups

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

Generalized Squared Distance to type

From type	Binge	Control	Steady
Binge	0.41775	6.15085	28.22743
Control	24.98332	1.67654	26.17993
Steady	34.15288	6.46586	-0.20048

Classification Summary for Calibration Data: AMUL.GAMBLEGRP
Resubstitution Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j| - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each type

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

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	32	1	0	33
	96.97	3.03	0.00	100.00
Control	7	39	2	48
	14.58	81.25	4.17	100.00
Steady	0	1	18	19
	0.00	5.26	94.74	100.00
Total	39	41	20	100
	39.00	41.00	20.00	100.00
Priors	0.33	0.48	0.19	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.0303	0.1875	0.0526	0.1100
Priors	0.3300	0.4800	0.1900	

Classification Summary for Test Data: AMUL.GAMBLEGRP2
Classification Summary using Quadratic Discriminant Function

Generalized Squared Distance Function

$$D_j^2(X) = (X-\bar{X}_j)' \, COV_j^{-1} \, (X-\bar{X}_j) + \ln |COV_j| - 2 \ln PRIOR_j$$

Posterior Probability of Membership in Each type

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

Number of Observations and Percent Classified into type

From type	Binge	Control	Steady	Total
Binge	31	4	0	35
	88.57	11.43	0.00	100.00
Control	7	33	2	42
	16.67	78.57	4.76	100.00
Steady	3	5	15	23
	13.04	21.74	65.22	100.00
Total	41	42	17	100
	41.00	42.00	17.00	100.00
Priors	0.33	0.48	0.19	

Error Count Estimates for type

	Binge	Control	Steady	Total
Rate	0.1143	0.2143	0.3478	0.2067
Priors	0.3300	0.4800	0.1900	