9.7 Complete Example of Discriminant Analysis

The example of direct discriminant analysis in this section explores how role-dissatisfied housewives, role-satisfied housewives, and employed women differ in attitudes. The sample of 465 women is described in Appendix B.1. The grouping variable is role-dissatisfied housewives (UNHOUSE), role-satisfied housewives (HAPHOUSE), and working women (WORKING). Data are in DISCRIM.*.

Predictors are internal versus external locus of control (CONTROL), satisfaction with current marital status (ATTMAR), attitude toward women's role (ATTROLE), and attitude toward housework (ATTHOUSE). Scores are scaled so that low values represent more positive or "desirable" attitudes. A fifth attitudinal variable, attitude toward paid work, was dropped from analysis because data were available only for women who had been employed within the past 5 years and use of this predictor would have involved nonrandom missing values (cf. Chapter 4). The example of DISCRIM, then, involves prediction of group membership from the four attitudinal variables.

The direct discriminant analysis allows us to evaluate the distinctions among the three groups on the basis of attitudes. We explore the dimensions on which the groups differ, the predictors contributing to differences among groups on these dimensions, and the degree to which we can accurately classify members into their own groups. We also evaluate efficiency of classification with a cross-validation sample.

9.7.1 Evaluation of Assumptions

The data are first evaluated with respect to practical limitations of DISCRIM.

9.7.1.1 Unequal Sample Sizes and Missing Data

In a screening run through SAS MEANS (cf. Section 4.2.2.1), seven cases had missing values among the four attitudinal predictors. Missing data were scattered over predictors and groups in apparently random fashion, so that deletion of the cases was deemed appropriate.[11] The full data set includes 458 cases, once cases with missing values are deleted.

[11] Alternative strategies for dealing with missing data are discussed in Chapter 4.

During classification, unequal sample sizes are used to modify the probabilities with which cases are classified into groups. Because the sample is randomly drawn from predictors contributing to differences among groups on these dimensions, and the degree to which we can accurately classify members into their own groups. We also evaluate efficiency of classification with a cross-validation sample.

9.7.1 Evaluation of Assumptions

The data are first evaluated with respect to practical limitations of DISCRIM.

9.7.1.1 Unequal Sample Sizes and Missing Data

In a screening run through SAS MEANS (cf. Section 4.2.2.1), seven cases had missing values among the four attitudinal predictors. Missing data were scattered over predictors and groups in apparently random fashion, so that deletion of the cases was deemed appropriate.[11] The full data set includes 458 cases, once cases with missing values are deleted.

[11] Alternative strategies for dealing with missing data are discussed in Chapter 4.

During classification, unequal sample sizes are used to modify the probabilities with which cases are classified into groups. Because the sample is randomly drawn from the population of interest, sample sizes in groups are believed to represent some real process in the population that should be reflected in classification. For example, knowledge that over half the women are employed implies that greater weight should be given to the WORKING group.

9.7.1.2 Multivariate Normality

After deletion of cases with missing data, there are still over 80 cases per group. Although a SAS MEANS run (not shown) revealed skewness in ATTMAR, sample sizes are large enough to suggest normality of sampling distributions of means. Therefore, there is no reason to expect distortion of results due to failure of multivariate normality.

9.7.1.3 Linearity

Although ATTMAR is skewed, there is no expectation of curvilinearity between this and the remaining predictors. At worst, ATTMAR in conjunction with the remaining continuous, well-behaved predictors may contribute to a mild reduction in association.

9.7.1.4 Outliers

To identify univariate outliers, *z*-scores associated with minimum and maximum values on each of the four predictors are investigated through SAS MEANS for each group separately, as per Section 4.2.2 (not shown). There were some questionable values on ATTHOUSE, with a few exceptionally positive (low) scores. These values were about 4.5 standard deviations below their group means, making them candidates for deletion or alteration. However, the cases are retained for the search for multivariate outliers.

Multivariate outliers are sought through SAS REG by subsets (groups) and a request for an output table containing leverage statistics, as seen in Table 9.5. Data first are sorted by WORKSTAT, which then becomes the `by` variable in the `proc reg` run. Leverage values (Leverage) are saved in a file labeled DISC_OUT. Table 9.5 shows a portion of the output data file for the working women (WORKSTAT=1).

Table 9.5 Identification of Multivariate Outliers (SAS SORT and REG Syntax and Selected Portion of Output File from SAS REG)

```
proc sort data = Sasuser.Discrim;
   by WORKSTAT;
run;
proc reg data = Sasuser.Discrim;
   by WORKSTAT;
    model CASESEQ= CONTROL ATTMAR ATTROLE ATTHOUSE/ selection=
       RSQUARE COLLIN;
    output out=SASUSER.DISC_OUT H=H;
run;
```

VIEWTABLE: Sasuser.Disc_out

	Case sequence	Current employment status	Locus-of-control	Attitude toward current marital status	Attitudes toward role of women	Attitudes toward housework	Leverage
136	345	1	8	19	38	19	0.0144003602
137	346	1	5	20	41	2	0.0941107394
138	347	1	6	28	34	26	0.0079273826
139	348	1	7	23	26	24	0.0103006485
140	349	1	5	25	31	30	0.0210723785
141	355	1	6	25	27	30	0.0147929944
142	357	1	6	17	38	22	0.0087458208
143	358	1	8	42	36	26	0.0276384488
144	359	1	7	21	22	30	0.0203930567
145	362	1	8	24	35	24	0.0088289495
146	365	1	5	14	36	18	0.0185871895
147	369	1	7	23	27	34	0.0249177763
148	372	1	7	18	45	13	0.0284221267
149	378	1	5	26	23	21	0.0284008123
150	380	1	7	.	29	28	.
151	381	1	6	35	30	20	0.0224070311
152	383	1	7	30	44	25	0.0176472831
153	384	1	7	25	41	27	0.0136302175
154	386	1	7	20	30	25	0.0066955586
155	387	1	7	23	29	22	0.0081856807
156	397	1	5	16	35	15	0.0244622406
157	398	1	6	30	23	33	0.0266531324
158	399	1	9	12	25	24	0.0360983343
159	400	1	7	25	23	29	0.016045569
160	401	1	5	42	35	21	0.0393409292
161	403	1	6	35	27	29	0.0190822473
162	404	1	7	20	30	21	0.0087346339
163	406	1	7	39	35	18	0.0305524425
164	407	1	6	20	42	2	0.0898029757
165	425	1	6	30	33	22	0.0098487137

Outliers are identified as cases with too large a Mahalanobis' D^2 for their own group, evaluated as χ^2 with degrees of freedom equal to the number of predictors. Critical χ^2 with 4 df at $\alpha = .001$ is 18.467; any case with $D^2 > 18.467$ is an outlier. Translating this critical value to leverage h_{ii} for the first group using the variation on Equation 4.3:

$$h_{ii} = \frac{\text{Mahalanobis distance}}{N-1} + \frac{1}{N} = \frac{18.467}{240} + \frac{1}{241} = .081$$

In Table 9.5, CASESEQ 346 ($\mathrm{H} = .0941$) and CASESEQ 407 ($\mathrm{H} = .0898$) are identified as outliers in the group of WORKING women. No additional outliers were found.

The multivariate outliers are the same cases that have extreme univariate scores on ATTHOUSE. Because transformation is questionable for ATTHOUSE (where it seems unreasonable to transform the predictor for only two cases), it is decided to delete the outliers.

Therefore, of the original 465 cases, 7 are lost due to missing values and 2 are both univariate and multivariate outliers, leaving a total of 456 cases for analysis.

9.7.1.5 Homogeneity of Variance–Covariance Matrices

A SAS DISCRIM run, Table 9.6, deletes the outliers in order to evaluate homogeneity of variance–covariance

The multivariate outliers are the same cases that have extreme univariate scores on ATTHOUSE. Because transformation is questionable for ATTHOUSE (where it seems unreasonable to transform the predictor for only two cases), it is decided to delete the outliers.

Therefore, of the original 465 cases, 7 are lost due to missing values and 2 are both univariate and multivariate outliers, leaving a total of 456 cases for analysis.

9.7.1.5 Homogeneity of Variance–Covariance Matrices

A SAS DISCRIM run, **Table 9.6**, deletes the outliers in order to evaluate homogeneity of variance–covariance matrices. Most output has been omitted here. The instruction to produce the test of homogeneity of variance–covariance matrices is `pool = test`.

Table 9.6 Syntax and Selected Output from SAS DISCRIM to Check Homogeneity of Variance–Covariance Matrices

```
proc discrim data=Sasuser.Discrim short noclassify
          pool=test slpool=.001;
          class workstat;
     var CONTROL ATTMAR ATTROLE ATTHOUSE;
     priors proportional;
     where CASESEQ^=346 and CASESEQ^=407;
run;
```

The DISCRIM Procedure
Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
50.753826	20	0.0002

Since the Chi-Square value is significant at the 0.001 level, the within covariance matrices will be used in the discriminant function.
Reference: Morrison, D.F. (1976) Multivariate Statistical Methods p252.

This test shows significant heterogeneity of variance–covariance matrices. The program uses separate matrices in the classification phase of discriminant analysis if `pool = test` is specified and the test shows significant heterogeneity.

9.7.1.6 Multicollinearity and Singularity

Because SAS DISCRIM, used for the major analysis, protects against multicollinearity through checks of tolerance, no formal evaluation is necessary (cf. **Chapter 4**). However, the SAS REG syntax of **Table 9.5** that evaluates multivariate outliers also requests collinearity information, shown in **Table 9.7**. No problems with multicollinearity are noted.

Table 9.7 SAS REG Output Showing Collinearity Information for All Groups Combined (Syntax Is in Table 9.5)

Collinearity Diagnostics								
			Proportion of Variation					
Number	Eigenvalue	Condition Index	Intercept	CONTROL	ATTMAR	ATTROLE	ATTHOUSE
1	4.83193	1.00000	0.00036897	0.00116	0.00508	0.00102	0.00091481
2	0.10975	6.63518	0.00379	0.00761	0.94924	0.02108	0.00531
3	0.03169	12.34795	0.00188	0.25092	0.04175	0.42438	0.10031
4	0.02018	15.47559	0.00266	0.61843	0.00227	0.01676	0.57008
5	0.00645	27.37452	0.99129	0.12189	0.00166	0.53676	0.32339

9.7.2 Direct Discriminant Analysis

Direct DISCRIM is performed through SAS DISCRIM with the four attitudinal predictors all forced into the equation. The program instructions and some of the output appear in Table 9.8. Simple statistics are requested to provide predictor means, helpful in interpretation. The `anova` and `manova` instructions request univariate statistics on group differences separately for each of the predictors and a multivariate test for the difference among groups; `pcorr` requests the pooled within-groups correlation matrix, and `crossvalidate` requests jackknifed classification. The `priors proportional` instruction specifies prior probabilities for classification proportional to sample sizes. The `stdmean` instruction requests canonical centroids.

Table 9.8 Syntax and Partial Output from SAS DISCRIM Analysis of Four Attitudinal Variables

```
proc discrim data=SASUSER.DISCRIM simple anova manova pcorr can
     crossvalidate stdmean pool=test;
   class workstat;
   var CONTROL ATTMAR ATTROLE ATTHOUSE;
   priors proportional;
   where CASESEQ^=346 and CASESEQ^=407;
run;
```

The DISCRIM Procedure

Pooled Within-Class Correlation Coefficients / Pr > |r|				
Variable	CONTROL	ATTMAR	ATTROLE	ATTHOUSE
CONTROL Locus-of-control	1.0000	0.1717 0.0002	0.0091 0.8463	0.1550 0.0009
ATTMAR Attitude toward current marital status	0.1717 0.0002	1.0000	-0.0701 0.1359	0.2823 <.0001
ATTROLE Attitudes toward role of women	0.0091 0.8463	-0.0701 0.1359	1.0000	-0.2914 <.0001
ATTHOUSE Attitudes toward housework	0.1550 0.0009	0.2823 <.0001	-0.2914 <.0001	1.0000

The DISCRIM Procedure
Simple Statistics

Total-Sample						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
CONTROL	Locus-of-control	456	3078	6.75000	1.60769	1.2679
ATTMAR	Attitude toward current marital status	456	10469	22.95833	72.73892	8.5287
ATTROLE	Attitudes toward role of women	456	16040	35.17544	45.68344	6.7590
ATTHOUSE	Attitudes toward housework	456	10771	23.62061	18.30630	4.2786

WORKSTAT = 1						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
CONTROL	Locus-of-control	239	1605	6.71548	1.53215	1.2378
ATTMAR	Attitude toward current marital status	239	5592	23.39749	72.76151	8.5300
ATTROLE	Attitudes toward role of women	239	8093	33.86192	48.38842	6.9562
ATTHOUSE	Attitudes toward housework	239	5691	23.81172	19.85095	4.4554

WORKSTAT = 2						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
CONTROL	Locus-of-control	136	902.00000	6.63235	1.71569	1.3098
ATTMAR	Attitude toward current marital status	136	2802	20.60294	43.87081	6.6235
ATTROLE	Attitudes toward role of women	136	5058	37.19118	41.71133	6.4584

ATTROLE	Attitudes toward role of women	136	5058	37.19118	41.71133	6.4584
ATTHOUSE	Attitudes toward housework	136	3061	22.50735	15.08143	3.8835

WORKSTAT = 3						
Variable	Label	N	Sum	Mean	Variance	Standard Deviation
CONTROL	Locus-of-control	81	571.00000	7.04938	1.57253	1.2540
ATTMAR	Attitude toward current marital status	81	2075	25.61728	106.03920	10.2975
ATTROLE	Attitudes toward role of women	81	2889	35.66667	33.17500	5.7598
ATTHOUSE	Attitudes toward housework	81	2019	24.92593	15.66944	3.9585

Univariate Test Statistics

F Statistics, Num DF=2, Den DF=453

Variable	Label	Total Standard Deviation	Pooled Standard Deviation	Between Standard Deviation	R-Square	R-Square / (1-RSq)	F Value	Pr > F
CONTROL	Locus-of-control	1.2679	1.2625	0.1761	0.0129	0.0131	2.96	0.0530
ATTMAR	Attitude toward current marital status	8.5287	8.3683	2.1254	0.0415	0.0433	9.81	<.0001
ATTROLE	Attitudes toward role of women	6.7590	6.6115	1.7996	0.0474	0.0497	11.26	<.0001
ATTHOUSE	Attitudes toward housework	4.2786	4.2061	1.0184	0.0379	0.0393	8.91	0.0002

Average R-Square	
Unweighted	0.0348993
Weighted by Variance	0.0426177

Multivariate Statistics and F Approximations

S=2 M=0.5 N=224

Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.89715033	6.27	8	900	<.0001
Pillai's Trace	0.10527259	6.26	8	902	<.0001
Hotelling-Lawley Trace	0.11193972	6.29	8	640.54	<.0001
Roy's Greatest Root	0.07675307	8.65	4	451	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

The DISCRIM Procedure
Canonical Discriminant Analysis

	Canonical Correlation	Adjusted Canonical Correlation	Approximate Standard Error	Squared Canonical Correlation	Eigenvalues of Inv(E)*H = CanRsq/(1-CanRsq) Eigenvalue	Difference	Proportion	Cumulative
1	0.266987	0.245497	0.043539	0.071282	0.0768	0.0416	0.6857	0.6857
2	0.184365	0.182794	0.045287	0.033991	0.0352		0.3143	1.0000

The DISCRIM Procedure
Canonical Discriminant Analysis

Test of H0: The canonical correlations in the current row and all that follow are zero

Test of H0: The canonical correlations in the current row and all that follow are zero

Likelihood Ratio	Approximate F Value	Num DF	Den DF	Pr > F
0.89715033	6.27	8	900	<.0001
0.96600937	5.29	3	451	0.0014

Pooled Within Canonical Structure			
Variable	Label	Can1	Can2
CONTROL	Locus-of-control	0.281678	0.444939
ATTMAR	Attitude toward current marital status	0.718461	0.322992
ATTROLE	Attitudes toward role of women	-0.639249	0.722228
ATTHOUSE	Attitudes toward housework	0.679447	0.333315

Class Means on Canonical Variables		
WORKSTAT	Can1	Can2
1	0.1407162321	-.1505321835
2	-.4160079128	0.0539321812
3	0.2832826750	0.3536100644

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

Number of Observations and Percent Classified into WORKSTAT				
From WORKSTAT	1	2	3	Total
1	184 76.99	48 20.08	7 2.93	239 100.00
2	73 53.68	59 43.38	4 2.94	136 100.00
3	59 72.84	12 14.81	10 12.35	81 100.00
Total	316 69.30	119 26.10	21 4.61	456 100.00
Priors	0.52412	0.29825	0.17763	

Error Count Estimates for WORKSTAT				
	1	2	3	Total
Rate	0.2301	0.5662	0.8765	0.4452
Priors	0.5241	0.2982	0.1776	

The DISCRIM Procedure
Classification Summary for Calibration Data: SASUSER.DISCR
Cross-validation Summary using Quadratic Discriminant Funct

Number of Observations and Percent Classified

	into WORKSTAT			
From WORKSTAT	1	2	3	Total
1	179 74.90	50 20.92	10 4.18	239 100.00
2	78 57.35	53 38.97	5 3.68	136 100.00
3	60 74.07	13 16.05	8 9.88	81 100.00
Total	317 69.52	116 25.44	23 5.04	456 100.00
Priors	0.52412	0.29825	0.17763	

Error Count Estimates for WORKSTAT				
	1	2	3	Total
Rate	0.2510	0.6103	0.9012	0.4737
Priors	0.5241	0.2982	0.1776	

When all four predictors are used, the *F* of 6.27 (with 8 and 900 df based on Wilks' lambda) is highly significant. That is, there is statistically significant separation of the three groups basedon all four predictors combined, as discussed in **Section 9.6.1.1**. Partial η^2 and associated 95%confidence limits are found through **Smithson's (2003)** NoncF2.sas procedure (as in **Table 8.16**), yielding $\eta^2 = .05$ with limits from .02 to .08.

Canonical correlations (in the section of output following multivariate analysis) for each discriminant function (.267 and .184), although small, are relatively equal for the two discriminant functions. The adjusted values are not very much different with this relatively large sample. The "peel down" test shows that both functions significantly discriminate among the groups. That is, even after the first function is removed, there remains significant discrimination, Pr > F = 0.0014. Because there are only two possible discriminant functions, this is a test of the second one. Applying Smithson's procedure, $\eta^2 = .03$ with confidence limits from .01 to .07.

The loading matrix (correlations between predictors and discriminant functions) appears in the section of output labeled Pooled Within Canonical Structure. Class means on canonical variables are centroids on the discriminant functions for the groups, discussed in **Sections 9.4.1** and **9.6.3.1**.

A plot of the placement of the centroids for the three groups on the two discriminant functions (canonical variables) as axes appears in **Figure 9.3**. The points that are plotted are given in **Table 9.8** as Class means on Canonical Variables.

Figure 9.3 Plots of three group centroids on two discriminant functions derived from four attitudinal variables.

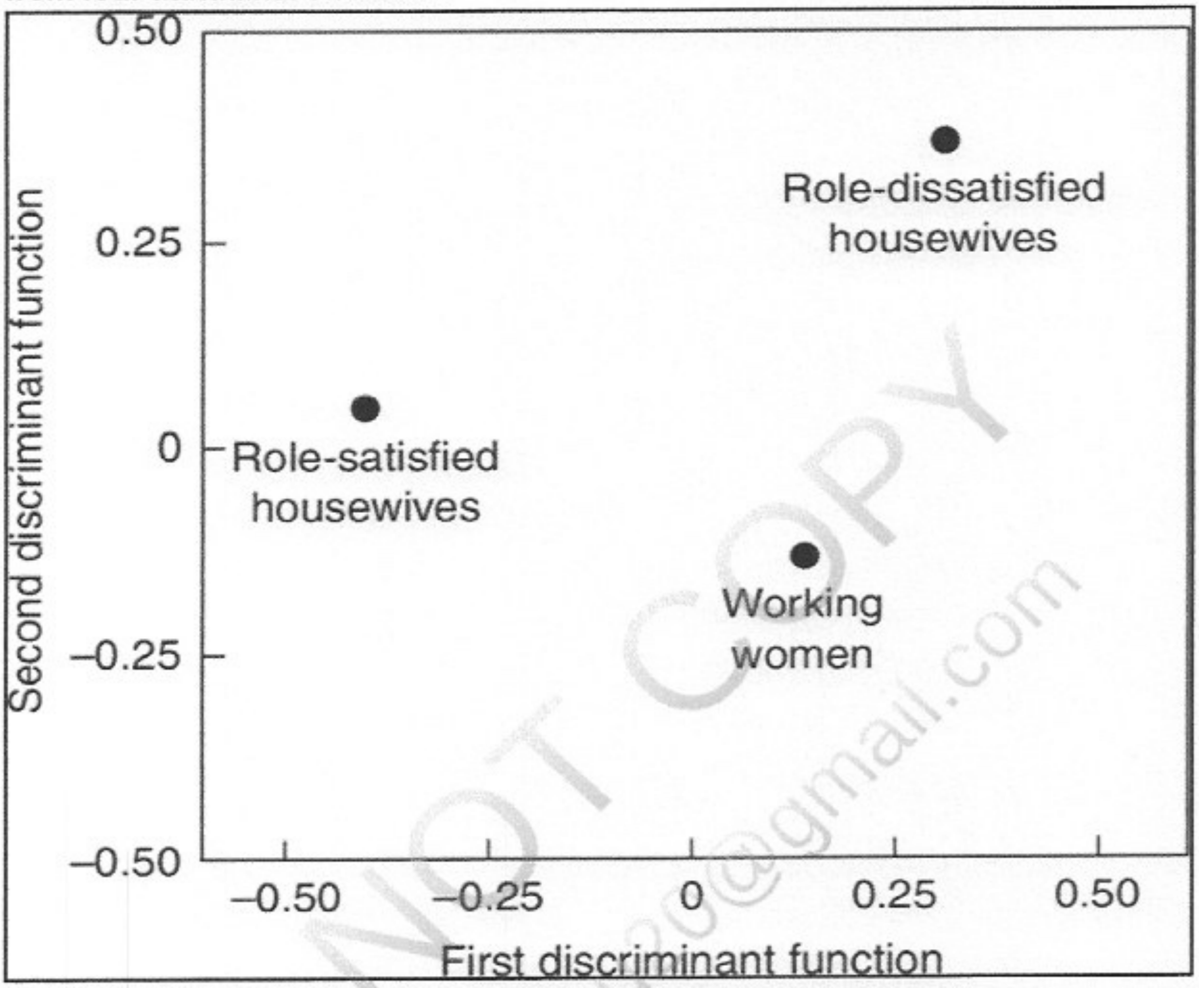

Table 9.8 shows the classification functions used to classify cases into the three groups (see Equation 9.3) and the results of that classification, with and without jackknifing (see Section 9.6.7). In this case, classification is made on the basis of a modified equation in which unequal prior probabilities are used to reflect unequal group sizes by the use of `prior proportional` in the syntax. Classification is based on the quadratic discriminant function to compensate for heterogeneity of various covariance matrices.

A total of 55% (1–Error Count Estimates for WORKSTAT of 0.4452) of cases are correctly classified by normal procedures, and 52% by jackknifed procedures. How do these compare with random assignment? Prior probabilities, specified as .52 (WORKING), .30 (HAPHOUSE), and .18 (UNHOUSE), put 237 cases (.52 × 456) in the WORKING group, 137 in the HAPHOUSE group, and 82 in the UNHOUSE group. If cases are randomly assigned to the WORKING group, 123 ($.52 \times 237$) should be correct, while 41.1 ($.30 \times 137$) and 14.8 ($.18 \times 82$) should be correct by chance in the HAPHOUSE and UNHOUSE groups, respectively. Over all three groups, 178.9 out the HAPHOUSE and UNHOUSE groups, respectively. Over all three groups, 178.9 out of the 456 cases, or 39%, should be correct by chance alone. Both classification procedures correctly classify substantially more than that.

An additional SAS DISCRIM run for cross-validation is shown in Table 9.9. SAS DISCRIM has no direct procedure of forming and using a cross-validation sample. Instead, other procedures must be used to split the file into the "training" cases, used to develop (calibrate) the classification equations, and the "testing" cases, used to validate the classification.

Table 9.9 Cross-Validation of Classification of Cases by Four Attitudinal Variables (Syntax for SAS DATA: Syntax and Selected Output from SAS DISCRIM)

```
data Sasuser.Discrimx;
     set SASUSER.DISCRIM;
     if ATTHOUSE=2 or ATTHOUSE=. or ATTMAR=. or ATTROLE=.
               or CONTROL=. then delete;
     TEST1=0;
     if uniform(11738) <= .25 then TEST1=1;
run;
data Sasuser.Disctrng;
     set Sasuser.Discrimx;
     where TEST1=0;
data Sasuser.Disctest;
     set Sasuser.Discrimx;
     where TEST1=1;
run;
proc discrim data=SASUSER.Disctrng outstat=INFO pool=test;
     class WORKSTAT;
     var CONTROL ATTMAR ATTROLE ATTHOUSE;
     priors proportional;
run;
proc discrim data=INFO testdata=SASUSER.Disctest pool=test;
     class WORKSTAT;
     var CONTROL ATTMAR ATTROLE ATTHOUSE;
     priors proportional;
run;
```

The DISCRIM Procedure
Classification Summary for Calibration Data: SASUSER.DISCTRNG
Resubstitution Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into WORKSTAT				
From WORKSTAT	1	2	3	Total
1	129 75.00	32 18.60	11 6.40	172 100.00
2	49 48.04	46 45.10	7 6.86	102 100.00
3	45 69.23	9 13.85	11 16.92	65 100.00
Total	223 65.78	87 25.66	29 8.55	339 100.00
Priors	0.50737	0.30088	0.19174	

Error Count Estimates for WORKSTAT				
	1	2	3	Total
Rate	0.2500	0.5490	0.8308	0.4513
Priors	0.5074	0.3009	0.1917	

The DISCRIM Procedure
Classification Summary for Test Data: SASUSER.DISCTEST
Classification Summary using Quadratic Discriminant Function

Observation Profile for Test Data	
Number of Observations Read	117
Number of Observations Used	117

Number of Observations and Percent Classified into WORKSTAT				
From WORKSTAT	1	2	3	Total
1	40 59.70	16 23.88	11 16.42	67 100.00
2	17 50.00	15 44.12	2 5.88	34 100.00
3	10 62.50	2 12.50	4 25.00	16 100.00
Total	67 57.26	33 28.21	17 14.53	117 100.00
Priors	0.50737	0.30088	0.19174	

Error Count Estimates for WORKSTAT				
	1	2	3	Total
Rate	0.4030	0.5588	0.7500	0.5164
Priors	0.5074	0.3009	0.1917	

First a new data set is created: data SASUSER.DISCRIMX. The original data set is identified as set SASUSER.DISCRIM. Then outliers and cases with missing data are omitted. Finally, a variable is created on which to split the data set, here called TEST1, which is set to zero, and then changed to 1 for 25% of the cases. Then an additional two files are created on the basis of `TEST1` with set `SASUSER.DISCRIMX:` a calibration (training) file, through data `SASUSER.DISCTRNG`, and a cross-validation (test) file through data `SASUSER.DISCTEST`. Finally, a discriminant analysis on the training file (with 339 cases) is run which saves the calibration information in a file called `INFO`, and then applies the calibration information to the test file (with 117 cases). Again, the quadratic classification procedure is used.

A summary of information appropriate for publication appears in Table 9.10. In the table are the loadings, univariate *F* for each predictor, and pooled within-group correlations among predictors.

Table 9.10 Results of Discriminant Analysis of Attitudinal Variables Predictor Variable

Predictor Variable	Correlations of Predictor Variables with Discriminant Functions		Univariate $F(2, 453)$	Pooled Within-Group Correlations Among Predictors		
	1	*2*		*ATTMAR*	*ATTROLE*	*ATTHOUSE*
CONTROL	.28	.44	2.96	.17	.01	.16
ATTMAR	.72	.32	9.81		$-.07$	.28
ATTROLE	$-.64$	.72	11.26			$-.29$
ATTHOUSE	.68	.33	8.91			
Canonical R	.27	.18				
Eigenvalue	.08	.04				

SAS DISCRIM has no contrast procedure, nor does it provide F or t ratios for predictor variables adjusted for all other variables. However, the information is available to produce contrasts with separate analyses of covariance for each predictor variable in GLM. In each

SASUSER.DISCTRNG, and a cross-validation (test) file through data SASUSER.DISCTEST. Finally, a discriminant analysis on the training file (with 339 cases) is run which saves the calibration information in a file called INFO, and then applies the calibration information to the test file (with 117 cases). Again, the quadratic classification procedure is used.

A summary of information appropriate for publication appears in Table 9.10. In the table are the loadings, univariate *F* for each predictor, and pooled within-group correlations among predictors.

Table 9.10 Results of Discriminant Analysis of Attitudinal Variables Predictor Variable

Predictor Variable	Correlations of Predictor Variables with Discriminant Functions		Univariate $F(2, 453)$	Pooled Within-Group Correlations Among Predictors		
	1	*2*		*ATTMAR*	*ATTROLE*	*ATTHOUSE*
CONTROL	.28	.44	2.96	.17	.01	.16
ATTMAR	.72	.32	9.81		$-.07$	.28
ATTROLE	$-.64$	.72	11.26			$-.29$
ATTHOUSE	.68	.33	8.91			
Canonical *R*	.27	.18				
Eigenvalue	.08	.04				

SAS DISCRIM has no contrast procedure, nor does it provide *F* or *t* ratios for predictor variables adjusted for all other variables. However, the information is available to produce contrasts with separate analyses of covariance for each predictor variable in GLM. In each analysis of covariance, the variable of interest is declared the DV and the remaining variables are declared covariates. The process is demonstrated for the 12 contrast runs needed in Tables 9.11 through 9.13; means on each predictor adjusted for all other predictors for each group are contrasted with the pooled means for the other two groups. WORKING women are contrasted with the pooled means for HAPHOUSE and UNHOUSE to determine which predictors distinguish WORKING women from others in Table 9.11. Table 9.12 has the HAPHOUSE group contrasted with the other two groups; Table 13 shows the UNHOUSE group contrasted with the other two groups. Note that df for error $= N - k - c - 1 = 450$.

Table 9.11 Syntax and Highly Abbreviated Output of SAS GLM Contrasting the WORKING Group with the Other Two Groups

```
proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTHOUSE = WORKSTAT CONTROL ATTMAR ATTROLE ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'WORKING VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTROLE = WORKSTAT ATTHOUSE CONTROL ATTMAR ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'WORKING VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTMAR = WORKSTAT CONTROL ATTHOUSE ATTROLE ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'WORKING VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model CONTROL = WORKSTAT ATTROLE ATTHOUSE ATTMAR ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'WORKING VS. OTHERS' WORKSTAT 1 -2 1 ;
run;
```

Dependent Variable: ATTHOUSE Attitudes toward housework

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
WORKING VS. OTHERS	1	60.74947570	60.74947570	4.09	0.0436

Dependent Variable: ATTROLE Attitudes toward role of women

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
WORKING VS. OTHERS	1	218.5434340	218.5434340	5.45	0.0201

Dependent Variable: ATTMAR Attitude toward current marital status

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
WORKING VS. OTHERS	1	615.1203307	615.1203307	9.66	0.0020

Dependent Variable: CONTROL Locus-of-control

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
WORKING VS. OTHERS	1	1.18893484	1.18893484	0.78	0.3789

Table 9.12 Syntax and Highly Abbreviated Output of SAS GLM Contrasting the HAPHOUSE Group with the Other Two Groups

```
proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTHOUSE = WORKSTAT CONTROL ATTMAR ATTROLE ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'HAPHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTROLE = WORKSTAT ATTHOUSE CONTROL ATTMAR ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast ' HAPHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTMAR = WORKSTAT CONTROL ATTHOUSE ATTROLE ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast ' HAPHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model CONTROL = WORKSTAT ATTROLE ATTHOUSE ATTMAR ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast ' HAPHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;
```

Dependent Variable: ATTHOUSE Attitudes toward housework

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
HAPHOUSE VS. OTHERS	1	60.74947570	60.74947570	4.09	0.0436

Dependent Variable: ATTROLE Attitudes toward role of women

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
HAPHOUSE VS. OTHERS	1	218.5434340	218.5434340	5.45	0.0201

Dependent Variable: ATTMAR Attitude toward current marital status

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
HAPHOUSE VS. OTHERS	1	615.1203307	615.1203307	9.66	0.0020

Dependent Variable: CONTROL Locus-of-control

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
HAPHOUSE VS. OTHERS	1	1.18893484	1.18893484	0.78	0.3789

Table 9.13 Syntax and Highly Abbreviated Output of SAS GLM Contrasting the UNHOUSE Group with the Other Two Groups

```
proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTHOUSE = WORKSTAT CONTROL ATTMAR ATTROLE ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'UNHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTROLE = WORKSTAT ATTHOUSE CONTROL ATTMAR ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'UNHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model ATTMAR = WORKSTAT CONTROL ATTHOUSE ATTROLE ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'UNHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;

proc glm data=SASUSER.DISCRIM;
   class WORKSTAT;
   model CONTROL = WORKSTAT ATTROLE ATTHOUSE ATTMAR ;
      where CASESEQ^=346 and CASESEQ^=407;
   contrast 'UNHOUSE VS. OTHERS' WORKSTAT 1 -2 1 ;
run;
```

Dependent Variable: ATTHOUSE Attitudes toward housework

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
UNHOUSE VS. OTHERS	1	60.74947570	60.74947570	4.09	0.0436

Dependent Variable: ATTROLE Attitudes toward role of women

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
UNHOUSE VS. OTHERS	1	218.5434340	218.5434340	5.45	0.0201

Dependent Variable: ATTMAR Attitude toward current marital status

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
UNHOUSE VS. OTHERS	1	615.1203307	615.1203307	9.66	0.0020

Dependent Variable: CONTROL Locus of control

Contrast	DF	Contrast SS	Mean Square	F Value	Pr > F
UNHOUSE VS. OTHERS	1	1.18893484	1.18893484	0.78	0.3789

Based on familywise $\alpha = .05, \alpha_i = .0125$, the predictor that most clearly distinguishes the WORKING group from the other two is ATTROLE after adjustment for the other predictors. The HAPHOUSE group differs from the other two groups on the basis of ATTMAR after adjustment for the remaining predictors. The UNHOUSE group does not differ from the other two when each predictor is adjusted for all others. Separate runs without covariates would be needed if there is interest in which predictors separate each group from the others *without* adjustment for the other predictors. Table 9.14 summarizes the results of Smithson's procedure for finding effect sizes and 98.75% confidence limits for all 12 runs.

Table 9.14 Effect Sizes and 98.75% Confidence Limits for Contrasts of Each Group with the Two Other Groups Pooled for Each Predictor Adjusted for the Three Other Predictors

Contrast		Predictor (Adjusted for All Others)			
		Attitude toward housework	*Attitude toward role of women*	*Attitude toward marriage*	*Locus-of-control*
Working women vs. others	Effect Size	.00.	.04	.00	.00
	98.75% CL	.00–.03	.01–.09	.00–.02	.00–.02
Role-satisfied housewives vs. others	Effect Size	.01	.01	.02	.01
	98.75% CL	.00–.04	.00–.05	.00–.07	.00–.02
Role-dissatisfied housewives vs. others	Effect Size	.01	.01	.01	.01
	98.75% CL	.00–.05	.00–.03	.00–.05	.00–.03

A checklist for a direct discriminant function analysis appears in Table 9.15. It is followed by an example of a Results section, in journal format, for the analysis just described.

Table 9.15 Checklist for Direct Discriminant Analysis

1. Issues
    a. Unequal sample sizes and missing data
    b. Normalityof sampling distributions
    c. Outliers
    d. Linearity
    e. Homogeneity of variance–covariance matrices

f. Multicollinearity and singularity

2. Major analysis
   a. Significance of discriminant functions. If significant:
      1. Variance accounted for and confidence limits for each significant function
      2. Plot(s) of discriminant functions
      3. Structure matrix
   b. Effect size and confidence limits for solution
   c. Predictor variables separating each group with effect sizes and confidence limits
3. Additional analyses
   a. Group means and standard deviations for high-loading predictors
   b. Pooled within-group correlations among predictor variables
   c. Classification results
      1. Jackknifed classification
      2. Cross-validation
   d. Change in Rao's *V* (or stepdown *F*) plus univariate *F* for predictors

Results

A direct discriminant analysis evaluated four attitudinal variables as predictors of membership in three groups using SAS GLM Version 9.4. Predictors were locus of control, attitude toward marital status, attitude toward role of women, and attitude toward housework. Groups were working women, role-satisfied housewives, and role-dissatisfied housewives.

Of the original 465 cases, 7 were dropped from analysis because of missing data. Missing data appeared to be randomly scattered throughout groups and predictors. Two additional cases were identified as multivariate outliers with $p <. 001$ and were also deleted. Both of the outlying cases were in the working group; they were women with extraordinarily favorable attitudes toward housework. For the remaining 456 cases (239 working women, 136 role-satisfied housewives, and 81 role-dissatisfied housewives), evaluation of assumptions of linearity, normality, multicollinearity, or singularity were satisfactory. Statistically significant heterogeneity of variance–covariance matrices $(p <. 10)$ was observed, however, so a quadratic procedure was used by SAS PROC DISCRIM for analysis.

Two discriminant functions were calculated, with a combined $F(8,\ 900) = 6.27,\ p <.01,\ \eta^2 = .05$ with 95% confidence limits from .02 to .08. After removal of the first function, there was still statistically significant association between groups and predictors, $F(3,\ 451) = 5.29,\ p < .01,\ \eta^2 =.03$ with 95% confidence limits from .01 to .07. Canonical $R^2 = .07$ for the first discriminant function and .03 for the second discriminant function. Thus, the two functions accounted for about 7% and 3% of the total relationship between predictors and groups. The two discriminant functions account for 69% and 31%, respectively, of the between-group variability. [*F values, squared canonical correlations, and percents of variance are from **Table 9.8**; cf. **Section 9.6.2**.*] As shown in **Figure 9.3**, the first discriminant function maximally separates role-satisfied housewives from the other two groups. The second discriminant function discriminates role-dissatisfied housewives from working women, with role satisfied housewives falling between these two groups.

The structure (loading) matrix of correlations between predictors and discriminant functions, as seen in **Table 9.10**, suggests that the best predictors for distinguishing between role-satisfied housewives and the other two groups (first function) are attitudes toward current marital status, toward women's role, and toward housework. Role-satisfied housewives had more favorable attitudes toward marital status $(M = 20.60,\ SD = 6.62)$ than working women $(M = 23.40, SD = 8.53)$ or role-dissatisfied housewives $(M = 25.62, SD = 10.30)$, and more conservative attitudes toward women's role $(M = 37.19, SD = 6.46)$ than working women

$(\text{mean} = 33.86, SD = 6.96)$ or dissatisfied housewives $(M = 35.67, SD = 5.76)$. Role-satisfied housewives were more favorable toward housework $(M = 22.51, SD = 3.88)$ than either working women $(M = 23.81, SD = 4.55)$ or role-dissatisfied housewives $(M = 24.93, SD = 3.96)$. [*Group means and standard deviations are shown in Table 9.8*.] Loadings less than .50 are not interpreted.

One predictor, attitudes toward women's role, had a loading in excess of .50 on the second discriminant function, which separated