Exercise 1

(a) The p-value for the test of homogeneity of within covariance is less than .0001. Thus we can conclude that the two types of wine have significantly different covariances and quadratic discriminant analysis needs to be implemented.

The MANOVA tests show p-values less than 0.05, thus we can conclude that there are significant differences in some wine characteristics between red and white wines. This implies that discrimination between alcohol types based on these variables should be a reasonable approach and provide some separation between alcohol types.

Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
9819.681585	78	<.0001

Multivariate Statistics and Exact F Statistics					
S=1 M=5 N=3241					
Statistic	**Value**	**F Value**	**Num DF**	**Den DF**	**Pr > F**
Wilks' Lambda	0.13755855	3387.69	12	6484	<.0001
Pillai's Trace	0.86244145	3387.69	12	6484	<.0001
Hotelling-Lawley Trace	6.26963171	3387.69	12	6484	<.0001
Roy's Greatest Root	6.26963171	3387.69	12	6484	<.0001

(b) The cross-validation estimated overall error rate is 0.0146 based on proportional-prior discriminant analysis and the individual group error rate estimates are all around 1-2%. The misclassified observations are highlighted by green in the table. 17 red wines are assigned to white wine group and 78 white wines are classified as red wine. All others are correctly classified. The discrimination matches the groups with pretty good performance.

The classification results from the discriminant analysis show good performance for both types with just less than 100 misclassified cases overall.

Classification Summary for Calibration Data: WORK.WINE
Cross-validation Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into type			
From type	**R**	**W**	**Total**
R	1582 98.94	17 1.06	1599 100.00
W	78 1.59	4820 98.41	4898 100.00
Total	1660 25.55	4837 74.45	6497 100.00
Priors	0.24611	0.75389	

Error Count Estimates for type			
	R	**W**	**Total**
Rate	0.0106	0.0159	0.0146
Priors	0.2461	0.7539	

Exercise 2

(a) From stepwise selection, 12 predictors are chosen to construct the discriminant functions and only first 8 variables have partial R-square greater than .02. Thus 8 variables are kept with the additional R-square constraint. The variables kept are: **total_sulfur_dioxide** , **density**, **residual_sugar**, **volatile_acidity**, **alcohol**, **free_sulfur_dioxide**, **fixed_acidity**, and **chlorides.**

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	total_sulfur_dioxide		0.4905	6252.80	<.0001	0.50949986	<.0001	0.49050014	<.0001
2	2	density		0.3357	3281.26	<.0001	0.33847617	<.0001	0.66152383	<.0001
3	3	residual_sugar		0.3172	3016.57	<.0001	0.23110681	<.0001	0.76889319	<.0001
4	4	volatile_acidity		0.1908	1531.18	<.0001	0.18700138	<.0001	0.81299862	<.0001
5	5	alcohol		0.1411	1066.43	<.0001	0.16061361	<.0001	0.83938639	<.0001
6	6	free_sulfur_dioxide		0.0512	350.56	<.0001	0.15238271	<.0001	0.84761729	<.0001
7	7	fixed_acidity		0.0413	279.76	<.0001	0.14608453	<.0001	0.85391547	<.0001
8	8	chlorides		0.0258	171.50	<.0001	0.14232238	<.0001	0.85767762	<.0001
9	9	pH		0.0107	70.30	<.0001	0.14079658	<.0001	0.85920342	<.0001
10	10	sulphates		0.0084	54.77	<.0001	0.13961752	<.0001	0.86038248	<.0001
11	11	citric_acid		0.0086	56.17	<.0001	0.13841851	<.0001	0.86158149	<.0001
12	12	quality		0.0062	40.54	<.0001	0.13755855	<.0001	0.86244145	<.0001

(b) The followings are results of discriminant analysis for alcohol as a function of 8 predictors.
The p-value for the test of homogeneity of within covariance is still less than .0001. Thus we can conclude that the red and white wines have significantly different covariances and quadratic discriminant analysis is implemented.

The cross-validation estimated overall error rate is 0.0168 and there are 109 misclassified observations. Each individual group shows error estimate as about 1-2%. In detail, 18 red wines are assigned to the group white and 91 white wines are classified as type red. The quality of the two discrimination models is very similar to each other and the slight increase in estimated error rate is easily offset by the simplification of the model in Exercise 2.

Test of Homogeneity of Within Covariance Matrices

Chi-Square	DF	Pr > ChiSq
8456.802303	36	<.0001

Classification Summary for Calibration Data: WORK.WINE
Cross-validation Summary using Quadratic Discriminant Function

Number of Observations and Percent Classified into type			
From type	R	W	Total
R	1581 98.87	18 1.13	1599 100.00
W	91 1.86	4807 98.14	4898 100.00
Total	1672 25.73	4825 74.27	6497 100.00
Priors	0.24611	0.75389	

Error Count Estimates for type			
	R	W	Total
Rate	0.0113	0.0186	0.0168
Priors	0.2461	0.7539	

Exercise 3

The following are results from a quadratic discriminant analysis based on training and test set. Among 1000 observations assigned to the test set, only 15 observations are misclassified and the total error rate is observed as 0.0151, which is slightly less than the crossvalidation error rate of 0.0168 in Exercise 2. Thus we can say that its performance is still good, and it may be closer to the realistic performance. In practice, we use a discriminant model to predict the class of observations from new data set. The highest individual error rate estimate is for type white, at just under 2%. Based on the overall and individual error rate estimates, this looks like a very good model for discriminating between the two wine types.

Classification Summary for Test Data: WORK.TEST
Classification Summary using Quadratic Discriminant Function

Observation Profile for Test Data	
Number of Observations Read	1000
Number of Observations Used	1000

Number of Observations and Percent Classified into type			
From type	**R**	**W**	**Total**
R	255 99.22	2 0.78	257 100.00
W	13 1.75	730 98.25	743 100.00
Total	268 26.80	732 73.20	1000 100.00
Priors	0.24413	0.75587	

Error Count Estimates for type			
	R	**W**	**Total**
Rate	0.0078	0.0175	0.0151
Priors	0.2441	0.7559	