Discriminant Analysis I

Mike Strube

November 6, 2018

1 Preliminaries

In this section, the RStudio workspace and console panes are cleared of old output, variables, and other miscellaneous debris. Packages are loaded and any required data files are retrieved.

```
options(replace.assign = TRUE, width = 65, digits = 4, scipen = 4, fig.width = 4,
    fig.height = 4)
# Clear the workspace and console.
rm(list = ls(all = TRUE))
cat("\f")
```

```
# Turn off showing of significance asterisks.
options(show.signif.stars = F)
# Set the contrast option; important for ANOVAs.
options(contrasts = c("contr.sum", "contr.poly"))
how_long <- Sys.time()
set.seed(123)
library(knitr)
```

```
library(psych)

## Warning:  package 'psych' was built under R version 3.5.1

library(ggplot2)

## Warning:  package 'ggplot2' was built under R version 3.5.1
##
## Attaching package:  'ggplot2'
## The following objects are masked from 'package:psych':
##
##     %+%, alpha

library(MASS)
library(sciplot)
library(plyr)
library(dawai)

## Warning:  package 'dawai' was built under R version 3.5.1
## Loading required package:  mvtnorm
## Loading required package:  ibdreg
## Loading required package:  boot
```

```
##
## Attaching package: 'boot'
## The following object is masked from 'package:psych':
##
##     logit

library(candisc)

## Loading required package: car
## Warning: package 'car' was built under R version 3.5.1
## Loading required package: carData
##
## Attaching package: 'car'
## The following object is masked from 'package:boot':
##
##     logit
## The following object is masked from 'package:psych':
##
##     logit
## Loading required package: heplots
##
## Attaching package: 'candisc'
## The following object is masked from 'package:stats':
##
##     cancor

library(biotools)

## Loading required package: rpanel
## Loading required package: tcltk
## Package `rpanel', version 1.1-4: type help(rpanel) for summary information
##
## Attaching package: 'rpanel'
## The following object is masked from 'package:boot':
##
##     poisons
## Loading required package: tkrplot
## Loading required package: lattice
##
## Attaching package: 'lattice'
## The following object is masked from 'package:boot':
##
##     melanoma
## Loading required package: SpatialEpi
## Loading required package: sp

## ---
## biotools version 3.1

##
##
## Attaching package: 'biotools'
## The following object is masked from 'package:heplots':
##
##     boxM

library(DiscriMiner)
```

```
## Warning:  package 'DiscriMiner' was built under R version 3.5.1
library(ade4)
## Warning:  package 'ade4' was built under R version 3.5.1
library(MVN)
## sROC 0.1-2 loaded
library(biotools)
library(klaR)
## Warning:  package 'klaR' was built under R version 3.5.1
library(GGally)
library(reshape2)
library(MVN)
library(qqplotr)
##
## Attaching package:  'qqplotr'
## The following objects are masked from 'package:ggplot2':
##
##    stat_qq_line, StatQqLine
library(flipMultivariates)
```

1.1 Data

```
setwd("C:\\Courses\\Psychology 516\\PowerPoint\\2018")

Iris <- read.table("iris.csv", sep = ",", header = TRUE)
Iris <- as.data.frame(Iris)
Iris$Species_Num <- Iris$Species
Iris$Species[Iris$Species == "1"] <- "Setosa"
Iris$Species[Iris$Species == "2"] <- "Versicolor"
Iris$Species[Iris$Species == "3"] <- "Virginica"
# Centered versions of the predictors.
Iris$SL_C <- scale(Iris$Sepal_Length, center = TRUE, scale = FALSE)
Iris$SW_C <- scale(Iris$Sepal_Width, center = TRUE, scale = FALSE)
Iris$PL_C <- scale(Iris$Petal_Length, center = TRUE, scale = FALSE)
Iris$PW_C <- scale(Iris$Petal_Width, center = TRUE, scale = FALSE)
# Residuals
Iris$SL_R <- lm(Sepal_Length ~ as.factor(Species_Num), data = Iris)$residuals
Iris$SW_R <- lm(Sepal_Width ~ as.factor(Species_Num), data = Iris)$residuals
Iris$PL_R <- lm(Petal_Length ~ as.factor(Species_Num), data = Iris)$residuals
Iris$PW_R <- lm(Petal_Width ~ as.factor(Species_Num), data = Iris)$residuals
```

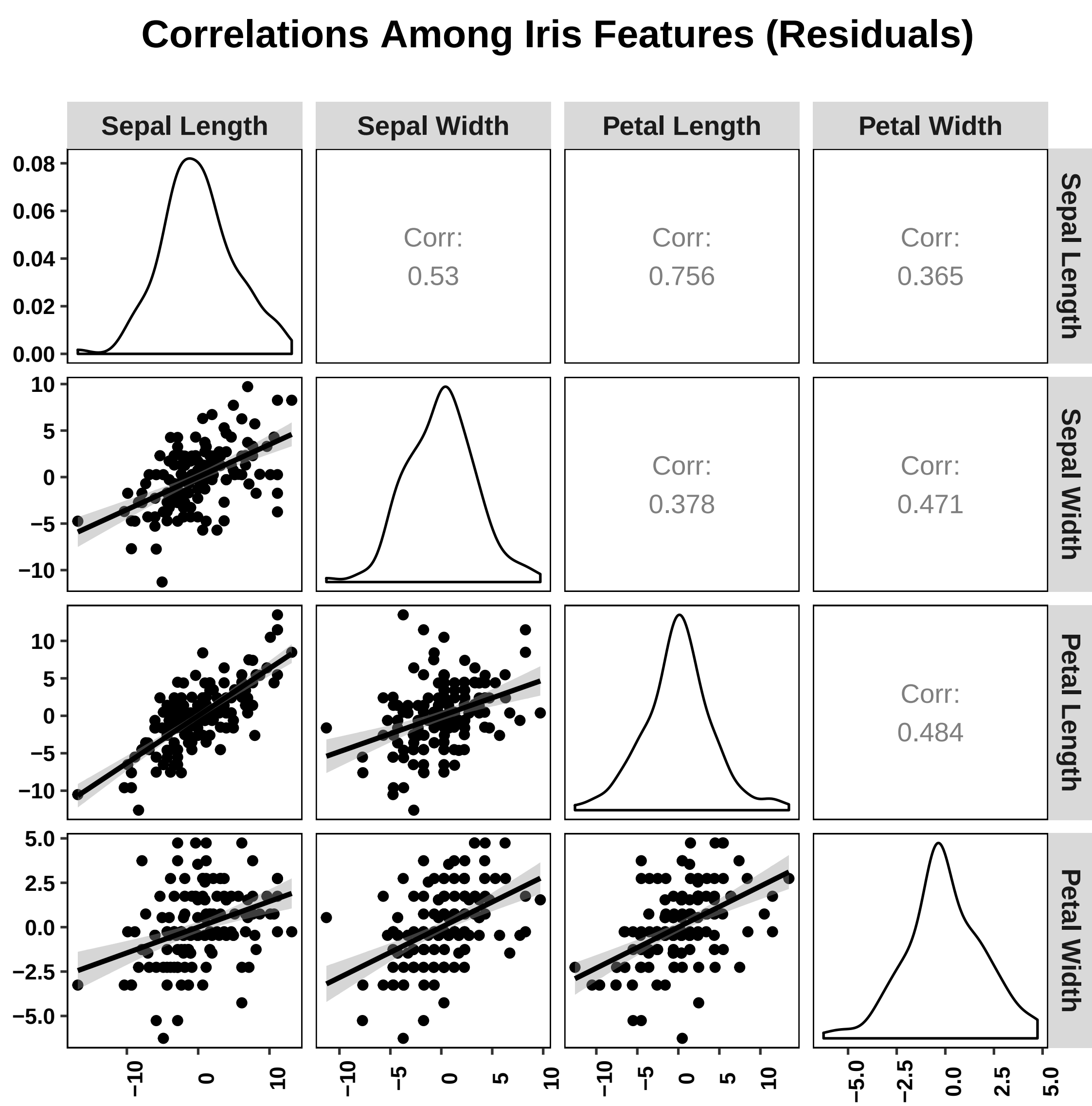
2 The Iris Data

The Fisher iris data provides familiar territory for introducing basic concepts. We know there are three species in the data set and we know that four measures were taken on each flower. What is the best linear combination of those four measures for maximizing the separation of the three species? How many linear combinations do we need to produce good separation?

2.1 Basic Visualization

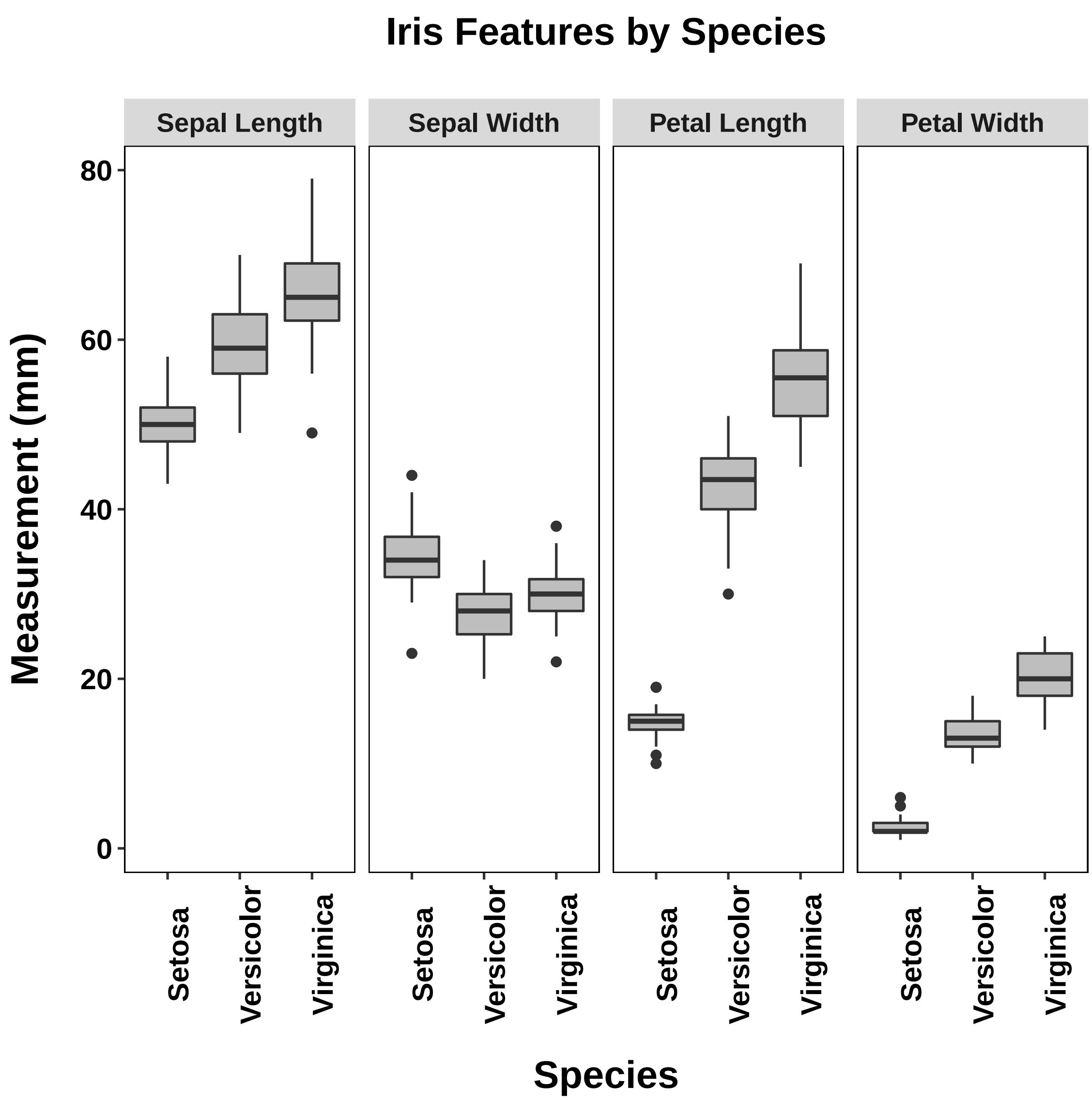
The basic nature of the data is easily viewed with some simple graphics.

```r
ggpairs(Iris[11:14], lower = list(continuous = "smooth"), upper = list(continuous = "cor"),
    columnLabels = c("Sepal Length", "Sepal Width", "Petal Length",
        "Petal Width")) + theme(text = element_text(size = 14, family = "sans",
    color = "black", face = "bold"), axis.text.y = element_text(colour = "black",
    size = 9, face = "bold"), axis.text.x = element_text(colour = "black",
    size = 9, face = "bold", angle = 90), axis.title.x = element_text(margin = margin(15,
    0, 0, 0), size = 16), axis.title.y = element_text(margin = margin(0,
    15, 0, 0), size = 16), axis.line.x = element_blank(), axis.line.y = element_blank(),
    plot.title = element_text(size = 16, face = "bold", margin = margin(0,
        0, 20, 0), hjust = 0.5), panel.background = element_rect(fill = "white",
        linetype = 1, color = "black"), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), plot.background = element_rect(fill = "white"),
    plot.margin = unit(c(1, 1, 1, 1), "cm"), legend.position = "bottom",
    legend.title = element_blank()) + ggtitle("Correlations Among Iris Features (Residuals)")
```




```
plot_data <- melt(data = Iris[, 1:5], id.vars = "Species", measure.vars = c("Sepal_Length",
    "Sepal_Width", "Petal_Length", "Petal_Width"))
plot_data <- as.data.frame(plot_data)
plot_data$variable <- factor(plot_data$variable, levels = c("Sepal_Length",
    "Sepal_Width", "Petal_Length", "Petal_Width"), labels = c("Sepal Length",
    "Sepal Width", "Petal Length", "Petal Width"))

p <- ggplot(plot_data, aes(x = Species, y = value)) + geom_boxplot(fill = "gray") +
    ylab("Measurement (mm)") + theme(text = element_text(size = 14,
    family = "sans", color = "black", face = "bold"), axis.text.y = element_text(colour = "black",
    size = 12, face = "bold"), axis.text.x = element_text(colour = "black",
    size = 12, face = "bold", angle = 90), axis.title.x = element_text(margin = margin(15,
    0, 0, 0), size = 16), axis.title.y = element_text(margin = margin(0,
    15, 0, 0), size = 16), axis.line.x = element_blank(), axis.line.y = element_blank(),
```

```
    plot.title = element_text(size = 16, face = "bold", margin = margin(0,
        0, 20, 0), hjust = 0.5), panel.background = element_rect(fill = "white",
        linetype = 1, color = "black"), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), plot.background = element_rect(fill = "white"),
    plot.margin = unit(c(1, 1, 1, 1), "cm"), legend.position = "bottom",
    legend.title = element_blank()) + ggtitle("Iris Features by Species")
p + facet_grid(~variable)
```



2.2 Group Differences

A univariate look at the data will provide some clues about likely variables of influence in the discriminant analysis.

```
Iris_MANOVA <- manova(as.matrix(Iris[, 1:4]) ~ Iris$Species)
Iris_Wilks <- summary(Iris_MANOVA, test = c("Pillai", "Wilks", "Hotelling-Lawley",
    "Roy"))
summary(Iris_MANOVA)

##               Df Pillai approx F num Df den Df Pr(>F)
## Iris$Species   2   1.19     53.5      8    290 <2e-16
## Residuals    147

summary.aov(Iris_MANOVA)

##  Response Sepal_Length :
##               Df Sum Sq Mean Sq F value Pr(>F)
## Iris$Species   2   6321    3161     119 <2e-16
## Residuals    147   3896      27
##
##  Response Sepal_Width :
##               Df Sum Sq Mean Sq F value Pr(>F)
## Iris$Species   2   1134     567    49.2 <2e-16
## Residuals    147   1696      12
##
##  Response Petal_Length :
##               Df Sum Sq Mean Sq F value Pr(>F)
## Iris$Species   2  43710   21855    1180 <2e-16
## Residuals    147   2722      19
##
##  Response Petal_Width :
##               Df Sum Sq Mean Sq F value Pr(>F)
## Iris$Species   2   8041    4021     960 <2e-16
## Residuals    147    616       4
```

2.3 Basic Function

There are several packages that provide discriminant analysis, but they vary in the specific results they can produce. The two most common functions are lda( ) from the MASS package and candisc( ) from the candisc package. They are illustrated here; some others that provide a few additional useful results are shown later.

```
Iris_LDA <- lda(Species_Num ~ Sepal_Length + Sepal_Width + Petal_Length +
    Petal_Width, data = Iris)
```

Note that candisc( ) uses a multivariate linear model object from the lm( ) function.

```
Iris_MLM <- lm(cbind(Sepal_Length, Sepal_Width, Petal_Length, Petal_Width) ~
    as.factor(Species), data = Iris)
Iris_CDA <- candisc(Iris_MLM, data = Iris)
```

2.4 Standardized and Unstandardized Discriminant Function

The nature of the functions produced by lda( ) is a source of considerable confusion. The documentation describes them as standardized coefficients that would be applied to the centered but not standardized variables. That would usually be the definition for unstandardized weights and the call for raw coefficients from candisc( ) confirms that

interpretation. The candisc( ) also produces true standardized coefficients that allow determining the relative contribution of the variables to group discrimination, even if the variables are on different scales. The candisc( ) function produces the structure matrix as well representing the correlations between the original discriminating variables and the discriminant functions. Note a potential problem if you use several packages on the same data. They may use slightly different estimation algorithms and even if they do not, they may reflect some functions (reversing the signs).

```
Iris_LDA$scaling

##                    LD1      LD2
## Sepal_Length  0.08294 -0.00241
## Sepal_Width   0.15345 -0.21645
## Petal_Length -0.22012  0.09319
## Petal_Width  -0.28105 -0.28392

Iris_CDA$coeffs.raw

##                   Can1     Can2
## Sepal_Length -0.08294 -0.00241
## Sepal_Width  -0.15345 -0.21645
## Petal_Length  0.22012  0.09319
## Petal_Width   0.28105 -0.28392

Iris_CDA$coeffs.std

##                  Can1     Can2
## Sepal_Length -0.4270 -0.01241
## Sepal_Width  -0.5212 -0.73526
## Petal_Length  0.9473  0.40104
## Petal_Width   0.5752 -0.58104

Iris_CDA$structure

##                  Can1     Can2
## Sepal_Length  0.7919 -0.21759
## Sepal_Width  -0.5308 -0.75799
## Petal_Length  0.9850 -0.04604
## Petal_Width   0.9728 -0.22290
```

2.5 Eigenvalues, Lambda, and Canonical Correlations

Classic discriminant analysis is a special case of canonical correlation analysis and within that context it can be useful to examine the canonical correlations and eigenvalues. These are most useful in estimating the magnitude of discrimination of the functions by calculating the percentage of total discrimination that is due to each function. A closely related issue is the statistical significant of the functions.

The candisc( ) function readily provides the eigenvalues and squared canonical correlations. They are related to eachother. The ratio of the squared canonical correlation to (1-squared canonical correlation) is the eigenvalue for that function. Wilks' lambda, used in tests of significance, is the product of (1-squared canonical correlation) for a set of functions. Tests of significance in discriminant analysis are made in a step-wise fashion. First, the entire set is tested for significance by calculating Wilks' lambda (the produce of all 1-squared canonical correlations) and estimating an F ratio (or sometimes

a chi-square). If this test is significant, then we can conclude that there is significant discrimination possible. Then the first (and most important) function is excluded and the remainder are tested. If this is not significant, then the first function was the only source of discrimination. If this is significant, then the first and at least the the second are significant sources of discrimination. The second is next excluded and the inferences follow in the same fashion. If the remainder are not significant, the the first and second were the only sources of significance.

```
Iris_CDA$rank

## [1] 2

Iris_CDA$eigenvalues

## [1]  3.219e+01  2.854e-01  4.081e-15 -2.331e-15

Iris_CDA$canrsq

## [1] 0.9699 0.2220

Iris_CDA$pct

## [1]  9.912e+01  8.787e-01  1.257e-14 -7.179e-15

Iris_CDA

##
## Canonical Discriminant Analysis for as.factor(Species):
##
##   CanRsq Eigenvalue Difference Percent Cumulative
## 1  0.970     32.192       31.9  99.121       99.1
## 2  0.222      0.285       31.9   0.879      100.0
##
## Test of H0: The canonical correlations in the
## current row and all that follow are zero
##
##   LR test stat approx F numDF denDF     Pr(> F)
## 1        0.023    199.1     8   288     < 2e-16
## 2        0.778     13.8     3   145 0.000000058
```

2.6 Homogeneity Assumption

We assume in discriminant analysis that the separate group variance-covariance matrices are homogeneous. Box's test can be used to test this assumption. Note, however, that it is also sensitive to violations of multivariate normality.

```
boxM(Iris[, 1:4], Iris$Species)

##
##  Box's M-test for Homogeneity of Covariance Matrices
##
## data:  Iris[, 1:4]
## Chi-Sq (approx.) = 140, df = 20, p-value <2e-16

boxM(Iris[, 1:4], Iris$Species)$cov
```

```
## $Setosa
##              Sepal_Length Sepal_Width Petal_Length Petal_Width
## Sepal_Length       12.425      9.9216       1.6355      1.0331
## Sepal_Width         9.922     14.3690       1.1698      0.9298
## Petal_Length        1.636      1.1698       3.0159      0.6069
## Petal_Width         1.033      0.9298       0.6069      1.1106
##
## $Versicolor
##              Sepal_Length Sepal_Width Petal_Length Petal_Width
## Sepal_Length       26.643       8.518       18.290       5.578
## Sepal_Width         8.518       9.847        8.265       4.120
## Petal_Length       18.290       8.265       22.082       7.310
## Petal_Width         5.578       4.120        7.310       3.911
##
## $Virginica
##              Sepal_Length Sepal_Width Petal_Length Petal_Width
## Sepal_Length       40.434       9.376       30.329       4.909
## Sepal_Width         9.376      10.400        7.138       4.763
## Petal_Length       30.329       7.138       30.459       4.882
## Petal_Width         4.909       4.763        4.882       7.543
```

```
boxM(Iris[, 1:4], Iris$Species)$pooled
```

```
##              Sepal_Length Sepal_Width Petal_Length Petal_Width
## Sepal_Length       26.501       9.272       16.751       3.840
## Sepal_Width         9.272      11.539        5.524       3.271
## Petal_Length       16.751       5.524       18.519       4.267
## Petal_Width         3.840       3.271        4.267       4.188
```

2.7 Multivariant Normality Assumption

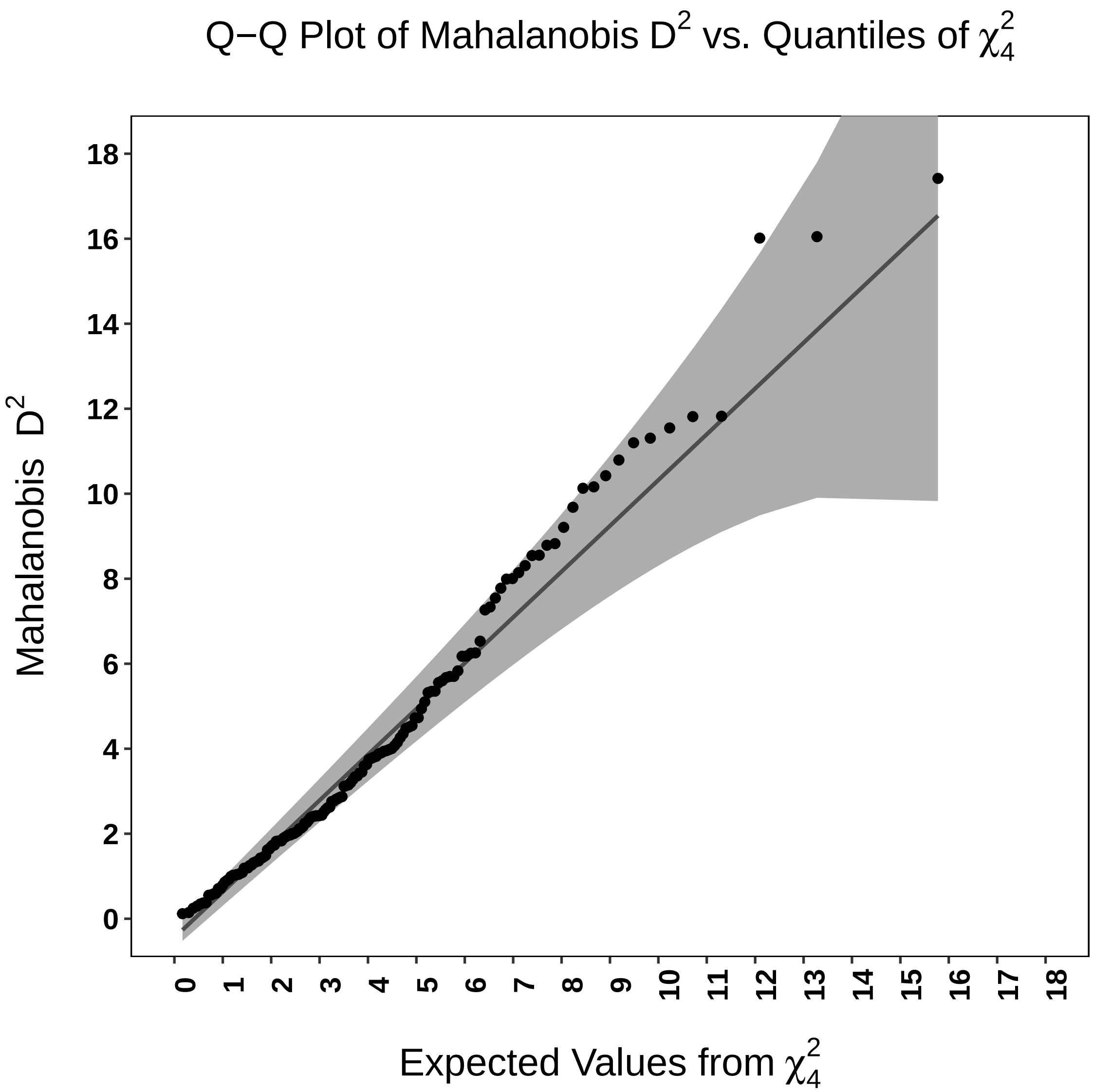
The classification part of discriminant analysis (as well as any significance tests for the discriminant functions) rely on the multivariate normality assumption. The tests are performed on the residualized data so that species differences do not affect the results. Note that a violation of multivariate normality will also affect the test of homogeneity of covariance matrices.

```
mvn(Iris[, 11:14], mvnTest = "mardia")
```

```
## $multivariateNormality
##              Test        Statistic             p value Result
## 1 Mardia Skewness 31.8480630655843  0.0449444331703922     NO
## 2 Mardia Kurtosis 3.28196485281224 0.00103086453506429     NO
## 3             MVN             <NA>                <NA>     NO
##
## $univariateNormality
##           Test  Variable Statistic   p value Normality
## 1 Shapiro-Wilk   SL_R       0.9879    0.2189    YES
## 2 Shapiro-Wilk   SW_R       0.9895    0.3230    YES
## 3 Shapiro-Wilk   PL_R       0.9811    0.0368    NO
## 4 Shapiro-Wilk   PW_R       0.9722    0.0039    NO
##
## $Descriptives
```

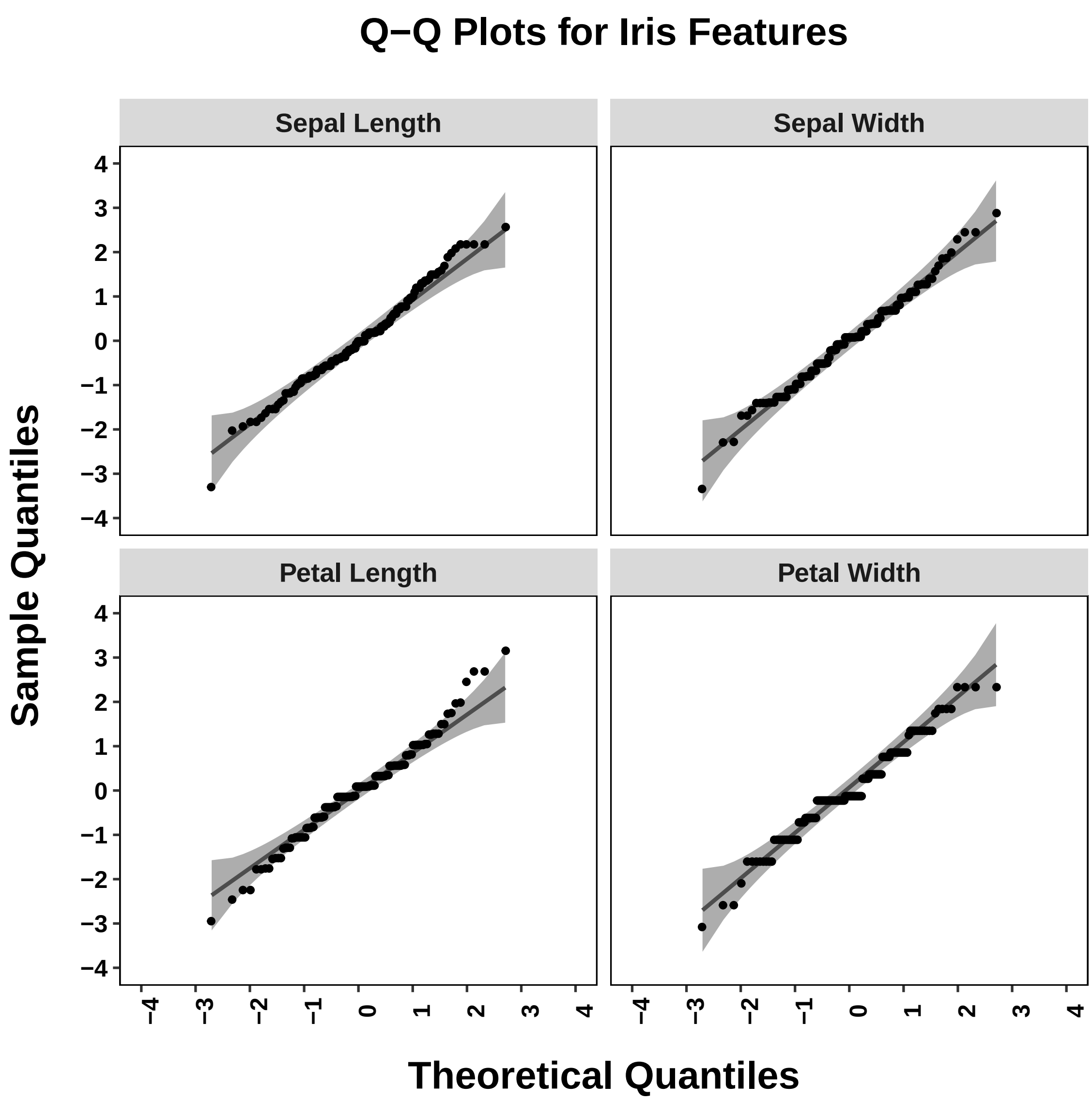
```
##         n       Mean Std.Dev Median    Min   Max   25th 75th
## SL_R 150  1.016e-17   5.113  -0.06 -16.88 13.12 -3.285 3.12
## SW_R 150  7.893e-19   3.374   0.26 -11.28  9.72 -2.280 2.26
## PL_R 150  1.515e-16   4.274   0.38 -12.60 13.48 -2.580 2.40
## PW_R 150 -9.076e-17   2.033  -0.26  -6.26  4.74 -1.260 1.54
##          Skew Kurtosis
## SL_R  0.11751   0.2199
## SW_R  0.02698   0.4614
## PL_R  0.12041   0.8424
## PW_R -0.05436   0.3181
```

```
CV <- cov(Iris[, 11:14])
D2_1 <- mahalanobis(Iris[, 11:14], center = colMeans(Iris[, 11:14]),
    cov = CV)
D2_1 <- as.data.frame(D2_1)
ggplot(D2_1, aes(sample = D2_1)) + stat_qq_band(distribution = "chisq",
    dparams = list(df = 4)) + stat_qq_line(distribution = "chisq",
    dparams = list(df = 4)) + stat_qq(distribution = "qchisq", dparams = list(df = 4)) +
    scale_y_continuous(breaks = seq(0, 18, 2)) + scale_x_continuous(breaks = seq(0,
    18, 1)) + coord_cartesian(xlim = c(0, 18), ylim = c(0, 18)) +
    xlab(expression("Expected Values from" * ~chi[4]^2)) + ylab(expression("Mahalanobis " *
    ~D^2)) + theme(text = element_text(size = 14, family = "sans",
    color = "black", face = "bold"), axis.text.y = element_text(colour = "black",
    size = 12, face = "bold"), axis.text.x = element_text(colour = "black",
    size = 12, face = "bold", angle = 90), axis.title.x = element_text(margin = margin(15,
    0, 0, 0), size = 16), axis.title.y = element_text(margin = margin(0,
    15, 0, 0), size = 16), axis.line.x = element_blank(), axis.line.y = element_blank(),
    plot.title = element_text(size = 16, face = "bold", margin = margin(0,
        0, 20, 0), hjust = 0.5), panel.background = element_rect(fill = "white",
        linetype = 1, color = "black"), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), plot.background = element_rect(fill = "white"),
    plot.margin = unit(c(1, 1, 1, 1), "cm"), legend.position = "bottom",
    legend.title = element_blank()) + ggtitle(expression("Q-Q Plot of Mahalanobis" *
    ~D^2 * " vs. Quantiles of" * ~chi[4]^2))
```




```
Iris_QQ <- scale(Iris[11:14])
Data_long <- melt(Iris_QQ)
Data_long <- as.data.frame(Data_long)
names(Data_long) <- c("Index", "feature", "value")
Data_long$feature_F <- factor(Data_long$feature, levels = c("SL_R",
    "SW_R", "PL_R", "PW_R"), labels = c("Sepal Length", "Sepal Width",
    "Petal Length", "Petal Width"))
p <- ggplot(Data_long, aes(sample = value)) + stat_qq_band() + stat_qq_line() +
    stat_qq(distribution = qnorm, size = 1) + scale_y_continuous(breaks = seq(-4,
    4, 1)) + scale_x_continuous(breaks = seq(-4, 4, 1)) + coord_cartesian(xlim = c(-4,
    4), ylim = c(-4, 4)) + xlab("Theoretical Quantiles") + ylab("Sample Quantiles") +
    theme(text = element_text(size = 14, family = "sans", color = "black",
        face = "bold"), axis.text.y = element_text(colour = "black",
        size = 10, face = "bold"), axis.text.x = element_text(colour = "black",
```

```
        size = 10, face = "bold", angle = 90), axis.title.x = element_text(margin = margin(15,
        0, 0, 0), size = 16), axis.title.y = element_text(margin = margin(0,
        15, 0, 0), size = 16), axis.line.x = element_blank(), axis.line.y = element_blank(),
        plot.title = element_text(size = 16, face = "bold", margin = margin(0,
            0, 20, 0), hjust = 0.5), panel.background = element_rect(fill = "white",
            linetype = 1, color = "black"), panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), plot.background = element_rect(fill = "white"),
        plot.margin = unit(c(1, 1, 1, 1), "cm"), legend.position = "bottom",
        legend.title = element_blank()) + ggtitle("Q-Q Plots for Iris Features")
p + facet_wrap(~feature_F)
```



2.8 Predicted Group Membership

Originally, Fisher proposed the construction of as many classification functions are there were groups, with classification determined by the group with the highest classification score. The Fisher classification functions are available in the DiscriMiner package as a value that can be requested from the linDA( ) function.

```
Iris_Fisher <- linDA(Iris[, 1:4], Iris$Species, prior = NULL, validation = NULL,
    learn = NULL, test = NULL, prob = FALSE)
Iris_Fisher$functions

##               Setosa Versicolor Virginica
## constant     -86.308   -72.8526 -104.3683
## Sepal_Length   2.354     1.5698    1.2446
## Sepal_Width    2.359     0.7073    0.3685
## Petal_Length  -1.643     0.5211    1.2767
## Petal_Width   -1.740     0.6434    2.1079

Iris_Fisher$scores

##        Setosa Versicolor Virginica
## 1    82.7698      37.56   -7.8887
## 2     0.1322      90.76  103.4706
## 3    31.0961      82.61   77.1933
## 4    10.7921      98.88  112.5258
## 5    18.1719      82.08   81.0874
## 6    73.9710      32.63  -10.3906
## 7    25.4561      98.77  106.5237
## 8    11.5232      73.14   69.9717
## 9    17.8991      79.00   80.0769
## 10   87.0007      31.32  -16.8680
## 11   28.1358      77.10   70.8441
## 12    7.0101      77.31   79.0930
## 13   17.2561      90.37   96.1298
## 14   21.2908      60.14   47.5182
## 15   15.8066      90.65   95.7439
## 16   -2.9183      76.10   82.9276
## 17   12.3159     102.07  115.8609
## 18   74.9755      42.62    3.5096
## 19   17.3835      68.25   61.7441
## 20   11.1223      91.46  102.7472
## 21   36.2537     123.97  137.3787
## 22   34.7983      84.17   79.9313
## 23   12.1268     101.46  116.6474
## 24   18.4119     115.58  129.8013
## 25  -15.4876      56.14   59.1135
## 26  100.9025      46.30   -2.2053
## 27   16.7454      95.44  104.9427
## 28   52.4001      93.17   84.0590
## 29   39.8199      83.35   76.1461
## 30   35.0165      71.92   60.3392
## 31   70.0572      34.05   -8.5616
## 32    8.2523      79.14   83.1698
## 33   18.2207      57.34   44.6284
## 34    5.7793      82.01   87.1368
```

```
## 35   12.7567      92.67  103.2227
## 36   89.8374      41.41   -5.0310
## 37   89.2315      37.47  -10.1356
## 38   11.5582      66.24   62.9632
## 39   49.3714     124.26  131.8221
## 40   67.9275      26.91  -17.0014
## 41   19.0861      98.88  108.2157
## 42   77.2418      42.59    3.8333
## 43   24.9755      65.15   53.7604
## 44   63.2100      25.50  -17.7384
## 45   16.1455     115.61  129.4776
## 46   13.8798      82.26   84.4893
## 47   70.0616      33.19   -9.4376
## 48   11.3400      62.53   55.1333
## 49   14.0449      45.54   29.5458
## 50   28.7897     105.66  111.5764
## 51   69.2447      32.94   -9.3755
## 52   93.6320      43.71   -2.2481
## 53   16.2473      81.24   83.1057
## 54   72.2044      37.73   -3.6260
## 55   72.4073      36.48   -6.4409
## 56   83.6971      35.81  -10.8105
## 57    2.2701      79.49   82.1365
## 58    1.8720      90.11  101.3627
## 59   65.8815      22.24  -23.6442
## 60  121.4027      54.03    2.0944
## 61   85.2231      46.56    5.7977
## 62   47.9073      86.19   76.1268
## 63   15.5272      80.88   82.3365
## 64   73.3390      33.87  -10.2388
## 65   89.8418      40.54   -5.9070
## 66   16.2670      69.38   65.4524
## 67   29.1709      65.09   51.1898
## 68   78.3627      40.60   -0.7510
## 69   70.9933      30.57  -13.2355
## 70   23.3259      74.95   70.0622
## 71   34.3161      59.69   41.9168
## 72  120.0746      58.14   10.3697
## 73   89.8461      39.68   -6.7831
## 74   25.5020     115.83  128.8784
## 75    8.4508      94.72  108.6530
## 76   -2.9183      76.10   82.9276
## 77   24.6715      67.39   58.2827
## 78   30.8380     101.91  104.0702
## 79   89.4235      46.36    2.9507
## 80  106.4457      46.24   -3.2825
## 81   18.1413     104.08  114.6418
## 82   15.8109      89.79   94.8679
## 83   15.5359      79.15   80.5844
## 84    7.0959      87.32   94.1629
## 85   14.0493      44.67   28.6697
## 86   17.5995      64.41   56.3010
## 87   28.7701      67.45   56.5433
## 88  115.7709      51.77    1.6510
```

```
## 89   74.0547      35.10   -8.5937
## 90    2.9223     117.17  137.6176
## 91   -1.4010      72.60   73.8658
## 92   97.9313      50.93    7.3466
## 93   39.2890      83.60   74.5905
## 94   11.3712      57.21   49.6006
## 95   33.3800      79.13   74.0128
## 96   83.4855      38.79   -6.2435
## 97   59.2081      25.31  -16.8303
## 98    3.6825      44.46   30.9935
## 99   21.6036      56.18   41.2438
## 100  28.9774      65.34   52.8523
## 101  74.9908      31.62  -13.2676
## 102  66.9915      30.39  -12.3274
## 103  17.9565     102.60  114.5522
## 104  34.8005      75.76   65.7822
## 105  20.6805     107.13  115.9760
## 106  28.2594      72.52   65.3562
## 107  85.8399      40.36   -4.9989
## 108  87.3906      39.10   -6.3204
## 109  -3.7221      73.26   80.3615
## 110  23.7289      64.18   50.5597
## 111  19.1386     106.67  115.5453
## 112   3.7923      91.81  100.2204
## 113  75.7945      34.45  -10.7016
## 114  41.2382      88.40   82.0647
## 115  21.3591      73.61   66.0923
## 116  99.9795      47.19   -0.1596
## 117  26.3189      66.01   56.1300
## 118  14.3817      78.91   77.4285
## 119  25.2549      74.92   67.1678
## 120  39.5093      78.90   68.2714
## 121  24.3063      47.60   29.8054
## 122  23.9558      66.16   56.6375
## 123   3.9180      94.79  108.0055
## 124  24.0066      99.04  106.1379
## 125 104.5035      48.85    2.2400
## 126  88.1019      41.19   -3.7991
## 127  24.4029     113.51  129.0825
## 128  23.6167      91.26   95.5902
## 129  24.1340      76.92   71.7522
## 130  32.1755      62.69   49.3782
## 131  42.6609      92.58   87.1071
## 132   7.9221      86.56   92.9484
## 133  17.6503      97.29  105.8014
## 134  13.8136      60.38   51.7661
## 135  77.1336      36.17   -7.4560
## 136  72.7244      31.66  -13.5913
## 137  47.3130      22.76  -16.9656
## 138 -10.0871      73.23   81.7772
## 139 104.3756      54.29    8.6039
## 140  93.5352      43.83   -1.4169
## 141   9.0961      58.96   51.0290
## 142  43.1941      83.91   74.5137
```

```
## 143  36.6131      87.72   81.3724
## 144  89.6170      46.11    1.2882
## 145  89.4366      43.77    0.3225
## 146  90.5531      42.64   -3.3858
## 147  -9.2587      80.02   93.8356
## 148  30.4706      90.54   89.7419
## 149  -2.2201      96.74  115.4990
## 150  97.6251      45.62   -1.4041
```

```
Iris_Fisher$classification
```

```
##   [1] Setosa     Virginica  Versicolor Virginica  Versicolor
##   [6] Setosa     Virginica  Versicolor Virginica  Setosa
##  [11] Versicolor Virginica  Virginica  Versicolor Virginica
##  [16] Virginica  Virginica  Setosa     Versicolor Virginica
##  [21] Virginica  Versicolor Virginica  Virginica  Virginica
##  [26] Setosa     Virginica  Versicolor Versicolor Versicolor
##  [31] Setosa     Virginica  Versicolor Virginica  Virginica
##  [36] Setosa     Setosa     Versicolor Virginica  Setosa
##  [41] Virginica  Setosa     Versicolor Setosa     Virginica
##  [46] Virginica  Setosa     Versicolor Versicolor Virginica
##  [51] Setosa     Setosa     Virginica  Setosa     Setosa
##  [56] Setosa     Virginica  Virginica  Setosa     Setosa
##  [61] Setosa     Versicolor Virginica  Setosa     Setosa
##  [66] Versicolor Versicolor Setosa     Setosa     Versicolor
##  [71] Versicolor Setosa     Setosa     Virginica  Virginica
##  [76] Virginica  Versicolor Virginica  Setosa     Setosa
##  [81] Virginica  Virginica  Virginica  Virginica  Versicolor
##  [86] Versicolor Versicolor Setosa     Setosa     Virginica
##  [91] Virginica  Setosa     Versicolor Versicolor Versicolor
##  [96] Setosa     Setosa     Versicolor Versicolor Versicolor
## [101] Setosa     Setosa     Virginica  Versicolor Virginica
## [106] Versicolor Setosa     Setosa     Virginica  Versicolor
## [111] Virginica  Virginica  Setosa     Versicolor Versicolor
## [116] Setosa     Versicolor Versicolor Versicolor Versicolor
## [121] Versicolor Versicolor Virginica  Virginica  Setosa
## [126] Setosa     Virginica  Virginica  Versicolor Versicolor
## [131] Versicolor Virginica  Virginica  Versicolor Setosa
## [136] Setosa     Setosa     Virginica  Setosa     Setosa
## [141] Versicolor Versicolor Versicolor Setosa     Setosa
## [146] Setosa     Virginica  Versicolor Virginica  Setosa
## Levels: Setosa Versicolor Virginica
```

Another approach available in the biotools package calculates the Mahalanobis distances of an object from the group centroids for each group and classifies it in the nearest.

```
Iris_Mahal <- D2.disc(Iris[, 1:4], iris[, 5])
Iris_Mahal$D2
```

```
##      setosa versicolor virginica   grouping       pred misclass
## 1    2.8442     1.5821    1.7829     setosa versicolor        *
## 2    1.8747     3.6195    2.6761     setosa     setosa
## 3    2.2986     2.7451    1.4686     setosa  virginica        *
## 4    5.2060     7.6918    5.8462     setosa     setosa
```

```
## 5    2.3827    2.2101    2.8775     setosa versicolor        *
## 6    3.3301    2.2389    3.3905     setosa versicolor        *
## 7   12.1806   14.9397   11.5396     setosa  virginica        *
## 8    7.9112    8.4830    6.3460     setosa  virginica        *
## 9    2.5555    3.5056    4.2156     setosa     setosa
## 10   4.1096    3.0626    3.9609     setosa versicolor        *
## 11   1.1180    1.0216    1.7099     setosa versicolor        *
## 12   2.2802    2.3606    3.2322     setosa     setosa
## 13   1.3017    2.7831    1.6995     setosa     setosa
## 14   2.5224    2.0660    2.1801     setosa versicolor        *
## 15   1.5963    2.0602    2.5903     setosa     setosa
## 16   2.3189    3.3607    3.6166     setosa     setosa
## 17   2.2956    4.1262    3.4638     setosa     setosa
## 18   2.8105    2.2989    1.8901     setosa  virginica        *
## 19   3.2763    2.8332    4.2673     setosa versicolor        *
## 20   6.4954    8.5952    8.6692     setosa     setosa
## 21  12.0097   12.8390   13.7169     setosa     setosa
## 22   0.9363    1.4341    1.7747     setosa     setosa
## 23   6.4480    9.1275    7.6411     setosa     setosa
## 24   5.8155    6.5446    5.9286     setosa     setosa
## 25   9.3542   10.0547   11.4357     setosa     setosa
## 26   6.5181    5.5119    4.5235     setosa  virginica        *
## 27   8.6266   11.2226    8.4181     setosa  virginica        *
## 28   4.8239    4.9196    3.8702     setosa  virginica        *
## 29   0.5455    0.9746    0.6367     setosa     setosa
## 30   1.6676    1.9202    0.8580     setosa  virginica        *
## 31   2.8730    1.3611    2.2007     setosa versicolor        *
## 32   2.8503    3.5186    4.5931     setosa     setosa
## 33   3.1633    2.7991    2.6760     setosa  virginica        *
## 34   3.9245    5.3100    3.4360     setosa  virginica        *
## 35   4.9678    7.2965    6.0695     setosa     setosa
## 36   3.4922    2.2846    2.1681     setosa  virginica        *
## 37   4.2641    2.5519    4.1239     setosa versicolor        *
## 38   5.7374    5.8275    7.7819     setosa     setosa
## 39  12.5967   13.1363   13.4056     setosa     setosa
## 40   3.9232    2.5618    3.9067     setosa versicolor        *
## 41   2.1498    3.4453    3.5166     setosa     setosa
## 42   3.9626    3.8235    3.4935     setosa  virginica        *
## 43   1.6686    1.5043    1.1899     setosa  virginica        *
## 44   3.8661    2.5486    3.4232     setosa versicolor        *
## 45   8.7835    9.1401    8.4453     setosa  virginica        *
## 46   1.6356    2.7399    1.4473     setosa  virginica        *
## 47   3.2412    1.6691    3.0201     setosa versicolor        *
## 48   5.0341    4.3208    5.9483     setosa versicolor        *
## 49   4.8244    4.4873    4.5084     setosa versicolor        *
## 50   5.6649    5.8103    6.2682     setosa     setosa
## 51   3.8320    2.9051    2.5635 versicolor  virginica        *
## 52   5.0903    3.5397    5.1966 versicolor versicolor
## 53   0.6308    1.5927    1.5581 versicolor     setosa        *
## 54   6.0610    4.1578    6.5377 versicolor versicolor
## 55   3.5338    2.1204    2.1723 versicolor versicolor
## 56   4.4500    3.4268    2.8360 versicolor  virginica        *
## 57  12.9539   11.8600   14.3904 versicolor versicolor
## 58   1.0701    2.4689    1.9149 versicolor     setosa        *
```

```
## 59   4.4034     2.9186    3.9024 versicolor versicolor
## 60   9.7458     8.8534    7.9338 versicolor  virginica        *
## 61   4.8634     3.6792    5.3921 versicolor versicolor
## 62   3.1232     3.4515    2.1609 versicolor  virginica        *
## 63   1.1689     2.3214    1.2620 versicolor     setosa        *
## 64   3.9024     2.6677    2.4394 versicolor  virginica        *
## 65   2.9804     1.7127    2.1076 versicolor versicolor
## 66   3.0658     3.2325    4.6344 versicolor     setosa        *
## 67   3.7627     2.7759    3.7538 versicolor versicolor
## 68   2.3315     1.5223    1.7360 versicolor versicolor
## 69   3.0802     1.6869    2.7576 versicolor versicolor
## 70   0.1672     0.5089    0.5706 versicolor     setosa        *
## 71   2.6674     2.3155    1.6648 versicolor  virginica        *
## 72  10.0097     9.3577    9.7109 versicolor versicolor
## 73   3.1808     1.8529    2.7592 versicolor versicolor
## 74   9.6908    11.6927    8.6302 versicolor  virginica        *
## 75   6.9598     9.2268    9.1917 versicolor     setosa        *
## 76   2.3189     3.3607    3.6166 versicolor     setosa        *
## 77   1.1131     0.9736    1.8081 versicolor versicolor
## 78   7.1931     6.9073    7.1173 versicolor versicolor
## 79   5.4997     4.9522    3.7825 versicolor  virginica        *
## 80   9.0443     7.2288    9.4261 versicolor versicolor
## 81   2.1146     3.4120    2.2293 versicolor     setosa        *
## 82   2.7838     3.1875    4.2290 versicolor     setosa        *
## 83   0.9214     1.9536    1.9170 versicolor     setosa        *
## 84   3.1196     3.4204    4.5461 versicolor     setosa        *
## 85   4.4449     4.0476    4.5802 versicolor versicolor
## 86   1.4617     1.3278    1.9682 versicolor versicolor
## 87   2.6882     2.1808    3.6399 versicolor versicolor
## 88   7.8219     6.5538    7.4898 versicolor versicolor
## 89   2.9472     1.5821    1.8671 versicolor versicolor
## 90   7.0806     8.3020    6.5251 versicolor  virginica        *
## 91   4.5290     4.4816    4.2886 versicolor  virginica        *
## 92   4.3518     3.4775    4.0403 versicolor versicolor
## 93   3.1546     2.9255    2.3868 versicolor  virginica        *
## 94   2.8216     3.2062    3.7274 versicolor     setosa        *
## 95   2.2645     2.8426    3.6906 versicolor     setosa        *
## 96   2.6710     1.2784    1.9927 versicolor versicolor
## 97   4.1451     2.7409    3.6587 versicolor versicolor
## 98   8.0575     7.5692    7.4414 versicolor  virginica        *
## 99   3.6091     3.0074    2.9955 versicolor  virginica        *
## 100  1.1178     1.0401    0.6826 versicolor  virginica        *
## 101  2.8690     1.6225    2.1387  virginica versicolor        *
## 102  3.2254     1.7454    2.8593  virginica versicolor        *
## 103  3.3689     5.4549    3.9596  virginica     setosa        *
## 104  0.8726     0.8160    0.5475  virginica  virginica
## 105  5.7777     6.3247    5.1784  virginica  virginica
## 106  0.2159     0.8230    0.5613  virginica     setosa        *
## 107  2.7944     1.4401    1.8781  virginica versicolor        *
## 108  3.2121     2.3606    2.3640  virginica versicolor        *
## 109  4.4753     5.9817    6.0793  virginica     setosa        *
## 110  8.4765     7.7844    6.7558  virginica  virginica
## 111  7.5308     7.4546    7.7656  virginica versicolor        *
## 112  3.2883     3.6128    3.2494  virginica  virginica
```

```
## 113  3.4728     1.7617    2.4362  virginica versicolor         *
## 114  1.9421     2.2908    1.4456  virginica  virginica
## 115  7.2637     7.2684    5.5910  virginica  virginica
## 116  4.1722     2.8669    3.1816  virginica versicolor         *
## 117  1.5026     1.4115    2.4791  virginica versicolor         *
## 118  2.5885     2.6987    2.1041  virginica  virginica
## 119  4.7543     3.9012    5.1764  virginica versicolor         *
## 120  1.9746     1.9945    1.1734  virginica  virginica
## 121  5.0063     5.3352    4.3261  virginica  virginica
## 122  0.6233     0.6143    0.9353  virginica versicolor         *
## 123  1.5757     3.0979    2.9051  virginica     setosa         *
## 124  3.1358     4.8773    3.0909  virginica  virginica
## 125  6.2914     5.8512    5.3008  virginica  virginica
## 126  3.0408     2.1190    2.1245  virginica versicolor         *
## 127  6.3339     8.7049    7.6557  virginica     setosa         *
## 128  1.6849     3.2310    2.3394  virginica     setosa         *
## 129  1.5236     1.3406    2.0720  virginica versicolor         *
## 130  0.9998     1.4732    0.9143  virginica  virginica
## 131  3.2414     3.4496    2.6084  virginica  virginica
## 132  0.9373     1.9917    1.1466  virginica     setosa         *
## 133  2.0281     3.6447    2.1777  virginica     setosa         *
## 134  2.0491     2.1378    2.0276  virginica  virginica
## 135  3.5013     1.9237    3.4687  virginica versicolor         *
## 136  3.7182     2.0993    2.5366  virginica versicolor         *
## 137 12.5652    11.7854   10.3483  virginica  virginica
## 138  4.0792     5.5811    4.9751  virginica     setosa         *
## 139  5.3615     4.2781    4.1663  virginica  virginica
## 140  3.9472     2.8511    3.8403  virginica versicolor         *
## 141  3.7753     3.9078    3.3273  virginica  virginica
## 142  2.6718     2.9838    1.7343  virginica  virginica
## 143  4.6924     4.6907    3.5314  virginica  virginica
## 144  3.7119     2.2553    2.4211  virginica versicolor         *
## 145  3.3683     2.6402    3.0048  virginica versicolor         *
## 146  3.0967     1.7587    2.1557  virginica versicolor         *
## 147 10.3041    13.0539   11.5978  virginica     setosa         *
## 148  1.3882     2.1253    1.1009  virginica  virginica
## 149  7.6235     9.8243   10.2777  virginica     setosa         *
## 150  3.8327     2.4892    3.0801  virginica versicolor         *
```

The most common approach uses a Bayesian model, takes prior probabilities into account, and calculates the posterior probabilities for group membership. Cases are classified into the group for which they have the highest posterior probability. This information can be provided by both the lda( ) and candisc( ) functions. The lda( ) function provides the posterior probabilities. The candisc ( ) function shows the classification group and can also provide function means for the groups. The linDA( ) function from the DiscriMiner package directly gives the misclassification rate and a confusion table.

```
Iris_Predicted <- predict(Iris_LDA)
Iris_Predicted$class

##   [1] 1 3 2 3 2 1 3 2 3 1 2 3 3 2 3 3 3 1 2 3 3 2 3 3 3 1 3 2 2 2
##  [31] 1 3 2 3 3 1 1 2 3 1 3 1 2 1 3 3 1 2 2 3 1 1 3 1 1 1 3 3 1 1
##  [61] 1 2 3 1 1 2 2 1 1 2 2 1 1 3 3 3 2 3 1 1 3 3 3 3 2 2 2 1 1 3
##  [91] 3 1 2 2 2 1 1 2 2 2 1 1 3 2 3 2 1 1 3 2 3 3 1 2 2 1 2 2 2 2
## [121] 2 2 3 3 1 1 3 3 2 2 2 3 3 2 1 1 1 3 1 1 2 2 2 1 1 1 3 2 3 1
## Levels: 1 2 3

Iris_Predicted$posterior

##             1         2         3
## 1   1.000e+00 2.322e-20 4.242e-40
## 2   1.320e-45 3.014e-06 1.000e+00
## 3   4.214e-23 9.956e-01 4.410e-03
## 4   6.571e-45 1.181e-06 1.000e+00
## 5   1.284e-28 7.294e-01 2.706e-01
## 6   1.000e+00 1.113e-18 2.303e-37
## 7   6.203e-36 4.276e-04 9.996e-01
## 8   1.670e-27 9.596e-01 4.043e-02
## 9   7.408e-28 2.532e-01 7.468e-01
## 10  1.000e+00 6.569e-25 7.769e-46
## 11  5.399e-22 9.981e-01 1.907e-03
## 12  4.242e-32 1.434e-01 8.566e-01
## 13  5.549e-35 3.146e-03 9.969e-01
## 14  1.342e-17 1.000e+00 3.296e-06
## 15  1.910e-35 6.084e-03 9.939e-01
## 16  5.214e-38 1.078e-03 9.989e-01
## 17  1.074e-45 1.029e-06 1.000e+00
## 18  1.000e+00 8.912e-15 9.179e-32
## 19  8.127e-23 9.985e-01 1.498e-03
## 20  1.614e-40 1.257e-05 1.000e+00
## 21  1.208e-44 1.504e-06 1.000e+00
## 22  3.550e-22 9.858e-01 1.417e-02
## 23  4.048e-46 2.525e-07 1.000e+00
## 24  4.209e-49 6.656e-07 1.000e+00
## 25  3.798e-33 4.862e-02 9.514e-01
## 26  1.000e+00 1.940e-24 1.663e-45
## 27  4.970e-39 7.473e-05 9.999e-01
## 28  1.970e-18 9.999e-01 1.106e-04
## 29  1.243e-19 9.993e-01 7.425e-04
## 30  9.399e-17 1.000e+00 9.345e-06
## 31  1.000e+00 2.300e-16 7.183e-35
## 32  2.858e-33 1.754e-02 9.825e-01
## 33  1.022e-17 1.000e+00 3.008e-06
```

```
## 34  4.617e-36 5.899e-03 9.941e-01
## 35  5.142e-40 2.605e-05 1.000e+00
## 36  1.000e+00 9.272e-22 6.298e-42
## 37  1.000e+00 3.310e-23 7.005e-44
## 38  1.725e-24 9.636e-01 3.644e-02
## 39  1.556e-36 5.198e-04 9.995e-01
## 40  1.000e+00 1.541e-18 1.306e-37
## 41  1.956e-39 8.837e-05 9.999e-01
## 42  1.000e+00 8.941e-16 1.316e-32
## 43  3.560e-18 1.000e+00 1.129e-05
## 44  1.000e+00 4.190e-17 6.991e-36
## 45  6.033e-50 9.510e-07 1.000e+00
## 46  1.951e-31 9.712e-02 9.029e-01
## 47  1.000e+00 9.665e-17 2.978e-35
## 48  5.879e-23 9.994e-01 6.144e-04
## 49  2.104e-14 1.000e+00 1.135e-07
## 50  1.109e-36 2.679e-03 9.973e-01
## 51  1.000e+00 1.715e-16 7.172e-35
## 52  1.000e+00 2.083e-22 2.290e-42
## 53  7.965e-30 1.342e-01 8.658e-01
## 54  1.000e+00 1.071e-15 1.168e-33
## 55  1.000e+00 2.497e-16 5.710e-35
## 56  1.000e+00 1.597e-21 9.034e-42
## 57  1.927e-35 6.602e-02 9.340e-01
## 58  6.191e-44 1.304e-05 1.000e+00
## 59  1.000e+00 1.117e-19 1.317e-39
## 60  1.000e+00 5.487e-30 1.531e-52
## 61  1.000e+00 1.616e-17 3.206e-35
## 62  2.370e-17 1.000e+00 4.268e-05
## 63  7.842e-30 1.884e-01 8.116e-01
## 64  1.000e+00 7.218e-18 5.042e-37
## 65  1.000e+00 3.896e-22 2.611e-42
## 66  8.429e-24 9.806e-01 1.935e-02
## 67  2.505e-16 1.000e+00 9.152e-07
## 68  1.000e+00 3.968e-17 4.379e-35
## 69  1.000e+00 2.793e-18 2.630e-37
## 70  3.774e-23 9.925e-01 7.483e-03
## 71  9.555e-12 1.000e+00 1.911e-08
## 72  1.000e+00 1.262e-27 2.269e-48
## 73  1.000e+00 1.637e-22 1.083e-42
## 74  1.271e-45 2.153e-06 1.000e+00
## 75  3.039e-44 8.882e-07 1.000e+00
## 76  5.214e-38 1.078e-03 9.989e-01
## 77  2.802e-19 9.999e-01 1.108e-04
## 78  1.406e-32 1.037e-01 8.963e-01
## 79  1.000e+00 1.975e-19 2.788e-38
## 80  1.000e+00 7.100e-27 2.216e-48
## 81  1.231e-42 2.593e-05 1.000e+00
## 82  4.606e-35 6.166e-03 9.938e-01
## 83  4.539e-29 1.925e-01 8.075e-01
## 84  1.537e-38 1.068e-03 9.989e-01
## 85  5.007e-14 1.000e+00 1.120e-07
## 86  4.701e-21 9.997e-01 3.020e-04
## 87  1.584e-17 1.000e+00 1.826e-05
```

```
## 88  1.000e+00 1.610e-28 2.744e-50
## 89  1.000e+00 1.205e-17 1.277e-36
## 90  3.181e-59 1.317e-09 1.000e+00
## 91  1.599e-33 2.208e-01 7.792e-01
## 92  1.000e+00 3.883e-21 4.567e-40
## 93  5.683e-20 9.999e-01 1.219e-04
## 94  1.241e-20 9.995e-01 4.973e-04
## 95  1.345e-20 9.940e-01 5.960e-03
## 96  1.000e+00 3.878e-20 1.074e-39
## 97  1.000e+00 1.903e-15 9.483e-34
## 98  1.957e-18 1.000e+00 1.421e-06
## 99  9.648e-16 1.000e+00 3.267e-07
## 100 1.616e-16 1.000e+00 3.778e-06
## 101 1.000e+00 1.464e-19 4.676e-39
## 102 1.000e+00 1.269e-16 3.567e-35
## 103 1.119e-42 6.452e-06 1.000e+00
## 104 1.627e-18 1.000e+00 4.640e-05
## 105 4.108e-42 1.442e-04 9.999e-01
## 106 5.969e-20 9.992e-01 7.706e-04
## 107 1.000e+00 1.769e-20 3.542e-40
## 108 1.000e+00 1.063e-21 2.004e-41
## 109 3.038e-37 8.273e-04 9.992e-01
## 110 2.716e-18 1.000e+00 1.220e-06
## 111 1.352e-42 1.395e-04 9.999e-01
## 112 1.323e-42 2.235e-04 9.998e-01
## 113 1.000e+00 1.112e-18 2.724e-38
## 114 3.305e-21 9.982e-01 1.778e-03
## 115 2.035e-23 9.995e-01 5.443e-04
## 116 1.000e+00 1.185e-23 3.237e-44
## 117 5.807e-18 9.999e-01 5.137e-05
## 118 7.675e-29 8.155e-01 1.845e-01
## 119 2.683e-22 9.996e-01 4.277e-04
## 120 7.814e-18 1.000e+00 2.421e-05
## 121 7.638e-11 1.000e+00 1.867e-08
## 122 4.679e-19 9.999e-01 7.306e-05
## 123 6.243e-46 1.813e-06 1.000e+00
## 124 2.140e-36 8.291e-04 9.992e-01
## 125 1.000e+00 6.754e-25 3.868e-45
## 126 1.000e+00 4.224e-21 1.224e-40
## 127 3.453e-46 1.727e-07 1.000e+00
## 128 5.453e-32 1.305e-02 9.869e-01
## 129 1.184e-23 9.943e-01 5.673e-03
## 130 5.575e-14 1.000e+00 1.649e-06
## 131 2.088e-22 9.958e-01 4.193e-03
## 132 1.183e-37 1.674e-03 9.983e-01
## 133 5.204e-39 2.006e-04 9.998e-01
## 134 5.981e-21 9.998e-01 1.817e-04
## 135 1.000e+00 1.622e-18 1.833e-37
## 136 1.000e+00 1.459e-18 3.263e-38
## 137 1.000e+00 2.174e-11 1.214e-28
## 138 1.270e-40 1.949e-04 9.998e-01
## 139 1.000e+00 1.775e-22 2.552e-42
## 140 1.000e+00 2.593e-22 5.792e-42
## 141 2.199e-22 9.996e-01 3.577e-04
```

```
## 142 2.073e-18 9.999e-01 8.291e-05
## 143 6.358e-23 9.983e-01 1.746e-03
## 144 1.000e+00 1.275e-19 4.358e-39
## 145 1.000e+00 1.466e-20 1.987e-39
## 146 1.000e+00 1.548e-21 1.595e-41
## 147 1.685e-45 1.000e-06 1.000e+00
## 148 5.639e-27 6.892e-01 3.108e-01
## 149 7.503e-52 7.127e-09 1.000e+00
## 150 1.000e+00 2.598e-23 9.821e-44
```

```
Iris_Predicted$x
```

```
##         LD1       LD2
## 1    7.6720  0.134894
## 2   -6.8002 -0.580895
## 3   -2.5487  0.472205
## 4   -6.6531 -1.805320
## 5   -3.8152  0.942986
## 6    7.2126 -0.355836
## 7   -5.1056 -1.992182
## 8   -3.4981  1.684956
## 9   -3.7159 -1.044514
## 10   8.6810 -0.877590
## 11  -2.2925  0.332860
## 12  -4.4985  0.882750
## 13  -4.9677 -0.821141
## 14  -1.0904  1.626583
## 15  -5.0660  0.026273
## 16  -5.5075  0.035814
## 17  -6.7960 -1.460687
## 18   6.2514 -0.439696
## 19  -2.4300  0.966132
## 20  -5.8861 -2.345091
## 21  -6.6088 -1.751636
## 22  -2.4485 -0.795962
## 23  -6.8474 -2.428951
## 24  -7.4182  0.173118
## 25  -4.6780  0.499095
## 26   8.6137 -0.403254
## 27  -5.6450 -1.677717
## 28  -1.4593 -0.028544
## 29  -1.7977 -0.484386
## 30  -0.9976  0.490531
## 31   6.7590  0.759003
## 32  -4.6832 -0.332034
## 33  -1.1067  1.752254
## 34  -5.1796  0.363475
## 35  -5.8070 -2.010199
## 36   7.9913 -0.086379
## 37   8.3304 -0.228134
## 38  -2.9340 -0.027379
## 39  -5.2205 -1.468195
## 40   7.2410  0.272615
## 41  -5.7232 -1.293276
```

```
## 42   6.4144 -1.247301
## 43  -1.2723  1.214584
## 44   6.9341  0.705519
## 45  -7.5812  0.980723
## 46  -4.3715  0.121297
## 47   6.8295  0.544961
## 48  -2.4016  1.594583
## 49  -0.2932  1.798715
## 50  -5.2792  0.042458
## 51   6.7647  0.505152
## 52   8.0819 -0.763393
## 53  -4.0770 -0.523238
## 54   6.5589  0.389223
## 55   6.7714  0.970634
## 56   7.9588  0.164962
## 57  -5.1075  2.130590
## 58  -6.5191 -0.296976
## 59   7.5725  0.805464
## 60   9.8498 -1.585937
## 61   6.8594 -1.051654
## 62  -1.2012 -0.084437
## 63  -4.0809 -0.185937
## 64   7.1287  0.786660
## 65   8.0618 -0.300421
## 66  -2.7681 -0.032200
## 67  -0.7769  1.659162
## 68   6.8231 -0.463012
## 69   7.1868  0.360987
## 70  -2.5898  0.174612
## 71   0.3075  1.318871
## 72   9.1582 -2.737596
## 73   8.1323 -0.514463
## 74  -6.7967 -0.863090
## 75  -6.5245 -2.445035
## 76  -5.5075  0.035814
## 77  -1.6162  0.470104
## 78  -4.5837  0.856816
## 79   7.3750 -0.565845
## 80   9.1263 -1.224433
## 81  -6.2920 -0.467176
## 82  -4.9955 -0.187769
## 83  -3.9399 -0.614020
## 84  -5.6055  0.340738
## 85  -0.2227  1.584673
## 86  -2.0060  0.905418
## 87  -1.1817  0.537570
## 88   9.4677 -1.825226
## 89   7.0620  0.663400
## 90  -9.1715  0.748255
## 91  -4.7645  2.155737
## 92   7.7019 -1.461721
## 93  -1.7502  0.821180
## 94  -1.9584  0.351564
## 95  -2.1036 -1.191568
```

```
## 96   7.6053  0.011634
## 97   6.5606  1.015164
## 98  -1.1938  2.634456
## 99  -0.6055  1.942980
## 100 -0.8987  0.904940
## 101  7.4898  0.265384
## 102  6.8132  0.670631
## 103 -6.2728 -1.649481
## 104 -1.4216  0.551245
## 105 -6.2282  0.712720
## 106 -1.8589 -0.319007
## 107  7.6882  0.009224
## 108  7.9179 -0.675121
## 109 -5.3607 -0.646121
## 110 -1.1581  2.643410
## 111 -6.3169  0.968981
## 112 -6.3277  1.383290
## 113  7.3431  0.947319
## 114 -2.1426 -0.088780
## 115 -2.4795  1.940739
## 116  8.3974 -0.647363
## 117 -1.3255  0.162870
## 118 -3.8353  1.405958
## 119 -2.2574  1.426794
## 120 -1.2557  0.546424
## 121  0.4760  0.799905
## 122 -1.5495  0.593364
## 123 -6.8506 -0.829825
## 124 -5.2038 -1.144768
## 125  8.5824 -1.834489
## 126  7.7808 -0.584339
## 127 -6.8528 -2.717590
## 128 -4.4407 -1.347237
## 129 -2.6661  0.642505
## 130 -0.3784 -0.086639
## 131 -2.4169  0.092784
## 132 -5.4501  0.207737
## 133 -5.6603 -0.832714
## 134 -1.9556  1.154348
## 135  7.2193  0.109646
## 136  7.3268  1.072989
## 137  5.6619  1.934355
## 138 -5.9582  0.094018
## 139  8.0784 -0.968581
## 140  8.0210 -1.140504
## 141 -2.2625  1.587253
## 142 -1.4376  0.134425
## 143 -2.4591  0.935277
## 144  7.4968  0.188377
## 145  7.5865 -1.207970
## 146  7.9246 -0.209639
## 147 -6.7593 -1.600232
## 148 -3.5185 -0.160589
## 149 -7.8395 -2.139733
```

```
## 150  8.3144 -0.644953
```

```
Iris_CDA$scores
```

```
##     as.factor(Species)    Can1      Can2
## 1               Setosa -7.6720  0.134894
## 2            Virginica  6.8002 -0.580895
## 3           Versicolor  2.5487  0.472205
## 4            Virginica  6.6531 -1.805320
## 5            Virginica  3.8152  0.942986
## 6               Setosa -7.2126 -0.355836
## 7            Virginica  5.1056 -1.992182
## 8           Versicolor  3.4981  1.684956
## 9           Versicolor  3.7159 -1.044514
## 10              Setosa -8.6810 -0.877590
## 11          Versicolor  2.2925  0.332860
## 12          Versicolor  4.4985  0.882750
## 13           Virginica  4.9677 -0.821141
## 14          Versicolor  1.0904  1.626583
## 15           Virginica  5.0660  0.026273
## 16           Virginica  5.5075  0.035814
## 17           Virginica  6.7960 -1.460687
## 18              Setosa -6.2514 -0.439696
## 19          Versicolor  2.4300  0.966132
## 20           Virginica  5.8861 -2.345091
## 21           Virginica  6.6088 -1.751636
## 22          Versicolor  2.4485 -0.795962
## 23           Virginica  6.8474 -2.428951
## 24           Virginica  7.4182  0.173118
## 25           Virginica  4.6780  0.499095
## 26              Setosa -8.6137 -0.403254
## 27           Virginica  5.6450 -1.677717
## 28          Versicolor  1.4593 -0.028544
## 29          Versicolor  1.7977 -0.484386
## 30          Versicolor  0.9976  0.490531
## 31              Setosa -6.7590  0.759003
## 32           Virginica  4.6832 -0.332034
## 33          Versicolor  1.1067  1.752254
## 34           Virginica  5.1796  0.363475
## 35           Virginica  5.8070 -2.010199
## 36              Setosa -7.9913 -0.086379
## 37              Setosa -8.3304 -0.228134
## 38          Versicolor  2.9340 -0.027379
## 39           Virginica  5.2205 -1.468195
## 40              Setosa -7.2410  0.272615
## 41           Virginica  5.7232 -1.293276
## 42              Setosa -6.4144 -1.247301
## 43          Versicolor  1.2723  1.214584
## 44              Setosa -6.9341  0.705519
## 45           Virginica  7.5812  0.980723
## 46           Virginica  4.3715  0.121297
## 47              Setosa -6.8295  0.544961
## 48          Versicolor  2.4016  1.594583
## 49          Versicolor  0.2932  1.798715
```

```
## 50          Virginica  5.2792  0.042458
## 51             Setosa -6.7647  0.505152
## 52             Setosa -8.0819 -0.763393
## 53          Virginica  4.0770 -0.523238
## 54             Setosa -6.5589  0.389223
## 55             Setosa -6.7714  0.970634
## 56             Setosa -7.9588  0.164962
## 57          Virginica  5.1075  2.130590
## 58          Virginica  6.5191 -0.296976
## 59             Setosa -7.5725  0.805464
## 60             Setosa -9.8498 -1.585937
## 61             Setosa -6.8594 -1.051654
## 62         Versicolor  1.2012 -0.084437
## 63          Virginica  4.0809 -0.185937
## 64             Setosa -7.1287  0.786660
## 65             Setosa -8.0618 -0.300421
## 66         Versicolor  2.7681 -0.032200
## 67         Versicolor  0.7769  1.659162
## 68             Setosa -6.8231 -0.463012
## 69             Setosa -7.1868  0.360987
## 70         Versicolor  2.5898  0.174612
## 71         Versicolor -0.3075  1.318871
## 72             Setosa -9.1582 -2.737596
## 73             Setosa -8.1323 -0.514463
## 74          Virginica  6.7967 -0.863090
## 75          Virginica  6.5245 -2.445035
## 76          Virginica  5.5075  0.035814
## 77         Versicolor  1.6162  0.470104
## 78          Virginica  4.5837  0.856816
## 79             Setosa -7.3750 -0.565845
## 80             Setosa -9.1263 -1.224433
## 81          Virginica  6.2920 -0.467176
## 82          Virginica  4.9955 -0.187769
## 83          Virginica  3.9399 -0.614020
## 84          Virginica  5.6055  0.340738
## 85         Versicolor  0.2227  1.584673
## 86         Versicolor  2.0060  0.905418
## 87         Versicolor  1.1817  0.537570
## 88             Setosa -9.4677 -1.825226
## 89             Setosa -7.0620  0.663400
## 90          Virginica  9.1715  0.748255
## 91          Virginica  4.7645  2.155737
## 92             Setosa -7.7019 -1.461721
## 93         Versicolor  1.7502  0.821180
## 94         Versicolor  1.9584  0.351564
## 95         Versicolor  2.1036 -1.191568
## 96             Setosa -7.6053  0.011634
## 97             Setosa -6.5606  1.015164
## 98         Versicolor  1.1938  2.634456
## 99         Versicolor  0.6055  1.942980
## 100        Versicolor  0.8987  0.904940
## 101            Setosa -7.4898  0.265384
## 102            Setosa -6.8132  0.670631
## 103         Virginica  6.2728 -1.649481
```

```
## 104         Versicolor  1.4216  0.551245
## 105          Virginica  6.2282  0.712720
## 106         Versicolor  1.8589 -0.319007
## 107             Setosa -7.6882  0.009224
## 108             Setosa -7.9179 -0.675121
## 109          Virginica  5.3607 -0.646121
## 110         Versicolor  1.1581  2.643410
## 111          Virginica  6.3169  0.968981
## 112          Virginica  6.3277  1.383290
## 113             Setosa -7.3431  0.947319
## 114         Versicolor  2.1426 -0.088780
## 115         Versicolor  2.4795  1.940739
## 116             Setosa -8.3974 -0.647363
## 117         Versicolor  1.3255  0.162870
## 118         Versicolor  3.8353  1.405958
## 119         Versicolor  2.2574  1.426794
## 120         Versicolor  1.2557  0.546424
## 121         Versicolor -0.4760  0.799905
## 122         Versicolor  1.5495  0.593364
## 123          Virginica  6.8506 -0.829825
## 124          Virginica  5.2038 -1.144768
## 125             Setosa -8.5824 -1.834489
## 126             Setosa -7.7808 -0.584339
## 127          Virginica  6.8528 -2.717590
## 128          Virginica  4.4407 -1.347237
## 129         Versicolor  2.6661  0.642505
## 130         Versicolor  0.3784 -0.086639
## 131         Versicolor  2.4169  0.092784
## 132          Virginica  5.4501  0.207737
## 133          Virginica  5.6603 -0.832714
## 134         Versicolor  1.9556  1.154348
## 135             Setosa -7.2193  0.109646
## 136             Setosa -7.3268  1.072989
## 137             Setosa -5.6619  1.934355
## 138          Virginica  5.9582  0.094018
## 139             Setosa -8.0784 -0.968581
## 140             Setosa -8.0210 -1.140504
## 141         Versicolor  2.2625  1.587253
## 142         Versicolor  1.4376  0.134425
## 143         Versicolor  2.4591  0.935277
## 144             Setosa -7.4968  0.188377
## 145             Setosa -7.5865 -1.207970
## 146             Setosa -7.9246 -0.209639
## 147          Virginica  6.7593 -1.600232
## 148         Versicolor  3.5185 -0.160589
## 149          Virginica  7.8395 -2.139733
## 150             Setosa -8.3144 -0.644953
```

```
Iris_CDA$means
```

```
##              Can1    Can2
## Setosa     -7.608 -0.2151
## Versicolor  1.825  0.7279
## Virginica   5.783 -0.5128
```

```
Iris_Fisher$error_rate

## [1] 0.02

Iris_Fisher$confusion

##              predicted
## original      Setosa Versicolor Virginica
##   Setosa          50          0         0
##   Versicolor       0         48         2
##   Virginica        0          1        49
```

2.9 Cross-Validation

The basic predict( ) function when applied to the entire data set will simply provide the classification from the model and the ability to determine the percentage correct classification. The jackknife procedure [use CV=TRUE in lda()] will leave each case out in turn, estimate the discriminant analysis with the remaining cases, and then use that information to classify the left out case. This approach insures that each case is classified with information it did not contribute to in the estimation. A traditional cross-validation uses part of the sample (or a separate sample) to estimate the discriminant functions and then applies that solution to the remaining cases. Each is illustrated here.

2.9.1 Simple Prediction

```
table(Original = Iris$Species_Num, Predicted = predict(Iris_LDA)$class)

##         Predicted
## Original  1  2  3
##        1 50  0  0
##        2  0 48  2
##        3  0  1 49

Proportion_of_Correct_Classification <- sum(diag(table(Original = Iris$Species_Num,
    Predicted = predict(Iris_LDA)$class)))/sum(table(Original = Iris$Species_Num,
    Predicted = predict(Iris_LDA)$class))
Proportion_of_Correct_Classification

## [1] 0.98
```

2.9.2 Leave-One-Out

```
Iris_Jack <- lda(Species ~ Sepal_Length + Sepal_Width + Petal_Length +
    Petal_Width, data = Iris, CV = TRUE)
table(Original = Iris$Species_Num, Predicted = Iris_Jack$class)

##         Predicted
## Original Setosa Versicolor Virginica
##        1     50          0         0
##        2      0         48         2
##        3      0          1        49
```

```r
Proportion_of_Correct_Classification <- sum(diag(table(Original = Iris$Species_Num,
    Predicted = Iris_Jack$class)))/sum(table(Original = Iris$Species_Num,
    Predicted = Iris_Jack$class))
Proportion_of_Correct_Classification

## [1] 0.98
```

2.9.3 Two-Sample Cross-Validation

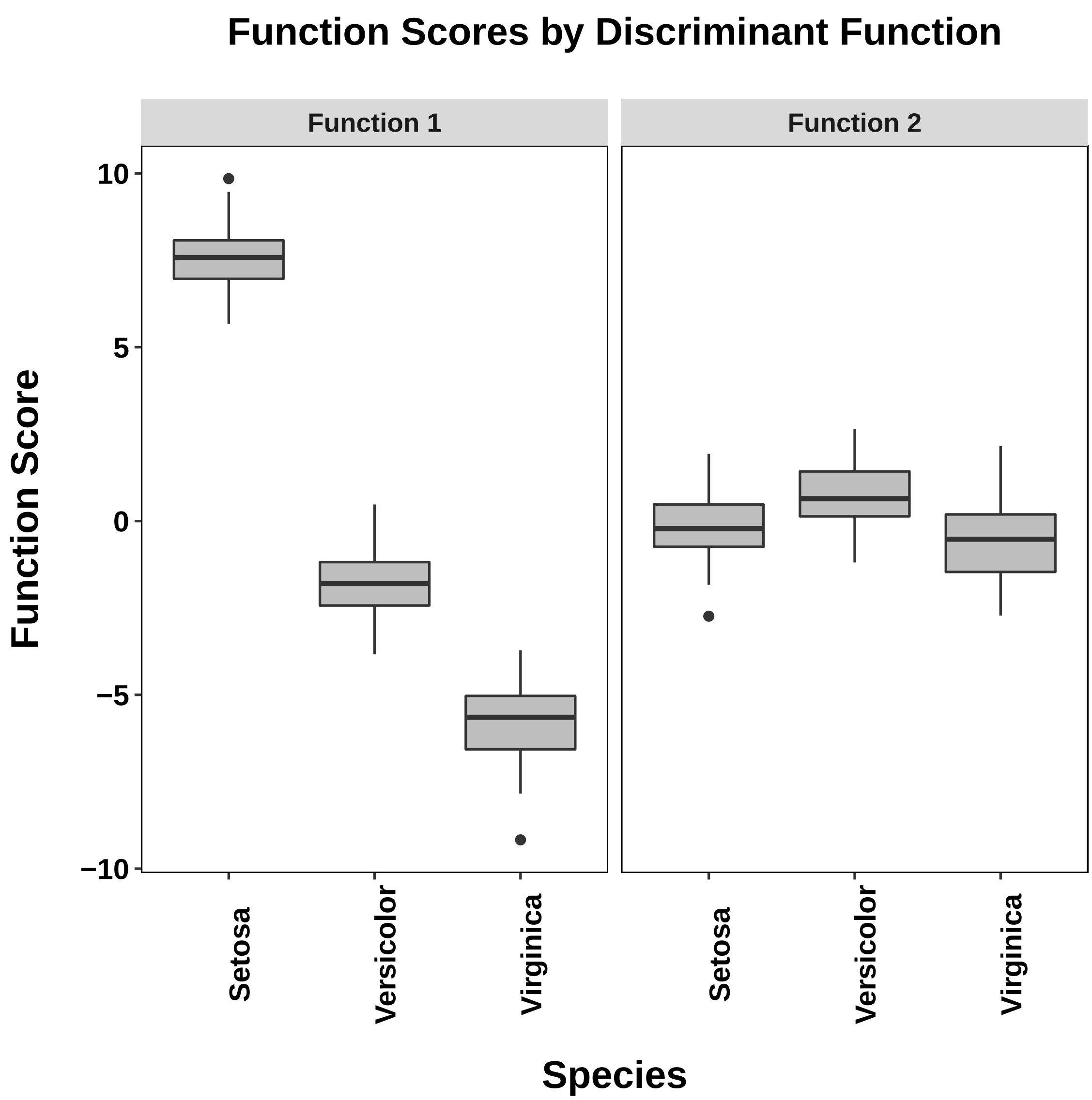
```r
training_sample <- sample(1:150, 75)

Iris_Train <- lda(Species ~ Sepal_Length + Sepal_Width + Petal_Length +
    Petal_Width, data = Iris, CV = FALSE, subset = training_sample)

Iris_Predict <- predict(Iris_Train, newdata = Iris[-training_sample,
    ])
Iris_Original <- as.data.frame(Iris[-training_sample, 5])
Iris_Cross <- cbind(Iris_Original, Iris_Predict$class)
names(Iris_Cross) <- c("Original_Species", "Predicted_Species")
table(Original = Iris_Cross$Original_Species, Predicted = Iris_Cross$Predicted_Species)

##              Predicted
## Original      Setosa Versicolor Virginica
##   Setosa          25          0         0
##   Versicolor       0         24         0
##   Virginica        0          1        25

Proportion_of_Correct_Classification <- sum(diag(table(Original = Iris_Cross$Original_Species,
    Predicted = Iris_Cross$Predicted_Species)))/sum(table(Original = Iris_Cross$Original_Species,
    Predicted = Iris_Cross$Predicted_Species))
Proportion_of_Correct_Classification

## [1] 0.9867
```

2.10 Visualization

```r
Iris_LDA_Values <- predict(Iris_LDA)
LDA_Means <- as.data.frame(aggregate(Iris_LDA_Values$x, by = list(Iris_LDA_Values$class),
    FUN = mean, na.rm = TRUE))
```

```r
plot_data <- rbind(cbind(Iris_LDA_Values$x[, 1], Iris_LDA_Values$class),
    cbind(Iris_LDA_Values$x[, 2], Iris_LDA_Values$class))
plot_data <- as.data.frame(plot_data)
names(plot_data) <- c("Score", "Class")
plot_data$Class_F <- factor(plot_data$Class, levels = c(1, 2, 3),
    labels = c("Setosa", "Versicolor", "Virginica"))
plot_data$Function <- c(rep("Function 1", 150), rep("Function 2",
    150))

p <- ggplot(plot_data, aes(x = Class_F, y = Score)) + geom_boxplot(fill = "gray") +
```

```
    ylab("Function Score") + xlab("Species") + theme(text = element_text(size = 14,
    family = "sans", color = "black", face = "bold"), axis.text.y = element_text(colour = "black",
    size = 12, face = "bold"), axis.text.x = element_text(colour = "black",
    size = 12, face = "bold", angle = 90), axis.title.x = element_text(margin = margin(15,
    0, 0, 0), size = 16), axis.title.y = element_text(margin = margin(0,
    15, 0, 0), size = 16), axis.line.x = element_blank(), axis.line.y = element_blank(),
    plot.title = element_text(size = 16, face = "bold", margin = margin(0,
        0, 20, 0), hjust = 0.5), panel.background = element_rect(fill = "white",
        linetype = 1, color = "black"), panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(), plot.background = element_rect(fill = "white"),
    plot.margin = unit(c(1, 1, 1, 1), "cm"), legend.position = "bottom",
    legend.title = element_blank()) + ggtitle("Function Scores by Discriminant Function")
p + facet_grid(~Function)
```

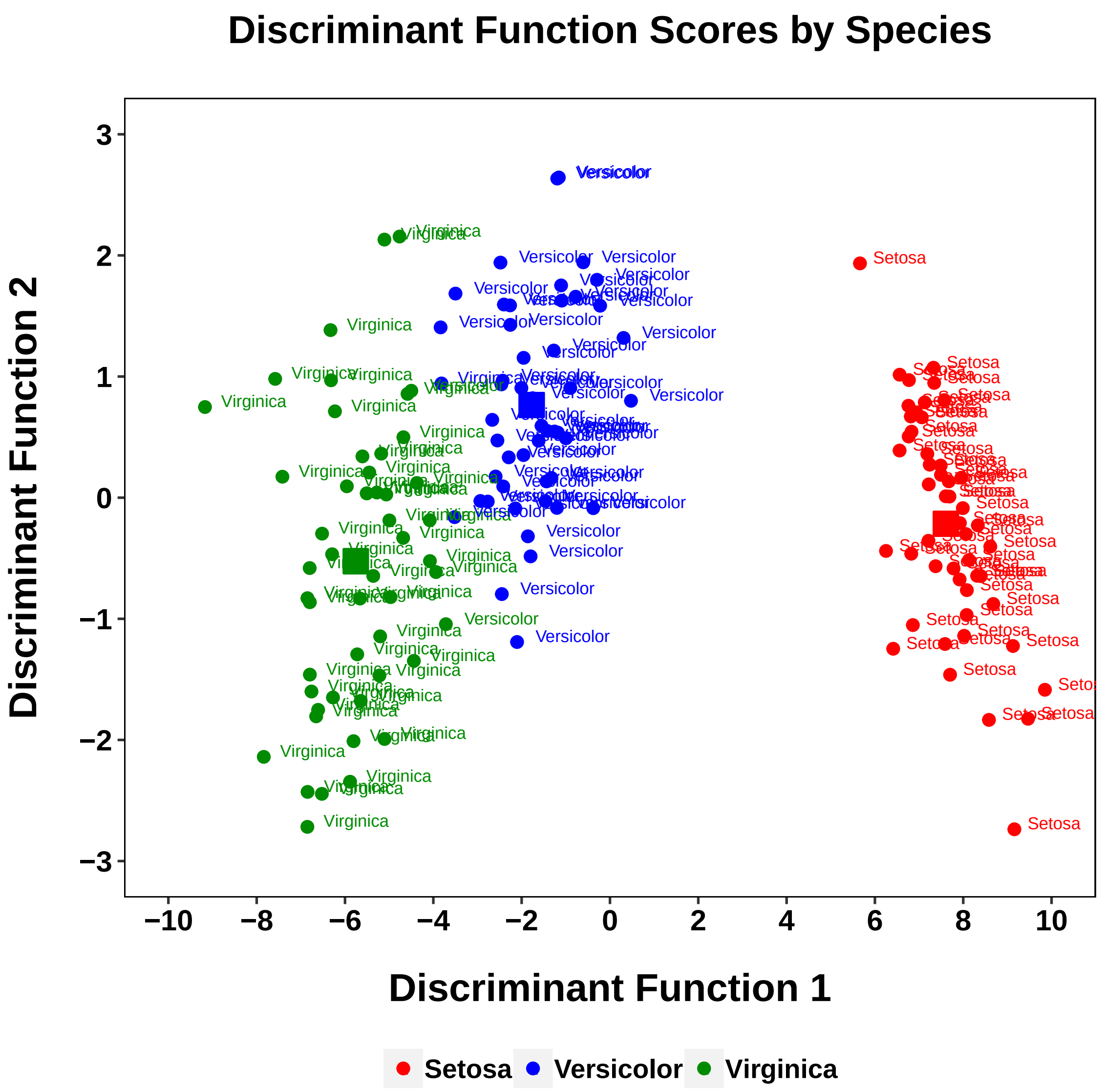


The different discrimination is evident when the discriminant function scores are exam-

ined in analyses of variance.

```
summary(aov(Iris_LDA_Values$x[, 1] ~ Iris$Species))

##                Df Sum Sq Mean Sq F value Pr(>F)
## Iris$Species    2   4732    2366    2366 <2e-16
## Residuals     147    147       1

summary(aov(Iris_LDA_Values$x[, 2] ~ Iris$Species))

##                Df Sum Sq Mean Sq F value  Pr(>F)
## Iris$Species    2     42      21      21 9.7e-09
## Residuals     147    147       1
```

```
plot_data <- cbind(Iris_LDA_Values$x[, 1], Iris_LDA_Values$x[, 2],
    Iris_LDA_Values$class, Iris$Species_Num)
plot_data <- as.data.frame(plot_data)
names(plot_data) <- c("DS1", "DS2", "Class", "Species")
plot_data$Class_F <- factor(plot_data$Class, levels = c(1, 2, 3),
    labels = c("Setosa", "Versicolor", "Virginica"))
plot_data$Species_F <- factor(plot_data$Species, levels = c(1, 2,
    3), labels = c("Setosa", "Versicolor", "Virginica"))

ggplot(plot_data, aes(x = DS1, y = DS2, color = Class_F)) + geom_point(shape = 19,
    size = 2, na.rm = TRUE) + geom_point(LDA_Means, mapping = aes(x = LDA_Means[1,
    2], y = LDA_Means[1, 3]), size = 5, color = "red", fill = "red",
    shape = 22) + geom_point(LDA_Means, mapping = aes(x = LDA_Means[2,
    2], y = LDA_Means[2, 3]), size = 5, color = "blue", fill = "blue",
    shape = 22) + geom_point(LDA_Means, mapping = aes(x = LDA_Means[3,
    2], y = LDA_Means[3, 3]), size = 5, color = "green4", fill = "green4",
    shape = 22) + scale_y_continuous(breaks = c(seq(-3, 3, 1))) +
    scale_x_continuous(breaks = c(round(seq(-10, 10, 2), 2))) + scale_color_manual(values = c("red",
    "blue", "green4")) + geom_text(aes(label = Species_F), hjust = -0.25,
    vjust = 0, size = 2.5) + coord_cartesian(xlim = c(-10, 10), ylim = c(-3,
    3)) + xlab("Discriminant Function 1") + ylab("Discriminant Function 2") +
    theme(text = element_text(size = 14, family = "sans", color = "black",
        face = "bold"), axis.text.y = element_text(colour = "black",
        size = 12, face = "bold"), axis.text.x = element_text(colour = "black",
        size = 12, face = "bold", angle = 0), axis.title.x = element_text(margin = margin(15,
        0, 0, 0), size = 16), axis.title.y = element_text(margin = margin(0,
        15, 0, 0), size = 16), axis.line.x = element_blank(), axis.line.y = element_blank(),
        plot.title = element_text(size = 16, face = "bold", margin = margin(0,
            0, 20, 0), hjust = 0.5), panel.background = element_rect(fill = "white",
            linetype = 1, color = "black"), panel.grid.major = element_blank(),
        panel.grid.minor = element_blank(), plot.background = element_rect(fill = "white"),
        plot.margin = unit(c(1, 1, 1, 1), "cm"), legend.position = "bottom",
        legend.title = element_blank()) + ggtitle("Discriminant Function Scores by Species")
```




```
Sys.time() - how_long

## Time difference of 15.84 secs
```