

Chapter 15

Multivariate Analysis of Variance

Jolicouer and Mosimann studied the relationship between the size and shape of painted turtles. The table below gives the length, width, and height (all in mm) for 24 males and 24 females.

```
#### Example: Painted turtle shells
fn.data <- "http://statacumen.com/teach/ADA2/ADA2_notes_Ch15_shells_mf.dat"
shells <- read.table(fn.data, header = TRUE)
str(shells)

## 'data.frame': 48 obs. of 4 variables:
## $ sex : Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
## $ length: int 98 103 103 105 109 123 123 133 133 133 ...
## $ width : int 81 84 86 86 88 92 95 99 102 102 ...
## $ height: int 38 38 42 42 44 50 46 51 51 51 ...

#head(shells)
```

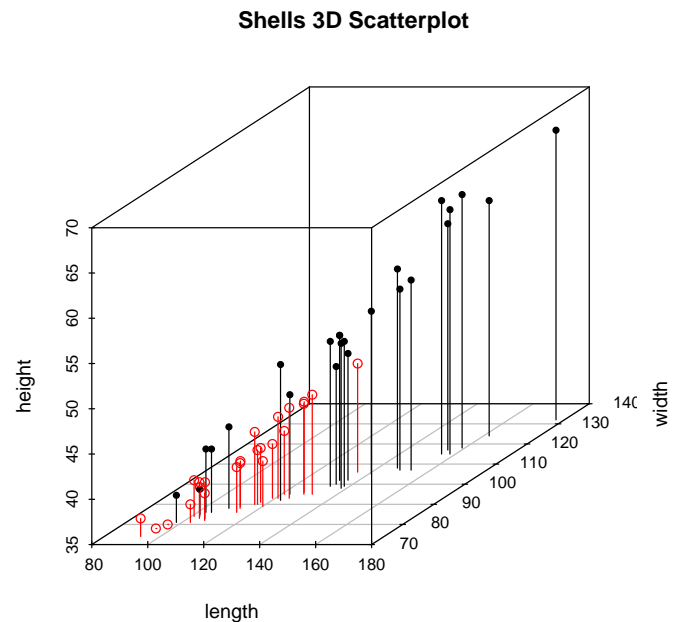
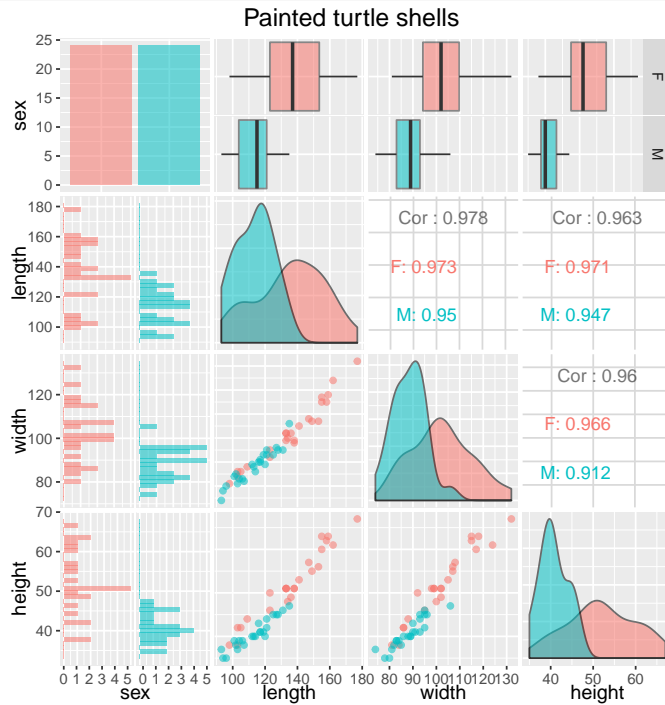
	sex	length	width	height		sex	length	width	height
1	F	98	81	38	25	M	93	74	37
2	F	103	84	38	26	M	94	78	35
3	F	103	86	42	27	M	96	80	35
4	F	105	86	42	28	M	101	84	39
5	F	109	88	44	29	M	102	85	38
6	F	123	92	50	30	M	103	81	37
7	F	123	95	46	31	M	104	83	39
8	F	133	99	51	32	M	106	83	39
9	F	133	102	51	33	M	107	82	38
10	F	133	102	51	34	M	112	89	40
11	F	134	100	48	35	M	113	88	40
12	F	136	102	49	36	M	114	86	40
13	F	138	98	51	37	M	116	90	43
14	F	138	99	51	38	M	117	90	41
15	F	141	105	53	39	M	117	91	41
16	F	147	108	57	40	M	119	93	41
17	F	149	107	55	41	M	120	89	40
18	F	153	107	56	42	M	121	93	44
19	F	155	115	63	43	M	121	95	42
20	F	155	117	60	44	M	125	93	45
21	F	158	115	62	45	M	127	96	45
22	F	159	118	63	46	M	128	95	45
23	F	162	124	61	47	M	131	95	46
24	F	177	132	67	48	M	135	106	47

```
## Scatterplot matrix
library(ggplot2)
#suppressMessages(suppressWarnings(library(GGally)))
library(GGally)
# color by sex
p <- ggpairs(shells
  , mapping = ggplot2::aes(colour = sex, alpha = 0.5)
  , title = "Painted turtle shells")
print(p)
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
# detach package after use so reshape2 works (old reshape (v.1) conflicts)
#detach("package:GGally", unload=TRUE)
#detach("package:reshape", unload=TRUE)

## 3D scatterplot
library(scatterplot3d)
with(shells, {
  scatterplot3d(x = length
    , y = width
    , z = height
    , main = "Shells 3D Scatterplot"
    , type = "h" # lines to the horizontal xy-plane
    , color = as.integer(sex) # color by group
    , pch = as.integer(sex)+19 # plotting character by group
    #, highlight.3d = TRUE # makes color change with z-axis value
    , angle = 40 # viewing angle (seems hard to control)
  )
})
```

```
})
```

```
#### Try this!
#### For a rotatable 3D plot, use plot3d() from the rgl library
### This uses the R version of the OpenGL (Open Graphics Library)
# library(rgl)
# with(shells, { plot3d(x = length, y = width, z = height, col = sex) })
```



MANOVA considers the following two questions:

- Are the population mean length, width, and height the same for males and females?
- If not, then what combination of features is most responsible for the differences?

To describe MANOVA, suppose you measure p features on independent random samples from k strata, groups, or populations. Let

$$\mu'_i = [\mu_{i1} \ \mu_{i2} \ \cdots \ \mu_{ip}]'$$

be the vector of population means for the i^{th} population, where μ_{ij} is the i^{th} population mean on the j^{th} feature. For the turtles, $p = 3$ features and $k = 2$ strata (sexes).

A one-way MANOVA tests the hypothesis that the population **mean vectors** are identical: $H_0 : \mu_1 = \mu_2 = \cdots = \mu_k$ against $H_A : \text{not } H_0$. For the

carapace data, you are simultaneously testing that the sexes have equal population mean lengths, equal population mean widths, and equal population mean heights.

Assume that the sample sizes from the different groups are n_1, n_2, \dots, n_k . The total sample size is $n = n_1 + n_2 + \dots + n_k$. Let

$$X'_{ij} = [X_{ij1} \ X_{ij2} \ \dots \ X_{ijp}]'$$

be the vector of responses for the j^{th} individual from the i^{th} sample. Let

$$\bar{X}'_i = [\bar{X}_{i1} \ \bar{X}_{i2} \ \dots \ \bar{X}_{ip}]'$$

and S_i be the mean vector and variance-covariance matrix for the i^{th} sample. Finally, let

$$\bar{X}' = [\bar{X}_1 \ \bar{X}_2 \ \dots \ \bar{X}_p]'$$

be the vector of means ignoring samples (combine all the data across samples and compute the average on each feature), and let

$$S = \frac{\sum_i (n_i - 1) S_i}{n - k}$$

be the pooled variance-covariance matrix. The pooled variance-covariance matrix is a weighted average of the variance-covariance matrices from each group.

To test H_0 , construct the following MANOVA table, which is the multivariate analog of the ANOVA table:

Source	df	SS	MS
Between	$k - 1$	$\sum_i n_i (\bar{X}_i - \bar{X})(\bar{X}_i - \bar{X})'$	
Within	$n - k$	$\sum_i (n_i - 1) S_i$	
Total	$n - 1$	$\sum_{ij} (X_{ij} - \bar{X})(X_{ij} - \bar{X})'$	

where all the MSs are SS/df.

The expressions for the SS have the same form as SS in univariate analysis of variance, except that each SS is a $p \times p$ symmetric matrix. The diagonal elements of the **SS matrices** are the SS for one-way ANOVAs on the individual

features. The off-diagonal elements are SS between features. The Error MS matrix is the pooled variance-covariance matrix S .

The standard MANOVA assumes that you have independent samples from multivariate normal populations with identical variance-covariance matrices. This implies that each feature is normally distributed in each population, that a feature has the same variability across populations, and that the correlation (or covariance) between two features is identical across populations. The Error MS matrix estimates the common population variance-covariance matrix when the population variance-covariance matrices are identical.

The H_0 of equal population mean vectors should be rejected when the difference among mean vectors, as measured by the Between MS matrix, is large relative to the variability within groups, as measured by the Error MS matrix. Equivalently, H_0 is implausible if a significant portion of the total variation in the data, as measured by the Total SS matrix, is due to differences among the groups. The same idea is used in a one-way ANOVA to motivate the F -test of no differences in population means. However, some care is needed to quantify these ideas in a MANOVA because there are several natural matrix definitions for comparing the Between MS matrix to the Error MS matrix. As a result, several MANOVA tests of H_0 have been proposed.

Graphical summaries for the carapace data are given above, with numerical summaries below.

```
# summary statistics for each sex
by(shells, shells$sex, summary)

## shells$sex: F
## sex      length      width      height
## F:24  Min.    : 98.0  Min.    : 81.00  Min.    :38.00
## M: 0   1st Qu.:123.0  1st Qu.: 94.25  1st Qu.:47.50
##       Median  :137.0  Median  :102.00  Median  :51.00
##       Mean    :136.0  Mean    :102.58  Mean    :52.04
##       3rd Qu.:153.5  3rd Qu.:109.75  3rd Qu.:57.75
##       Max.    :177.0  Max.    :132.00  Max.    :67.00
## -----
## shells$sex: M
## sex      length      width      height
## F: 0   Min.    : 93.0  Min.    : 74.00  Min.    :35.00
## M:24  1st Qu.:103.8  1st Qu.: 83.00  1st Qu.:38.75
##       Median  :115.0  Median  : 89.00  Median  :40.00
```

```

##           Mean    :113.4   Mean    : 88.29   Mean    :40.71
##           3rd Qu.:121.0   3rd Qu.: 93.00   3rd Qu.:43.25
##           Max.    :135.0   Max.    :106.00   Max.    :47.00

# standard deviations
by(shells[, 2:4], shells$sex, apply, 2, sd)

## shells$sex: F
##   length    width    height
## 21.24900 13.10465  8.04595
## -----
## shells$sex: M
##   length    width    height
## 11.806103  7.074013  3.355452

# correlation matrix (excluding associated p-values testing "H0: rho == 0")
library(Hmisc)
rcorr(as.matrix(shells[, 2:4]))[[1]]           # all
##           length    width    height
## length 1.0000000 0.9778962 0.9628010
## width  0.9778962 1.0000000 0.9599055
## height 0.9628010 0.9599055 1.0000000

rcorr(as.matrix(shells[shells$sex == "F", 2:4]))[[1]] # females
##           length    width    height
## length 1.0000000 0.9731162 0.9706748
## width  0.9731162 1.0000000 0.9659029
## height 0.9706748 0.9659029 1.0000000

rcorr(as.matrix(shells[shells$sex == "M", 2:4]))[[1]] # males
##           length    width    height
## length 1.0000000 0.9501287 0.9470730
## width  0.9501287 1.0000000 0.9122648
## height 0.9470730 0.9122648 1.0000000

```

The features are positively correlated within each sex. The correlations between pairs of features are similar for males and females. Females tend to be larger on each feature. The distributions for length, width, and height are fairly symmetric within sexes. No outliers are present. Although females are more variable on each feature than males, the MANOVA assumptions do not appear to be grossly violated here. (Additionally, you could consider transforming the each dimension of the data in hopes to make the covariances between sexes more similar, though it may not be easy to find a good transformation to use.)

```

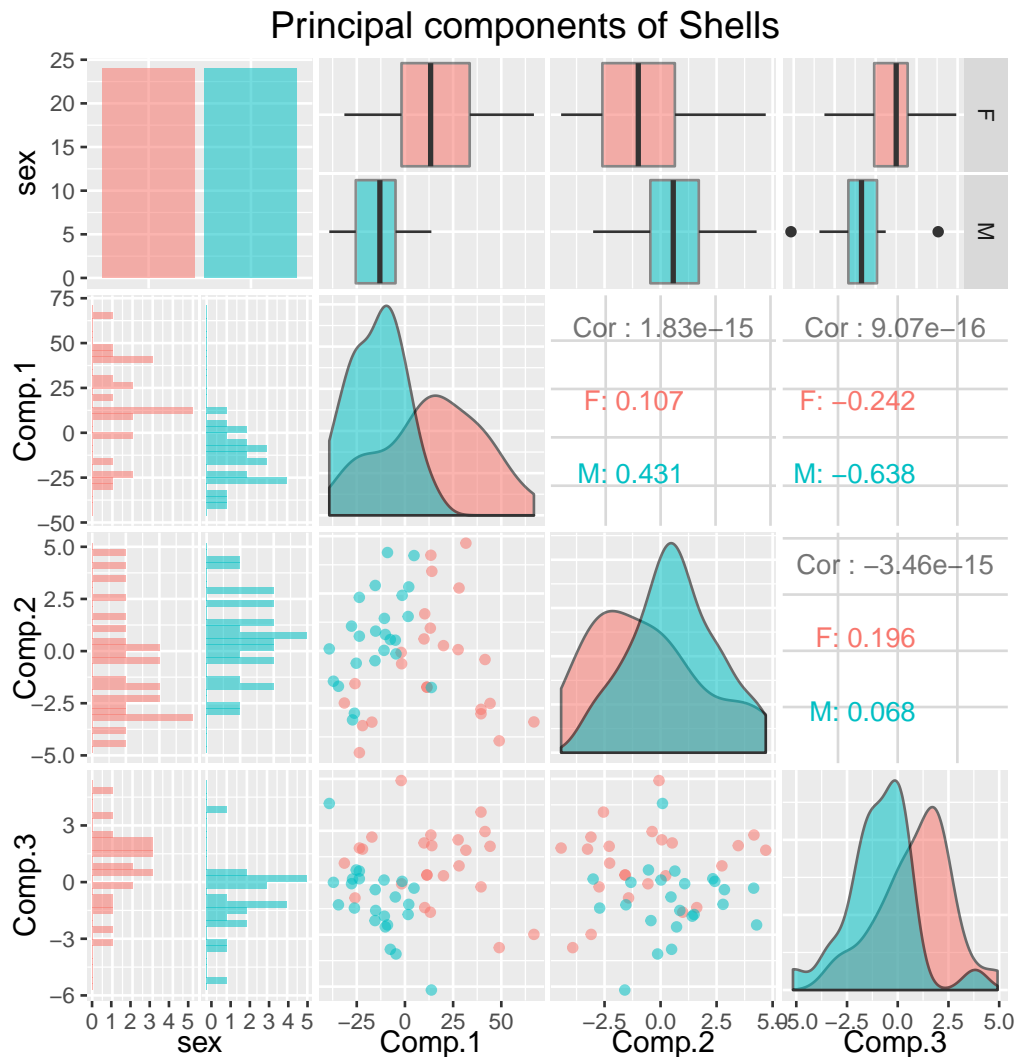
pca.sh <- princomp(shells[, 2:4])
df.pca.sh <- data.frame(sex = shells$sex, pca.sh$scores)
str(df.pca.sh)

```

```
## 'data.frame': 48 obs. of 4 variables:
## $ sex : Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
## $ Comp.1: num -31.4 -25.9 -23.6 -22 -17.1 ...
## $ Comp.2: num -2.27 -1.43 -4.44 -3.26 -3.11 ...
## $ Comp.3: num 0.943 -0.73 1.671 1.618 2.2 ...

## Scatterplot matrix
library(ggplot2)
#suppressMessages(suppressWarnings(library(GGally)))
library(GGally)
# put scatterplots on top so y axis is vertical
p <- ggpairs(df.pca.sh
, mapping = ggplot2::aes(colour = sex, alpha = 0.5)
, title = "Principal components of Shells")
print(p)

## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
# detach package after use so reshape2 works (old reshape (v.1) conflicts)
#detach("package:GGally", unload=TRUE)
#detach("package:reshape", unload=TRUE)
```



For comparison with the MANOVA below, here are the univariate ANOVAs for each feature. For the carapace data, the univariate ANOVAs indicate significant differences between sexes on length, width, and height. Females are larger on average than males on each feature.

```
# Univariate ANOVA tests, by each response variable
lm.sh <- lm(cbind(length, width, height) ~ sex, data = shells)
summary(lm.sh)

## Response length :
##
## Call:
## lm(formula = length ~ sex, data = shells)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -38.042 -10.667  1.271  11.927  40.958
##
## Coefficients:
```



```

##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 136.042      3.509  38.77 < 2e-16 ***
## sexM        -22.625      4.962  -4.56 3.79e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17.19 on 46 degrees of freedom
## Multiple R-squared:  0.3113, Adjusted R-squared:  0.2963
## F-statistic: 20.79 on 1 and 46 DF,  p-value: 3.788e-05
##
##
## Response width :
##
## Call:
## lm(formula = width ~ sex, data = shells)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -21.5833  -5.5417  -0.4375   4.8854  29.4167
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 102.583      2.149  47.725 < 2e-16 ***
## sexM        -14.292      3.040  -4.701 2.38e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10.53 on 46 degrees of freedom
## Multiple R-squared:  0.3246, Adjusted R-squared:  0.3099
## F-statistic: 22.1 on 1 and 46 DF,  p-value: 2.376e-05
##
##
## Response height :
##
## Call:
## lm(formula = height ~ sex, data = shells)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.0417  -2.7917  -0.7083   4.0417  14.9583
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept)  52.042      1.258  41.360 < 2e-16 ***
## sexM        -11.333      1.779  -6.369 8.09e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##

```

```
## Residual standard error: 6.164 on 46 degrees of freedom
## Multiple R-squared:  0.4686, Adjusted R-squared:  0.457
## F-statistic: 40.56 on 1 and 46 DF,  p-value: 8.087e-08
# Alternatively, for many ANOVAs at once, it may be easier
# to select by column number, but you won't get the column names in the output.
# Also, the left-hand side needs to be a matrix data type.
# lm.sh <- lm(as.matrix(shells[, 2:4]) ~ shells[, 1])
```

A few procedures can be used for one-way MANOVA; two are `manova()` and the `car` package's `Manova()`. First we check the assumption of multivariate normality.

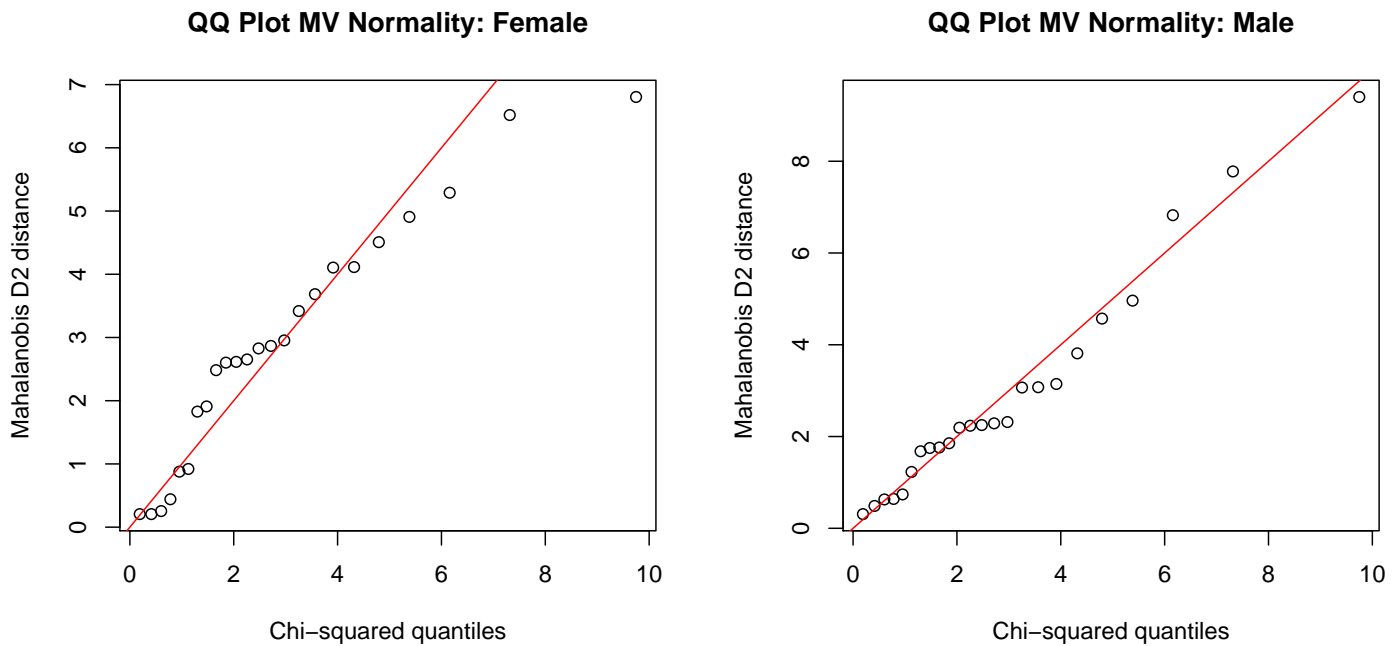
```
# Test multivariate normality using the Shapiro-Wilk test for multivariate normality
library(mvnormtest)
# The data needs to be transposed t() so each variable is a row
# with observations as columns.

mshapiro.test(t(shells[shells$sex == "F", 2:4]))
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.89324, p-value = 0.01551
mshapiro.test(t(shells[shells$sex == "M", 2:4]))
##
##  Shapiro-Wilk normality test
##
## data:  Z
## W = 0.93602, p-value = 0.1329
# Graphical Assessment of Multivariate Normality
f.mnv.norm.qqplot <- function(x, name = "") {
  # creates a QQ-plot for assessing multivariate normality

  x <- as.matrix(x)          # n x p numeric matrix
  center <- colMeans(x)     # centroid
  n <- nrow(x);
  p <- ncol(x);
  cov <- cov(x);
  d <- mahalanobis(x, center, cov) # distances
  qqplot(qchisq(ppoints(n), df=p), d
    , main=paste("QQ Plot MV Normality:", name)
    , ylab="Mahalanobis D2 distance"
    , xlab="Chi-squared quantiles")
  abline(a = 0, b = 1, col = "red")
}

f.mnv.norm.qqplot(shells[shells$sex == "F", 2:4], "Female")
```

```
f.mnv.norm.qqplot(shells[shells$sex == "M", 2:4], "Male")
```



The curvature in the Female sample cause us to reject normality, while the males do not deviate from normality. We'll proceed anyway since this deviation from normality in the female sample will largely increase the variability of the sample and not displace the mean greatly, and the sample sizes are somewhat large.

Multivariate test statistics These four multivariate test statistics are among the most common to assess differences across the levels of the categorical variables for a linear combination of responses. In general Wilks' lambda is recommended unless there are problems with small total sample size, unequal sample sizes between groups, violations of assumptions, etc., in which case Pillai's trace is more robust.

Wilks' lambda, (λ)

- Most commonly used statistic for overall significance
- Considers differences over all the characteristic roots
- The smaller the value of Wilks' lambda, the larger the between-groups dispersion

Pillai's trace

- Considers differences over all the characteristic roots
- More robust than Wilks'; should be used when sample size decreases, unequal cell sizes or homogeneity of covariances is violated

Hotelling's trace

- Considers differences over all the characteristic roots

Roy's greatest characteristic root

- Tests for differences on only the first discriminant function (Chapter 16)
- Most appropriate when responses are strongly interrelated on a single dimension
- Highly sensitive to violation of assumptions, but most powerful when all assumptions are met.

```
# Multivariate MANOVA test
# the specific test is specified in summary()
# test = c("Pillai", "Wilks", "Hotelling-Lawley", "Roy")
man.sh <- manova(cbind(length, width, height) ~ sex, data = shells)
summary(man.sh, test="Wilks")

##           Df  Wilks approx F num Df den Df   Pr(>F)
## sex         1 0.38695   23.237     3    44 3.622e-09 ***
## Residuals 46
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# I prefer the output from the car package
library(car)
lm.man <- lm(cbind(length, width, height) ~ sex, data = shells)
man.sh <- Manova(lm.man)
summary(man.sh)

##
## Type II MANOVA Tests:
##
## Sum of squares and products for error:
##           length  width  height
## length 13590.792 8057.500 4679.875
## width   8057.500 5100.792 2840.458
## height  4679.875 2840.458 1747.917
##
## -----
##
## Term: sex
##
## Sum of squares and products for the hypothesis:
##           length  width  height
## length  6142.688 3880.188 3077.000
```

```
## width 3880.188 2451.021 1943.667
## height 3077.000 1943.667 1541.333
##
## Multivariate Tests: sex
##           Df test stat approx F num Df den Df Pr(>F)
## Pillai    1 0.6130506 23.23665      3   44 3.622e-09 ***
## Wilks     1 0.3869494 23.23665      3   44 3.622e-09 ***
## Hotelling-Lawley 1 1.5843173 23.23665      3   44 3.622e-09 ***
## Roy       1 1.5843173 23.23665      3   44 3.622e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The four MANOVA tests of no differences between sexes are all highly significant. These tests reinforce the univariate analyses. Of the four tests, I prefer Roy's test because it has an intuitive interpretation. I will mostly ignore the other three tests for discussion.

Roy's test locates the linear combination of the features that produces the most significant one-way ANOVA test for no differences among groups. If the groups are not significantly different on the linear combination that best separates the groups in a one-way ANOVA sense, then there is no evidence that the population mean vectors are different. The critical value for Roy's test accounts for the linear combination being suggested by the data. That is, the critical value for Roy's test is not the same critical value that is used in a one-way ANOVA. The idea is similar to a Bonferroni-type correction with multiple comparisons.

Roy's method has the ability to locate linear combinations of the features on which the groups differ, even when the differences across groups are not significant on any feature. This is a reason for treating multivariate problems using multivariate methods rather than through individual univariate analyses on each feature.

```
## For Roy's characteristic Root and vector
#str(man.sh)
H <- man.sh$SSP$sex # H = hypothesis matrix
#   man.sh$df      #   hypothesis df
E <- man.sh$SSE     # E = error matrix
#   man.sh$error.df #   error df

# characteristic roots of (E inverse * H)
```

```

EinvH <- solve(E) %*% H # solve() computes the matrix inverse
ev <- eigen(EinvH)      # eigenvalue/eigenvectors
ev
## $values
## [1]  1.584317e+00  1.401030e-15 -1.315838e-15
##
## $vectors
##          [,1]      [,2]      [,3]
## [1,]  0.2151877  0.3563618  0.02975353
## [2,]  0.1014317  0.1651253  0.59173018
## [3,] -0.9712908 -0.9196412 -0.80558682

```

The **characteristic vector** (eigenvector) output gives one linear combination of the features for each variable in the data set. By construction, the linear combinations are uncorrelated (adjusting for groups). In general, the first $a = \text{minimum}(p, k - 1)$ linear combinations contain information in decreasing amounts for distinguishing among the groups. The first linear combination is used by Roy's test.

The three linear combinations for the carapace data are (reading down the columns in the matrix of eigenvectors)

$$D1 = 0.2152 \text{ Length} + 0.1014 \text{ Width} + -0.9713 \text{ Height}$$

$$D2 = 0.3564 \text{ Length} + 0.1651 \text{ Width} + -0.9196 \text{ Height}$$

$$D3 = 0.02975 \text{ Length} + 0.5917 \text{ Width} + -0.8056 \text{ Height.}$$

Here $p = 3$ and $k = 2$ gives $a = \min(p, k - 1) = \min(3, 2 - 1) = 1$ so only D1 from (D1, D2, and D3) contains information for distinguishing between male and female painted turtles.

As in PCA, the linear combinations should be interpreted. However, do not discount the contribution of a feature with a small loading (though, in the shells example, D2 and D3 have 0 for loadings). In particular, the most important feature for distinguishing among the groups might have a small loading because of the measurement scale.

The following output shows the differences between sexes on D1. The separation between sexes on D1 is greater than on any single feature. Females typically have much larger D1 scores than males.

Since the matrix of eigenvectors are a rotation matrix¹ we can create the D linear combinations by matrix multiplication of the eigenvector (rotation) matrix with the original data (being careful about dimensions).

```
# linear combinations of features
D <- as.matrix(shells[,2:4]) %*% ev$vector
colnames(D) <- c("D1", "D2", "D3")
df.D <- data.frame(sex = shells$sex, D)
str(df.D)

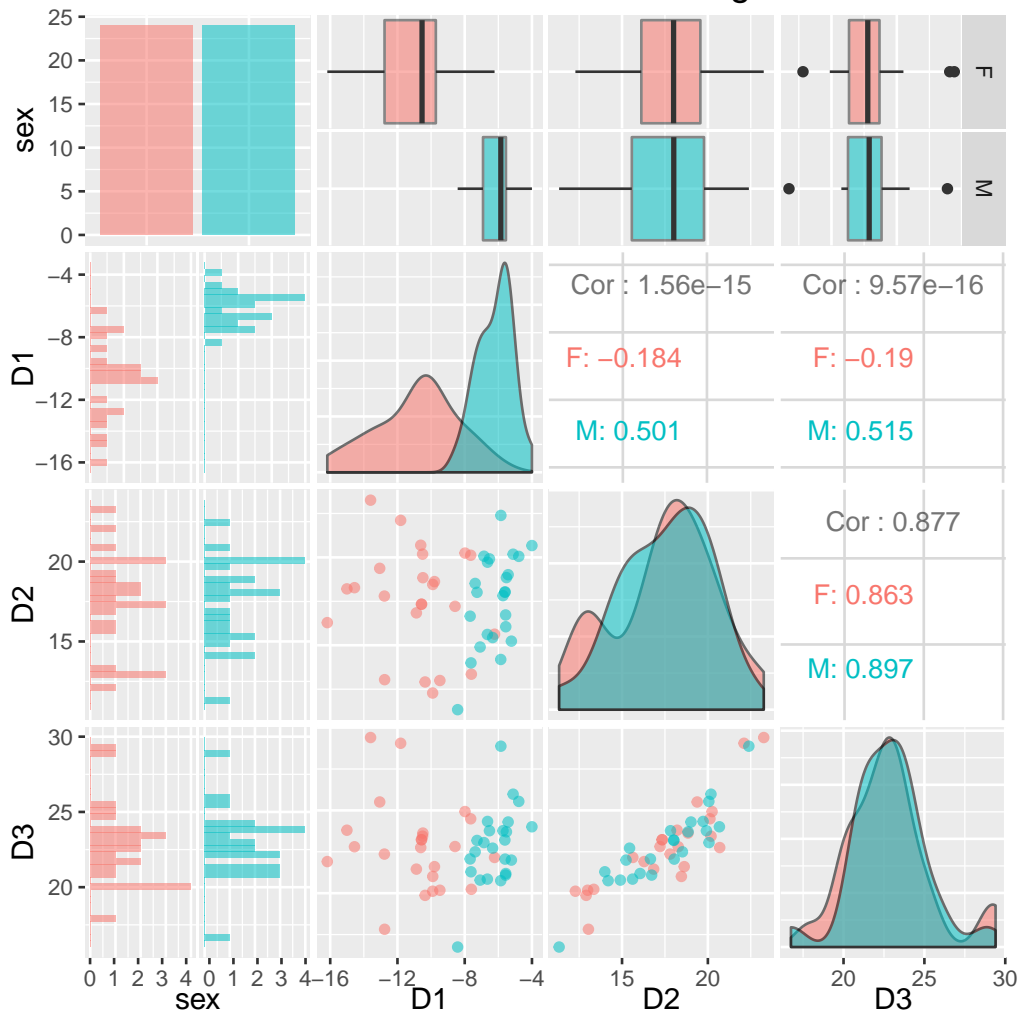
## 'data.frame': 48 obs. of 4 variables:
## $ sex: Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
## $ D1 : num -7.6 -6.22 -9.91 -9.48 -10.36 ...
## $ D2 : num 13.4 15.6 12.3 13 12.9 ...
## $ D3 : num 20.2 22.2 20.1 20.2 19.9 ...

## Scatterplot matrix
library(ggplot2)
#suppressMessages(suppressWarnings(library(GGally)))
library(GGally)
# put scatterplots on top so y axis is vertical
p <- ggpairs(df.D
  , mapping = ggplot2::aes(colour = sex, alpha = 0.5)
  , title = "D1 is the linear combination that best distinguishes the sexes")
print(p)

## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
# detach package after use so reshape2 works (old reshape (v.1) conflicts)
#detach("package:GGally", unload=TRUE)
#detach("package:reshape", unload=TRUE)
```

¹http://en.wikipedia.org/wiki/Rotation_matrix

D1 is the linear combination that best distinguishes the sexes



```
# Univariate ANOVA tests, by D1 linear combination variable
```

```
lm.D.sh <- lm(D1 ~ sex, data = df.D)
```

```
summary(lm.D.sh)
```

```
##
```

```
## Call:
```

```
## lm(formula = D1 ~ sex, data = df.D)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max
```

```
## -5.3047 -0.9791  0.3260  0.9570  4.6434
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept) -10.8679    0.3884 -27.978 < 2e-16 ***
```

```
## sexM         4.6897    0.5493  8.537 4.84e-11 ***
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 1.903 on 46 degrees of freedom
```

```
## Multiple R-squared:  0.6131, Adjusted R-squared:  0.6046
```



```
## F-statistic: 72.88 on 1 and 46 DF, p-value: 4.841e-11
```