Chapter 6: Multivariate Analysis of Variance

DA 410

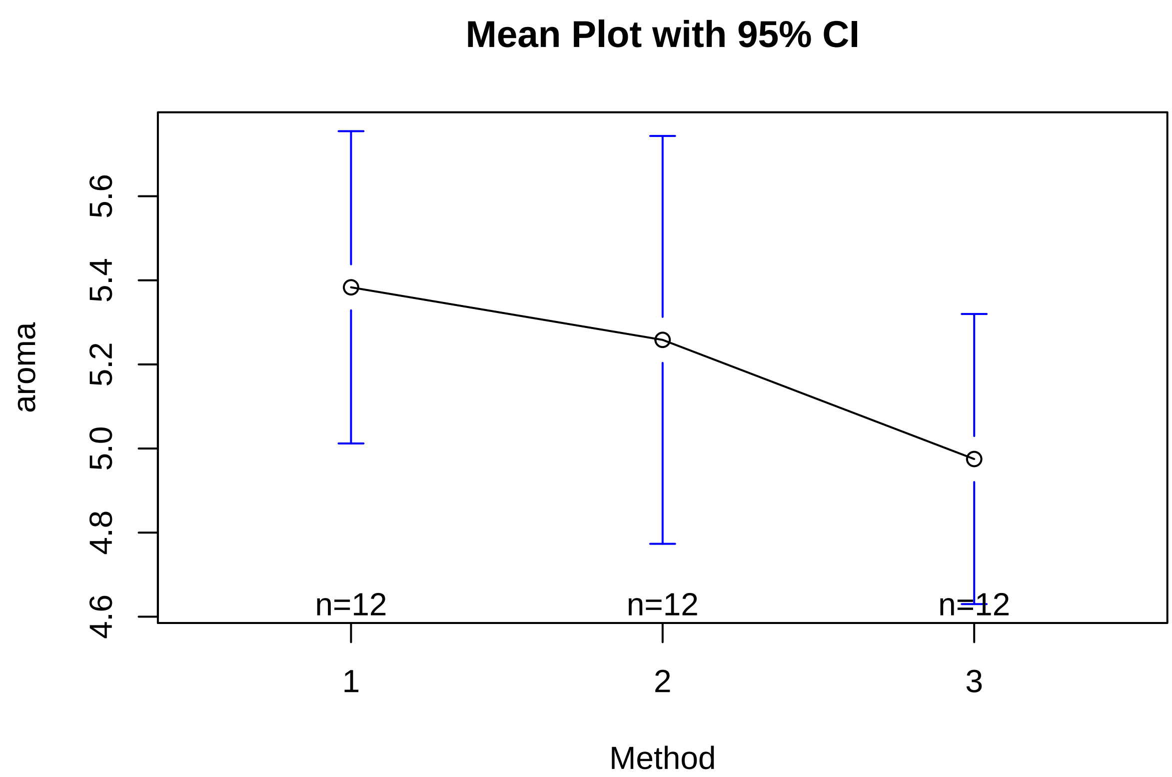
Marjorie Blanco

Problem 6.27

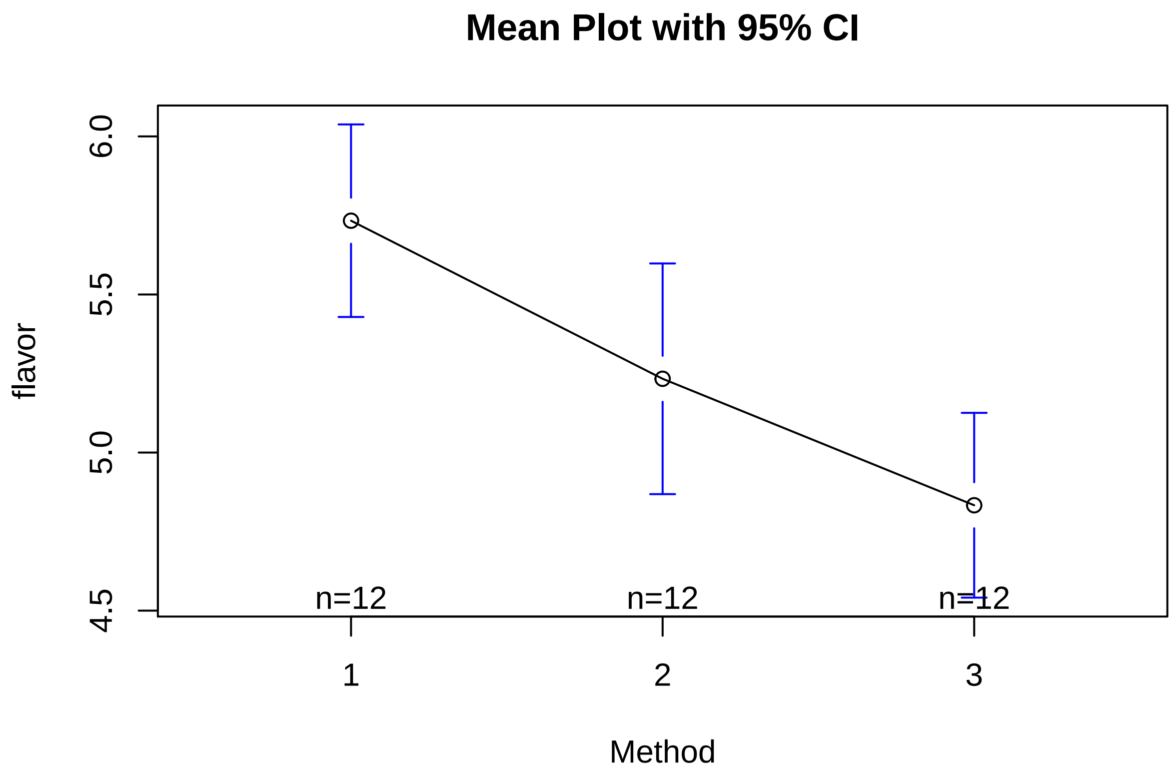
Baten, Tack, and Baeder (1958) compared judges' scores on fish prepared by three methods. Twelve fish were cooked by each method, and several judges tasted fish samples and rated each on four variables: y_1 = aroma, y_2 = flavor, y_3 = texture, and y_4 = moisture. The data are in Table 6.17. Each entry is an average score for the judges on that fish.

Method	y1	y2	y3	y4
1	5.4	6.0	6.3	6.7
1	5.2	6.5	6.0	5.8
1	6.1	5.9	6.0	7.0
1	4.8	5.0	4.9	5.0
1	5.0	5.7	5.0	6.5
1	5.7	6.1	6.0	6.6
1	6.0	6.0	5.8	6.0
1	4.0	5.0	4.0	5.0
1	5.7	5.4	4.9	5.0
1	5.6	5.2	5.4	5.8
1	5.8	6.1	5.2	6.4
1	5.3	5.9	5.8	6.0
2	5.0	5.3	5.3	6.5
2	4.8	4.9	4.2	5.6
2	3.9	4.0	4.4	5.0
2	4.0	5.1	4.8	5.8
2	5.6	5.4	5.1	6.2
2	6.0	5.5	5.7	6.0
2	5.2	4.8	5.4	6.0
2	5.3	5.1	5.8	6.4
2	5.9	6.1	5.7	6.0
2	6.1	6.0	6.1	6.2
2	6.2	5.7	5.9	6.0
2	5.1	4.9	5.3	4.8
3	4.8	5.0	6.5	7.0
3	5.4	5.0	6.0	6.4
3	4.9	5.1	5.9	6.5
3	5.7	5.2	6.4	6.4
3	4.2	4.6	5.3	6.3
3	6.0	5.3	5.8	6.4
3	5.1	5.2	6.2	6.5
3	4.8	4.6	5.7	5.7
3	5.3	5.4	6.8	6.6
3	4.6	4.4	5.7	5.6
3	4.5	4.0	5.0	5.9
3	4.4	4.2	5.6	5.5

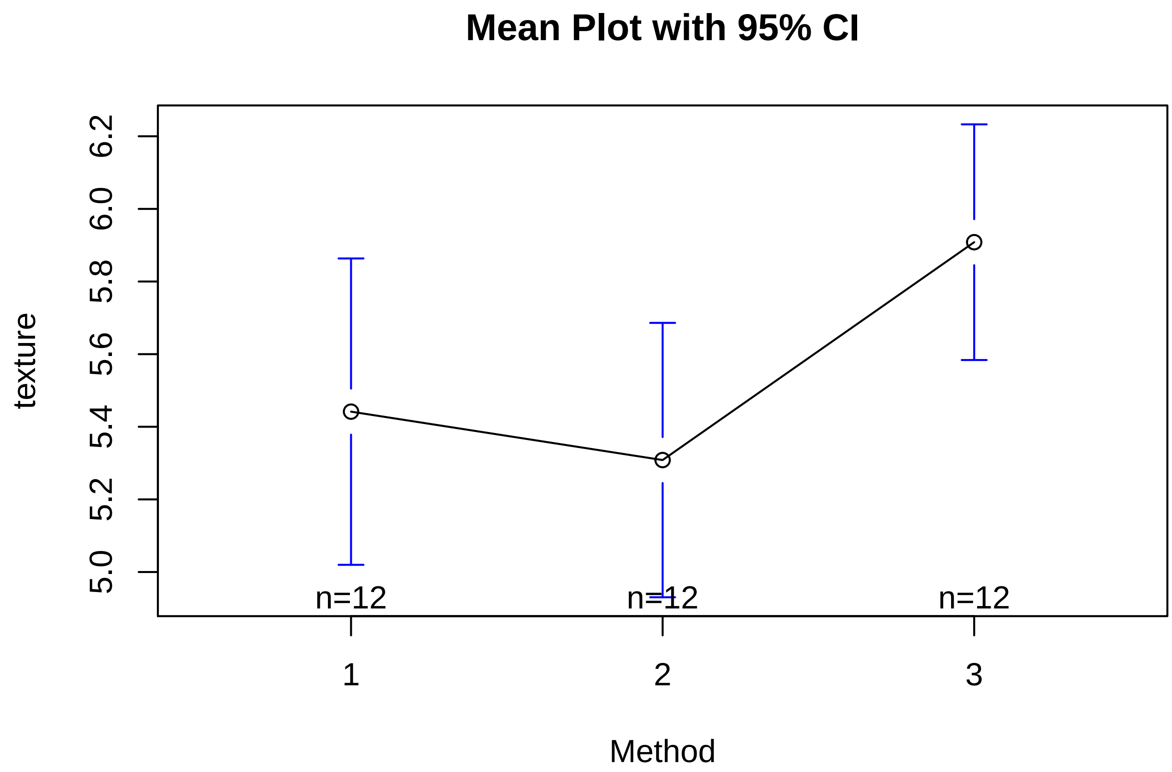
```r
plotmeans(y1 ~ Method, data = fish,
          xlab = "Method", ylab = "aroma",
          main="Mean Plot with 95% CI")
```



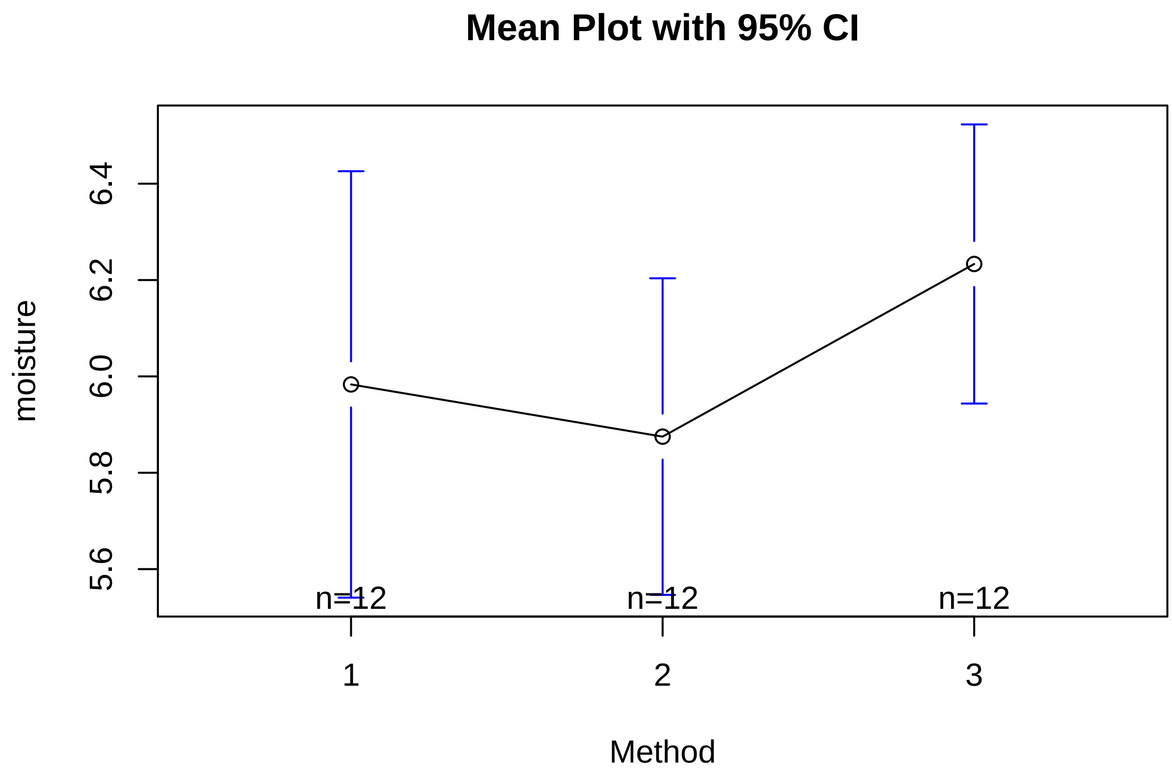

```r
plotmeans(y2 ~ Method, data = fish,
          xlab = "Method", ylab = "flavor",
          main="Mean Plot with 95% CI")
```




```
plotmeans(y3 ~ Method, data = fish,
          xlab = "Method", ylab = "texture",
          main="Mean Plot with 95% CI")
```




```r
plotmeans(y4 ~ Method, data = fish,
          xlab = "Method", ylab = "moisture",
          main="Mean Plot with 95% CI")
```



(a) Compare the three methods using all four MANOVA tests.

The overall mean vector:

	x
y1	5.205556
y2	5.266667
y3	5.552778
y4	6.030556

The mean vectors represent 3 points in four-dementional space. The three mean vectors:

	1	2	3
y1	5.383333	5.258333	4.975000
y2	5.733333	5.233333	4.833333
y3	5.441667	5.308333	5.908333
y4	5.983333	5.875000	6.233333

H =

1.0505556	2.173333	-1.375556	-0.7602778
2.1733333	4.880000	-2.373333	-1.2566667
-1.3755556	-2.373333	2.382222	1.3844444
-0.7602778	-1.256667	1.384444	0.8105556

E =

13.408333	7.723333	8.675000	5.864167
7.723333	8.480000	7.526667	6.213333
8.675000	7.526667	11.607500	7.037500
5.864167	6.213333	7.037500	10.565833

k = 3

p = 4

$_vH$ = 2

$_vE$ = 33

t = 2

df_1 = 8

df_2 = 60

F = 8.3294328

```
# MANOVA test
fish.manova <- manova(cbind(fish$y1, fish$y2, fish$y3, fish$y4) ~ Method, data = fish)
fish.summary <- summary(fish.manova)
fish.summary
```

```
##           Df  Pillai approx F num Df den Df    Pr(>F)
## Method     2 0.85987    5.845      8     62 1.465e-05 ***
## Residuals 33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Look to see which differ
summary.aov(fish.manova)
```

```
##  Response 1 :
##             Df  Sum Sq Mean Sq F value Pr(>F)
## Method       2  1.0506 0.52528  1.2928  0.288
## Residuals   33 13.4083 0.40631
##
##  Response 2 :
##             Df Sum Sq Mean Sq F value   Pr(>F)
## Method       2   4.88 2.44000  9.4953 0.000553 ***
## Residuals   33   8.48 0.25697
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##  Response 3 :
##             Df  Sum Sq Mean Sq F value  Pr(>F)
## Method       2  2.3822 1.19111  3.3863 0.04596 *
## Residuals   33 11.6075 0.35174
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
```

```
##  Response 4 :
##              Df  Sum Sq Mean Sq F value Pr(>F)
## Method       2  0.8106 0.40528  1.2658 0.2954
## Residuals   33 10.5658 0.32018
```

```r
fish.summary$SS
```

```
## $Method
##            [,1]      [,2]      [,3]       [,4]
## [1,]  1.0505556  2.173333 -1.375556 -0.7602778
## [2,]  2.1733333  4.880000 -2.373333 -1.2566667
## [3,] -1.3755556 -2.373333  2.382222  1.3844444
## [4,] -0.7602778 -1.256667  1.384444  0.8105556
##
## $Residuals
##           [,1]     [,2]      [,3]      [,4]
## [1,] 13.408333 7.723333  8.675000  5.864167
## [2,]  7.723333 8.480000  7.526667  6.213333
## [3,]  8.675000 7.526667 11.607500  7.037500
## [4,]  5.864167 6.213333  7.037500 10.565833
```

We would then like to test if the properties are the same across the three cooking methods.

$H_0 : \mu_1 = \mu_2 = \mu_3$

H_1 : The $\mu's$ are unequal

```r
summary(manova(cbind(fish$y1, fish$y2,
                     fish$y3, fish$y4) ~ fish$Method), test = "Wilks")
```

```
##             Df   Wilks approx F num Df den Df    Pr(>F)
## fish$Method  2 0.22449   8.3294      8     60 1.609e-07 ***
## Residuals   33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$\lambda = 0.224$

The Wilks's test rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

```r
summary(manova(cbind(fish$y1, fish$y2,
                     fish$y3, fish$y4) ~ fish$Method), test = "Roy")
```

```
##             Df    Roy approx F num Df den Df    Pr(>F)
## fish$Method  2 2.9515   22.874      4     31 7.077e-09 ***
## Residuals   33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Θ = 0.747+0i

The Roy's test also rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

```r
summary(manova(cbind(fish$y1, fish$y2,
                     fish$y3, fish$y4) ~ fish$Method), test = "Hotelling-Lawley")
```

```
##             Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## fish$Method  2           3.0788   11.161      8     58 2.161e-09 ***
## Residuals   33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$U^{(s)}$ = 3.079+0i

The Hotelling-Lawley's test also rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

```r
summary(manova(cbind(fish$y1, fish$y2,
                     fish$y3, fish$y4) ~ fish$Method), test = "Pillai")
```

```
##             Df  Pillai approx F num Df den Df    Pr(>F)
## fish$Method  2 0.85987    5.845      8     62 1.465e-05 ***
## Residuals   33
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$V^{(s)}$ = 0.86+0i

The Pillai's test also rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

Answer using the tip sheet

```r
method1 <- fish %>% filter(Method == 1)  %>% select(-Method)
method2 <- fish %>% filter(Method == 2)  %>% select(-Method)
method3 <- fish %>% filter(Method == 3)  %>% select(-Method)

method1.bar <- colMeans(method1)
method2.bar <- colMeans(method2)
method3.bar <- colMeans(method3)
method.all.bar <- (method1.bar+method2.bar+method3.bar)/3

method1.bar.diff <- method1.bar - method.all.bar
method2.bar.diff <- method2.bar - method.all.bar
method3.bar.diff <- method3.bar - method.all.bar

H <- 12 * unname(method1.bar.diff %*% t(method1.bar.diff) +
                 method2.bar.diff %*% t(method2.bar.diff) +
                 method3.bar.diff %*% t(method3.bar.diff))
```

H =

1.0505556	2.173333	-1.375556	-0.7602778
2.1733333	4.880000	-2.373333	-1.2566667
-1.3755556	-2.373333	2.382222	1.3844444
-0.7602778	-1.256667	1.384444	0.8105556

```r
"compute.within.matrix" <-function(data, mean) {
  ret <- matrix(as.numeric(0), nrow=4, ncol=4)
  for (i in 1:12) {
    diff <- as.numeric(unname(data[i,] - mean))
    ret <- ret + diff %*% t(diff)}
  return(ret)
  }
E <- compute.within.matrix(method1, method1.bar) + compute.within.matrix(method2, method2.bar) + comput
```

E =

13.408333	7.723333	8.675000	5.864167
7.723333	8.480000	7.526667	6.213333
8.675000	7.526667	11.607500	7.037500
5.864167	6.213333	7.037500	10.565833

```r
Lambda <-det(E) / det(E + H)
V.s <- tr(solve(E + H) %*% H)
U.s <-tr(solve(E) %*% H)
lambda.1 <-eigen(solve(E) %*% H)$values[1]
theta <- lambda.1 / (1 + lambda.1)
```

$\lambda = 0.224$

The Wilks's test rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

$\Theta = 0.747$

The Roy's test also rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

$U^{(s)} = 3.079$

The Hotelling-Lawley's test also rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

$V^{(s)} = 0.86$

The Pillai's test also rejects the hypothesis H_0 that the mean vector for the three cooking methods are equal.

(b) Compute the following measures of multivariate association from Section 6.1.8 : η^2_Λ, η^2_Θ, A_Λ, A_{LH}, A_P.

$\eta^2_\Lambda = 0.776$

$\eta^2_\Theta = \Theta = 0.747{+}0i$

$A_\Lambda = 0.526$

$A_{LH} = 0.606{+}0i$

$A_P = 0.43{+}0i$

(c) Based on the eigenvalues, is the essential dimensionality of the space containing the mean vectors equal to 1 or 2?

The eigenvalues of E^{-1} H are

```r
eigen(inv(E) %*% H)
```

```
## eigen() decomposition
## $values
## [1]  2.951475e+00  1.273244e-01 -3.132799e-17  1.734154e-17
##
## $vectors
##             [,1]        [,2]       [,3]        [,4]
## [1,] -0.03181703 -0.63526645  0.6817038 -0.37126023
## [2,] -0.81967776  0.59729861 -0.1930297  0.08261924
## [3,]  0.53294807  0.48673083  0.4941324 -0.55987018
## [4,]  0.20756300 -0.05257385 -0.5038379  0.73612858
```

The essential dimensionality of the space containing the mean vectors is equal to 1.

Problem 6.28

Table 6.18, from Keuls, Martakis, and Magid (1984), gives data from a two-way (fixed-effects) MANOVA on snap beans showing the results of four variables: y_1 = yield earliness, y_2 = specific leaf area (SLA) earliness, y_3 = total yield, and y_4 = average SLA. The factors are sowing date (S) and variety (V).

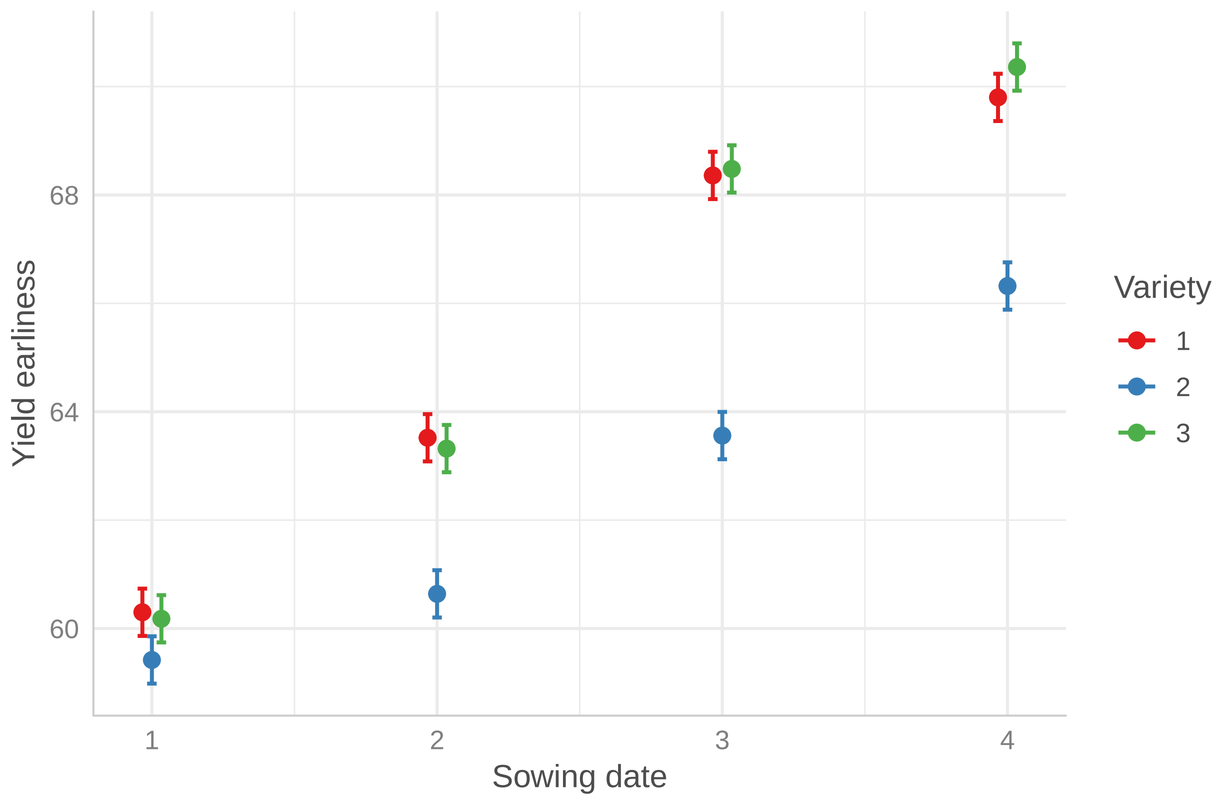
S	V	record	y1	y2	y3	y4
1	1	1	59.3	4.5	38.4	295
1	1	2	60.3	3.5	38.6	302
1	1	3	60.9	5.3	37.2	318
1	1	4	60.6	5.8	38.1	345
1	1	5	60.4	6.0	38.8	325
1	2	1	59.3	6.7	37.9	275
1	2	2	59.4	4.8	36.6	290
1	2	2	60.0	5.1	38.7	295
1	2	4	58.9	5.8	37.5	296
1	2	5	59.5	4.8	37.0	330

(a) Test for main effects and interaction using all four MANOVA statistics.

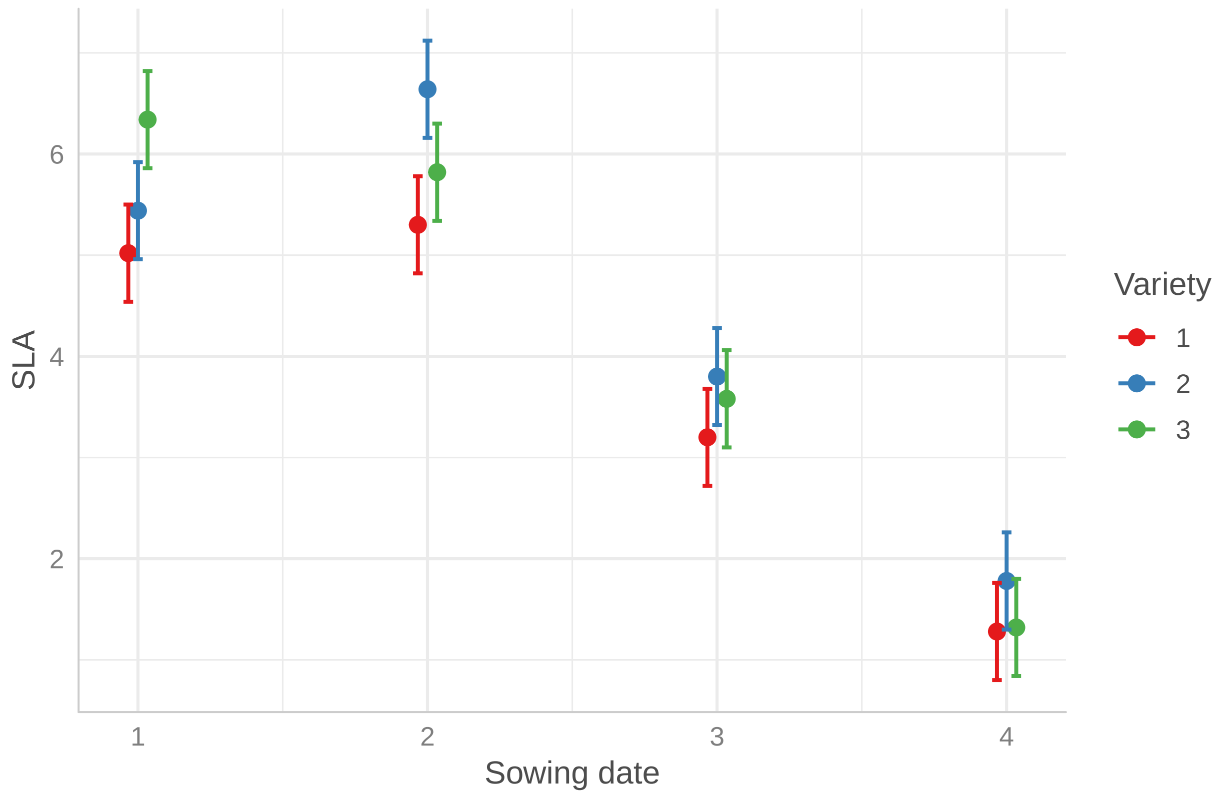
```r
snapbean.manova <- manova(cbind(snapbean$y1, snapbean$y2,
                                snapbean$y3, snapbean$y4) ~ snapbean$S * snapbean$V,
                          data = snapbean)
snapbean.summary <- summary(snapbean.manova)
snapbean.summary
```

```
##                        Df Pillai approx F num Df den Df    Pr(>F)
## snapbean$S              3 2.3568   43.052     12    141 < 2.2e-16 ***
## snapbean$V              2 1.1070   14.256      8     92 2.564e-13 ***
## snapbean$S:snapbean$V   6 1.3213    3.946     24    192 3.912e-08 ***
## Residuals              48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

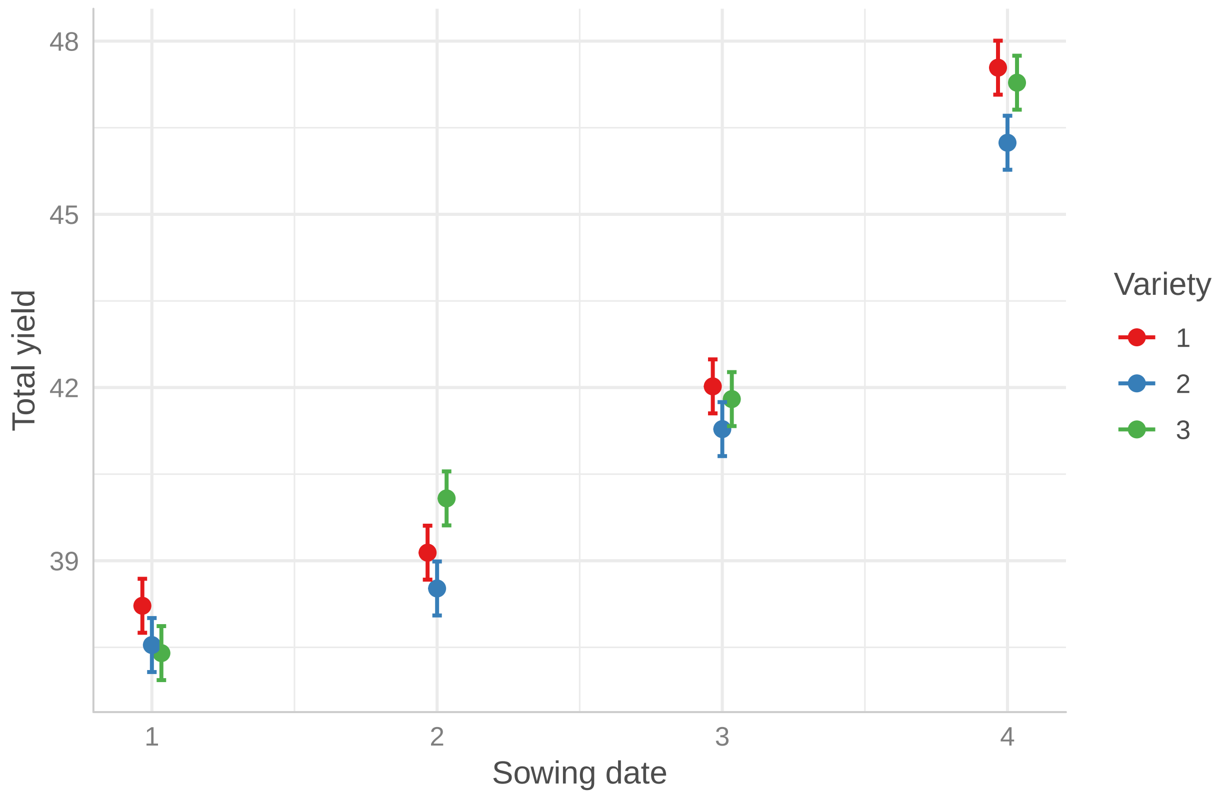
```
plot_model(fit1, type = "pred", terms = c("S", "V") ,
           legend.title = "Variety", axis.title = c("Sowing date", "Yield earliness"), title = "")
```



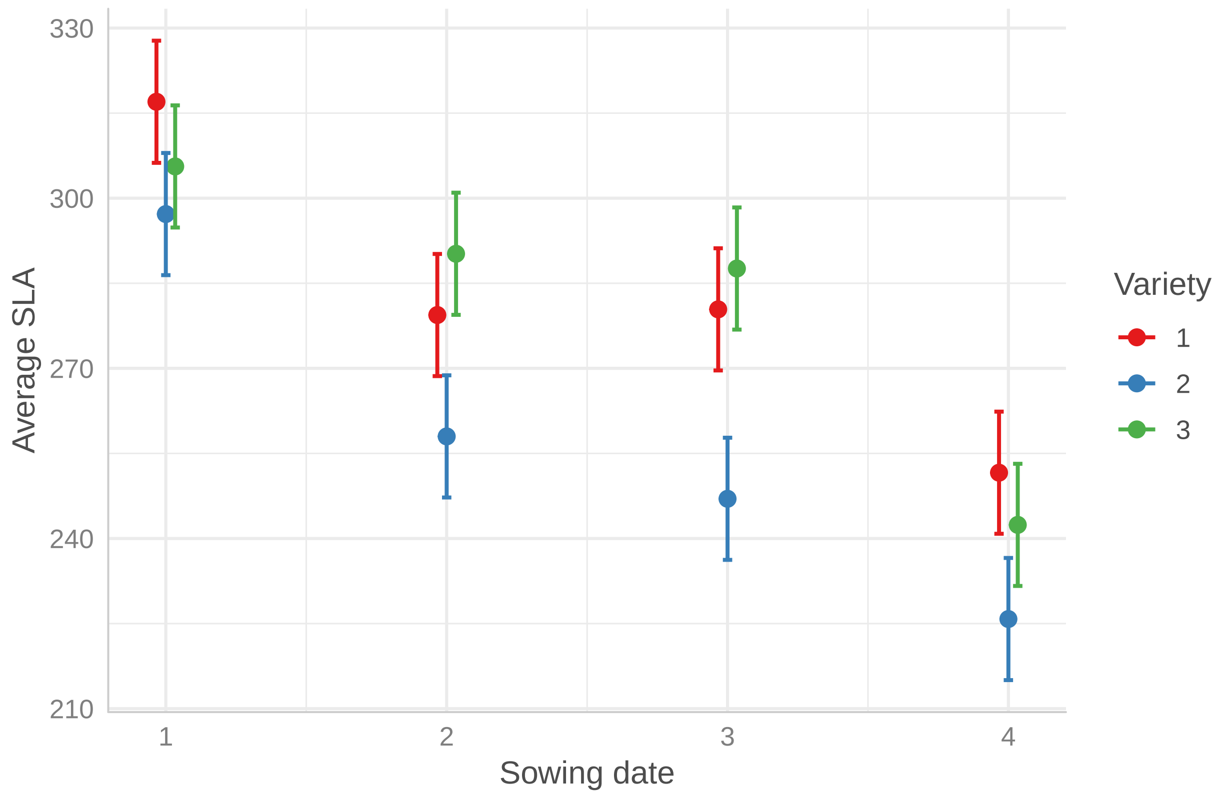

```
plot_model(fit2, type = "pred", terms = c("S", "V") ,
           legend.title = "Variety", axis.title = c("Sowing date","SLA"), title = "")
```




```
plot_model(fit3, type = "pred", terms = c("S", "V") ,
           legend.title = "Variety", axis.title = c("Sowing date","Total yield"), title = "")
```




```
plot_model(fit4, type = "pred", terms = c("S", "V") ,
           legend.title = "Variety", axis.title = c("Sowing date", "Average SLA"), title = "")
```



Error

$E =$

11.896	0.054	-0.108	45.62
0.054	14.404	-1.982	8.98
-0.108	-1.982	13.656	41.02
45.620	8.980	41.020	7245.60

S effect

$H_S =$

728.7898	-347.9327	690.1152	-4563.785
-347.9327	192.8680	-366.7527	2154.340
690.1152	-366.7527	747.7765	-4741.505
-4563.7850	2154.3400	-4741.5050	33469.383

$V^{(s)} = 2.3567703$ $U^{(s)} = 142.3042339$ $\Theta = 0.9927624$

The main effect and interaction are statistically significant.

V effect

H_V =

124.52133	-17.108667	32.098333	1008.5833
-17.10867	5.686333	-5.064167	-146.5417
32.09833	-5.064167	8.402333	261.5483
1008.58333	-146.541667	261.548333	8188.2333

$\Lambda = 0.0653001$ $V^{(s)} = 1.1070066$ $U^{(s)} = 11.6752251$ $\Theta = 0.9196443$

The main effect and interaction are statistically significant.

SV effect

H_{SV} =

30.294667	-5.387333	2.956333	130.710
-5.387333	5.119000	-3.095833	-40.805
2.956333	-3.095833	5.867000	59.665
130.710000	-40.805000	59.665000	1887.767

$\Lambda = 0.1379474$ $V^{(s)} = 1.3212987$ $U^{(s)} = 3.4504638$ $\Theta = 0.725937$

The main effect and interaction are statistically significant.

```
summary(manova(snapbean.manova), test = "Wilks")
```

```
##                         Df    Wilks approx F num Df den Df    Pr(>F)
## snapbean$S               3 0.000645  149.831     12 119.35 < 2.2e-16 ***
## snapbean$V               2 0.065300   32.775      8  90.00 < 2.2e-16 ***
## snapbean$S:snapbean$V    6 0.137947    5.039     24 158.20 1.611e-10 ***
## Residuals               48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The main effect and interaction are statistically significant.

```
summary(manova(snapbean.manova), test = "Pillai")
```

```
##                         Df Pillai approx F num Df den Df    Pr(>F)
## snapbean$S               3 2.3568   43.052     12    141 < 2.2e-16 ***
## snapbean$V               2 1.1070   14.256      8     92 2.564e-13 ***
## snapbean$S:snapbean$V    6 1.3213    3.946     24    192 3.912e-08 ***
## Residuals               48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The main effect and interaction are statistically significant.

```r
summary(manova(snapbean.manova), test = "Hotelling-Lawley")
```

```
##                        Df Hotelling-Lawley approx F num Df den Df    Pr(>F)
## snapbean$S              3          142.304   517.83     12    131 < 2.2e-16
## snapbean$V              2           11.675    64.21      8     88 < 2.2e-16
## snapbean$S:snapbean$V   6            3.450     6.25     24    174 8.671e-14
## Residuals              48
##
## snapbean$S             ***
## snapbean$V             ***
## snapbean$S:snapbean$V ***
## Residuals
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The main effect and interaction are statistically significant.

```r
summary(manova(snapbean.manova), test = "Roy")
```

```
##                        Df     Roy approx F num Df den Df    Pr(>F)
## snapbean$S              3 137.168  1611.72      4     47 < 2.2e-16 ***
## snapbean$V              2  11.445   131.61      4     46 < 2.2e-16 ***
## snapbean$S:snapbean$V   6   2.649    21.19      6     48 5.691e-12 ***
## Residuals              48
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The main effect and interaction are statistically significant.