A Handbook of Statistical Analyses Using R

Brian S. Everitt and Torsten Hothorn
4.1 Introduction

4.2 Analysis of Variance

4.3 Analysis Using R

4.3.1 Weight Gain in Rats

Before applying analysis of variance to the data in Table ?? we should try to summarise the main features of the data by calculating means and standard deviations and by producing some hopefully informative graphs. The data is available in the data.frame weightgain. The following R code produces the required summary statistics

```R
R> data("weightgain", package = "HSAUR")
R> tapply(weightgain$weightgain, +   list(weightgain$source, weightgain$type), mean)

  High Low
Beef 100.0 79.2
Cereal 85.9 83.9

R> tapply(weightgain$weightgain, +   list(weightgain$source, weightgain$type), sd)

  High Low
Beef 15.13642 13.88684
Cereal 15.02184 15.70881
```

To apply analysis of variance to the data we can use the `aov` function in R and then the `summary` method to give us the usual analysis of variance table. The model formula specifies a two-way layout with interaction terms, where the first factor is source, and the second factor is type.

```R
R> wg_aov <- aov(weightgain ~ source * type, data = weightgain)

R> coef(wg_aov)

(Intercept) sourceCereal sourceCereal:typeLow typeLow
100.0 -14.1 -20.8
```

The estimates of the intercept and the main and interaction effects can be extracted from the model fit by R> coef(wg_aov)
4 ANALYSIS OF VARIANCE

R> plot.design(weightgain)

Figure 4.1  Plot of mean weight gain for each level of the two factors.

Note that the model was fitted with the restrictions $\gamma_1 = 0$ (corresponding to Beef) and $\beta_1 = 0$ (corresponding to High) because treatment contrasts were used as default as can be seen from

R> options("contrasts")

$contrasts
  unordered         ordered
"contr.treatment"   "contr.poly"

Thus, the coefficient for source of $-14.1$ can be interpreted as an estimate of the difference $\gamma_2 - \gamma_1$. Alternatively, we can use the restriction $\sum \gamma_i = 0$ by

R> coef(aov(weightgain ~ source + type + source:type, + data = weightgain, contrasts = list(source = contr.sum)))

(Intercept) source1 typeLow
ANALYSIS USING R

\begin{verbatim}
R> summary(wg_aov)

Call: aov(formula = weight ~ type, data = wg)

Residuals:   Df Sum Sq Mean Sq F value Pr(>F)
source      1    221 220.91 0.9883 0.3269
type        1   1299 1299.68 5.8120 0.0211 *
source:type 1    883 883.68 3.9520 0.0545 .
Residuals   36 8049 223.65

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
\end{verbatim}

Figure 4.2  R output of the ANOVA fit for the weightgain data.

\begin{verbatim}
92.95 7.05 -11.40
source:typeLow -9.40
\end{verbatim}

4.3.2 Foster Feeding of Rats of Different Genotype

As in the previous subsection we will begin the analysis of the foster feeding data in Table ?? with a plot of the mean litter weight for the different genotypes of mother and litter (see Figure 4.4). The data are in the data.frame foster

\begin{verbatim}
R> data("foster", package = "HSAUR")
\end{verbatim}

We can derive the two analyses of variance tables for the foster feeding example by applying the R code

\begin{verbatim}
R> summary(aov(weight ~ litgen * motgen, data = foster))

Call: aov(formula = weight ~ litgen * motgen, data = foster)

Residuals:   Df Sum Sq Mean Sq  F value Pr(>F)
litgen       3   60.2  20.05 0.3700 0.7752
motgen       3  775.1 258.36 4.7629 0.0057 **
litgen:motgen 9  824.1  91.56 1.6878 0.1201
Residuals    45 2440.8  54.24

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
\end{verbatim}

There are (small) differences in the sum of squares for the two main effects and, consequently, in the associated $F$-tests and $p$-values. This would not be true if in the previous example in Subsection 4.3.1 we had used the code

\begin{verbatim}
R> summary(aov(weight ~ motgen * litgen, data = foster))

Call: aov(formula = weight ~ motgen * litgen, data = foster)

Residuals:   Df Sum Sq Mean Sq  F value Pr(>F)
motgen       3  771.6 257.20 4.7420 0.0058 **
litgen       3   63.6 21.21 0.3910 0.7600
motgen:litgen 9  824.1  91.56 1.6880 0.1201
Residuals    45 2440.8  54.24

---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
\end{verbatim}
R> interaction.plot(weightgain$type, weightgain$source, +
    weightgain$weightgain)

Figure 4.3 Interaction plot of type × source.

R> summary(aov(weightgain ~ type * source, data = weightgain))

instead of the code which produced Figure 4.2 (readers should confirm that
this is the case).

We can investigate the effect of genotype B on litter weight in more detail
by the use of multiple comparison procedures (see Everitt, 1996). Such proce-
dures allow a comparison of all pairs of levels of a factor whilst maintaining
the nominal significance level at its selected value and producing adjusted
confidence intervals for mean differences. One such procedure is called Tukey
honest significant differences suggested by Tukey (1953), see Hochberg and
Tamhane (1987) also. Here, we are interested in simultaneous confidence in-
tervals for the weight differences between all four genotypes of the mother.
First, an ANOVA model is fitted
ANALYSIS USING R

R> plot.design(foster)

Figure 4.4  Plot of mean litter weight for each level of the two factors for the foster data.

R> foster_aov <- aov(weight ~ litgen * motgen, data = foster)
which serves as the basis of the multiple comparisons, here with allpair differences by

R> foster_hsd <- TukeyHSD(foster_aov, "motgen")
R> foster_hsd

Tukey multiple comparisons of means
95% family-wise confidence level

Fit: aov(formula = weight ~ litgen * motgen, data = foster)

$motgen
diff  lwr  upr  p adj
B-A  3.330369 -3.859729 10.5204672 0.6078581
8 ANALYSIS OF VARIANCE

R> plot(foster_hsd)

![Graphical presentation of multiple comparison results for the foster feeding data.](image)

**Figure 4.5** Graphical presentation of multiple comparison results for the foster feeding data.

<table>
<thead>
<tr>
<th></th>
<th>I-A</th>
<th>J-A</th>
<th>I-B</th>
<th>J-B</th>
<th>J-I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.895574</td>
<td>-6.566168</td>
<td>-5.225943</td>
<td>-9.896537</td>
<td>-4.670593</td>
</tr>
<tr>
<td>3</td>
<td>5.0507207</td>
<td>0.4949498</td>
<td>1.9641552</td>
<td>-2.5954489</td>
<td>2.3905240</td>
</tr>
<tr>
<td>4</td>
<td>0.8853702</td>
<td>0.0767540</td>
<td>0.2266493</td>
<td>0.0040509</td>
<td>0.3035490</td>
</tr>
</tbody>
</table>

A convenient plot method exists for this object and we can get a graphical representation of the multiple confidence intervals as shown in Figure 4.5. It appears that there is only evidence for a difference in the B and J genotypes.

### 4.3.3 Water Hardness and Mortality

The water hardness and mortality data for 61 large towns in England and Wales (see Table 2.3) was analysed in Chapter 2 and here we will extend the
ANALYSIS USING R

Analysis by an assessment of the differences of both hardness and mortality in the North or South. The hypothesis that the two-dimensional mean-vector of water hardness and mortality is the same for cities in the North and the South can be tested by Hotelling-Lawley test in a multivariate analysis of variance framework. The R function `manova` can be used to fit such a model and the corresponding `summary` method performs the test specified by the `test` argument:

```R
R> data("water", package = "HSAUR")
R> summary(manova(cbind(hardness, mortality) ~ location, +
    data = water), test = "Hotelling-Lawley")
```

```
                  Df Hotelling-Lawley approx F num Df den Df Pr(>F)
location          1     0.90021     26.106     2   58 8.217e-09
Residuals         59

location ***
Residuals
---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The `cbind` statement in the left hand side of the formula indicates that a multivariate response variable is to be modelled. The p-value associated with the Hotelling-Lawley statistic is very small and there is strong evidence that the mean vectors of the two variables are not the same in the two regions. Looking at the sample means:

```R
R> tapply(water$hardness, water$location, mean)
   North  South
       30.40000 69.76923
R> tapply(water$mortality, water$location, mean)
   North  South
      1633.600 1376.808
```

we see large differences in the two regions both in water hardness and mortality, where low mortality is associated with hard water in the South and high mortality with soft water in the North (see Figure ?? also).

4.3.4 Male Egyptian Skulls

We can begin by looking at a table of mean values for the four measurements within each of the five epochs. The measurements are available in the `data.frame` `skulls` and we can compute the means over all epochs by:

```R
R> data("skulls", package = "HSAUR")
R> means <- aggregate(skulls[,c("mb", "bh", "bl", "nh")], +
    list(epoch = skulls$epoch), mean)
R> means
```
R> pairs(means[, -1],
+     panel = function(x, y) {
+         text(x, y, abbreviate(levels(skulls$epoch)))
+     })

**Figure 4.6** Scatterplot matrix of epoch means for Egyptian skulls data.

```
epoch   mb   bh   bl   nh
1 c4000BC 131.3667 133.6000 99.16667 50.53333
2 c3300BC 132.3667 132.7000 99.06667 50.23333
3 c1850BC 134.4667 133.8000 96.03333 50.56667
4 c200BC  135.5000 132.3000 94.53333 51.96667
5 cAD150  136.1667 130.3333 93.50000 51.36667
```

It may also be useful to look at these means graphically and this could be done in a variety of ways. Here we construct a scatterplot matrix of the means using the code attached to Figure 4.6.

There appear to be quite large differences between the epoch means, at
ANALYSIS USING R

least on some of the four measurements. We can now test for a difference more formally by using MANOVA with the following R code to apply each of the four possible test criteria mentioned earlier:

```R
R> skulls_manova <- manova(cbind(mb, bh, bl, nh) ~ epoch,
+                           data = skulls)
R> summary(skulls_manova, test = "Pillai")

Df Pillai approx F num Df den Df  Pr(>F)
epoch 4 0.35331 3.512 16 580 4.675e-06 ***
Residuals 145
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R> summary(skulls_manova, test = "Wilks")

Df Wilks approx F num Df den Df  Pr(>F)
epoch 4 0.66359 3.9009 16 434.45 7.01e-07 ***
Residuals 145
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R> summary(skulls_manova, test = "Hotelling-Lawley")

Df Hotelling-Lawley approx F num Df den Df
epoch 4 0.48182 4.231 16 562
Residuals 145
```

Pr(>F)
epoch 8.278e-08 ***
Residuals
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R> summary(skulls_manova, test = "Roy")

Df Roy approx F num Df den Df  Pr(>F)
epoch 4 0.4251 15.41 4 145 1.588e-10 ***
Residuals 145
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

The p-value associated with each four test criteria is very small and there is strong evidence that the skull measurements differ between the five epochs. We might now move on to investigate which epochs differ and on which variables. We can look at the univariate \( F \)-tests for each of the four variables by using the code

```R
R> summary.aov(skulls_manova)

Response mb :

Df Sum Sq Mean Sq F value Pr(>F)
epoch 4 502.83 125.707 5.9546 0.0001826 ***
Residuals 145 3061.07 21.111
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```
We see that the results for the maximum breadths (mb) and basialveolar length (bl) are highly significant, with those for the other two variables, in particular for nasal heights (nh), suggesting little evidence of a difference. To look at the pairwise multivariate tests (any of the four test criteria are equivalent in the case of a one-way layout with two levels only) we can use the summary method and manova function as follows:

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls, +   subset = epoch %in% c("c4000BC", "c3300BC")))

Df Pillai approx F num Df den Df Pr(>F)
epoch 1 0.027674 0.39135 4 55 0.8139
Residuals 58

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls, +   subset = epoch %in% c("c4000BC", "c1850BC")))

Df Pillai approx F num Df den Df Pr(>F)
epoch 1 0.18757 3.1744 4 55 0.02035 *
Residuals 58

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls, +   subset = epoch %in% c("c4000BC", "c200BC")))

Df Pillai approx F num Df den Df Pr(>F)
epoch 1 0.30297 5.9766 4 55 0.0004564 ***
Residuals 58

R> summary(manova(cbind(mb, bh, bl, nh) ~ epoch, data = skulls, +   subset = epoch %in% c("c4000BC", "cAD150")))

Df Pillai approx F num Df den Df Pr(>F)
epoch 1 0.027674 0.39135 4 55 0.8139
Residuals 58

---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

---
Signif. codes: 0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
<table>
<thead>
<tr>
<th></th>
<th>Pillai approx F</th>
<th>num Df</th>
<th>den Df</th>
<th>Pr(&gt;F)</th>
</tr>
</thead>
<tbody>
<tr>
<td>epoch</td>
<td>0.36182</td>
<td>7.7956</td>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>Residuals</td>
<td>58</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

To keep the overall significance level for the set of all pairwise multivariate tests under some control (and still maintain a reasonable power), Stevens (2001) recommends setting the nominal level \( \alpha = 0.15 \) and carrying out each test at the \( \alpha/m \) level where \( m \) is the number of tests performed. The results of the four pairwise tests suggest that as the epochs become further separated in time the four skull measurements become increasingly distinct.

