A Handbook of Statistical Analyses Using R — 2nd Edition

Brian S. Everitt and Torsten Hothorn

CHAPTER 5

Analysis of Variance: Weight Gain, Foster Feeding in Rats, Water Hardness and Male Egyptian Skulls

5.1 Introduction

5.2 Analysis of Variance

5.3 Analysis Using R

5.3.1 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> data("weightgain", package = "HSAUR2")
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   15.1 13.9
Cereal 15.0 15.7
```

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

The estimates of the intercept and the main and interaction effects can be extracted from the model fit by

```
R> coef(wg_aov)
```

```
         (Intercept)         sourceCereal              typeLow 
               100.0                -14.1                -20.8 
sourceCereal:typeLow 
                18.8 
```

```
R> plot.design(weightgain)
```

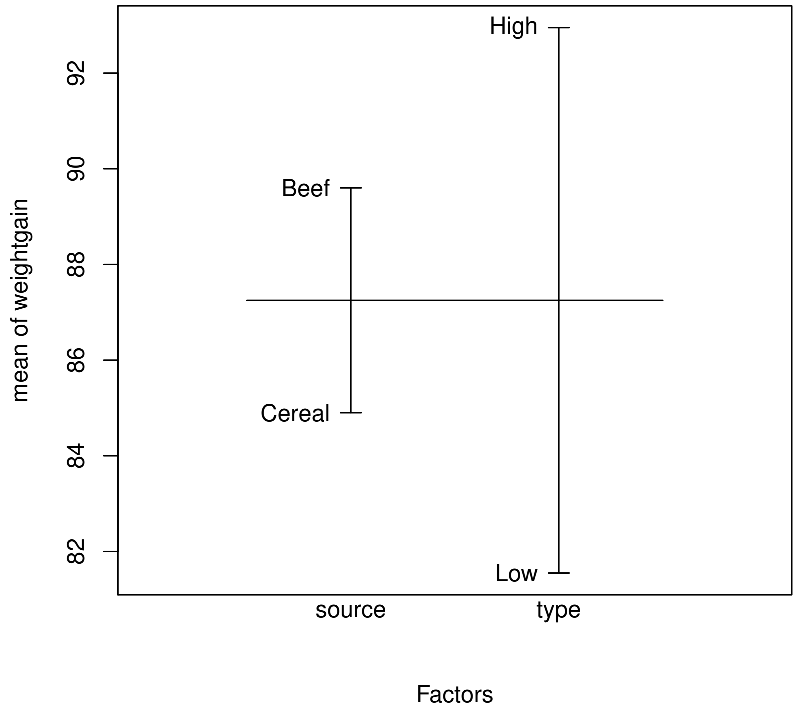

Figure 5.1 Plot of mean weight gain for each level of the two factors.

```
R> summary(wg_aov)

            Df Sum Sq Mean Sq F value Pr(>F)
source       1    221     221    0.99  0.327
type         1   1300    1300    5.81  0.021
source:type  1    884     884    3.95  0.054
Residuals   36   8049     224
```

Figure 5.2 R output of the ANOVA fit for the `weightgain` data.

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
```

```
R> interaction.plot(weightgain$type, weightgain$source,
+                   weightgain$weightgain)
```

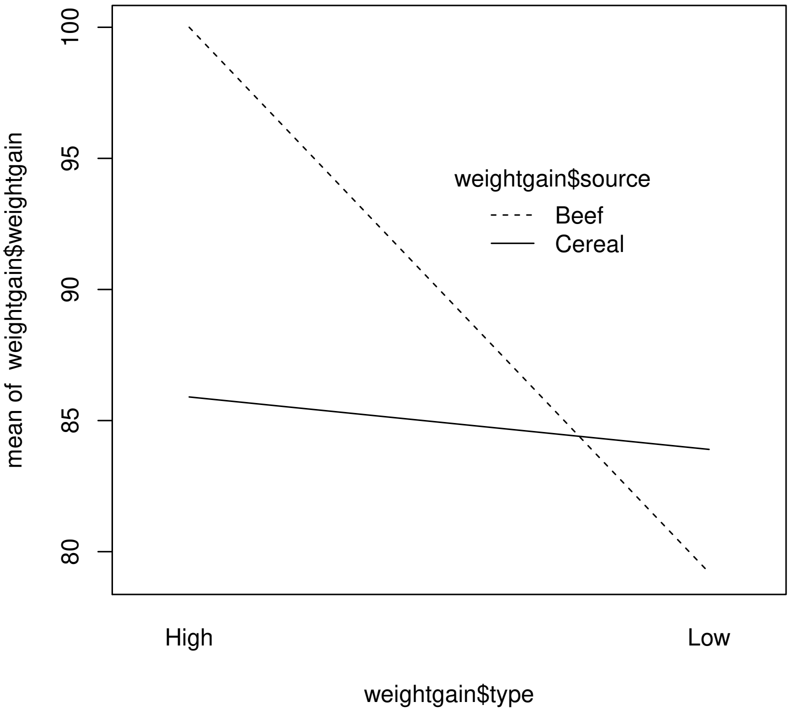

Figure 5.3 Interaction plot of type and source.

```
        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 $\sum_i \gamma_i = 0$ by

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

```
    (Intercept)         source1          typeLow
          92.95            7.05           -11.40
source1:typeLow
          -9.40
```

```
R> plot.design(foster)
```

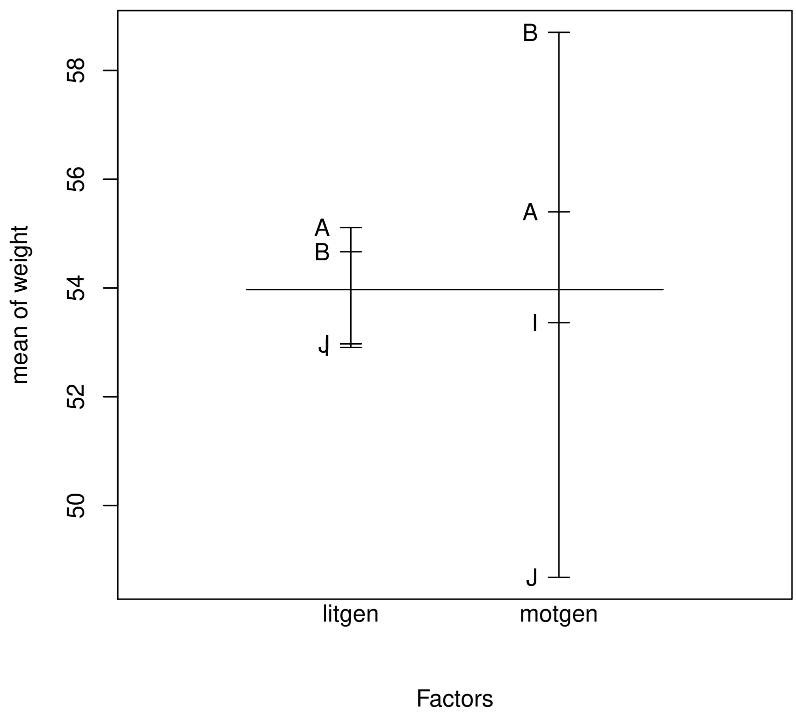

Figure 5.4 Plot of mean litter weight for each level of the two factors for the `foster` data.

5.3.2 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 5.4). The data are in the *data.frame* `foster`

```
R> data("foster", package = "HSAUR2")
```

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

```
R> summary(aov(weight ~ litgen * motgen, data = foster))
```

to give

```
            Df Sum Sq Mean Sq F value Pr(>F)
```

```
litgen         3     60    20.1    0.37 0.7752
motgen         3    775   258.4    4.76 0.0057
litgen:motgen  9    824    91.6    1.69 0.1201
Residuals     45   2441    54.2
```

and then the code

```
R> summary(aov(weight ~ motgen * litgen, data = foster))
```

to give

```
              Df Sum Sq Mean Sq F value Pr(>F)
motgen         3    772   257.2    4.74 0.0059
litgen         3     64    21.2    0.39 0.7600
motgen:litgen  9    824    91.6    1.69 0.1201
Residuals     45   2441    54.2
```

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 5.3.1 we had used the code

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

instead of the code which produced Figure 5.2 (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, and Chapter 14). Such procedures allow a comparison of all pairs of levels of a factor whilst maintaining the nominal significance level at its specified 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 intervals for the weight differences between all four genotypes of the mother. First, an ANOVA model is fitted

```
R> foster_aov <- aov(weight ~ litgen * motgen, data = foster)
```

which serves as the basis of the multiple comparisons, here with all pair-wise 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  3.33  -3.86 10.520 0.608
I-A -1.90  -8.84  5.051 0.885
J-A -6.57 -13.63  0.495 0.077
I-B -5.23 -12.42  1.964 0.227
J-B -9.90 -17.20 -2.595 0.004
J-I -4.67 -11.73  2.391 0.304
```

```
R> plot(foster_hsd)
```

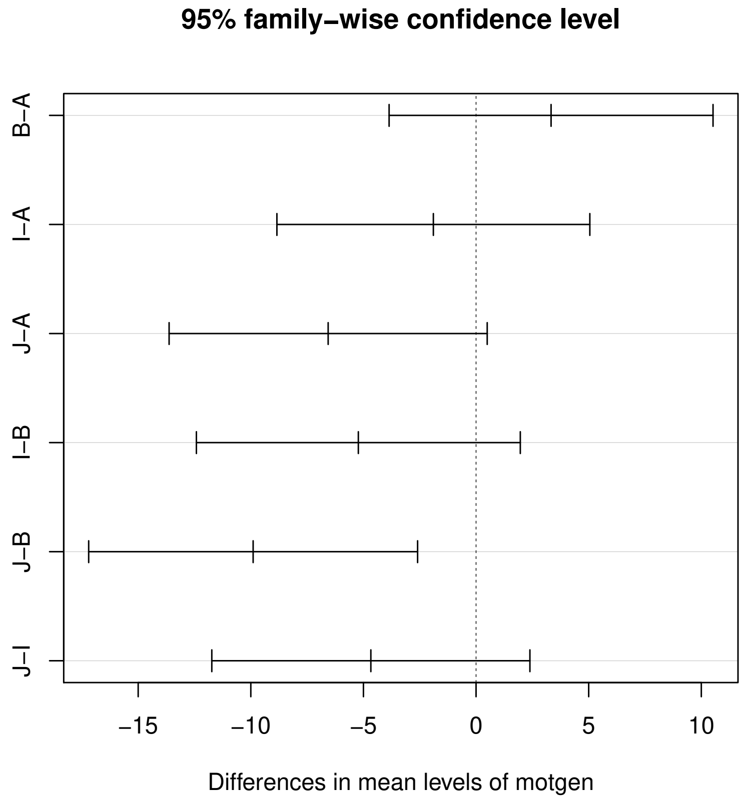

Figure 5.5 Graphical presentation of multiple comparison results for the `foster` feeding data.

A convenient `plot` method exists for this object and we can get a graphical representation of the multiple confidence intervals as shown in Figure 5.5. It appears that there is only evidence for a difference in the B and J genotypes. Note that the particular method implemented in `TukeyHSD` is applicable only to balanced and mildly unbalanced designs (which is the case here). Alternative approaches, applicable to unbalanced designs and more general research questions, will be introduced and discussed in Chapter 14.

5.3.3 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 3 and here we will extend the 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> data("water", package = "HSAUR2")
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              0.9     26.1      2     58 8.2e-09
Residuals 59
```

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> tapply(water$hardness, water$location, mean)
```

```
North South
 30.4  69.8
```

```
R> tapply(water$mortality, water$location, mean)
```

```
North South
 1634  1377
```

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).

5.3.4 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> data("skulls", package = "HSAUR2")
R> means <- aggregate(skulls[,c("mb", "bh", "bl", "nh")],
+                     list(epoch = skulls$epoch), mean)
R> means
```

```
    epoch  mb  bh   bl   nh
1 c4000BC 131 134 99.2 50.5
2 c3300BC 132 133 99.1 50.2
3 c1850BC 134 134 96.0 50.6
4  c200BC 136 132 94.5 52.0
5  cAD150 136 130 93.5 51.4
```

```
R> pairs(means[,-1],
+      panel = function(x, y) {
+          textplot(x, y, levels(skulls$epoch), new = FALSE, cex = 0.8)
+      })
```

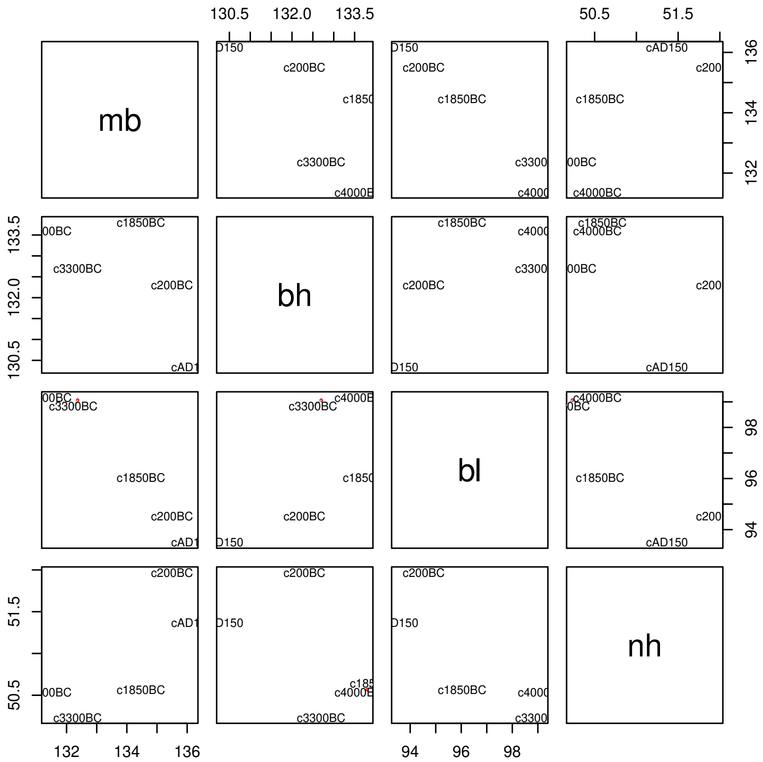

Figure 5.6 Scatterplot matrix of epoch means for Egyptian `skulls` data.

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 5.6.

There appear to be quite large differences between the epoch means, at 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> 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  0.353     3.51     16    580 4.7e-06
Residuals 145
```

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

```
           Df Wilks approx F num Df den Df Pr(>F)
epoch       4 0.664      3.9     16    434  7e-07
Residuals 145
```

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

```
           Df Hotelling-Lawley approx F num Df den Df  Pr(>F)
epoch       4            0.482     4.23     16    562 8.3e-08
Residuals 145
```

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

```
           Df   Roy approx F num Df den Df  Pr(>F)
epoch       4 0.425     15.4      4    145 1.6e-10
Residuals 145
```

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> summary.aov(skulls_manova)
```

```
 Response mb :
             Df Sum Sq Mean Sq F value  Pr(>F)
epoch         4    503   125.7    5.95 0.00018
Residuals   145   3061    21.1

 Response bh :
             Df Sum Sq Mean Sq F value Pr(>F)
epoch         4    230    57.5    2.45  0.049
Residuals   145   3405    23.5

 Response bl :
             Df Sum Sq Mean Sq F value  Pr(>F)
epoch         4    803   200.8    8.31 4.6e-06
Residuals   145   3506    24.2

 Response nh :
             Df Sum Sq Mean Sq F value Pr(>F)
epoch         4     61    15.3    1.51    0.2
Residuals   145   1472    10.2
```

We see that the results for the maximum breadths (`mb`) and basialiveolar 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 0.0277    0.391      4     55   0.81
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  0.188     3.17      4     55   0.02
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  0.303     5.98      4     55 0.00046
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  0.362      7.8      4     55 4.7e-05
Residuals 58
```

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 α/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.

Bibliography

Everitt, B. S. (1996), *Making Sense of Statistics in Psychology: A Second-Level Course*, Oxford, UK: Oxford University Press.

Hochberg, Y. and Tamhane, A. C. (1987), *Multiple Comparison Procedures*, New York, USA: John Wiley & Sons.

Stevens, J. (2001), *Applied Multivariate Statistics for the Social Sciences*, Mahwah, New Jersey, USA: Lawrence Erlbaum, 4th edition.

Tukey, J. W. (1953), "The problem of multiple comparisons (unpublished manuscript)," in *The Collected Works of John W. Tukey VIII. Multiple Comparisons: 1948-1983*, New York, USA: Chapman & Hall.