Chapter 7
Multiple Comparisons

In Exercise 3.13 we discover that the probability of simultaneously making three correct inferences, when each of the three individually has P(correct inference) = $1 - \alpha = 0.95$, is only $(1 - \alpha)^3 = .95^3 = 0.857$. Alternatively, the probability of making at least one incorrect inference is $1 - -0.857 = 0.143 \approx 3\alpha$. In general, the more simultaneous inferences we make at one time, the smaller the probability that all are correct. In this chapter we learn how to control the probability that all inferences are simultaneously correct. We usually phrase the goal as controlling the probability of making at least one incorrect inference.

We consider all inferences in a related *family* of inferences. Such a family is typically a natural and coherent collection; for example, all inferences resulting from a single experiment. The inferences can be individual tests of hypotheses or confidence intervals. In the context of a family of hypothesis tests, if we control the Type I error probability for each test at level α, the probability of committing at least one Type I error in the family will be much larger than α. For example, if the tests are independent and $\alpha = .05$, then the probability of at least one Type I error is $1 - (1 - .05)^6 \approx .26$, which seems an unacceptably large error threshold. The way to control the probability of at least one Type I error in the family is to choose a smaller α for each individual test. For example, with a single two-sided test with $\alpha = .05$ from a standard normal the critical value is 1.96. The intention of all multiple comparison procedures is to provide a larger critical value than the default value.

A way to avoid such errors when conducting many related inferences simultaneously is to employ a *multiple comparison procedure*. Such a procedure for *simultaneous hypothesis testing* may seek to (strongly) control the familywise error rate (FWE), defined as P(reject at least one true hypothesis under any configuration of true and false hypotheses). A procedure for *simultaneous confidence intervals* should control the probability that at least one member of the family of confidence

© Springer Science+Business Media New York 2015

R.M. Heiberger, B. Holland, *Statistical Analysis and Data Display*,
Springer Texts in Statistics, DOI 10.1007/978-1-4939-2122-5_7

intervals does not contain the parameter being estimated by the interval. When a multiple comparison procedure is used, it is said that the analyst is *controlling for multiplicity*.

In order to exert FWE control over a family of related hypothesis tests, it is necessary to have a reduced probability of rejecting any particular null hypothesis in the family. As explained in Section 3.7, reducing the probability of rejecting particular hypotheses results in an increased probability of retaining them, and therefore reduced power for tests of these hypotheses. This implies that, as compared with testing hypotheses in isolation from one another, a multiple comparison procedure has a diminished ability to reject false null hypotheses. In other words, a test of a particular hypothesis using a multiple comparison procedure will be less powerful than the test of the same hypothesis in isolation. In deciding whether to use a multiple comparison procedure, the protection against the possibility of an excessive number of incorrect hypothesis rejections must be weighted against this loss of power. An analogous statement holds for simultaneous versus isolated confidence intervals.

In general, the choice of multiple comparison procedure to be used depends on the structure of the *family* of related inferences and the nature of the collection of statistics from which the confidence intervals or tests will be calculated.

Section 7.1 summarizes the most frequently used multiple comparisons procedures. Section 7.2 presents a graphical procedure for looking at the results of the multiple comparisons procedures.

7.1 Multiple Comparison Procedures

7.1.1 Bonferroni Method

A very general way to control the FWE is based on the Bonferroni inequality, $P(\bigcup E_i) \leq \sum_i P(E_i)$, where the E_i are arbitrary events. If the family consists of m related tests, conducting each test at level $\frac{\alpha}{m}$ ensures that FWE $\leq \alpha$. If the family consists of m related confidence intervals, maintaining confidence $100(1 - \frac{\alpha}{m})\%$ for each interval will ensure that the overall confidence of all m intervals will be at least $100(1 - \alpha)\%$. The Bonferroni method should be considered for use when the family of related inferences is unstructured (e.g., not like the structured families required for the procedures discussed in Sections 7.1.2–7.1.4), or when the statistics used for inference about each family member have nonidentical probability distributions.

The Bonferroni inequality is very blunt in the sense that its right side is typically much larger than its left. One reason for this is that it does not seek to take into account information about the intersections of the events E_i. As a result, the Bonferroni approach is very conservative in the sense of typically guar-

anteeing an FWE substantially less than its nominal value of α, and the extent of this conservativeness increases with m. The value of this approach is that it is very generally applicable, for example, when the pivotal statistics associated with the m inferences have nonidentical probability distributions. Hochberg (1988) provides an easy-to-understand improvement to the Bonferroni approach for hypothesis testing that tends to reject more false null hypotheses than Bonferroni. Hochberg's procedure has been proven to be applicable to a wide variety of testing situations; see Sarkar (1998).

7.1.2 Tukey Procedure for All Pairwise Comparisons

Often a family of inferences has a special structure that allows us to use available information about the joint distributions of the pivotal statistics, thus enabling the use of a less conservative approach than Bonferroni. An example of this, discussed in Section 6.3, is the family consisting of all $m = \binom{k}{2}$ comparisons among all pairs of means of k populations. For this family, Tukey's Studentized range test is usually recommended.

7.1.3 The Dunnett Procedure for Comparing One Mean with All Others

The Dunnett procedure is used when the family of inferences of interest is the comparisons of the mean of one designated population with each of the means of the remaining populations, all populations being at least approximately normal with approximately the same variance. Often in practice the designated population is a control and the others are active treatments. The Dunnett procedure uses the percentiles of a multivariate t distribution rather than a univariate t distribution discussed in Section 5.4.3.

For purposes of illustration of the Dunnett procedure, we use weightloss data. A random sample of 50 men who were matched for pounds overweight was randomly separated into 5 equal groups. Each group was given exactly one of the weight loss agents `A`, `B`, `C`, `D`, or `E`. After a fixed period of time, each man's weight loss was recorded. The data, taken from Ott (1993), are accessed as `data(weightloss)` and shown in Figure 7.1.

The F-statistic tests the null hypothesis that the five groups have identical mean weight loss vs the alternative that the groups do not have identical mean weight loss. The small p-value from the F test in the basic ANOVA in Table 7.1 suggests that the agents have differing impacts on weight loss.

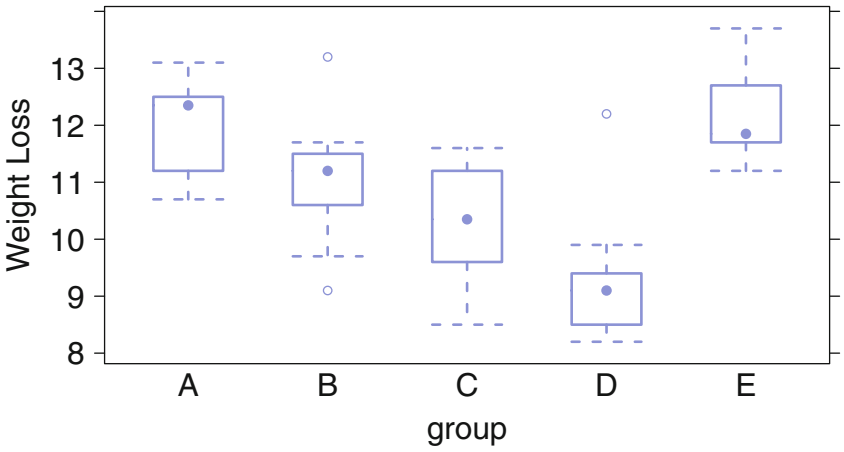

Fig. 7.1 Weightloss data: Boxplots of weight loss for each group.

Table 7.1 Weightloss ANOVA

```
> weightloss.aov <- aov(loss ~ group, data=weightloss)

> summary(weightloss.aov)
            Df Sum Sq Mean Sq F value  Pr(>F)
group        4   59.9   14.97    15.1 6.9e-08 ***
Residuals   45   44.7    0.99
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

When we regard agent D as the control, we seek to investigate whether any of the other four agents appear to promote significantly greater weight loss than agent D. From Figure 7.1 we see that the five populations are approximately normal with approximately the same variance. Therefore, we may proceed with the Dunnett procedure. Since we are investigating whether the other agents *improve* on D, we display infinite upper one-sided confidence intervals against D in Table 7.2 and Figures 7.2 and 7.3.

The (default) 95% confidence level in Table 7.2 applies simultaneously to all four confidence statements. The fact that all four confidence intervals lie entirely above zero suggests that D is significantly inferior to the other four weightloss agents.

Figure 7.3 is a mean–mean display of Dunnett's multiple comparison procedure applied to the weightloss data. Tabular results are shown in Table 7.3. Figure 7.3 is analogous to Figure 6.2 in Section 6.3. The mean–mean display technique is discussed in detail in Section 7.2. In Figure 7.3, reflecting the results for upper one-sided Dunnett confidence intervals, all horizontal lines except that for comparing

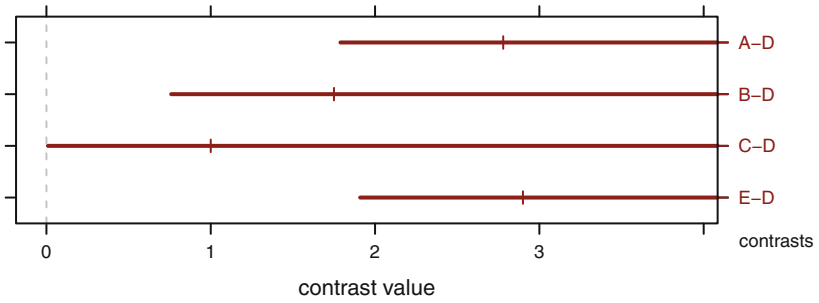

Fig. 7.2 Weightloss data: Standard display of one-sided multiple comparisons using the Dunnett method against the control treatment `D`.

groups `D` and `C` fall to the right of zero. The line for `C-D` sits on the boundary with lower limit .009. Consistent with the boxplots in Figure 7.1, we conclude that all weightloss agents (except possibly `C`) provide superior mean weight loss to that provided by agent `D`.

The Dunnett procedure is used in Exercises 7.7 and 12.4.

7.1.3.1 Computing Note—Specifying the Alternative Hypothesis

There are at least three conventions for indicating the alternative hypothesis. Be very clear which you are using.

As shown here, the `glht` function in R uses the argument `alternative` to indicate the alternative hypothesis. `glht` uses `alternative="greater"` to indicate an infinite upper bound, `alternative="less"` for an infinite lower bound, and defaults to `alternative="two-sided"`.

The S-Plus function `multicomp` uses the argument `bounds="lower"` to indicate a finite lower bound, implying an infinite upper bound. `multicomp` uses `bounds="upper"` for a finite upper bound, implying an infinite lower bound. For two-sided intervals `multicomp` defaults to `bounds="both"`.

SAS `PROC ANOVA` specifies the alternative hypothesis by using a different option name for each. SAS uses the option `dunnettu`, with the suffix "`u`" to indicate an infinite upper interval, the option `dunnettl` with the suffix "`l`" to indicate an infinite lower bound, and the option `dunnett` with no suffix for two-sided intervals.

Table 7.2 Weight loss using the Dunnett procedure.

```
> weightloss.dunnett <-
+ glht(weightloss.aov,
+      linfct=mcp(group=
+                 contrMat(table(weightloss$group), base=4)),
+      alternative = "greater")

> confint(weightloss.dunnett)

 Simultaneous Confidence Intervals

Multiple Comparisons of Means: User-defined Contrasts


Fit: aov(formula = loss ~ group, data = weightloss)

Quantile = -2.222
95% family-wise confidence level


Linear Hypotheses:
           Estimate lwr     upr
A - D <= 0 2.78000  1.78949     Inf
B - D <= 0 1.75000  0.75949     Inf
C - D <= 0 1.00000  0.00949     Inf
E - D <= 0 2.90000  1.90949     Inf
```

Table 7.3 MMC calculations for weightloss using the Dunnett procedure.

```
> weightloss.mmc <-
+   mmc(weightloss.aov,
+       linfct=mcp(group=
+                  contrMat(table(weightloss$group), base=4)),
+       alternative = "greater")

> weightloss.mmc
Dunnett contrasts
Fit: aov(formula = loss ~ group, data = weightloss)
Estimated Quantile = -2.222
95% family-wise confidence level
$mca
    estimate stderr    lower upper height
E-D     2.90   -Inf 1.909477   Inf  10.72
A-D     2.78   -Inf 1.789477   Inf  10.66
B-D     1.75   -Inf 0.759477   Inf  10.14
C-D     1.00   -Inf 0.009477   Inf   9.77
$none
  estimate stderr lower upper height
E    12.17   -Inf 11.47   Inf  12.17
A    12.05   -Inf 11.35   Inf  12.05
B    11.02   -Inf 10.32   Inf  11.02
C    10.27   -Inf  9.57   Inf  10.27
D     9.27   -Inf  8.57   Inf   9.27
```

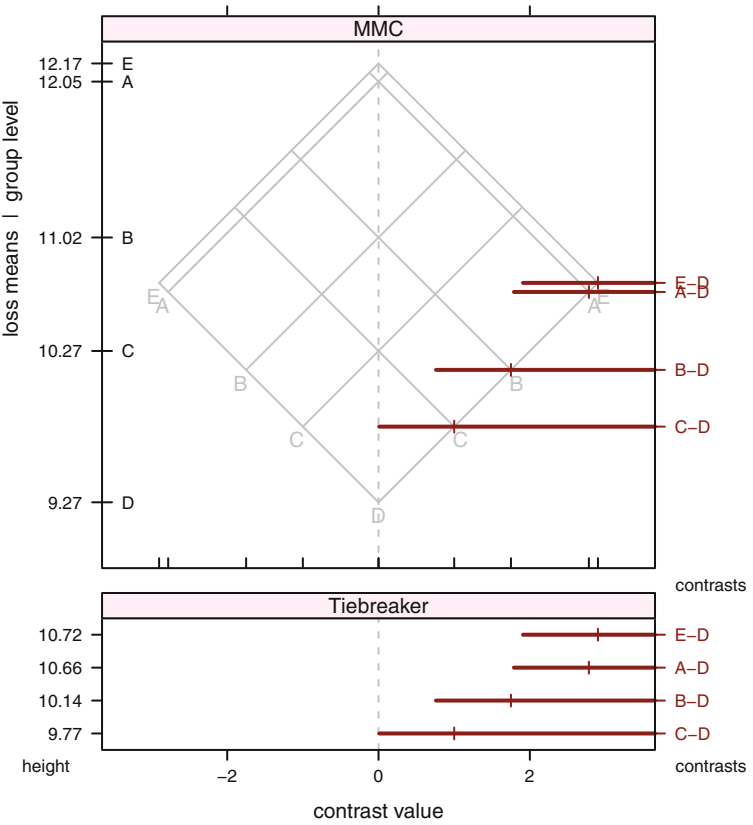

Fig. 7.3 Weightloss data: Mean–mean display of one-sided multiple comparisons using the Dunnett method against the control treatment `D`. The Tiebreaker panel is needed in this example because the `E-D` and `A-D` contrasts are at almost the same height in the top panel and are therefore overprinted. The similar heights for these two contrasts follow from the similar means for the `E` and `A` levels of the loss factor. Please see the discussion of the mean–mean display in Section 7.2.

7.1.4 Simultaneously Comparing All Possible Contrasts Scheffé and Extended Tukey

7.1.4.1 The Scheffé Procedure

In the context of comparing the means of a populations, the Scheffé multiple comparison procedure controls the familywise error rate over the infinite-sized family consisting of all possible contrasts $\sum_{j=1}^{a} c_j \mu_j$ involving the population means. The Scheffé procedure is therefore appropriate for exerting simultaneous error control

over the set of four contrasts in our analysis of the turkey data `data(turkey)` from Section 6.8. In exchange for maintaining familywise error control over so large a family, the Scheffé method gives rise to wide confidence limits and relatively unpowerful tests. Therefore, we recommend its use only in the narrowly defined situation of simultaneously inferring about mean contrasts more complex than a comparison of two means. The Scheffé procedure uses a percentile of an F distribution, derived as the distribution of the most significant standardized contrast among the sample means.

The confidence interval formula by the Scheffé procedure is

$$\text{CI}\left(\sum_{j=1}^{a} c_j \mu_j\right) = \sum_{j=1}^{a} c_j \bar{y}_j \pm \sqrt{(a-1)F_{.05,a-1,N-a}}\; s \sqrt{\sum_{j=1}^{a} \frac{c_j^2}{n_j}} \tag{7.1}$$

This provides the set of $100(1-\alpha)\%$ simultaneous confidence intervals for all possible contrasts among the population means. In this equation $N = \sum_{j=1}^{a} n_j$.

For R `glht`, we must manually calculate the critical value with, for example in the turkey data,

```
scheffe.quantile <- sqrt(4*qf(.95, 4, 25))
```

The Scheffé test is one of the methods available in the S-Plus `multicomp` function and is one of the options for the `MEANS` statement in SAS `PROC ANOVA`.

7.1.4.2 Scheffé Intervals with the Turkey Data

Table 6.8 provides F-tests of the hypotheses that the members of a basis set of four contrasts are zero. These four tests do not control for multiplicity. The finding in Table 6.8 is that three of these contrasts differ significantly from zero. We do not declare the fourth contrast significantly different from zero because its p-value exceeds 0.05.

The Scheffé procedure allows us to make inferences about these same contrasts while controlling for multiplicity. The confidence interval and testing results are shown in Table 7.4 and in Figure 7.4. An additional advantage of the Scheffé analysis is that the results specify the direction of contrasts' significant difference from zero. For example, in Table 7.4, the fact that the confidence interval on `A.vs.B` lies entirely below zero implies that, on average, the mean weight gain from diet `B` exceeds that from diet `A`. The F-statistics in Table 6.8 are essentially squared t-statistics, and this obscures information on directionality unless the definitions of the contrasts being tested are carefully examined alongside the test results.

We may use the results of the Scheffé analysis to assess the extent to which, if any, of the Scheffé simultaneous confidence intervals cause us to modify our previous conclusions about the contrasts. When doing so it is important to observe the contrast codings, that is, the numerical values defining the contrast. Observing that

Table 7.4 Scheffé Test for Turkey Data Contrasts. See also Figure 7.4.

```
> data(turkey)

> turkey.aov <- aov(wt.gain ~ diet, data=turkey)

> scheffe.quantile <- sqrt(4*qf(.95, 4, 25))

> turkey.lmat <-
+   cbind(control.vs.treatment=c(1,-.25,-.25,-.25,-.25),
+         A.vs.B              =c(0, .5,  .5, -.5, -.5 ),
+         amount              =c(0, .5, -.5,  .5, -.5 ),
+         A.vs.B.by.amount    =c(0, .5, -.5, -.5,  .5 ))

> row.names(turkey.lmat) <- row.names(contrasts(turkey$diet))

> turkey.mmc <- mmc(turkey.aov, calpha=scheffe.quantile, focus="diet",
+                   focus.lmat=turkey.lmat,
+                   estimate.sign=0, order.contrasts=FALSE)

> turkey.mmc$lmat
                     estimate stderr   lower  upper height
control.vs.treatment   -3.433 0.2563 -4.2849 -2.582  5.500
A.vs.B                 -1.950 0.2293 -2.7116 -1.188  7.217
amount                 -1.933 0.2293 -2.6950 -1.172  7.217
A.vs.B.by.amount        0.450 0.2293 -0.3116  1.212  7.217
```

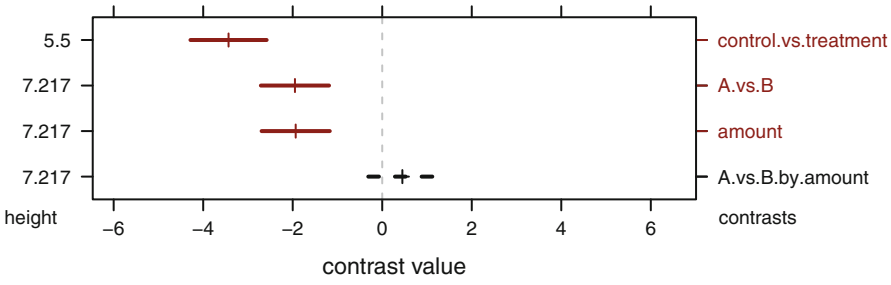

Fig. 7.4 Scheffé plot for turkey data. See also Table 7.4.

the first three of the four Scheffé intervals exclude 0 while the last one includes 0, the Scheffé results reinforce our original impressions from the nonsimultaneous F-tests of these contrasts in Table 6.8.

In this example, examination of the Scheffé results did not cause us to revise our earlier results ignoring multiplicity. In general, use of a multiple comparison procedure is an appropriately conservative approach that may not declare a difference found by nonsimultaneous tests or confidence intervals.

Figure 7.5 is a graphic presentation of the Scheffé procedure applied to comparisons of all pairs of means. We use Scheffé intervals here because these pairwise comparisons are part of a larger family of contrasts that includes those displayed in Figure 7.6. There are $10 = \binom{5}{2}$ pairwise differences among the means of the 5 diet combinations studied. Figure 7.5 is a mean–mean display of Scheffé simultaneous confidence intervals on these mean differences.

Figures 7.5 and 7.6 contain overprinting of the confidence lines and labels for several of their comparisons of level means. The overprinting in Figure 7.5 is due to almost identical mean values for levels `B1` and `A2`. The overprinting in Figure 7.6 is a consequence of the same almost identical mean values, now reflected as identical heights for the contrasts because the interaction of the `A.vs.B` and the `amount` comparisons is not significant. In situations with such overprinting, we augment the mean–mean display with a traditional display of these same confidence intervals. This *Tiebreaker* plot lists the contrasts in the same vertical order as in the mean–mean plot. The conclusions here, based on the fact that 9 of the 10 intervals lie entirely above zero, are

- For both amount `1` and `2`, the mean weight gain from additive `B` is significantly greater than the mean weight gain from additive `A`.
- For both additive `A` or `B`, the mean weight gain from amount `2` significantly exceeds the mean weight gain from amount `1`.
- The weight gain from the control diet is significantly below that from any of the other 4 diets.

We graphically summarize these conclusions with the orthogonal contrasts in Figure 7.6. The 3 contrasts that differ significantly from zero do not cross the vertical $d = 0$ axis. The nonsignificant contrast does cross the $d = 0$ axis.

Table 7.4 and Figure 7.4 show three of the user-defined contrasts to have negative estimates. Figure 7.6 shows those contrasts to be reversed to have positive contrasts. We believe that multiple comparisons are most easily interpreted when the means are sequenced in numerical order (not lexicographic order), and consequently that all displayed contrasts should compare the larger value to the smaller value. That is, all displayed contrast values should be positive. Such reversal of the direction of a contrast creates no problem when assessing how contrasts relate to zero so long as the reversal is noted. We note the reversal by appending a "–" to the names of the reversed contrasts.

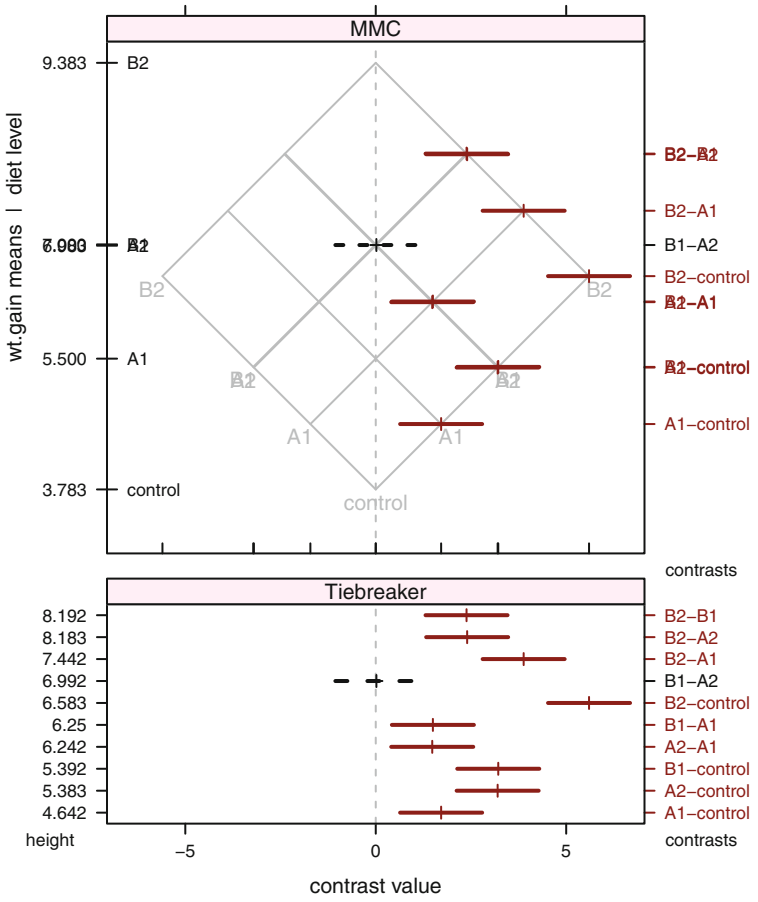

Fig. 7.5 MMC: mca plot for Turkey data. Overprinting of contrasts at the same height in the MMC panel are separated in the Tiebreaker panel by a standard multiple comparisons plot ordered to match the order of the MMC plot.

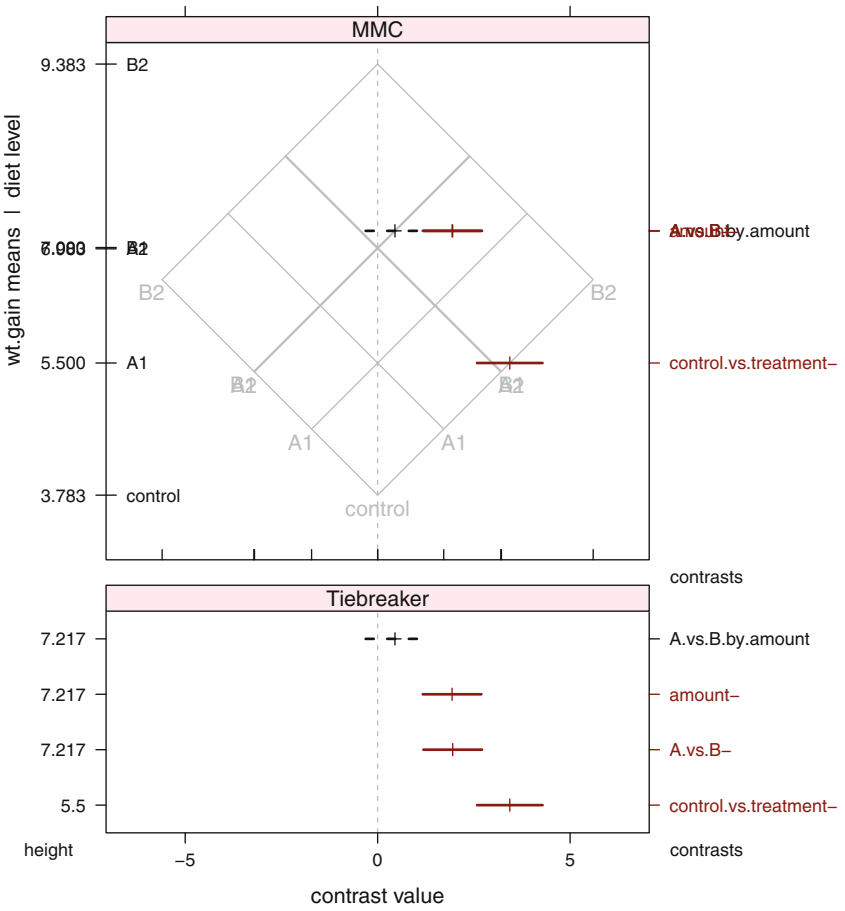

Fig. 7.6 MMC: Orthogonal basis set of contrasts for Turkey data. Overprinting of the confidence lines (for contrasts `A.vs.B-`, `amount-`, and `A.vs.B.by.amount`, in this example) and their labels in the right-axis labels of the MMC panel is a consequence of almost identical values for the group means in the left-axis labels (`A2` and `B1`). The overprinting is resolved in the Tiebreaker panel, a standard multiple comparisons plot (without information on the group means) ordered to match the order of the MMC plot. The contrasts in these panels are the same contrasts that appear in Figure 7.4, but negative estimates there have been reversed here. During the reversal a "`-`" was appended to contrast names for which it was not possible to figure out how to reverse the contrast name.

7.1.4.3 The Extended Tukey Procedure

The Tukey procedure can be extended to cover the family of all possible contrasts when the samples are of the same size n. Generalizing Equation (6.11) to any contrast vector (c_j) in the equal n case, we get

$$\mathrm{CI}\left(\sum_{j=1}^{a} c_j \mu_j\right) = \sum_{j=1}^{a} c_j \bar{y}_j \pm \frac{q_\alpha}{2} \frac{s}{\sqrt{n}} \sum_{j=1}^{a} |c_j| \tag{7.2}$$

as the set of $100(1-\alpha)\%$ simultaneous confidence intervals for all possible contrasts among the population means.

The q_α here is the same value used in Equation (6.11). Except for very simple contrasts, such as between pairs of means, these generalized Tukey intervals will be even wider than the analogous Scheffé intervals, Hochberg and Tamhane (1987). The generalized Tukey intervals (7.2) may be considered for use when interest lies in a family consisting of the union of all pairwise contrasts with a small number of more complicated contrasts.

As discussed in Hochberg and Tamhane (1987), the family encompassed by the generalized Tukey intervals also includes the set of individual intervals on each population mean,

$$\mathrm{CI}(\mu_j) = \bar{y}_j \pm q_\alpha \frac{s}{\sqrt{n_j}} \tag{7.3}$$

These intervals are illustrated for the artificial data in Figure 7.11.

7.2 The Mean–Mean Multiple Comparisons Display (MMC Plot)

7.2.1 Difficulties with Standard Displays

The conclusions from the application of the Tukey procedure to the catalyst data are not well conveyed by the standard tabular and graphical output shown in Table 7.5 and Figure 7.7. In both displays, the magnitudes of the sample means themselves are not shown. These displays are therefore not capable of depicting the relative distances between adjacent sorted sample means. Indeed, the standard display ignores the sample means entirely and instead sorts the contrasts alphabetically. Compare Table 7.5 to the `$mca` section of Table 6.3, and Figure 7.7 to the bottom panel of Figure 6.2.

Another standard display of results of a Tukey test, shown here in Figure 7.8, is often used to communicate results when sample sizes are equal. The sample means are listed in ascending magnitude. Straight-line segments are used to indicate significance according to the following rules. If two sample means are not covered by the same line segment, the corresponding population means are declared significantly different. If two sample means are covered by a common line segment, the corresponding population means are declared not significantly different.

Table 7.5 Tukey Multiple Comparisons for Catalyst Data—Standard Display (not showing means)

```
> catalystm.glht <-
+    glht(catalystm1.aov, linfct = mcp(catalyst = "Tukey"))

> confint(catalystm.glht)

 Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = concent ~ catalyst, data = catalystm)

Quantile = 2.966
95% family-wise confidence level


Linear Hypotheses:
           Estimate lwr    upr
B - A == 0 -1.125   -4.501  2.251
C - A == 0 -3.667   -7.342  0.009
D - A == 0 -5.775   -9.151 -2.399
C - B == 0 -2.542   -6.386  1.302
D - B == 0 -4.650   -8.209 -1.091
D - C == 0 -2.108   -5.952  1.736
```

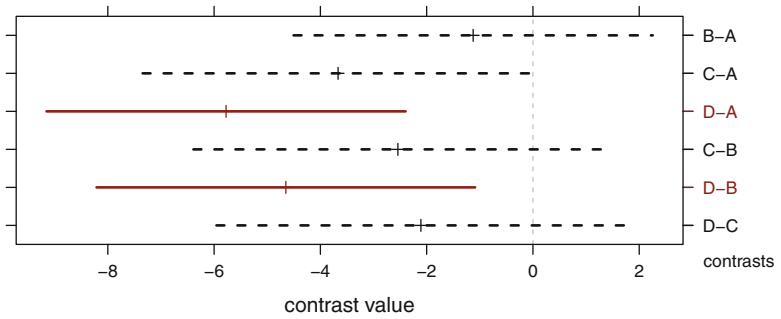

Fig. 7.7 All Pairwise Comparisons of Catalyst Means. In this standard display, the group means are not displayed. The contrasts are sorted alphabetically. Compare this figure to the bottom panel of Figure 6.2 where the contrasts are sorted by the values of the means being compared.

With this procedure it is difficult to depict correctly the relative distances between adjacent sorted sample means because the table is constrained by the limited resolution of a fixed-width typewriter font rather than the high resolution of a graphical display.

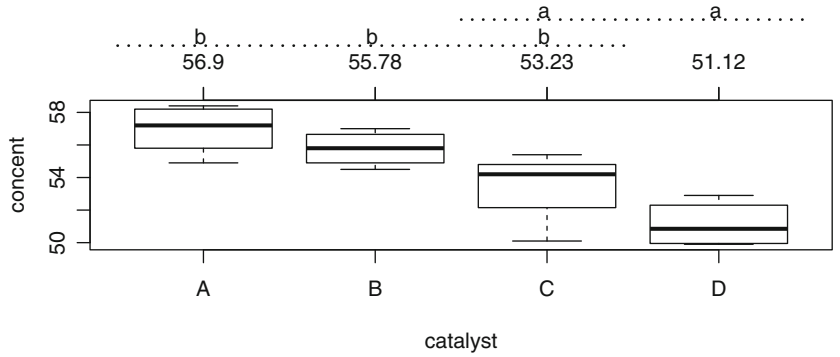

Fig. 7.8 This example is constructed from the `cld` function in the **multcomp** package. The function call `plot(cld(catalystm.glht))` draws the boxplots and the letter values. We manually (by supplementary code) placed the numerical values of the means and the underlines connecting the letters.

Further, the procedure cannot be used when sample sizes are unequal. Table 7.6 and Figure 7.9 illustrate this limitation using artificial data:

Group	N	Mean
A	5	2.0
B	100	2.1
C	100	2.8
D	5	3.0

The Tukey procedure shown in Table 7.6 uncovers a significant difference between the means of populations B and C, for which the sample sizes are large, but no significant difference between the means of populations A and D, for which the sample sizes are small. With the lines-type graph in Figure 7.9 the nonsignificant difference between the means of A and D requires that a common line covers the range from 2.0 to 3.0, including the location of the means of groups B and C. The presence of this line contradicts the finding of a significant difference between the means of groups B and C that is seen in the standard displays in Figures 7.10 and 7.11 and in the mean–mean display (described in Section 7.2.2) in Figure 7.12.

Table 7.6 Simultaneous confidence intervals on all pairs of mean differences. The means of samples B and C both lie between the means of samples A and D. This example is based on highly unbalanced artificial data. Sample sizes were 100 from populations B and C and 5 from populations A and D. The Tukey procedure finds a significant difference between the means of populations B and C but no significant difference between the means of populations A and D.

```
> group <- factor(LETTERS[1:4])

> n <- c(5,100,100,5)

> ybar <- c(2, 2.1, 2.8, 3)

> inconsistent.aov <- aovSufficient(ybar ~ group, weights=n, sd=.8)

> anova(inconsistent.aov)
Analysis of Variance Table

Response: ybar
           Df Sum Sq Mean Sq F value  Pr(>F)
group       3     27    9.01    14.1 2.2e-08 ***
Residuals 206    132    0.64
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

> inconsistent.glht <-
+    glht(inconsistent.aov, linfct=mcp(group="Tukey"),
+         vcov.=vcovSufficient, df=inconsistent.aov$df.residual)

> crit.point <- qtukey(.95, 4, 206)/sqrt(2)

> confint(inconsistent.glht, calpha=crit.point)

 Simultaneous Confidence Intervals

Multiple Comparisons of Means: Tukey Contrasts


Fit: aov(formula = formula, data = data, weights = weights, x = TRUE)

Quantile = 2.59
95% confidence level


Linear Hypotheses:
           Estimate lwr     upr
B - A == 0  0.1000  -0.8496  1.0496
C - A == 0  0.8000  -0.1496  1.7496
D - A == 0  1.0000  -0.3105  2.3105
C - B == 0  0.7000   0.4070  0.9930
D - B == 0  0.9000  -0.0496  1.8496
D - C == 0  0.2000  -0.7496  1.1496
```

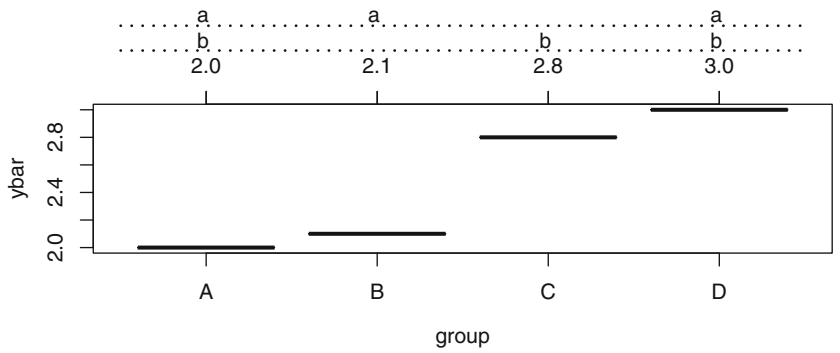

Fig. 7.9 Underlining of means that are not significantly different. Both the `a` and `b` lines, which are valid for the comparison of catalysts A and D (in this example based on the low precision test for small sample sizes), mask the significant difference between catalysts B and C (based on a much higher precision test for much larger sample sizes).

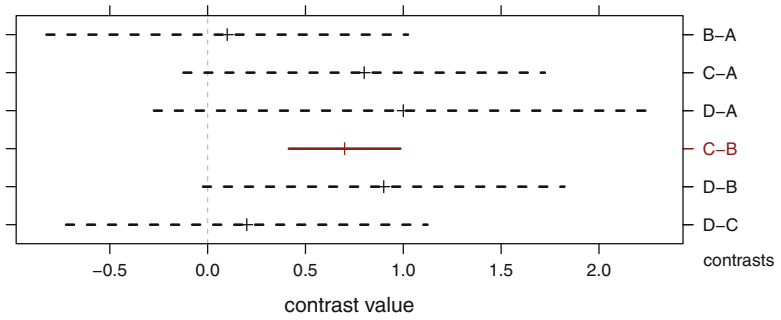

Fig. 7.10 Simultaneous confidence intervals on all pairs of mean differences. The means of samples B and C both lie between the means of samples A and D. Sample sizes were 100 from populations B and C and 5 from populations A and D. The Tukey procedure finds a significant difference between the means of populations B and C but no significant difference between the means of populations A and D. The short confidence interval for the `C-A` contrast reflects the higher precision of the contrasts based on the larger sample sizes.

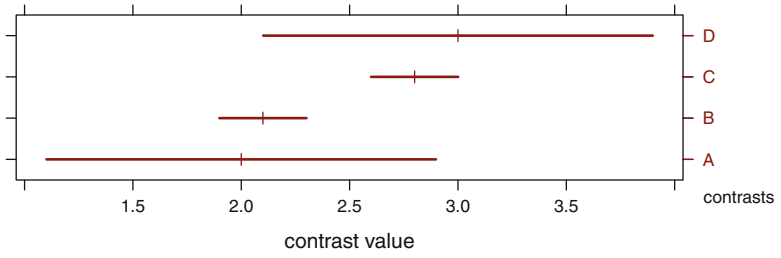

Fig. 7.11 The means of samples B and C both lie between the means of samples A and D. Sample sizes were 100 from populations B and C and 5 from populations A and D. The Tukey procedure finds a significant difference between the means of populations B and C but no significant difference between the means of populations A and D. The underlying formula for these intervals appears in Equation (7.3).

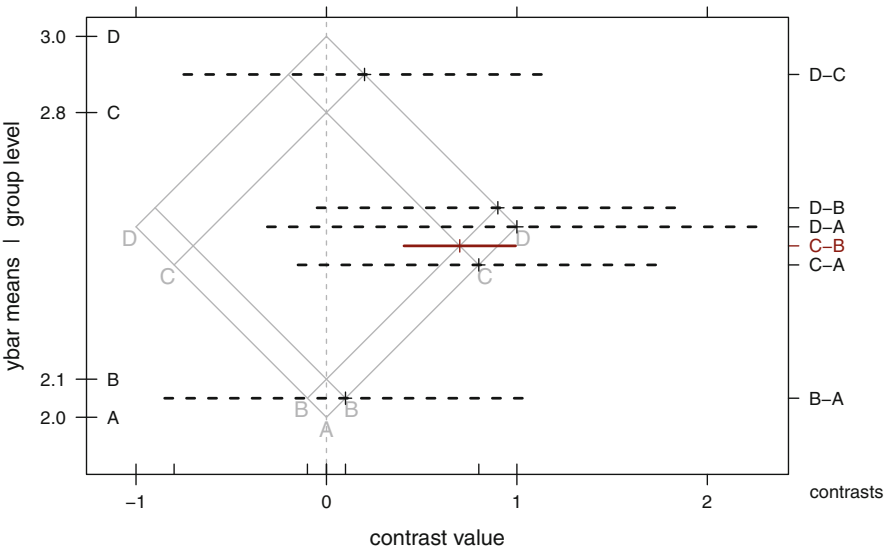

Fig. 7.12 A mean–mean display (MMC plot described in Section 7.2.2) of simultaneous confidences on the means from populations A, B, C, D in the artificial data. Each confidence interval on a mean difference is represented by a horizontal line. If and only if an horizontal line crosses the vertical "contrast value = 0" line, the corresponding population mean difference is declared nonsignificant. In this display we use dashed black lines for nonsignificant comparisons and solid red lines for significant comparisons. This display shows the relative differences between sample means and allows for unequal sample sizes. The short confidence interval for the `C-A` contrast reflects the higher precision of the contrasts based on the larger sample sizes.

7.2.2 Hsu and Peruggia's Mean–Mean Scatterplot

Hsu and Peruggia (1994) address the deficiencies in standard displays of multiple comparison procedures with their innovative graphical display of the Tukey procedure for all pairwise comparisons. In Section 7.2.2.1 we show the details of the construction of the MMC plot displayed in Figure 7.13. We postpone interpretation of Figure 7.13 until Section 7.2.2.2.

In Section 7.2.3 we extend their display to show other multiple comparison procedures for arbitrary sets of contrasts. Software for our extension is included in the **HH** package as function `mmc` and its related functions.

7.2.2.1 Construction of the Mean–Mean Scatterplot

We begin with data-oriented orthogonal h- and v-axes in Figures 7.14 and 7.15 and then move to rotated difference $(h - v)$ and mean $(h + v)/2$ axes in Figure 7.16. The rotations by 45° introduce factors of $\sqrt{2}$ that are there to maintain the orthogonality of h and v in the rotated coordinates.

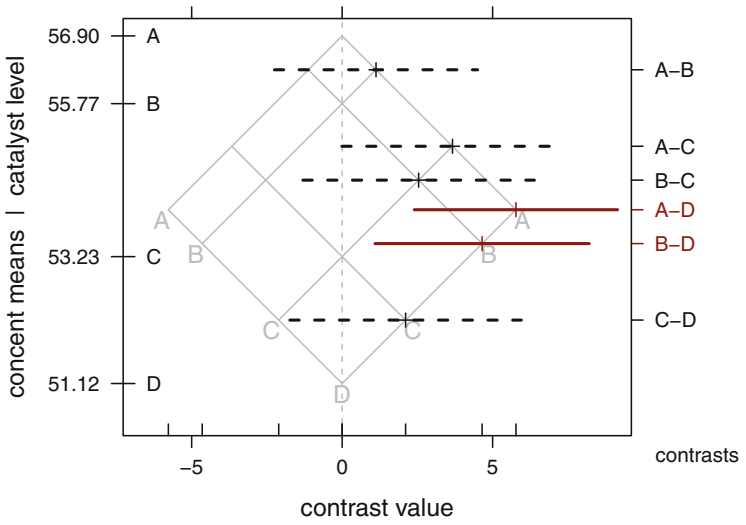

Fig. 7.13 Multiple comparisons of all pairwise comparisons of catalyst means with the MMC display. This is a repeat of the top panel of Figure 6.2.

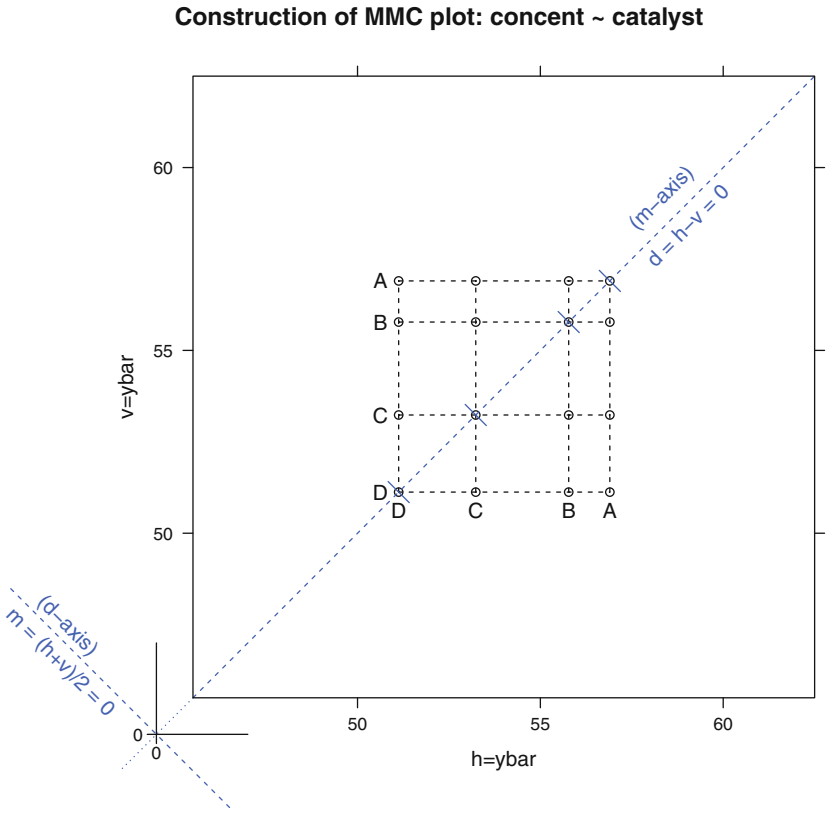

Fig. 7.14 Construction of mean–mean multiple comparisons plot for the catalyst data. Data-oriented axes and isomeans grid, steps 1–6 in the discussion in Section 7.2.

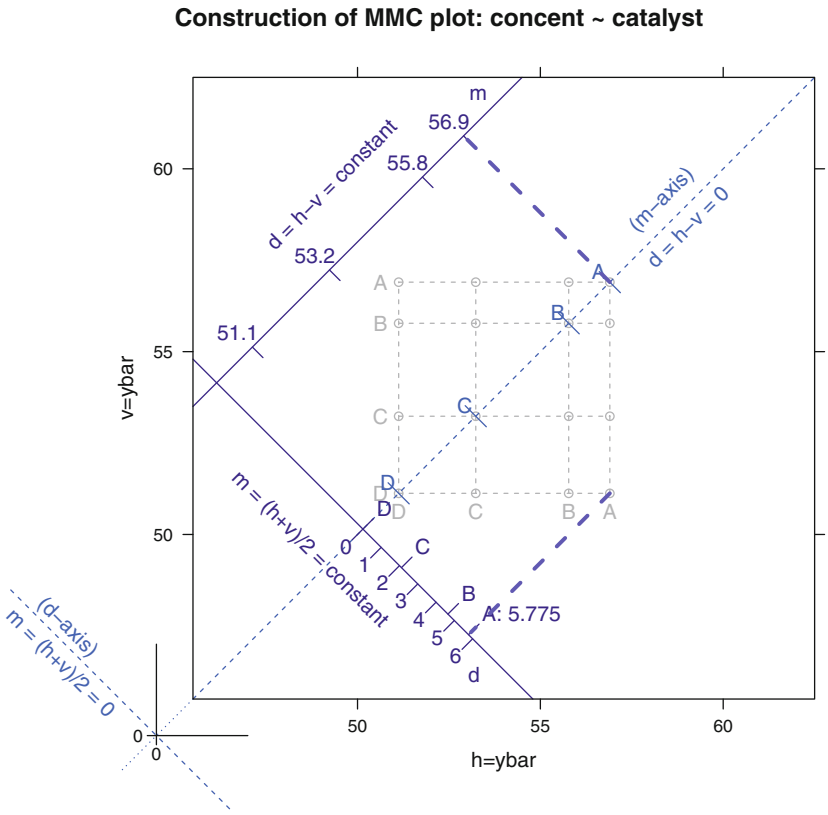

Fig. 7.15 Construction of mean–mean multiple comparisons plot for the catalyst data. Data-oriented axes, steps 7–9 in the discussion in Section 7.2.

1. Draw a square plot in Figure 7.14 on (h, v)-axes. Define (d, m)-axes at $\pm 45°$.
2. Plot each $\bar{y}_i$ against $\bar{y}_j$.
3. Connect the points with $h = \bar{y}_i$ and $v = \bar{y}_j$ lines. The lines are labeled with the level names of the group means. We call this the *isomeans grid*. It is used as the background reference for the MMC plot.
4. Draw the $45°$ line $h = v$. Define the value $d = h - v$, where the letter d indicates differences between group means. The line we just drew corresponds to $d = 0$. We will call this line the m-axis, where the name $m = (h + v)/2$ indicates the group means.
5. Place tick marks on the m-axis at the points $(\bar{y}_i, \bar{y}_i)$.
6. Draw the $-45°$ line through the origin $(h = 0, v = 0)$. The line we just drew corresponds to $m = 0$. We will call this line the d-axis, where the name d indicates the differences.

7. Copy Figure 7.14 to Figure 7.15.
8. Draw another m-axis parallel to the $d = 0$ line. Drop a perpendicular from the $(\bar{y}_A, \bar{y}_A)$ intersection on the $d = 0$ line to the new m-axis. Place a tick at that point and label it with the $m = \bar{y}_A$ value. Place similar tick marks at the heights $m = \bar{y}_i$. (The actual distances from the $m = 0$ line to the tick marks are $\bar{y}_i \sqrt{2}$.)
9. Draw another d-axis parallel to the line $m = 0$. We will place two sets of tick marks on the new d-axis: at the projections of the observed differences $(h, v) = (\bar{y}_i, \min_i(\bar{y}_i))$, and at unit intervals on the difference scale. Drop a perpendicular from the $(\bar{y}_A, \bar{y}_D)$ intersection to the new d-axis. Place a tick at that point and label it with the level name A and the value $\bar{y}_A - \bar{y}_D$. Place similar ticks at the distances $\bar{y}_i - \bar{y}_D$. (The actual distances from the $d = 0$ line to the tick marks are $(\bar{y}_i - \bar{y}_D)/\sqrt{2}$.) Place ticks below the d-axis at the distances $(0, 1, 2, 3, 4, 5, 6)/\sqrt{2}$ and label them $(0, 1, 2, 3, 4, 5, 6)$.
10. Rotate Figure 7.15 counterclockwise by 45° to get Figure 7.16.
11. Construct the confidence intervals. We show just one pairwise interval, the one centered on the point $\big(d = \bar{y}_B - \bar{y}_D,\ m = (\bar{y}_B + \bar{y}_D)/2\big)$. The confidence interval line is parallel to the d-axis at a height equal to the average of the two observed means. The interval is on the d-scale and covers all points $(\bar{y}_B - \bar{y}_D) \pm \hat{\sigma}\, q \sqrt{1/n_B + 1/n_D}$, where $\hat{\sigma}$ is the standard deviation from the ANOVA table and q is the critical value used for the comparison. In this example we use the critical value $q = q_{.05,4,12}/\sqrt{2} = 2.969141$ from the Studentized range distribution.
12. We show all $\binom{4}{2} = 6$ pairwise differences $\bar{y}_i - \bar{y}_j$ with their confidence intervals in Figure 7.13.

Figures 7.14, 7.15, and two additional intermediate figures were drawn with function `HH:::mmc.explain`, an unexported but accessible function in the **HH** package. Figure 7.16 is an ordinary MMC plot with an `lmat` matrix indicating exactly one contrast. The code for all figures is included in file `HHscriptnames(7)`.

7.2.2.2 Interpretation of the Mean–Mean Scatterplot

We construct the background of Figures 7.16 and 7.13 by rotating Figure 7.15 counterclockwise by 45° and suppressing the h- and v-axes. The horizontal d-axis shows the values of the contrasts and the vertical m-axis shows the average values of the two means being contrasted.

In Figure 7.13, each mean pair $(\bar{y}_i, \bar{y}_j)$ is plotted on the now-diagonal (h, v)-axes and can also be identified with its (d, m)-coordinates. In Figure 7.16, we focus on the pair of means $\bar{y}_B$ and $\bar{y}_D$. We begin with the (h, v)-system and identify the point as

$$(h, v) = (\bar{y}_B, \bar{y}_D) = (55.8,\ 51.1)$$

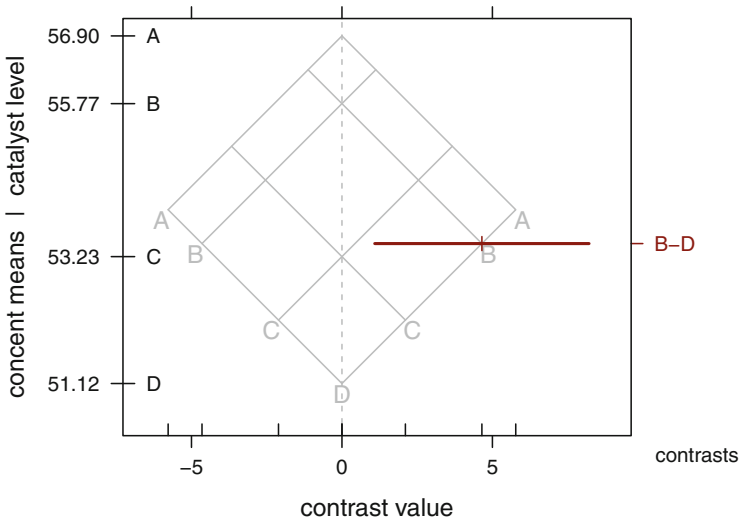

Fig. 7.16 Construction of mean–mean multiple comparisons plot for the catalyst data. Difference and mean-oriented axes. This figure shows steps 1–11 in the discussion in Section 7.2. This figure is essentially the same as Figure 7.15 with a single contrast and rotated 45° counterclockwise. This figure shows only one of the six pairwise contrasts. All six contrasts and the result of all 12 construction steps are shown in Figure 7.13.

The coordinates of the same pair of means $(\bar{y}_B, \bar{y}_D)$ in the (d, m)-system are

$$\begin{aligned}(d, m) &= \big(\bar{y}_B - \bar{y}_D,\ (\bar{y}_B + \bar{y}_D)/2\big) \\ &= \big((55.8 - 51.1),\ (55.8 + 51.1)/2\big) = (4.65,\ 53.45)\end{aligned}$$

We choose to label the ticks on the m-axis by the means because they are more easily interpreted: The confidence interval on $\bar{y}_B - \bar{y}_D$ is at the mean height $m = (\bar{y}_B + \bar{y}_D)/2$ in Figure 7.16. Hsu and Peruggia label the ticks on the m-axis by the sum $\bar{y}_B + \bar{y}_D = 2m$ because one unit on the $2m$-scale takes exactly the same number of inches as one unit on the d-scale.

Figure 7.13 is constructed from Figure 7.16 by including all of the $\binom{4}{2} = 6$ pairwise differences $\bar{y}_i - \bar{y}_j$, not just the single difference we use for the illustration.

Each of the confidence intervals for the $\binom{4}{2} = 6$ pairwise differences $\bar{y}_i - \bar{y}_j$ in Figure 7.13 is centered at a point whose height on the vertical m-axis is equal to the average of the corresponding means $\bar{y}_i$ and $\bar{y}_j$ and whose location along the horizontal d-axis is at distance $\bar{y}_i - \bar{y}_j$ from the vertical line $d = 0$. Horizontal lines are drawn at these heights so that the midpoints of these lines intersect their $(h = \bar{y}_i, v = \bar{y}_j)$ intersection. The width of each horizontal line is the width of a confidence interval estimating the difference $\bar{y}_i - \bar{y}_j$. By default the endpoints of the line are chosen to be the endpoints of the 95% two-sided confidence interval chosen by the Tukey procedure for all $\binom{4}{2}$ possible pairs.

If a horizontal confidence interval line crosses the vertical $d = 0$ line, the mean difference is declared not significant. Otherwise the mean difference is declared significant. If an end of a horizontal line is close to the vertical $d = 0$, this says that the declaration of significance was a close call.

When the critical value q is chosen by one of the standard multiple comparisons procedures (we illustrate with and default to the Tukey procedure), the widths of the horizontal confidence interval lines are the simultaneous confidence intervals for the six pairs of population mean differences. This depiction is not restricted to the case of equal sample sizes and hence equal interval widths.

The display in Figure 7.13 has several advantages over traditional presentations of Tukey procedure results. In a single graph we see

1. The means themselves, with correct relative distances,
2. The point and interval estimates of the $\binom{4}{2}$ pairwise differences,
3. The point and interval estimates for arbitrary contrasts of the level means,
4. Declarations of significance,
5. Confidence interval widths that are correct when the sample sizes are unequal.

7.2.3 Extensions of the Mean–Mean Display to Arbitrary Contrasts

Heiberger and Holland (2006) extend the mean–mean multiple comparisons plot to arbitrary contrasts, that is, contrasts that are not limited to the set of pairwise comparisons.

Two critical issues needed to be addressed. The first is the scaling of the contrast and the second is the set of contrasts selected for consideration.

7.2.3.1 Scaling

The standard definition of a contrast in Equation (6.20) requires that it satisfy the zero-sum constraint Equation (6.18). The variance of the contrast is calculated with Equation (6.21).

When we calculate sums of squares and F-tests, this definition is sufficient. When we wish to plot arbitrary contrasts on the mean–mean multiple comparisons plot described in Section 7.2.2, the contrasts must be comparably scaled. The heights must be in the range of the observed $\bar{y}_j$, and all confidence intervals must fall inside the range of the d-axis. To satisfy this additional requirement, we need to require the absolute-sum-2 scaling introduced in Section 6.9.2.1 and made explicit in Equation (6.22). Any other scaling makes it impossible to fit these values on the mean–mean plot.

With the absolute-sum-2 scaling we can think of any contrast as the comparison of two weighted averages of $\bar{y}_j$. Let us call them $\bar{y}_+ = \sum c_j^+ \bar{y}_j$ and $\bar{y}_- = \sum c_j^- \bar{y}_j$, where we use the superscript notation $a^+ = \max(a, 0)$ and $a^- = \max(-a, 0)$. We illustrate with the contrast comparing the average of means $\bar{y}_A$ and $\bar{y}_B$ with the mean $\bar{y}_D$.

	A	B	C	D	$\bar{y}_+$	$\bar{y}_-$
absolute-sum-2	.5	.5	0	-1	$(\bar{y}_A + \bar{y}_B)/2$	$\bar{y}_D$
integer	1	1	0	-2		
normalized	$1/\sqrt{6}$	$1/\sqrt{6}$	0	$-2/\sqrt{6}$		

We plot the contrast centered at the (h, v)-location $(\bar{y}_-, \bar{y}_+)$, where each term is at the correctly weighted average of the observed $\bar{y}_j$-values. The height on the m-axis of the MMC plot is $(\bar{y}_+ + \bar{y}_-)/2$ and the difference on the d-axis is $\bar{y}_+ - \bar{y}_-$. The confidence interval widths are proportional to the standard error of $\bar{y}_+ - \bar{y}_-$, which, from (6.21), is proportional to $\sqrt{\sum c_j^2/n_j}$.

7.2.3.2 Contrasts

The simplest set of contrasts is the set of all pairwise comparisons $\bar{y}_i - \bar{y}_j$ (as in Figure 6.2). Others sets include comparisons $\bar{y}_j - \bar{y}_0$ of all treatment values to a control (as in Figure 7.3) and a basis set of orthogonal contrasts that span all possible contrasts (as will be seen in Figure 7.17).

7.2.3.3 Labeling

Our presentation of the MMC plot, for example in Figure 6.2, has improved labeling compared to the Hsu and Peruggia presentation.

The left-axis ticks are the $\bar{y}_i$-values themselves, at the heights of the intersections of the 45° h- and v-lines with the vertical $d = 0$ line. The labels on the outside of the left axis are the $\bar{y}_i$-values. The labels on the inside of the left axis are the names of the factor levels.

The right-axis labels belong to the horizontal CI lines for the contrasts. The labels outside the right axis are the automatically generated contrasts, either pairwise $\bar{y}_i - \bar{y}_j$ or comparisons $\bar{y}_j - \bar{y}_0$ of all treatment values to a control. The labels inside the right axis are the requested contrasts from the explicitly specified `lmat` matrix. Each CI line is at the height corresponding to the average of the two $\bar{y}_*$ $\big((\bar{y}_i + \bar{y}_j)/2$ or $(\bar{y}_+ + \bar{y}_-)/2\big)$ values they are comparing. Each CI line is centered at the observed difference

$\left((\bar{y}_i - \bar{y}_j) \quad \text{or} \quad (\bar{y}_+ - \bar{y}_-)\right)$. The half-width of the (two-sided) CI line is $qs_{\bar{y}_i - \bar{y}_j}$, where q is calculated according to the specified multiple comparisons criterion.

The bottom axis is in the difference $\bar{y}_i - \bar{y}_j$ d-scale. The ticks and labels outside the bottom axis are regularly spaced values on the difference scale. The ticks inside the bottom axis, at distances $\pm|\bar{y}_j - \min_j \bar{y}_j|$, correspond to the horizontal d-axis positions of the foot of the 45° h- and v-lines. The names of the factor levels appear at the foot of each 45° line.

7.2.3.4 q Multipliers

Hypothesis test and confidence interval formulas, introduced in Chapter 3, depend on a multiple of the standard deviation. The multiplier is a quantile chosen from an appropriate distribution. When only one hypothesis is tested or only one interval is constructed, the multiplier is denoted z when the test statistic is normally distributed and t when the test statistic is from a t distribution. Multipliers denoted q, sometimes with a subscript, are used in many of this chapter's formulas for confidence intervals and rules for rejecting null hypotheses. In both Sections 7.1.2 and 7.1.4.3 discussing Tukey procedures, and in plots in Section 7.2 displaying results from these procedures, q refers to the Studentized range distribution. The multiplier used in the Dunnett procedure of Section 7.1.3 is a percentile of a marginal distribution of a multivariate t distribution. The multiplier for the Scheffé procedure is the square root of a percentile of an F distribution. For details, see Hochberg and Tamhane (1987).

7.2.4 Display of an Orthogonal Basis Set of Contrasts

The sum of squares associated with the factor A with a levels has $a - 1$ degrees of freedom. The missing degree of freedom is associated with the grand mean and is normally suppressed from the ANOVA table.

In Section 6.8.3 we note that it is always possible to construct an orthogonal set of contrasts that decompose the $a - 1$ df sum of squares for an effect into $a - 1$ independent single-df sums of squares. In this section we illustrate the mathematics for constructing an orthogonal basis set by constructing one from the set of pairwise contrasts. From this basis set, we show that we can construct any other set of contrasts. We also show that an orthogonal basis set, augmented with an additional contrast for the grand mean (not actually a contrast since it doesn't sum to 0), can be used to construct any linear combination of the group means.

This discussion uses all the matrix algebra results summarized in Appendix Section I.4. This section is placed here in Chapter 7 because it belongs to the discussion of the MMC plots. It might be more easily read after Section 10.3 where contrast matrices and their relation to dummy variables are discussed.

We illustrate the discussion with the catalyst data in `data(catalystm)`. We begin with the set of pairwise contrasts behind the construction of Figure 7.13. We isolate the contrasts implicit in the `"mmc"` object with the `lmatPairwise` function in Table 7.7.

Table 7.7 Contrast matrix for pairwise comparisons. There are three matrices displayed here. The first is the contrasts for the `catalyst` factor as used by the `aov` function. We show the default contrasts as defined by the `contr.treatment` function. The first level is omitted. Note that these are not 'contrasts' as defined in the standard theory for linear models as they are not orthogonal to the intercept. Then the contrast matrix for pairwise comparisons is displayed in two different structures. The `glht` function uses the `linfct` (linear function) format. Each row is the difference of two columns of the `contr.treatment` matrix. The last matrix, structured to be used in the `focus.lmat` argument to the `mmc` function, shows columns which are standard contrasts (each column sums to zero).

```
> ## aov contrast matrix for catalyst factor.  The columns are
> ## constructed by contr.treatment with the default base=1
> contrasts(catalystm$catalyst)
  B C D
A 0 0 0
B 1 0 0
C 0 1 0
D 0 0 1

> ## Linear function used internally by glht for pairwise contrasts.
> ## The rows of linfct are the differences of the columns
> ## of the contrast matrix.
> catalystm.mmc$mca$glht$linfct
    (Intercept) catalystB catalystC catalystD
A-B           0        -1         0         0
A-C           0         0        -1         0
B-C           0         1        -1         0
A-D           0         0         0        -1
B-D           0         1         0        -1
C-D           0         0         1        -1

> ## Contrasts in lmat format, each column sums to zero.
> ## The last three rows are the transpose of the last three columns
> ## of the linfct matrix.
> ## The first row is prepended to make the column sum be zero.
> catalyst.pairwise <- lmatPairwise(catalystm.mmc)

> catalyst.pairwise
  A-B A-C B-C A-D B-D C-D
A   1   1   0   1   0   0
B  -1   0   1   0   1   0
C   0  -1  -1   0   0   1
D   0   0   0  -1  -1  -1
```

Table 7.8 illustrates an orthogonal basis set of contrasts for the catalyst data. This examination of 3 linearly independent contrasts succinctly summarizes the information contained in the 3 degrees of freedom for comparing the means of the 4 levels of the fixed factor `catalyst`. For completeness we show that `catalystm.lmat` and `catalyst.pairwise` span the same subspace.

Table 7.8 The orthogonal contrast matrix `catalysm.lmat` contains three columns that decompose the 3-df `catalyst` sum of squares term into three single-df sums of squares. The `crossprod` shows that `catalystm.lmat` is an orthogonal rank-3 matrix. The zero residuals from the regression of `catalystm.lmat` on `catalyst.pairwise` shows that they span the same subspace.

```
> ## An orthogonal set of ($4-1$) contrasts for the catalyst factor.
> ## user-specified contrasts       A  B  C  D
> catalystm.lmat <- cbind("AB-D" =c(1, 1, 0,-2),
+                         "A-B"  =c(1,-1, 0, 0),
+                         "ABD-C"=c(1, 1,-3, 1))

> dimnames(catalystm.lmat)[[1]] <- levels(catalystm$catalyst)

> catalystm.lmat
  AB-D A-B ABD-C
A    1   1     1
B    1  -1     1
C    0   0    -3
D   -2   0     1

> crossprod(catalystm.lmat)
      AB-D A-B ABD-C
AB-D     6   0     0
A-B      0   2     0
ABD-C    0   0    12

> catalyst.pairwise
  A-B A-C B-C A-D B-D C-D
A   1   1   0   1   0   0
B  -1   0   1   0   1   0
C   0  -1  -1   0   0   1
D   0   0   0  -1  -1  -1

> resid(lm(catalystm.lmat ~ catalyst.pairwise))
  AB-D A-B ABD-C
A    0   0     0
B    0   0     0
C    0   0     0
D    0   0     0
```

In Table 7.9 and Figure 7.17 we use the orthogonal basis to construct an easily interpretable MMC plot on the `catalyst` levels. The principal conclusion from Figure 7.13 is that the means of both catalysts `A` and `B` significantly exceed the mean of catalyst `D`. Figure 7.17 reforces this conclusion with the finding that the average of the means of catalysts `A` and `B` significantly exceeds the mean of catalyst `D` because the confidence interval for this contrast lies entirely above 0. A second new conclusion from Figure 7.17 is that the average of the means of catalysts `A`, `B`, and `D` is not significantly different from the mean of catalyst `C` because the confidence interval for this contrast includes 0.

Table 7.9 We use `catalysm.lmat` as the `focus.lmat` argument to `mmc` leading to Figure 7.17.

```
> catalystm.mmc <-
+    mmc(catalystm1.aov,
+        linfct = mcp(catalyst = "Tukey"),
+        focus.lmat=catalystm.lmat)

> catalystm.mmc
Tukey contrasts
Fit: aov(formula = concent ~ catalyst, data = catalystm)
Estimated Quantile = 2.966
95% family-wise confidence level
$mca
    estimate stderr     lower upper height
A-B    1.125  1.138 -2.251228 4.501  56.34
A-C    3.667  1.239 -0.008905 7.342  55.07
B-C    2.542  1.296 -1.302338 6.386  54.50
A-D    5.775  1.138  2.398772 9.151  54.01
B-D    4.650  1.200  1.091143 8.209  53.45
C-D    2.108  1.296 -1.735671 5.952  52.18
$none
  estimate stderr lower upper height
A    56.90 0.7590 54.65 59.15  56.90
B    55.77 0.8485 53.26 58.29  55.77
C    53.23 0.9798 50.33 56.14  53.23
D    51.12 0.8485 48.61 53.64  51.12
$lmat
      estimate stderr  lower upper height
A-B      1.125  1.138 -2.251 4.501  56.34
ABD-C    1.367  1.088 -1.860 4.594  53.92
AB-D     5.212  1.022  2.182 8.243  53.73
```

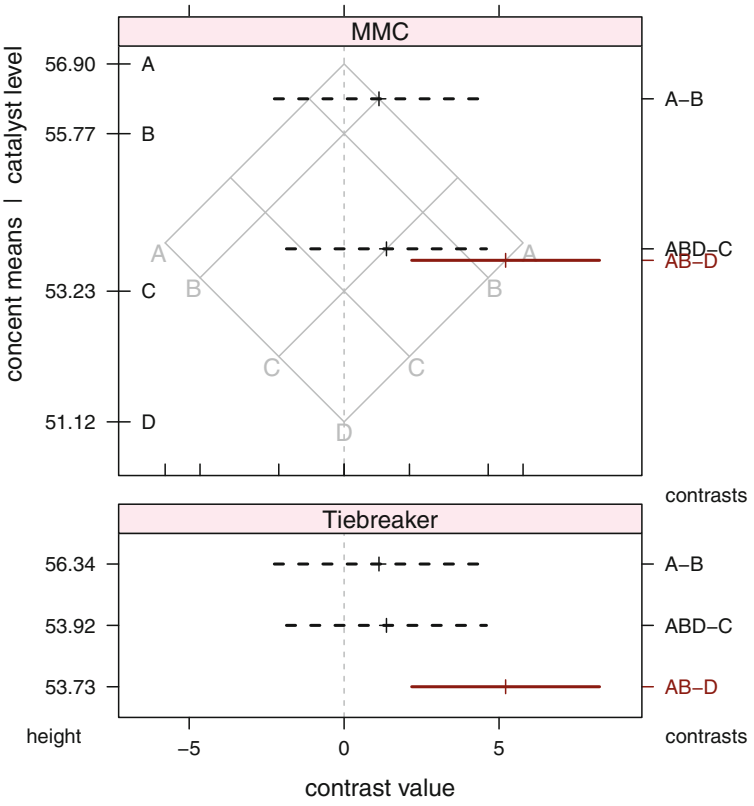

Fig. 7.17 MMC plot constructed with

```
mmcplot(catalystm.mmc, type="lmat", style="both")
```

using the orthogonal set of contrasts defined in Table 7.8 based on the pairwise set in Figures 7.13 and 6.2. The comparison between the average of $\bar{y}_A$ and $\bar{y}_B$ with the mean $\bar{y}_D$ is the only significant comparison. The other two confidence intervals include 0. The Tiebreaker panel is needed to respond to the overprinting of labels in the right axis of the MMC panel.

7.2.5 Hsu and Peruggia's Pulmonary Example

This is the example that Hsu and Peruggia (1994) use to introduce the mean–mean multiple comparisons plots. The response variable is `FVC`, forced vital capacity. The groups are levels of the `smoker` factor

Table 7.10 ANOVA table for pulmonary data.

```
> data(pulmonary)

> pulmonary
   smoker   n  FVC    s
NS     NS 200 3.35 0.63
PS     PS 200 3.23 0.46
NI     NI  50 3.19 0.52
LS     LS 200 3.15 0.39
MS     MS 200 2.80 0.38
HS     HS 200 2.55 0.38

> pulmonary.aov <-
+  aovSufficient(FVC ~ smoker, data=pulmonary,
+                weights=pulmonary$n, sd=pulmonary$s)

> summary(pulmonary.aov)
             Df Sum Sq Mean Sq F value Pr(>F)
smoker        5   89.3   17.85    83.9 <2e-16 ***
Residuals  1044  222.1    0.21
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

NS nonsmokers
PS passive smokers
NI noninhaling smokers
LS light smokers (1–10 cigarettes per day for at least the last 20 years)
MS moderate smokers (11–39 cigarettes per day for at least the last 20 years)
HS heavy smokers (≥40 cigarettes per day for at least the last 20 years)

There are six levels of the `smoker` factor, hence 5 df for comparing them. The means for the six groups are accessed as `data(pulmonary)`. The ANOVA table is in Table 7.10. The MMC plot is in Figure 7.18. The MMC plot of a set of orthogonal contrasts is in Figure 7.19.

Figure 7.18 shows that the three levels {PS, NI, and LS} are indistinguishable; we call this the low-smoker cluster. This comparison of three levels uses 2 df. There are only 3 df left. From the SW–NE HS line, we see that the MS-HS contrast is significant, that the comparisons between each of the three levels in the low-smoker cluster with MS is significant, and that the comparison of NS with HS and with MS are each significant. All three comparisons of NS with the low-smoker cluster have lower bounds close to zero, and one of the three comparisons is significant.

We can summarize these visual impressions by constructing an orthogonal set of contrasts that reflect them exactly. Figure 7.19 shows a basis set of five orthogonal contrasts. In the center, the p-nl and n-l contrasts show that the three levels in

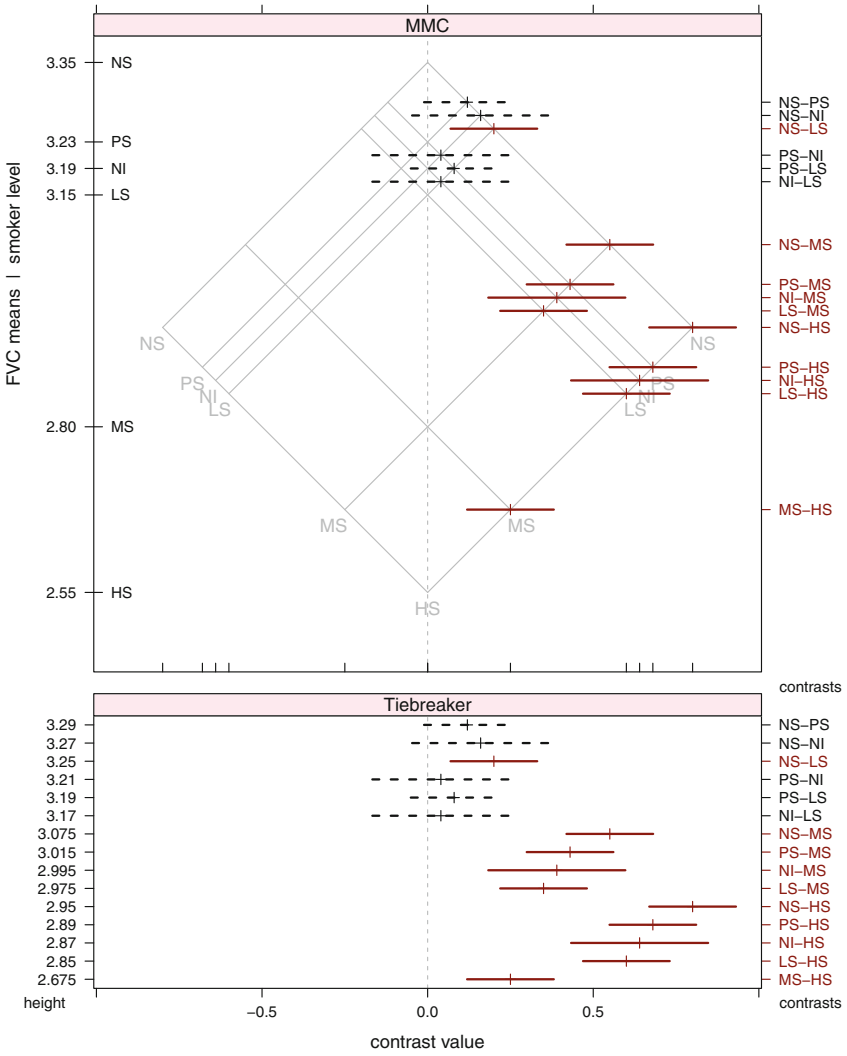

Fig. 7.18 Hsu and Peruggia's pulmonary example. The apparent clustering of the three groups `PS`, `NI`, `LS` suggests the set of contrasts we show in Figure 7.19.

the low-smoker cluster are indistinguishable. The other three lines show that the nonsmoker group is significantly different from the low-smoker cluster (`n-pnl`), that the moderate- and heavy-smoker groups are significantly different (`m-h`), and that the combined nonsmoker group and low-smoker cluster are significantly different from the combined moderate- and heavy-smoker groups (`npnl-mh`).

The center of the interval for each of the contrasts in Figure 7.19 is constructed by the linear combination of the means for the levels. For example, the `n-pnl` interval is on the NW–SE `NS` line and on the average of the NE–SW `PS`, `NI`, and `LS`

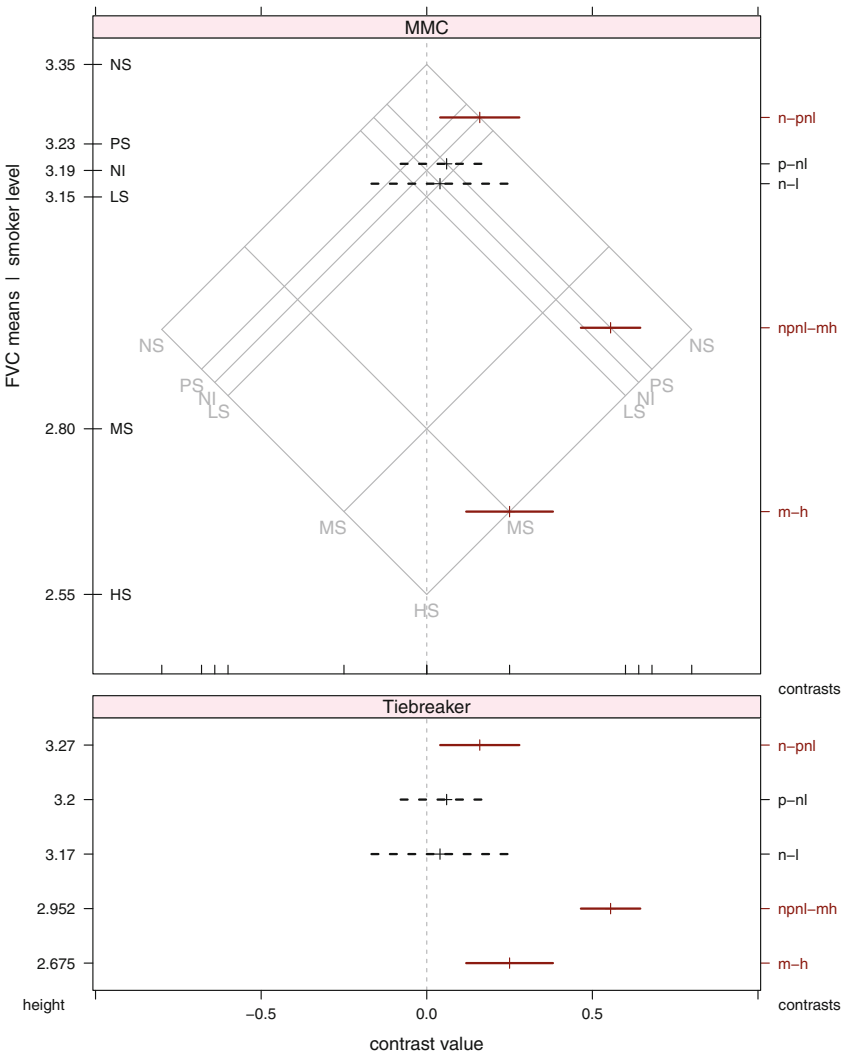

Fig. 7.19 Hsu and Peruggia's pulmonary example: An orthogonal set of contrasts. There are three significant contrasts and two not significant contrasts. The means for the three groups we discovered in Figure 7.18 are indistinguishable. The other differences are significant. The ability to display an arbitrary orthogonal set of contrasts is one of our enhancements to the mean–mean plot.

lines. The width of the interval is calculated from the algebra of the contrast. A simultaneous 95% coverage probability applies to the five confidence intervals in Figure 7.19 because they are constructed using the extended Tukey procedure. This procedure guarantees the coverage probability over the set of all possible contrasts. In exchange for this guarantee, these extended Tukey intervals are fairly wide. Having used the Tukey procedure to construct the intervals in Figure 7.18, it would be

incorrect to switch to the narrower Scheffé procedure simultaneous intervals for the basis set of contrasts. With such a switch we would have two competing analyses, and this would distort the claimed coverage probabilities for the now distinct analyses in the two figures.

7.3 Exercises

7.1. Use an MMC plot to display the results of the Tukey procedure in Exercise 6.2.

7.2. Use an MMC plot to display the results of the Tukey procedure in Exercise 6.4.

7.3. Use an MMC plot to display the results of the Tukey procedure applied to the log-transformed data discussed in Exercise 6.5.

7.4. Use an MMC plot to display the results of the Tukey procedure in Exercise 6.6.

7.5. Use an MMC plot to display the results of the Tukey procedure in Exercise 6.8.

7.6. Use an MMC plot to display the results of the Tukey procedure in Exercise 6.9.

7.7. The relative rotation angle between tangents to cervical vertebrae C3 and C4 is a standard musculoskeletal measurement. Figure 7.20 illustrates the measurement of relative rotation angles. Harrison et al. (2004) hypothesize that the value

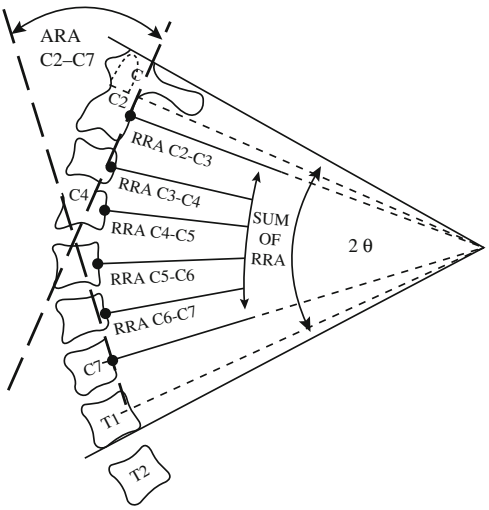

Fig. 7.20 Illustration of the relative rotation angles between the cervical vertebrae (neck area). Exercise 7.7 uses the C3–C4 angle.

of this angle, C3–C4, in persons complaining of neck pain tends to differ from that in healthy individuals. The dataset, accessed as `data(c3c4)`, contains the C3–C4 measurements of a random sample of 194 patients of which 72 had no complaints of neck pain, 52 complained of acute neck pain of recent origin, and 70 have had chronic neck pain. The pain `condition` is coded 0 for none, 1 for acute, and 2 for chronic. There is no implied ordering in this coding scheme. Perform an analysis of variance followed by Dunnett's procedure to determine if the mean C3–C4 value of persons with acute or chronic neck pain differs from the mean C3–C4 value of persons without neck pain.