One-way ANOVA

An alternative medical example for the same one-way design

The data from a one-way between-subjects design are shown in Table 2.2. Here we have a pilot study designed to investigate dosages for a new ACE-inhibitor in the treatment of hypertension. The new drug is believed to have fewer side effects than the currently favoured ACE-inhibitor. Fifty patients with systolic blood pressure (SBP) in the range 150 – 170 mm Hg are randomly allocated to one of five conditions. The IV is drug DOSAGE, with levels being 4mg, 6mg, 8mg, and 10mg for the new drug and 10mg for the old drug, which is known to be an effective level for that drug. The DV is the drop in systolic blood pressure (SBP) one week after administration.

The specific hypothesis was that SBPDROP (the DV), would show an upward trend with dosage of the new drug, equaling the drop achieved by 10mg dose of the currently favoured drug somewhere in the range 4mg to 10mg used in the study. Fifty participants are randomly assigned, ten to each condition, and the data are shown in Table 2.2. We note here that when we carried out a power analysis using the SamplePower software, we found that for a two-tailed test at $\alpha = 0.05$ and with a 'large' effect specified, 16 participants per condition would be required to achieve a power of 0.8. In this example experiment with fabricated data, the 10 participants per condition gave power equal to less than 0.6 for the same α level and effect size, which in a real experiment would be on the low side.

Table 2.2

SBP fall data from a one-way between-subjects design

new4mg	**new6mg**	**new8mg**	**new10mg**	**old10mg**
9	7	11	12	10
8	9	13	11	19
6	6	8	16	14
8	6	6	11	5
10	6	14	9	10
4	11	11	23	11
6	6	13	12	14
5	3	13	10	15
7	8	10	19	11
7	7	11	11	11

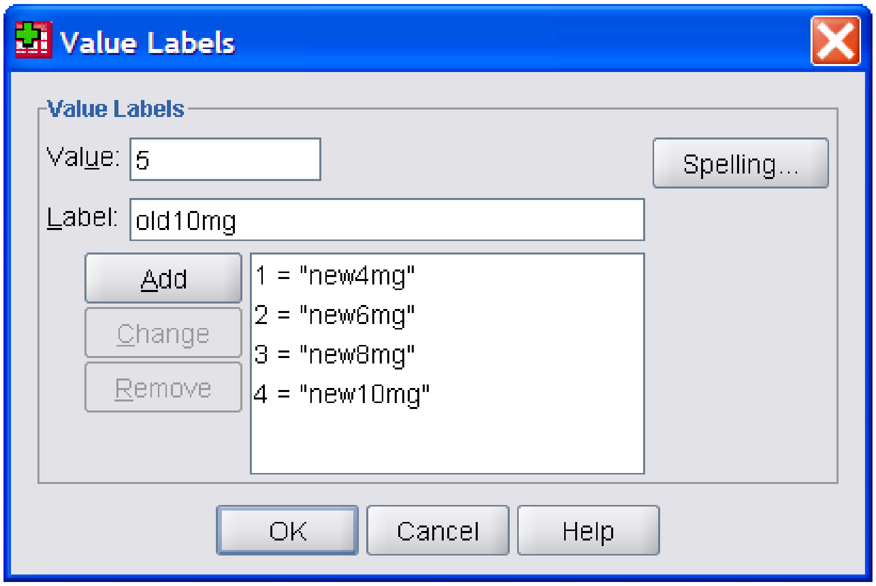
A one-way design: setting it up in SPSS

Because this is a between-subjects design, the data need to be entered in just two columns, one for the IV (DOSAGE) and one for the DV (SBPDROP), so that each participant occupies a single row. Thus, in the SBPDROP column, the data for 4mg of the new drug would be entered first and the data for 6, 8 and 10mg would be entered in turn below that, followed by the data for 10 mg of the old drug. The DOSAGE column would contain ten 1s, followed by ten 2s, ten 3s, ten 4s and ten 5s. The first and last few rows showing the data organized for entry into SPSS can be seen in Table 2.3 (the full dataset can be found on the book website as med.anova.oneway.sav).

Table 2.3
First and last few cases of data from a one-way between-subjects design set out for the SPSS datasheet (the full dataset can be found as med.anova.oneway.sav on the website)

dosage	**sbpdrop**
1	9
1	8
1	6
1	8
5	14
5	15
5	11
5	11

In fact, the output will be easier to read if the five dosage/drug conditions are given their names rather than the codes 1 to 5. We can easily arrange this once the data are in the datasheet. At the bottom of the datasheet, click the **Variable View** tab. Now we see each of the variables listed with their properties. Each of these properties can be altered: for instance we may want to specify that 0 decimal places are displayed (the default is 2) for our factor levels, and perhaps for other data when we have only whole numbers. Click in the **Values** cell for DOSAGE and a button appears: clicking this opens SPSS Dialog Box 2.1, and here we can assign labels to each of the five levels of DOSAGE. Type 1 in the **Value** box, type new4mg in the **Value Label** box and click **Add**. Repeat with 2 and new6mg, and so on to 5 and old10mg. The dialog box is shown just before we click **Add** for the last time. Then click **OK**. To see the labels displayed in the datasheet, return to **Data View** using the tab at the bottom, then click **View** on the menu bar, then **Value Labels,** which toggles between displaying code numbers and labels. If your output tables do not show the value labels, go to **Edit** in the menu bar, then **Options**. Click the **Output Labels** tab at the top, and make sure all of the boxes display **Values**. It may also be helpful to give the DV an extended label to be used in the output. To do this, select **Variable View** at the bottom tab, click **Label** in the SPBDROP row, and write 'fall in systolic blood pressure' in the highlighted cell.

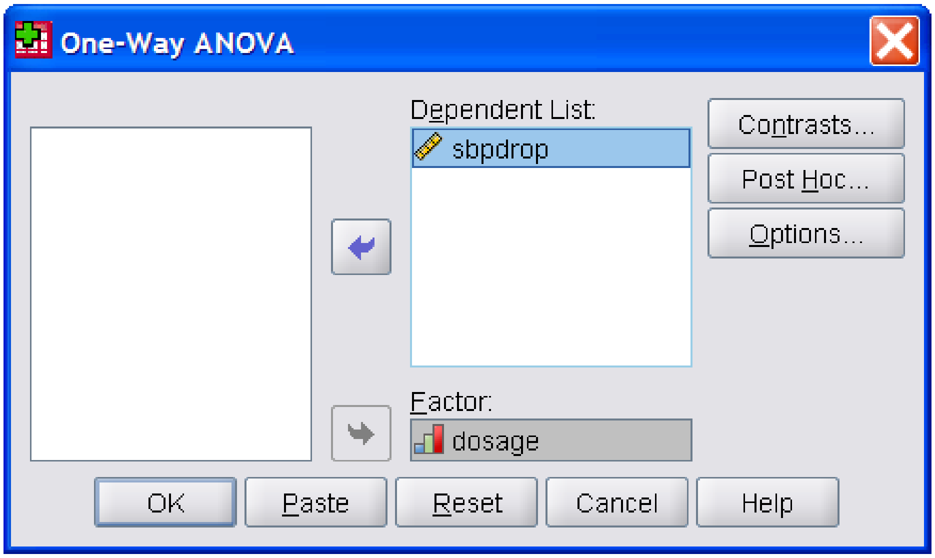


SPSS Dialog Box 2.1. Assigning factor level labels

SPSS offers great control over the way variables are defined, including different ways to deal with missing data. To see the full range of options click **Help** on the menu bar, then **Topics**. This gives an index, and you can scroll down or get to the right part of it more quickly by typing a word or part of a word in the box. From the index choose **variables** and the subentry **defining**. As well as a page of information there is an excellent tutorial that can be viewed by clicking **Show me**.

A one-way design: requesting the analysis in SPSS

Once the data are entered, select **Analyze** from the menu bar, then **Compare Means**, then **One-Way ANOVA**, to get SPSS Dialog Box 2.2. Select SBPDROP from the variable list and use the arrow to put it in the **Dependent List** box. Then put DOSAGE in the **Factor** box in the same way, so the dialog box appears as shown.



SPSS Dialog Box 2.2. Starting a one-way between-subjects ANOVA

Click the **Options button** to get a list of statistics for optional printing. Click the **Descriptives** box to get means etc., then **Homogeneity of variance test** for Levene's test of equality of group variances and **Means Plot**. Either the **Brown-Forsythe** or **Welch** test can be used instead of the *F* test if the assumption of equal variances is not met. The Options dialog box also offers a choice of methods for dealing with **Missing Values** but you will usually want to accept the default, **Exclude cases analysis by analysis**. We will ignore the Contrasts and Post Hoc buttons for the moment. When we have looked at the output for the main analysis, we will return to these buttons to carry out *follow-up* tests. **One-Way ANOVA** in SPSS does not offer an option to request a calculation of effect size or retrospective power, but the former is easily calculated from the output by hand (i.e., SS effect/SS total), and the latter can then be obtained using a power analysis package such as SamplePower. Click **Continue** and **OK** to get the main analysis.

A one-way design: understanding the output

If you are using version 16 of SPSS you will find that at the start of the output is a list of the syntax (SPSS commands) resulting from your dialog box choices. Version 15 and earlier did not provide this. (If you wish, you can turn it off. Click on **Edit** then **Options** and the **Viewer** tab. Untick **Display commands in the log** at the bottom left.) It can be useful, for example if you plan to carry out a series of similar analyses on different data sets, but we will not refer again to this part of the output except in the Appendix, where we provide the syntax for selected analyses.

SPSS Output 2.1 shows the output tables. The first table gives the means, standard deviations, etc. for each level of the IV. Next comes the test of homogeneity of variance. We may note, however, that some authors have questioned the legitimacy of Levene's test. In any case, ANOVA is quite robust to moderate departures from homogeneity unless treatment groups are small and unequal in size. In our example we see that the Levene statistic is not quite significant (the probability is 0.054, look at the Sig column), though if we had smaller and/or unequal group sizes, we might consider using the Brown-Forsythe or Welch test instead of the *F* test (these are available in the **Options** dialog box). Then we get the ANOVA summary table, with the *F* statistic quoted and its *df*, and we see that the difference among the five conditions is highly significant ($F(4,45) = 9.085$, $p < 0.001$). From the summary table we can easily compute effect size as $\eta^2 = 351.52/786.82 = 0.447$, which is equivalent to an *f* value of 0.90 (see formula above). This is a very large effect size and, when it is entered in the SamplePower package, together with $\alpha = 0.05$, two-tailed and $n = 10$ per cell, the power analysis indicates that retrospective power = 1.

Descriptives

sbpdrop

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean: Lower Bound	95% Confidence Interval for Mean: Upper Bound	Minimum	Maximum
new4mg	10	7.00	1.826	.577	5.69	8.31	4	10
new6mg	10	6.90	2.132	.674	5.38	8.42	3	11
new8mg	10	11.00	2.494	.789	9.22	12.78	6	14
new10mg	10	13.40	4.502	1.424	10.18	16.62	9	23
old10mg	10	12.00	3.742	1.183	9.32	14.68	5	19
Total	50	10.06	4.007	.567	8.92	11.20	3	23

Test of Homogeneity of Variances

sbpdrop

Levene Statistic	df1	df2	Sig.
2.529	4	45	.054

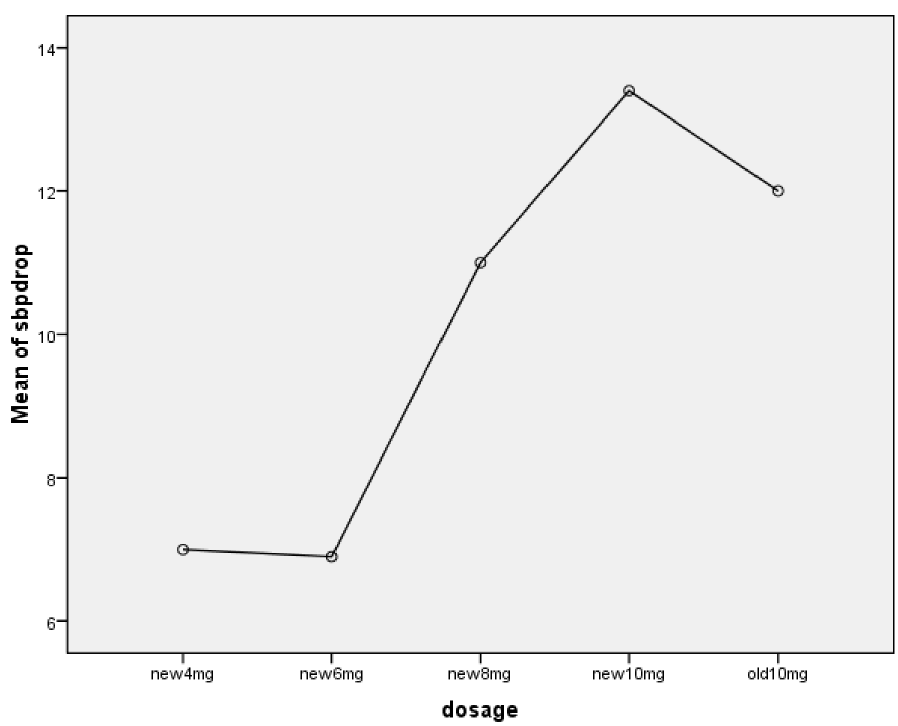
ANOVA

sbpdrop

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	351.520	4	87.880	9.085	.000
Within Groups	435.300	45	9.673		
Total	786.820	49			

SPSS Output 2.1. Results of the one-way between-subjects ANOVA

Finally, the plot we requested is shown in SPSS Output 2.2.



SPSS Output 2.2. Plot of means for the one-way between-subjects data

A one-way design: post hoc tests

It is obvious from the plot that there is not a steady increase with dosage of the new drug, and the drop in SBP with 10mg of the currently used drug is between those for 8mg and 10mg of the new drug. There are now several strategies available to us. We could carry out *post hoc* tests on the differences between all pairs of conditions, in which case we would need to deal with the problem of *multiple testing*. It would not be okay to just do a series of *t*-tests. Briefly, if we were to carry out 20 tests with alpha (the probability of a Type I error) set at 0.05 and the null hypothesis was true in every case, just by chance we might expect to find one difference significant at $p < 0.05$ (i.e., 1 in 20 Type I errors). There is a variety of procedures designed to set alpha for the *family of tests* at 0.05, and these differ in how conservative they are. One of the most commonly used is Tukey's Honestly Significant Difference (HSD) test. We will use that. We now re-do our analysis and ask for the Tukey HSD test at the same time. As before, select from the menu bar **Analyze**, then **Compare Means**, then **One-Way ANOVA**, but this time click the **Post Hoc** button to get a choice of post hoc tests. Select **Tukey**, then click **Continue** and **OK** to get the results of the Tukey test.

Another commonly used option, which is slightly less conservative than the Tukey test is the Student-Newman-Keuls (**S-N-K**) test. If you wanted to compare each experimental condition with a control condition, you could choose the **Dunnett** test and, if the assumption of homogeneity of variance is not met, several alternative post hoc tests are provided. Of these, we recommend the **Games-Howell** test. Further discussion of post hoc tests is beyond the scope of this book, but an excellent discussion is provided by Howell (2007).

The results of the post hoc tests are provided in two forms, of which the first is the simplest to follow. This is shown in SPSS Output 2.3. You can see that each level of DOSAGE, starting with new4mg, is compared with every other level. New4mg is first compared with new6mg, the mean difference between these levels was 0.100 with a *confidence interval* from -3.85 to 4.05. Since this confidence interval overlaps zero, the null hypothesis that the mean difference is zero would not be rejected. The probability (look in the Sig column) is 1.000, so it is virtually certain that the observed difference between these two levels is just random variation.

Multiple Comparisons

sbpdrop
Tukey HSD

(I) dosage	(J) dosage	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
new4mg	new6mg	.100	1.391	1.000	-3.85	4.05
	new8mg	-4.000*	1.391	.046	-7.95	-.05
	new10mg	-6.400*	1.391	.000	-10.35	-2.45
	old10mg	-5.000*	1.391	.007	-8.95	-1.05
new6mg	new4mg	-.100	1.391	1.000	-4.05	3.85
	new8mg	-4.100*	1.391	.039	-8.05	-.15
	new10mg	-6.500*	1.391	.000	-10.45	-2.55
	old10mg	-5.100*	1.391	.006	-9.05	-1.15
new8mg	new4mg	4.000*	1.391	.046	.05	7.95
	new6mg	4.100*	1.391	.039	.15	8.05
	new10mg	-2.400	1.391	.429	-6.35	1.55
	old10mg	-1.000	1.391	.951	-4.95	2.95
new10mg	new4mg	6.400*	1.391	.000	2.45	10.35
	new6mg	6.500*	1.391	.000	2.55	10.45
	new8mg	2.400	1.391	.429	-1.55	6.35
	old10mg	1.400	1.391	.851	-2.55	5.35
old10mg	new4mg	5.000*	1.391	.007	1.05	8.95
	new6mg	5.100*	1.391	.006	1.15	9.05
	new8mg	1.000	1.391	.951	-2.95	4.95
	new10mg	-1.400	1.391	.851	-5.35	2.55

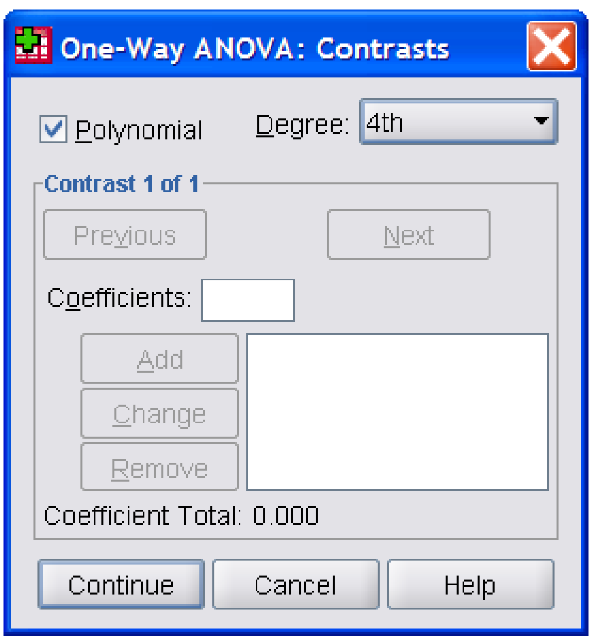
*. The mean difference is significant at the 0.05 level.

SPSS Output 2.3. Results of Tukey post hoc tests

This output tells us that new4mg and new6mg did not differ significantly and likewise, new8mg, new10mg and old10mg did not differ significantly from one another. On the other hand, each of new4mg and new6mg differed significantly ($p < 0.05$) from each of new8mg, new10mg and old10mg. The same information appears in the next output table, headed 'Homogeneous Subsets', which we have not shown.

A one-way design: planned comparisons

Another strategy would be to carry out planned comparisons (i.e., based on hypotheses that motivated the research). One such hypothesis might be that there would be a linear trend across the five conditions. This can be tested by re-doing the one-way ANOVA, but this time click the **Contrasts** button to get SPSS Dialog Box 2.3. Click the **Polynomial** box, and use the drop-down arrow to put **Linear** in the **Degree** box. We are selecting the first (linear) polynomial contrast or comparison. If we wanted to test for a quadratic trend (a single curve) we would tick **Polynomial** and select **Quadratic** in the **Degree** box. In the same fashion we could select a **Cubic** (a double curve) or **4th** polynomial in the **Degree** box. It is only possible to test up to a polynomial one less than the number of conditions (i.e., 5-1=4, in this case). In fact, if you select the 4th polynomial, you will get tests of all of the lower polynomials as well. We will do that because, as well as testing the linear trend, we can make a point about the cubic trend. So, the dialog box is as shown.



SPSS Dialog Box 2.3. Testing for trend: linear, quadratic, cubic and 4th polynomials requested

Click **Continue** and **OK** to see the results of the trend tests. The output is in SPSS Output 2.4.

ANOVA

sbpdrop

			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		351.520	4	87.880	9.085	.000
	Linear Term	Contrast	272.250	1	272.250	28.144	.000
		Deviation	79.270	3	26.423	2.732	.055
	Quadratic Term	Contrast	13.207	1	13.207	1.365	.249
		Deviation	66.063	2	33.031	3.415	.042
	Cubic Term	Contrast	64.000	1	64.000	6.616	.013
		Deviation	2.063	1	2.063	.213	.646
	4th-order Term	Contrast	2.063	1	2.063	.213	.646
	Within Groups		435.300	45	9.673		
	Total		786.820	49			

SPSS Output 2.4. Results of trend tests: linear, quadratic, cubic and 4th polynomials

In the first row, the results of the test of differences among the five conditions is repeated, then the results of the four trend tests are given. The one we were initially interested in is the *planned contrast*; the *a priori* hypothesis of a linear trend. We see that, even though the plot did not appear to be very close to a straight line, the linear trend is highly significant ($F(1,45) = 28.144$, $p < 0.001$). In the following row, we learn that the deviation from the linear trend; that is, the non-linear component of the trend remaining, approaches significance ($p = 0.055$). There are three *df* for the non-linear part of the trend, so the near significance of the *p* value suggests that a particular non-linear component of trend taking one of these *df*s may also be significant, which tells us that a particular non-linear component of trend may also exist. In fact, the cubic trend is significant ($p = 0.013$), which is not surprising given that the plot in SPSS Output 2.2 shows a double (S-shaped) curve. Even though the cubic trend is significant, we would not see any point in reporting it unless we could think of some plausible (post hoc) explanation for it. In this case, a possible explanation does exist. Neither 4mg nor 6mg of the new drug is sufficient to be effective in bringing down SPB, and it could be for that reason that SBPDROP does not differ between them. Once the drug is given at the higher level of 8mg, we do see a drop in SBP, and this effect is increased at 10mg. In fact at 10mg, the new drug exceeds the effect of 10mg of the currently favoured drug, so we see S shape that we observed on the graph in SPSS Output 2.2. Now, we need to be clear that, if we report the cubic effect, we would not be *confirming* a hypothesis – we would be *generating* a hypothesis from inspection of our data. This hypothesis would need to be tested in a new experiment. In fact for this study the most useful next step would be an investigation into just where between 8mg and 10mg of the new drug is the most useful dose. Also we would need to check on our belief that there are fewer side effects and that we have not missed an unexpected one.

A one-way design: complex post hoc comparisons

There is a further situation concerning follow-up tests that we will raise. This is when we look at our data and generate a complex post hoc hypothesis that requires more than testing differences between all pairs of means (Tukey) or testing each experimental mean against a control mean (Dunnett). For example, we might generate the hypothesis that there is a threshold effect in the drug dosage, and that below this level there is no effect on SBP. Specifically, we would be hypothesizing that at least 8mg is needed to see any effect on SBP whichever ACE inhibitor we use. We can use the *Scheffé* procedure to test complex post hoc hypotheses. For the example just suggested, we would compare the mean of the first two conditions with the mean of the last three. To do this we define a *contrast*, which multiplies each level mean by a suitably chosen *coefficient*, which is just a number. For our example we compare the mean of the first two levels (new4mg and new6mg) with the mean of the last three (new8mg, new10mg and old10mg). To find the difference we need to subtract one mean from the other. The steps for assigning coefficients to the levels are listed below.

1. Mean of first two levels = (new4mg + new6mg)/2 = $\frac{1}{2}$ new4mg + $\frac{1}{2}$ new6mg

2. Mean of last three levels

= (new8mg + new10mg + old10mg)/3 = $\frac{1}{3}$ new8mg + $\frac{1}{3}$ new10mg + $\frac{1}{3}$ old10mg

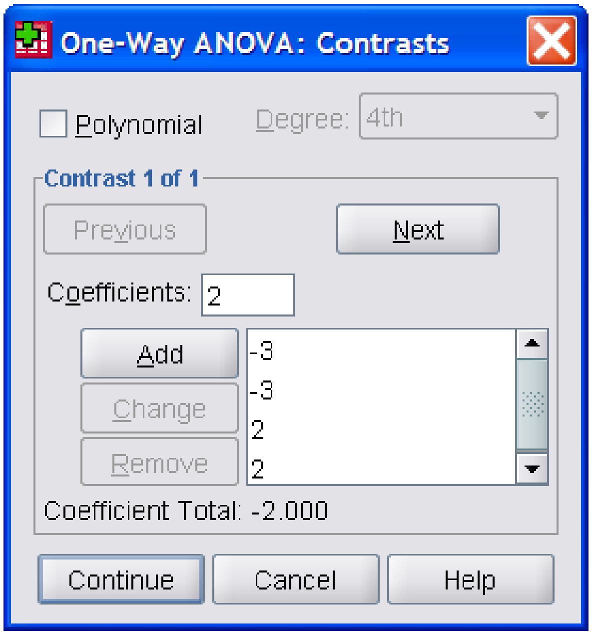
3. Contrast = $\frac{1}{3}$ new8mg + $\frac{1}{3}$ new10mg + $\frac{1}{3}$ old10mg - $\frac{1}{2}$ new4mg - $\frac{1}{2}$ new6mg

4. Coefficients for contrast (in the same order as the levels) $-\frac{1}{2}$, $-\frac{1}{2}$, $+\frac{1}{3}$, $+\frac{1}{3}$, $+\frac{1}{3}$

The coefficients must sum to zero, and those for the means in the first set that are assumed not to differ are identical, and those for the means in the other set are also identical. You can easily see that this holds in the above case. However, SPSS doesn't allow us to enter fractions, and many, such as 1/3 in our example, don't have an exact decimal version. So it's best to find a version of the contrast that uses small whole numbers. So there is a final step.

5. Multiply by the lowest common denominator (the smallest number that can be divided without a remainder by the two denominators 2 and 3, that is 6) to get all whole numbers -3, -3, 2, 2, 2

The contrast we end up with is six times the one we wanted, but we shall be testing whether it is zero, so six times the original is just as good.



SPSS Dialog Box 2.4. Weighting the means with coefficients to make the desired comparison

To do this in SPSS return to SPSS Dialog Box 2.2 and click the **Contrasts** button to get SPSS Dialog Box 2.4. Enter the first coefficient (-3) in the **Coefficients** box and click **Add**. This is repeated for successive coefficients. The dialog box is shown just before **Add** is clicked for the last time. Click **Continue** and **OK** to obtain the output. The result of the comparison is shown in SPSS Output 2.5.

Contrast Coefficients

Contrast	dosage				
	new4mg	new6mg	new8mg	new10mg	old10mg
1	-3	-3	2	2	2

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
sbpdrop	Assume equal variances	1	31.10	5.387	5.773	45	.000
	Does not assume equal variances	1	31.10	4.826	6.445	37.829	.000

SPSS Output 2.5. Coefficients creating the contrast and the result of the contrast

The first output table just displays the coefficients we entered. The second table gives the result of the comparison. We select the first or second row, depending on whether or not the Levene test indicated that we could assume equal variances. The Levene statistic was not significant (SPSS Output 2.1) so we look at the first row. We find a *t*-

value that is highly significant, but we do not accept the significance level given because we need to allow for the fact that we decided on the comparison after looking at our data, which is equivalent to testing all possible contrasts before looking at the data (a rather extreme form of multiple testing). Instead, we use the Scheffé correction. As the Scheffé correction works with F rather than t, we square the t-value to get $F = 33.33$ with 4 and 45 degrees of freedom. Now comes the adjustment. If we look up the critical value of $F(4,45)$ in a statistical table for alpha set at 0.001, we get $F_{crit} = 5.56$. The adjustment involves multiplying this critical value by the number of levels of the factor minus one (i.e., 4). So the adjusted critical value of F is $4 \times 5.56 = 22.24$, which is still less than our obtained value of $F = 33.33$, so the two sets of means differ significantly (adjusted $F(4,45) = 33.33$, $p < 0.001$) using a Scheffé correction for post hoc multiple testing.

A one-way design: multiple planned comparisons

Before leaving the topic of follow-up tests, we make one further point concerning the testing of planned comparisons (i.e., comparisons arising from a priori hypotheses). If you test more than two or three planned comparisons you run into the same issue of multiple testing that arose in connection with post hoc tests, especially if the various comparisons are not *orthogonal* (independent of one another). A convenient adjustment for multiple testing in this situation is provided by the *Bonferroni t*-test. We do not explain how to determine whether planned comparisons are independent, or why it matters, in this book. Neither do we give any details about the Bonferroni adjustment. There is an excellent chapter in Howell (2007) that we recommend if you want to understand more about these topics.

In SPSS the *p* value reported from the *Bonferroni* correction for multiple testing is the original *p* value multiplied by the number of possible tests (with a maximum reported value of 1.000). So if you had a factor with five levels, then the *p* value reported from the Bonferroni t-test would be multiplied by 10 (because of the ten pairs). Note that in general if you have k levels then there will be k(k-1)/2 pairs and so each Bonferroni p value will be multiplied by this value. This is unnecessarily conservative and would lead to possibly failing to reject a hypothesis if you only wanted to test, for example, two pairs.