

Chapter 13

Contrasts and Custom Hypotheses

Contrasts ask specific questions as opposed to the general ANOVA null vs. alternative hypotheses.

In a one-way ANOVA with a k level factor, the null hypothesis is $\mu_1 = \dots = \mu_k$, and the alternative is that at least one group (treatment) population mean of the outcome differs from the others. If $k = 2$, and the null hypothesis is rejected we need only look at the sample means to see which treatment is “better”. But if $k > 2$, rejection of the null hypothesis does not give the full information of interest. For some specific group population means we would like to know if we have sufficient evidence that they differ from certain other group population means. E.g., in a test of the effects of control and two active treatments to increase vocabulary, we might find that based on a the high value for the F-statistic we are justified in rejecting the null hypothesis $\mu_1 = \mu_2 = \mu_3$. If the sample means of the outcome are 50, 75 and 80 respectively, we need additional testing to answer specific questions like “Is the control population mean lower than the average of the two active treatment population means?” and “Are the two active treatment population means different?” To answer questions like these we frame “custom” hypotheses, which are formally expressed as **contrast hypothesis**.

Comparison and analytic comparison are other synonyms for contrast.

13.1 Contrasts, in general

A contrast null hypothesis compares two population means or combinations of population means. A **simple contrast hypothesis** compares two population means, e.g. $H_0 : \mu_1 = \mu_5$. The corresponding inequality is the alternative hypothesis: $H_1 : \mu_1 \neq \mu_5$.

A contrast null hypotheses that has multiple population means on either or both sides of the equal sign is called a **complex contrast hypothesis**. In the vast majority of practical cases, the multiple population means are combined as their mean, e.g., the custom null hypothesis $H_0 : \frac{\mu_1 + \mu_2}{2} = \frac{\mu_3 + \mu_4 + \mu_5}{3}$ represents a test of the equality of the average of the first two treatment population means to the average of the next three. An example where this would be useful and interesting is when we are studying five ways to improve vocabulary, the first two of which are different written methods and the last three of which are different verbal methods.

It is customary to rewrite the null hypothesis with all of the population means on one side of the equal sign and a zero on the other side. E.g., $H_0 : \mu_1 - \mu_5 = 0$ or $H_0 : \frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4 + \mu_5}{3} = 0$. This mathematical form, whose left side is checked for equality to zero is the standard form for a contrast. In addition to hypothesis testing, it is also often of interest to place a confidence interval around a contrast of population means, e.g., we might calculate that the 95% CI for $\mu_3 - \mu_4$ is [-5.0, +3.5].

As in the rest of classical statistics, we proceed by finding the null sampling distribution of the contrast statistic. A little bit of formalism is needed so that we can enter the correct custom information into a computer program, which will then calculate the contrast statistic (estimate of the population contrast), the standard error of the statistic, a corresponding t-statistic, and the appropriate p-value. As shown later, this process only works under the special circumstances called “planned comparisons”; otherwise it requires some modifications.

Let γ (gamma) represent the population contrast. In this section, will use an example from a single six level one-way ANOVA, and use subscripts 1 and 2 to distinguish two specific contrasts. As an example of a simple (population) contrast, define γ_1 to be $\mu_3 - \mu_4$, a contrast of the population means of the outcomes for the third vs. the fourth treatments. As an example of a complex contrast let γ_2 be $\frac{\mu_1 + \mu_2}{2} - \frac{\mu_3 + \mu_4 + \mu_5}{3}$, a contrast of the population mean of the outcome for the first two treatments to the population mean of the outcome for the third through fifth treatments. We can write the corresponding hypothesis as $H_{01} : \gamma_1 = 0$, $H_{A1} :$

$\gamma_1 \neq 0$ and $H_{02} : \gamma_2 = 0$, $H_{A2} : \gamma_2 \neq 0$.

If we call the corresponding estimates, g_1 and g_2 then the appropriate estimates are $g_1 = \bar{y}_3 - \bar{y}_4$ and $g_2 = \frac{\bar{y}_1 + \bar{y}_2}{2} - \frac{\bar{y}_3 + \bar{y}_4 + \bar{y}_5}{3}$. In the hypothesis testing situation, we are testing whether or not these estimates are consistent with the corresponding null hypothesis. For a confidence interval on a particular population contrast (γ), these estimates will be at the center of the confidence interval.

In the chapter on probability theory, we saw that the sampling distribution of any of the sample means from a (one treatment) sample of size n using the assumptions of Normality, equal variance, and independent errors is $\bar{y}_i \sim N(\mu_i, \sigma^2/n)$, i.e., across repeated experiments, a sample mean is Normally distributed with the “correct” mean and the variance equal to the common group variance reduced by a factor of n . Now we need to find the sampling distribution for some particular combination of sample means.

To do this, we need to write the contrast in “standard form”. The standard form involves writing a sum with one term for *each* population mean (μ), *whether or not it is in the particular contrast*, and with a single number, called a **contrast coefficient** in front of each population mean. For our examples we get:

$$\gamma_1 = (0)\mu_1 + (0)\mu_2 + (0)\mu_3 + (1)\mu_4 + (-1)\mu_5 + (0)\mu_6$$

and

$$\gamma_2 = (1/2)\mu_1 + (1/2)\mu_2 + (-1/3)\mu_3 + (-1/3)\mu_4 + (-1/3)\mu_5 + (0)\mu_6.$$

In a more general framing of the contrast we would write

$$\gamma = C_1\mu_1 + \cdots + C_k\mu_k.$$

In other words, each contrast can be summarized by specifying its k coefficients (C values). And it turns out that the k coefficients are what most computer programs want as input when you specify the contrast of a custom null hypothesis.

In our examples, the coefficients (and computer input) for null hypothesis H_{01} are $[0, 0, 1, -1, 0, 0]$, and for H_{02} they are $[1/2, 1/2, -1/3, -1/3, -1/3, 0]$. Note that the zeros *are* necessary. For example, if you just entered $[1, -1]$, the computer would not understand which pair of treatment population means you want it to compare. Also, note that any valid set of contrast coefficients must add to zero.

It is OK to multiply the set of coefficients by any (non-zero) number. E.g., we could also specify H_{02} as $[3, 3, -2, -2, -2, 0]$ and $[-3, -3, 2, 2, 2, 0]$. These alternate contrast coefficients give the same p-value, but they do give different estimates of γ , and that must be taken in to account when you interpret confidence intervals. If you really want to get a confidence interval on the difference in average group population outcome means for the first two vs. the next three treatments, it will be directly interpretable only in the fraction form.

A positive estimate for γ indicates higher means for the groups with positive coefficients compared to those with negative coefficients, while a negative estimate for γ indicates higher means for the groups with negative coefficients compared to those with positive coefficients

To get a computer program to test a custom hypothesis, you must enter the k coefficients that specify that hypothesis.

If you can handle a bit more math, read the theory behind contrast estimates provided here.

The simplest case is for two independent random variables Y_1 and Y_2 for which the population means are μ_1 and μ_2 and the variances are σ_1^2 and σ_2^2 . (We allow unequal variance, because even under the equal variance assumption, the sampling distribution of two means, depends on their sample sizes, which might not be equal.) In this case it is true that $E(C_1Y_1 + C_2Y_2) = C_1\mu_1 + C_2\mu_2$ and $\text{Var}(C_1Y_1 + C_2Y_2) = C_1^2\sigma_1^2 + C_2^2\sigma_2^2$. If in addition, the distributions of the random variables are Normal, we can conclude that the distribution of the linear combination of the random variables is also Normal. Therefore $Y_1 \sim N(\mu_1, \sigma_1^2)$, $Y_2 \sim N(\mu_2, \sigma_2^2)$, $\Rightarrow C_1Y_1 + C_2Y_2 \sim N(C_1\mu_1 + C_2\mu_2, C_1^2\sigma_1^2 + C_2^2\sigma_2^2)$.

We will also use the fact that if each of several independent random variables has variance σ^2 , then the variance of a sample mean of n of these has variance σ^2/n .

From these ideas (and some algebra) we find that in a one-way ANOVA with k treatments, where the group sample means are independent, if we let σ^2 be the common population variance, and n_i be the number of subjects sampled for treatment i , then $\text{Var}(g) = \text{Var}(C_1\bar{Y}_1 + \cdots + C_k\bar{Y}_k) = \sigma^2[\sum_{i=1}^k (C_i^2/n_i)]$.

In a real data analysis, we don't know σ^2 so we substitute its estimate, the within-group mean square. Then the square root of the estimated variance is the standard error of the contrast estimate, $\text{SE}(g)$.

For any normally distributed quantity, g , which is an estimate of a parameter, γ , we can construct a t-statistic, $(g - \gamma)/\text{SE}(g)$. Then the sampling distribution of that t-statistic will be that of the t-distribution with df equal to the number of degrees of freedom in the standard error ($\text{df}_{\text{within}}$).

From this we can make a hypothesis test using $H_0 : \gamma = 0$, or we can construct a confidence interval for γ , centered around g .

For two-way (or higher) ANOVA without interaction, main effects contrasts are constructed separately for each factor, where the population means represent setting a specific level for one factor and ignoring (averaging over) all levels of the other factor.

For two-way ANOVA with interaction, contrasts are a bit more complicated. E.g., if one factor is job classification (with k levels) and the other factor is incentive applied (with m levels), and the outcome is productivity, we might be interested in comparing any particular combination of factor levels to any other combination. In this case, a one-way ANOVA with $k \cdot m$ levels is probably the best way to go.

If we are only interested in comparing the size of the mean differences for two particular levels of one factor across two levels of the other factor, then we are more clearly in an "interaction framework", and contrasts written for the two-way ANOVA make the most sense. E.g., if the subscripts on μ represent the levels of the two factors, we might be interested in a confidence interval on the contrast

$$(\mu_{1,3} - \mu_{1,5}) - (\mu_{2,3} - \mu_{2,5}).$$

The contrast idea extends easily to two-way ANOVA with no interaction, but can be more complicated if there is an interaction.

13.2 Planned comparisons

The ANOVA module of most statistical computer packages allow entry of custom hypotheses through contrast coefficients, but the p-values produced are only valid under stringent conditions called **planned comparisons** or planned contrasts or planned custom hypotheses. Without meeting these conditions, the p-values will be smaller than 0.05 more than 5% of the time, often far more, when the null hypothesis is true (i.e., when you are studying ineffectual treatments). In other words, these requirements are needed to maintain the Type 1 error rate across the entire experiment.

Note that for some situations, such as genomics and proteomics, where k is very large, a better goal than trying to keep the chance of making any false claim at only 5% is to reduce the total fraction of positive claims that are false positive. This is called control of the false discovery rate (FDR).

The conditions needed for planned comparisons are:

1. The contrasts are selected *before* looking at the results, i.e., they are planned, not post-hoc (after-the-fact).
2. The tests are ignored if the overall null hypothesis ($\mu_1 = \dots = \mu_k$) is not rejected in the ANOVA.
3. The contrasts are orthogonal (see below). This requirement is often ignored, with relatively minor consequences.

4. The number of planned contrasts is no more than the corresponding degrees of freedom ($k - 1$, for one-way ANOVA).

The orthogonality idea is that each contrast should be based on independent information from the other contrasts. For the 36309 course, you can consider this paragraph optional. To test for orthogonality of two contrasts for which the contrast coefficients are $C_1 \cdots C_k$ and $D_1 \cdots D_k$, compute $\sum_{i=1}^k (C_i D_i)$. If the sum is zero, then the contrasts are orthogonal. E.g., if $k=3$, then $\mu_1 - 0.5\mu_2 - 0.5\mu_3$ is orthogonal to $\mu_2 - \mu_3$, but not to $\mu_1 - \mu_2$ because $(1)(0) + (-0.5)(1) + (-0.5)(-1) = 0$, but $(1)(1) + (-0.5)(-1) + (-0.5)(0) = 1.5$.

To reiterate the requirements of planned comparisons, let's consider the consequences of breaking each requirement. If you construct your contrasts after looking at your experimental results, you will naturally choose to compare the biggest and the smallest sample means, which suggests that you are implicitly comparing all of the sample means to find this interesting pair. Since each comparison has a 95% chance of correctly retaining the null hypothesis when it is true, after m independent tests you have a 0.95^m chance of correctly concluding that there are no significant differences when the null hypothesis is true. As examples, for $m=3, 5$, and 10 , the chance of correctly retaining all of the null hypotheses are 86%, 77% and 60% respectively. Put another way, choosing which groups to compare after looking at results puts you at risk of making a false claim 14, 23 and 40% of the time respectively. (In reality the numbers are often slightly better because of lack of independence of the contrasts.)

The same kind of argument applies to looking at your planned comparisons without first "screening" with the overall p-value of the ANOVA. Screening protects your Type 1 experiment-wise error rate, while lack of screening raises it.

Using orthogonal contrasts is also required to maintain your Type 1 experiment-wise error rate. Correlated null hypotheses tend to make the chance of having several simultaneous rejected hypotheses happen more often than should occur when the null hypothesis is really true.

Finally, making more than $k - 1$ planned contrasts (or $k - 1$ and $m - 1$ contrasts for a two-way $k \times m$ ANOVA without interaction) increases your Type 1 error

because each additional test is an additional chance to reject the null hypothesis incorrectly whenever the null hypothesis actually is true.

Many computer packages, including SPSS, assume that for any set of custom hypotheses that you enter you have already checked that these four conditions apply. Therefore, any p-value it gives you is wrong if you have not met these conditions.

It is up to you to make sure that your contrasts meet the conditions of “planned contrasts”; otherwise the computer package will give wrong p-values.

In SPSS, anything entered as “Contrasts” (in menus) or “LMATRIX” (in Syntax, see Section 5.1) is tested as if it is a planned contrast.

As an example, consider a trial of control vs. two active treatments ($k = 3$). Before running the experiment, we might decide to test if the average population means for the active treatments differs from the control, and if the two active treatments differ from each other. The contrast coefficients are $[1, -0.5, -0.5]$ and $[0, 1, -1]$. These are planned before running the experiment. We need to realize that we should only examine the contrast p-values if the overall (between-groups, 2 df) F test gives a p-value less than 0.05. The contrasts are orthogonal because $(1)(0) + (-0.5)(1) + (-0.5)(-1) = 0$. Finally, there are only $k-1=2$ contrasts, so we have not selected too many.

13.3 Unplanned or post-hoc contrasts

What should we do if we want to test more than $k - 1$ contrasts, or if we find an interesting difference that was not in our planned contrasts after looking at our results? These are examples of what is variously called unplanned comparisons, multiple comparisons, post-hoc (after-the-fact) comparisons, or data snooping. The answer is that we need to add some sort of penalty to preserve our Type 1 experiment-wise error rate. The penalty can either take the form of requiring a larger difference (g value) before an unplanned test is considered “statistically significant”, or using a smaller α value (or equivalently, using a bigger critical F-value or critical t-value).

How big of a penalty to apply is mostly a matter of considering the size of the “family” of comparisons within which you are operating. (Amount of dependence among the contrasts can also have an effect.) For example, if you pick out the biggest and the smallest means to compare, you are implicitly comparing all pairs of means. In the field of probability, the symbol $\binom{a}{b}$ (read *a* choose *b*) is used to indicate the number of different groups of size *b* that can be formed from a set of *a* objects. The formula is $\binom{a}{b} = \frac{a!}{b!(a-b)!}$ where $a! = a \cdot (a-1) \cdots (1)$ is read “a factorial”. The simplification for pairs, $b = 2$, is $\binom{a}{2} = \frac{a!}{2!(a-2)!} = a(a-1)/2$. For example, if we have a factor with 6 levels, there are $6(5)/2=15$ different paired comparisons we can make.

Note that these penalized procedures are designed to be applied *without* first looking at the overall p-value.

The simplest, but often overly conservative penalty is the Bonferroni correction. If *m* is the size of the family of comparisons you are making, the Bonferroni procedure says to reject any post-hoc comparison test(s) if $p \leq \alpha/m$. So for $k = 6$ treatment levels, you can make post-hoc comparisons of all pairs while preserving Type 1 error at 5% if you reject H_0 only when $p \leq \alpha/15 = 0.0033$.

The meaning of conservative is that this procedure is often more stringent than necessary, and using some other valid procedure might show a statistically significant result in some cases where the Bonferroni correction shows no statistical significance.

The Bonferroni procedure is completely general. For example, if we want to try all contrasts of the class “compare all pairs and compare the mean of any two groups to any other single group”, the size of this class can be computed, and the Bonferroni correction applied. If $k=5$, there are 10 pairs, and for each of these we can compare the mean of the pair to each of the three other groups, so the family has $10 \cdot 3 + 10 = 40$ possible comparisons. Using the Bonferroni correction with $m=40$ will ensure that you make a false positive claim no more than $100\alpha\%$ of the time.

Another procedure that is valid specifically for comparing pairs is the Tukey procedure. The mathematics will not be discussed here, but the procedure is commonly available, and can be used to compare any and all pairs of group population means after seeing the results. For two-way ANOVA without interaction, the Tukey procedure can be applied to each factor (ignoring or averaging over the other factor). For a $k \times m$ ANOVA with a significant interaction, if the desired

contrasts are between arbitrary cells (combinations of levels of the two factors), the Tukey procedure can be applied after reformulating the analysis as a one-way ANOVA with $k \times m$ distinct (arbitrary) levels. The Tukey procedure is more powerful (less conservative) than the corresponding Bonferroni procedure.

It is worth mentioning again here that none of these procedures is needed for $k = 2$. If you try to apply them, you will either get some form of “not applicable” or you will get no penalty, i.e., the overall $\mu_1 = \mu_2$ hypothesis p-value is what is applicable.

Another post-hoc procedure is Dunnett’s test. This makes the appropriate penalty correction for comparing one (control) group to all other groups.

The total number of available post-hoc procedures is huge. Whenever you see such an embarrassment of riches, you can correctly conclude that there is some lack of consensus on the matter, and that applies here. I recommend against using most of these, and certainly it is very bad practice to try as many as needed until you get the answer you want!

The final post-hoc procedure discussed here is the Scheffé procedure. This is a very general, but conservative procedure. It is applicable for the family of *all* possible contrasts! One way to express the procedure is to consider the usual uncorrected t-test for a contrast of interest. Square the t-statistic to get an F statistic. Instead of the usual F-critical value for the overall null hypothesis, often written as $F(1 - \alpha, k - 1, N - k)$, the penalized critical F value for a post-hoc contrast is $(k - 1)F(1 - \alpha, k - 1, N - k)$. Here, N is the total sample size for a one-way ANOVA, and $N - k$ is the degrees of freedom in the estimate of σ^2 .

The critical F value for a Scheffé penalized contrast can be obtained as $(k - 1) \times \text{qf}(0.95, k - 1, N - k)$ in R or from $(k - 1) \times \text{IDF.F}(0.95, k - 1, N - k)$ in SPSS.

Although Scheffé is a choice in the SPSS Post-Hoc dialog box, it doesn’t make much sense to choose this because it only compares all possible pairs, but applies the penalty needed to allow all possible contrasts. In practice, the Scheffé penalty makes sense when you see an interesting complex post-hoc contrast, and then want to see if you actually have good evidence

that it is “real” (statistically significant). You can either use the menu or syntax in SPSS to compute the contrast estimate (g) and its standard error ($SE(g)$), or calculate these manually. Then find $F = (g/SE(g))^2$ and reject H_0 only if this value exceeds the Scheffé penalized F cutoff value.

When you have both planned and unplanned comparisons (which should be most of the time), it is not worthwhile (re-)examining any planned comparisons that also show up in the list of unplanned comparisons. This is because the unplanned comparisons have a penalty, so if the contrast null hypothesis is rejected as a planned comparison we already know to reject it, whether or not it is rejected on the post-hoc list, and if it is retained as a planned comparison, there is no way it will be rejected when the penalty is added.

Unplanned contrasts should be tested only after applying an appropriate penalty to avoid a high chance of Type 1 error. The most useful post-hoc procedures are Bonferroni, Tukey, and Dunnett.

13.4 Do it in SPSS

SPSS has a Contrast button that opens a dialog box for specifying planned contrasts and a PostHoc button that opens a dialog box for specifying various post-hoc procedures. In addition, planned comparisons can be specified by using the Paste button to examine and extend the Syntax (see Section 5.1) of a command to include one or more contrast calculations.

13.4.1 Contrasts in one-way ANOVA

Here we will examine planned and post-hoc contrast analyses for an experiment with three levels of an independent variable called “additive” (which is a chemical additive to a reaction, and has nothing to do with additive vs. interaction model types). The outcome is the number of hours until the reaction completes.

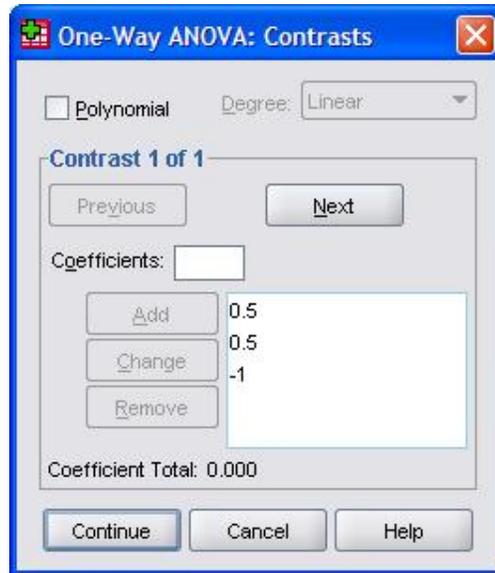


Figure 13.1: One-way ANOVA contrasts dialog box.

For a k -level one-way (between-subjects) ANOVA, accessed using Analyze/OneWayANOVA on the menus, the Contrasts button opens the “One-Way ANOVA: Contrasts” dialog box (see figure 13.1). From here you can enter the coefficients for each planned contrast. For a given contrast, enter the k coefficients that define any given contrast into the box labeled “Coefficients:” as a decimal number (no fractions allowed). Click the “Add” button after entering each of the coefficients. For a k -level ANOVA, you must enter all k coefficients, even if some are zero. Then you should check if the “Coefficient Total” equals 0.000. (Sometimes, due to rounding, this might be slightly above or below 0.000.) If you have any additional contrasts to add, click the Next button and repeat the process. Click the Continue button when you are finished. The figure shows a planned contrast for comparing the mean outcome (hours) for additives 1 and 2 to the mean outcome for additive 3.

When entering contrast coefficients in one-way ANOVA, SPSS will warn you and give no result if you enter more or less than the appropriate number of coefficients. It will not warn you if you enter more than $k - 1$ contrasts, if your coefficients do not add to 0.0, or if the contrasts are not orthogonal. Also, it will not prevent you from incorrectly analyzing post-hoc comparisons as planned comparisons.

The results for this example are given in Table 13.1. You should always look

Contrast Coefficients

	additive		
Contrast	1	2	3
1	0.5	0.5	-1
2	1	-1	0

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
hrs	Assume equal variance	1	-0.452	0.382	-1.18	47	0.243
		2	0.485	0.445	1.09	47	0.282
	Does not assume equal variance	1	-0.452	0.368	-1.23	35.58	0.228
		2	0.485	0.466	1.04	28.30	0.307

Table 13.1: Contrast results for one-way ANOVA.

at the Contrast Coefficients table to verify which contrasts you are testing. In this table, contrast 1, using coefficients (0.5, 0.5, -1) is testing $H_{01} : \frac{\mu_1 + \mu_2}{2} - \mu_3 = 0$. Contrast 2 with coefficients (1, -1, 0) is testing $H_{02} : \mu_1 - \mu_2 = 0$.

The Contrast Tests table shows the results. Note that “hrs” is the name of the outcome variable. The “Value of the Contrast” entry is the best estimate of the contrast. For example, the best estimate of $\mu_1 - \mu_2$ is 0.485. The standard error of this estimate (based on the equal variance section) is 0.445 giving a t-statistic of $0.485/0.445=1.09$, which corresponds to a p-value of 0.282 using the t-distribution with 47 df. So we retain the null hypothesis, and an approximate 95% CI for $\mu_1 - \mu_2$ is $0.485 \pm 2 \times 0.445 = [-0.405, 1.375]$. If you have evidence of unequal variance (violation of the equal variance assumption) you can use the lower section which is labeled “Does not assume equal variances.”

In SPSS, the two post-hoc tests that make the most sense are Tukey HSD and Dunnett. Tukey should be used when the only post-hoc testing is among all pairs of population means. Dunnett should be used when the only post-hoc testing is between a control and all other population means. Only one of these applies to a given experiment. (Although the Scheffé test is useful for allowing post-hoc testing of all combinations of population means, turning that procedure on in SPSS does not make sense because it still only tests all pairs, in which case Tukey is more appropriate.)

Multiple Comparisons

hrs
Tukey HSD

(I) additive	(J) additive	Mean Difference (I-J)	Std.Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
1	2	0.485	0.445	0.526	-0.593	1.563
	3	-0.209	0.445	0.886	-1.287	0.869
2	1	-0.485	0.445	0.526	-1.563	0.593
	3	-0.694	0.439	0.263	-1.756	0.367
3	1	0.209	0.445	0.886	-0.869	1.287
	2	0.694	0.439	0.263	-0.367	1.756

Homogeneous Subsets

hrs
Tukey HSD

additive	N	Subset for alpha=0.05
		1
2	17	16.76
1	16	17.244
3	17	17.453
Sig.		0.270

Table 13.2: Tukey Multiple Comparison results for one-way ANOVA.

Table 13.2 shows the Tukey results for our example. Note the two columns labeled I and J. For each combination of levels I and J, the “Mean Difference (I-J)” column gives the mean difference subtracted in that order. For example, the first mean difference, 0.485, tells us that the sample mean for additive 1 is 0.485 higher than the sample mean for additive 2, because the subtraction is I (level 1) minus J (level 2). The standard error of each difference is given. This standard error is used in the Tukey procedure to calculate the corrected p-value that is appropriate for post-hoc testing. For any contrast that is (also) a planned contrast, you should ignore the information given in the Multiple Comparisons table, and instead use the information in the planned comparisons section of the output. (The p-value for a planned comparison is smaller than for the corresponding post-hoc test.)

The Tukey procedure output also gives a post-hoc 95% CI for each contrast. Note again that if a contrast is planned, we use the CI from the planned contrasts section and ignore what is in the multiple comparisons section. Contrasts that are made post-hoc (or analyzed using post-hoc procedures because they do not meet the four conditions for planned contrasts) have appropriately wider confidence intervals than they would have if they were treated as planned contrasts.

The Homogeneous Subsets table presents the Tukey procedure results in a different way. The levels of the factor are presented in rows ordered by the sample means of the outcome. There are one or more numbered columns that identify “homogeneous subsets.” One way to read this table is to say that all pairs are significantly different except those that are in the same subset. In this example, with only one subset, no pairs have a significant difference.

You can alternately use the menu item Analyze/GeneralLinearModel/Univariate for one-way ANOVA. Then the Contrasts button does not allow setting arbitrary contrasts. Instead, there a fixed set of named planned contrasts. Figure 13.2 shows the “Univariate: Contrasts” dialog box. In this figure the contrast type has been changed from the default “None” to “Repeated”. Note the word “Repeated” under Factors confirms that the change of contrast type has actually been registered by pressing the Change button. *Be sure to also click the Change button whenever you change the setting of the Contrast choice, or your choice will be ignored!* The pre-set contrast choices include “Repeated” which compares adjacent levels, “Simple” which compares either the first or last level to all other levels, polynomial which looks for increasing orders of polynomial trends, and a few other less useful ones. These are all intended as planned contrasts, to be chosen before running the experiment.



Figure 13.2: Univariate contrasts dialog box.

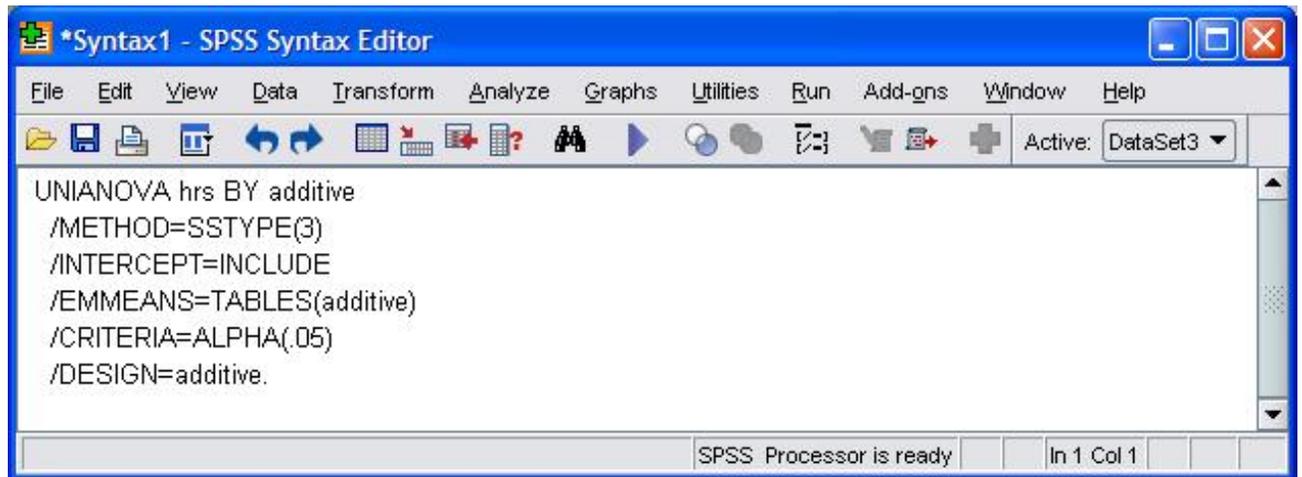


Figure 13.3: Univariate syntax window.

Custom Hypothesis Tests #1
Contrast Results (K Matrix)

Contrast		Dependent
		hrs
L1	Contrast Estimate	0.138
	Hypothesized Value	0
	Difference(Estimate-Hypothesized)	0.138
	Std. Error	0.338
	Sig.	0.724
	95% Confidence Interval Lower Bound	-0.642
	for Difference Upper Bound	0.918

Based on user-specified contrast coefficients: first vs. second and third

Table 13.3: Planned contrast in one-way ANOVA using LMATRIX syntax.

To make a *custom* set of planned contrasts in the Univariate procedure, click the Paste button of the Univariate dialog box. This brings up a syntax window with the SPSS native commands that are equivalent to the menu choices you have made so far (see Figure 13.3). You can now insert some appropriate subcommands to test your custom hypotheses. You can insert the extra lines anywhere between the first line and the final period. The lines that you would add to the Univariate syntax to test $H_{01} : \mu_1 - \frac{\mu_2 + \mu_3}{2} = 0$ and $H_{02} : \mu_2 - \mu_3 = 0$ are:

```
/LMATRIX = "first vs. second and third" additive 1 -1/2 -1/2
/LMATRIX = "second vs. third" additive 0 1 -1
```

Note that you can type any descriptive phrase inside the quotes, and SPSS will not (cannot) test if your phrase actually corresponds to the null hypothesis defined by your contrasts. Also note that fractions are allowed here. Finally, note that the name of the factor (additive) precedes its list of coefficients.

The output of the first of these LMATRIX subcommands is shown in Table 13.3. This gives the p-value and 95%CI appropriate for a planned contrast.

13.4.2 Contrasts for Two-way ANOVA

Contrasts in two-way (between-subjects) ANOVA *without* interaction work just like in one-way ANOVA, but with separate contrasts for each factor. Using the Univariate procedure on the Analyze/GeneralLinearModel menu, if one or both factors has more than two levels, then pre-defined planned contrasts are available with the Contrasts button, post-hoc comparisons are available with the Post-Hoc button, and arbitrary planned contrasts are available with Paste button and LMA-TRIX subcommands added to the Syntax.

For a $k \times m$ two-way ANOVA *with* interaction, two types of contrasts make sense. For planned comparisons, out of the km total treatment cells, you can test up to $(k-1)(m-1)$ pairs out of the $\binom{km}{2} = \frac{km(km-1)}{2}$ total pairs. With the LMA-TRIX subcommand you can only test a particular subset of these: comparisons between any two levels of one factor when the other factor is fixed at any particular level. To do this, you must first check the order of the two factors in the DESIGN line of the pasted syntax. If the factors are labeled A and B, the line will look either like

```
/DESIGN=A B A*B
```

or

```
/DESIGN=B A B*A
```

Let's assume that we have the "A*B" form with, say, 3 levels of factor A and 2 levels of factor B. Then a test of, say, level 1 vs. 3 of factor A when factor B is fixed at level 2 is performed as follows: Start the LMATRIX subcommand in the usual way:

```
/LMATRIX="compare A1B2 to A3B2"
```

Then add coefficients for the varying factor, which is A in this example:

```
/LMATRIX="compare A1B2 to A3B2" A 1 0 -1
```

Finally add the "interaction coefficients". There are km of these and the rule is "the first factor varies slowest". This means that if the interaction is specified as

A*B in the DESIGN statement then the first set of coefficients corresponds to all levels of B when A is set to level 1, then the next set is all levels of B when A is set to level 2, etc. For our example with we need to set A1B2 to 1 and A3B2 to -1, while setting everything else to 0. The correct subcommand is:

```
/LMATRIX="compare A1B2 to A3B2"  A 1 0 -1  A*B 0 1 0 0 0 -1
```

It is helpful to space out the A*B coefficients in blocks to see what is going on better. The first block corresponds to level 1 of factor A, the second block to level 2, and the third block to level 3. Within each block the first number is for B=1 and the second number for B=2. It is in this sense that B is changing quickly and A slowly as we move across the coefficients. To reiterate, position 2 in the A*B list corresponds to A=1 and B=2, while position 6 corresponds to A=3 and B=2. These two have coefficients that match those of the A block (1 0 -1) and the desired contrast ($\mu_{A1B2} - \mu_{A3B2}$).

To test other types of planned pairs or to make post-hoc tests of all pairs, you can convert the analysis to a one-way ANOVA by combining the factors using a calculation such as 10*A+B to create a single factor that encodes the information from both factors and that has km different levels. Then just use one-way ANOVA with either the specific planned hypotheses or the with the Tukey post-hoc procedure.

The other kind of hypothesis testing that makes sense in two-way ANOVA with interaction is to test the interaction effects directly with questions such as “is the effect of changing from level 1 to level 3 of factor A when factor B=1 the same or different from the effect of changing from level 1 to level 3 of factor A when factor B=2?” This corresponds to the null hypothesis: $H_0 : (\mu_{A3B1} - \mu_{A1B1}) - (\mu_{A3B2} - \mu_{A1B2}) = 0$. This can be tested as a planned contrast within the context of the two-way ANOVA with interaction by using the following LMATRIX subcommand:

```
/LMATRIX="compare A1 to A3 for B1 vs. B2"  A*B -1 1 0 0 1 -1
```

First note that we *only* have the interaction coefficients in the LMATRIX subcommand for this type of contrast. Also note that because the order is A then B in A*B, the A levels move change slowly, so the order of effects is A1B1 A1B2 A2B1 A2B2 A3B1 A3B2. Now you can see that the above subcommand matches the above null hypothesis. For an example of interpretation, assume that for fixed levels of both B=1 and B=2, A3-A1 is positive. Then a positive Contrast Estimate

for this contrast would indicate that the outcome difference with $B=1$ is greater than the difference with $B=2$.