1

Why contrasts?

The nature of contrasts

Contrasts are significance tests of focused questions in which specific predictions can be evaluated by comparing these predictions to the obtained data. By a focused test (as opposed to an omnibus test), we mean any statistical test that addresses precise questions, as in any 1 *df* *F* test or in any *t* test. Omnibus tests, on the other hand, are tests of significance that address diffuse (or unfocused) questions, as in *F* with numerator $df > 1$ or in χ^2 with $df > 1$.

To illustrate, suppose a developmental psychologist interested in psychomotor skills had children at five age levels (11, 12, 13, 14, and 15) play a new outer space shoot-'em-up video game called Spear-Man. The object of the game is to fly around advancing ranks of demon spaceships from the planet Rho and to try to "spear" them with space bullets while trying to avoid space garbage, meteorites, and rockets that are flying toward you from several different directions. Table 1.1 shows the mean performance of 10 children at each of these five age levels, and Table 1.2 shows the overall analysis of variance computed on these data.

Table 1.1. *Mean performance scores at five age levels*[a]

		Age levels		
11	12	13	14	15
25	30	40	50	55

[a] $n = 10$ at each age level.

Table 1.2. *Analysis of variance of performance scores*

Source	*SS*	*df*	*MS*	*F*	*p*
Age levels	6500	4	1625	1.03	.40
Within	70,875	45	1575		

We see in Table 1.2 that *F* for age levels is far from significant ($p = .40$). Should we conclude that age was not an effective variable in this investigation? If we did so, we would be making a mistake, for Figure 1.1 (which plots the means of Table 1.1) clearly shows a linear relationship between age and performance. When we compute the correlation between these variables, we indeed find $r(3) = .992$, $p < .001$ (two-tailed)!

Why were we led astray by the results of the analysis of variance telling us that age did not make a difference, when we see clear and obvious results once we plot the means and compute *r*? The answer is that our omnibus *F* test addressed the question of whether there were *any* differences among the five groups, disregarding entirely the arrangement of the ages that constitute

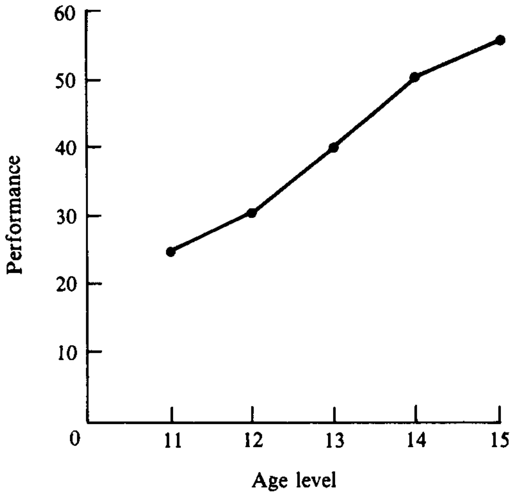

Figure 1.1. Means of Table 1.1 showing a linear relationship between age and performance.

the levels of the independent variable. We could rearrange the ages any way at all – 15, 14, 13, 12, 11 or 13, 12, 14, 15, 11 – and it would give us the same F as arranging them in the order previously shown. Our omnibus F test addresses a question actually of little interest to our researcher – the question is diffuse and unfocused. Our developmental researcher probably wanted to answer the question of whether psychomotor skills increased with age (i.e., showed a linear relationship) or first rose and then fell (i.e., showed a quadratic relationship), and so on. Our correlation addressed the specific question of whether performance increased linearly with age, and the answer was yes.

Comparing treatment means

To reiterate, contrast analysis permits us to ask focused questions of our data. Contrasts are comparisons employing two or more groups which we set up in such a way as to ensure that the results are compared to the predictions we make based on theory, hypothesis, or hunch. These predictions are expressed as lambda (λ) weights. The weights can take on any convenient numerical values as long as the sum of the weights ($\Sigma\lambda$) is zero for any given contrast. (We return to this idea in Chapter 2 and show how it applies in the example just discussed as well as in other illustrations of one-way analysis of variance.) What contrasts allow us to do is (a) to develop, at the outset, a number of focused questions we want to answer separately when analyzing our data (using *planned comparisons* instead of the overall analysis of variance and the omnibus F test) and (b) to use *incidental* or *post-hoc comparisons* to "snoop around" in our data after the overall analysis until we feel that we understand what the data actually show (Hays, 1963).

Previously, we noted as an example of a focused test any F with numerator $df = 1$ or any t test (and in this case, of course, $F = t^2$). Then we are comparing (or contrasting) only two means, $\bar{X}_1$ and $\bar{X}_2$. To be sure, we can also plan comparisons between means when there are more than two treatments (F with numerator $df > 1$), for example, three means:

$$\bar{X}_1 - \bar{X}_2 \quad \text{and} \quad \bar{X}_1 - \bar{X}_3 \quad \text{and} \quad \bar{X}_2 - \bar{X}_3$$

Suppose $\bar{X}_1$ is an experimental treatment group, $\bar{X}_2$ is a no-treatment control group, and $\bar{X}_3$ is a quasi-control group in which the subjects pretend (or simulate) being in the experimental group. We might well want to compare the average responses of each group with every other group. Alternatively, we might (after seeing there was no difference between the two control groups) want to combine the scores in $\bar{X}_2$ and $\bar{X}_3$ and compare the average responses in these two groups with those in $\bar{X}_1$; or else we might want to make up some other combination and average two of the three means and compare this average with the third mean, that is

$$\left(\frac{\bar{X}_1 + \bar{X}_2}{2}\right) - \bar{X}_3$$

and $$\left(\frac{\bar{X}_1 + \bar{X}_3}{2}\right) - \bar{X}_2$$

and $$\left(\frac{\bar{X}_2 + \bar{X}_3}{2}\right) - \bar{X}_1$$

In sum, the nature of the particular contrasts we choose to make – whether they be a series of comparisons between two treatment means or the average of two means compared with a third mean (as shown above) or a test for some trend in a given set of data – will depend on (a) our theory, hypothesis, or hunch and on (b) the type of data or research design we are using. In later chapters we shall examine a wide range of situations requiring different contrasts in the analysis of variance. Once we have mastered the use of contrasts to ask focused questions of our data (or of others' data, as discussed in Chapter 7), we will find that there are relatively few circumstances under which we will want to use omnibus F tests. What we will be getting in return for the small amount of computation required to employ contrasts is (a) very much greater statistical power and (b) very much greater clarity of substantive interpretation of research results.

Increased clarity of interpretation

It goes almost without saying that clarity of interpretation is essential in research, that is, to the extent that clarity is possible once we have subjected our data to proper examination (cf. Rosnow,

Table 1.3. *Mean scores of conservatives and liberals*[a]

Advertisement	Conservatives	Liberals	Means
Variation A	2.0	5.0	3.5
Variation B	2.0	1.0	1.5
Means	2.0	3.0	2.5

[a]n = 16 per cell.

1981). We will be talking more about interactions in later discussions of contrasts and a common example of the failure to subject data to proper examination occurs when researchers interpret interaction effects in the analysis of variance. Suppose we are comparing two variations of a political advertisement on a sample of 32 conservatives and 32 liberals. Table 1.3 shows the (post-treatment) opinion scores of the research participants, and Table 1.4 shows the analysis of variance of these data. In the published report, we could accurately state that there was a significant effect of the type of advertisement such that participants exposed to variation A expressed more positive opinions than participants exposed to B. We could also accurately state that there was no significant effect of the liberal versus conservative viewpoint of the participants on their expressed opinions regarding the advertisement. Finally, we might state – *but it would be wrong!* – that the significant interaction effect shown in Figure 1.2 (displaying the means of Table 1.3) demonstrates that liberals were more strongly influenced by advertisement A than by B while conservatives were unaffected by the type of advertisement.

Table 1.4. *Analysis of variance of opinion scores*

Source	*SS*	*df*	*MS*	*F*	*p*
Advertisements	64	1	64	4.0	.05
Respondents	16	1	16	1.0	–
Interaction	64	1	64	4.0	.05
Error term	960	60	16		

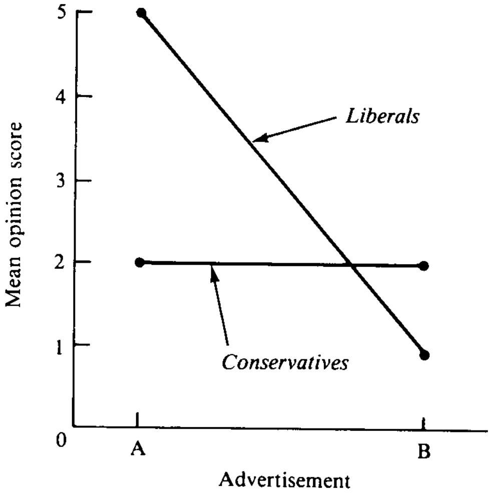

Figure 1.2. Overall results based on means of Table 1.3.

How have we erred? Figure 1.2 is a perfectly accurate display of the *overall* results of the study, in that it includes both main effects as well as the interaction. But it is not an accurate display of the interaction we believed was being depicted. To see what the interaction actually looks like, we would need to examine the data more closely. If we are to clarify the interaction accurately, then it is necessary for us to identify the residuals defining it. In this illustration all that is required is that we focus our data analysis by subtracting the grand mean along with the row and column effects from each condition of the experiment. Unfortunately, many researchers fail to do this, and instead base their interpretation (or rather misinterpretation) of the interaction on the overall means (i.e., the condition means) alone.

Table 1.5. *Row and column effects for means*

Advertisement	Conservatives	Liberals	Means	Row effects
Variation A	2.0	5.0	3.5	1.0
Variation B	2.0	1.0	1.5	−1.0
Means	2.0	3.0	2.5	
Column effects	−0.5	0.5		

Table 1.6. *Means "corrected for" row effects*

Advertisement	Conservatives	Liberals	Means	Row effects
Variation A	1.0	4.0	2.5	0
Variation B	3.0	2.0	2.5	0
Means	2.0	3.0	2.5	
Column effects	−0.5	0.5		

Since in Chapter 3 we return to the interpretation of interaction effects, it may be helpful if we review these calculations. Table 1.5 defines the particular row and column effects. Row effects are indicated for each row as the mean of that row minus the grand mean, and column effects are defined as the mean of each column minus the grand mean. Turning to Table 1.6, we see the original means "corrected for" row effects (that is, with the row effects removed), and in Table 1.7 we see these "corrected" means with the column effects now removed. In other words, we have decomposed the condition means shown in Table 1.3 to reveal the interaction effects to be the residual effects, or effects remaining after the lower-order effects of the rows and columns have been removed. The following summary table reviews what we have done so far:

	Interaction effect	=	Condition mean	−	Row effect	−	Column effect
Var A/Con	1.5	=	(2.0)	−	(1.0)	−	(−0.5)
Var B/Con	3.5	=	(2.0)	−	(−1.0)	−	(−0.5)
Var A/Lib	3.5	=	(5.0)	−	(1.0)	−	(0.5)
Var B/Lib	1.5	=	(1.0)	−	(−1.0)	−	(0.5)
Sums	10.0	=	10.0	−	0	−	0

Table 1.7. *Means "corrected for" row and column effects*

Advertisement	Conservatives	Liberals	Means	Row effects
Variation A	1.5	3.5	2.5	0
Variation B	3.5	1.5	2.5	0
Means	2.5	2.5	2.5	
Column effects	0	0		

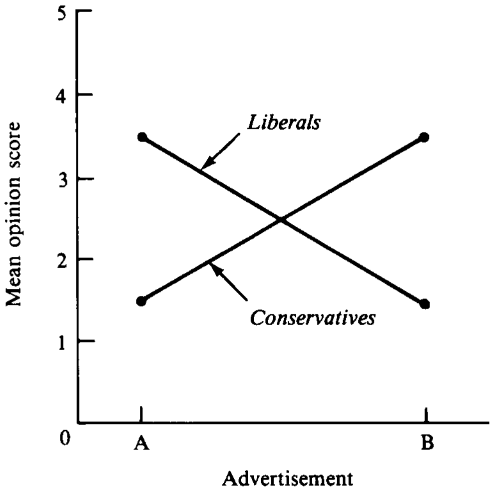

Figure 1.3. Interaction effect based on "corrected" means of Table 1.7.

Figure 1.3 displays the means shown in Table 1.7 (also in the column labeled "Interaction effect" above), that (with row and column effects removed) reveal the actual interaction effects. In this figure we see that the interaction actually shows that conservatives and liberals reacted in exactly opposite ways to the two types of propaganda. This is a very different conclusion than that based on the *overall* results depicted in Figure 1.2.

In most situations, we prefer to display the interaction effects freed of the grand mean as well as the row and column effects. To remove the grand mean from Table 1.7, we simply subtract 2.5 from every condition of the experiment. (For an extensive review of the computation and interpretation of interaction effects in the analysis of variance, see Chapter 21 in Rosenthal and Rosnow's *Essentials of Behavioral Research*, 1984.) Table 1.8 shows these results, which if plotted would be identical to those displayed in Figure 1.3 (except for changing the metric underlying the dependent variable. What we have done in subtracting the grand mean, row and column effects from the condition means can be summarized as follows:

	Interaction effect	=	Condition mean	−	Row effect	−	Column effect	−	Grand mean
Var A/Con	−1.0	=	(2.0)	−	(1.0)	−	(−0.5)	−	(2.5)
Var B/Con	1.0	=	(2.0)	−	(−1.0)	−	(−0.5)	−	(2.5)
Var A/Lib	1.0	=	(5.0)	−	(1.0)	−	(0.5)	−	(2.5)
Var B/Lib	−1.0	=	(1.0)	−	(−1.0)	−	(0.5)	−	(2.5)
Sums	0	=	10.0	−	0	−	0	−	10.0

Significance and effect size

In the example that began this chapter, we saw how an omnibus *F* test tells almost nothing except to emphasize overall significance of results. However, even when we do comparisons between treatment means, it is a good idea not to base our conclusions only on the statistical significance of our findings. Significance reveals only part of the overall picture. There will almost always be two kinds of information we want to have for each research question we address with planned or post-hoc comparisons: (a) the size of the effect and (b) its statistical significance. This theme can be expressed by a fundamental conceptual equation (Rosenthal and Rosnow, 1984):

$$\text{Significance test} = \text{size of effect} \times \text{size of study}$$

This equation tells us that, for any given size of effect (e.g., r, r^2, $r/\sqrt{1 - r^2}$, or $r^2/(1 - r^2)$) and size of study (e.g., N, df, $\sqrt{N}$, $\sqrt{df}$), there will be a corresponding test of significance.

Table 1.8. *Means "corrected for" row, column, and grand mean effects*

Advertisement	Conservatives	Liberals	Means	Row effects
Variation A	−1.0	1.0	0	0
Variation B	1.0	−1.0	0	0
Means	0	0	0	
Column effects	0	0		

Another way of saying this is that every test of significance (e.g., t or F) is made up of two components, the size of the effect and the size of the study. So, for example, when there are two means to be compared and the size of the effect of the independent variable is indexed by some correlation, the general relationship above can be rewritten as

$$t = \frac{r}{\sqrt{1 - r^2}} \times \sqrt{df}$$

or as

$$F = \frac{r^2}{1 - r^2} \times df$$

In the following chapters we shall refer both to estimates of the size of the relationship and its statistical significance as we explore various ways in which comparison procedures can be used in the analysis of variance. It is possible, of course, to do focused tests in the context of other analytic procedures, for example, within the framework of meta-analysis procedures that are used for comparing the significance levels or effect sizes of several research studies. When research studies are compared as to their significance or their effect sizes by focused tests, or contrasts, we learn whether the studies differ significantly among themselves in a theoretically predictable or meaningful way. Suppose we have a given set of p values for studies of teacher expectancy effects; we might want to know whether results from younger children show greater degrees of statistical significance than do results from older children (Rosenthal & Rubin, 1978a, 1978b). Meta-analytic procedures are extensively discussed elsewhere (Rosenthal, 1984), while in this volume we are concerned primarily with the essentials of comparison procedures in the analysis of variance.