Chapter 17—Factorial Analysis of Variance

17.1 Thomas and Wang (1996) study:

a) This design can be characterized as a 3 × 2 factorial, with 3 levels of Strategy and 2 levels of delay.
b) I would expect that recall would be better when subjects generated their own key words, and worse when subjects were in the rote learning condition. I would also expect better recall for the shorter retention interval. (But what do I know?)
c)

```
Summaries of     RECALL
By levels of     STRATEGY
                 DELAY
Variable        Value  Label              Mean    Std Dev    Cases

For Entire Population                11.602564   7.843170     78

STRATEGY       1.0000                 9.461538   6.906407     26
  DELAY        1.0000                14.923077   5.330127     13
  DELAY        2.0000                 4.000000   2.516611     13

STRATEGY       2.0000                11.269231   9.606488     26
  DELAY        1.0000                20.538462   1.983910     13
  DELAY        2.0000                 2.000000   1.471960     13

STRATEGY       3.0000                14.076923   6.183352     26
  DELAY        1.0000                15.384615   5.454944     13
  DELAY        2.0000                12.769231   6.796492     13
```

17.3 Analysis of variance on data in Exercise 17.1:

```
RECALL      by   STRATEGY
                 DELAY

                 UNIQUE sums of squares
                 All effects entered simultaneously
                           Sum of                  Mean               Sig
Source of Variation        Squares      DF        Square        F    of F

Main Effects              2510.603       3       836.868    42.992   .000
   STRATEGY                281.256       2       140.628     7.224   .001
   DELAY                  2229.346       1      2229.346   114.526   .000

2-Way Interactions         824.538       2       412.269    21.179   .000
   STRATEGY DELAY          824.538       2       412.269    21.179   .000

Explained                 3335.141       5       667.028    34.267   .000

Residual                  1401.538      72        19.466

Total                     4736.679      77        61.515
```

There are significant differences due to both Strategy and Delay, but, more importantly, there is a significant interaction. These effects are easily seen in the figure in Exercise 17.2.

17.5 Bonferroni tests to clarify simple effects for data in Exercise 17.4:

$$t = \frac{\overline{X}_i - \overline{X}_j}{\sqrt{\frac{MS_{error}}{n_i} + \frac{MS_{error}}{n_j}}}$$

For Data at 5 Minutes Delay:

For Generated versus Provided:

$$t = \frac{14.92 - 20.54}{\sqrt{\frac{20.7009}{13} + \frac{20.7009}{13}}} = \frac{-5.62}{1.784} = -3.15$$

For Generated versus Rote:

$$t = \frac{14.92 - 15.38}{\sqrt{\frac{20.7009}{13} + \frac{20.7009}{13}}} = \frac{-0.46}{1.784} = -0.26$$

For Provided versus Rote:

$$t = \frac{20.54 - 15.38}{\sqrt{\frac{18.2308}{13} + \frac{18.2308}{13}}} = \frac{5.16}{1.784} = 2.89$$

For Data at 2 Day Delay:

For Generated versus Provided:

$$t = \frac{4.00 - 2.00}{\sqrt{\frac{18.2308}{13} + \frac{18.2308}{13}}} = \frac{2.00}{1.674} = 1.19$$

For Generated versus Rote:

$$t = \frac{4 - 12.77}{\sqrt{\frac{18.2308}{13} + \frac{18.2308}{13}}} = \frac{-8.77}{1.674} = -5.24$$

For Provided versus Rote:

$$t = \frac{2 - 12.77}{\sqrt{\frac{20.7009}{13} + \frac{20.7009}{13}}} = \frac{-10.77}{1.674} = -6.43$$

For 6 comparisons with 36 *df*, the critical value of *t* is 2.80.

For the 5-minute delay, the condition with the key words provided by the experimenter is significantly better than both the condition in which the subjects generate their own key words and the rote learning condition. The latter two are not different from each other.

For the 2-day delay, the rote learning condition is better than either of the other two conditions, which do not differ between themselves.

We clearly see a different pattern of differences at the two delay conditions. The most surprising result (to me) in the superiority of rote learning with a 2 day interval.

17.7 The results in the last few exercises have suggested to me that if I were studying for a Spanish exam, I would fall back on rote learning, painful as it sounds and as much against common wisdom as it is.

17.9 In this experiment we have as many primiparous mothers as multiparous ones, which certainly does not reflect the population. Similarly, we have as many LBW infants as full-term ones, which is certainly not a reflection of reality. The mean for primiparous mothers is based on an equal number of LBW and full-term infants, which we know is not representative of the population of all primiparous births. Comparisons between groups are still legitimate, but it makes no sense to take the mean of all primiparous moms combined as a reflection of any meaningful population mean.

17.11 Simple effects versus *t* tests for Exercise 17.10.

a) If I had run a *t* test between those means my result would simply be the square root of the $F = 1.328$ that I obtained.

b) If I used MS_{error} for my estimated error term it would give me a *t* that is the square root of the *F* that I would have had if I had used the overall MS_{error}, instead of the MS_{error} obtained in computing the simple effect.

17.13 Analysis of variance for Spilich *et al*. Study:

Tests of Between-Subjects Effects

Dependent Variable: ERRORS

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	31744.726[a]	8	3968.091	36.798	.000
Intercept	45009.074	1	45009.074	417.389	.000
TASK	28661.526	2	14330.763	132.895	.000
SMKGRP	354.548	2	177.274	1.644	.197
TASK * SMKGRP	2728.652	4	682.163	6.326	.000
Error	13587.200	126	107.835		
Total	90341.000	135			
Corrected Total	45331.926	134			

a. R Squared = .700 (Adjusted R Squared = .681)

The main effect of Task and the interaction are significant. The main effect of Task is of no interest because there is no reason why different tasks should be equally difficult., We don't care about the main effect of Smoking either because it is created by large effects for two levels of Task and no effect for the third. What is important is the interaction.

17.15 Simple effects to clarify the Spilich *et al.* Example.

We have already seen these simple effects in Chapter 16, in Exercises 16.18, 16.19, and 16.21.

17.17 Factorial analysis of the data in Exercise 16.2:

Tests of Between-Subjects Effects

Dependent Variable: SCORE

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1059.800[a]	3	353.267	53.301	.000
Intercept	5017.600	1	5017.600	757.056	.000
AGE	115.600	1	115.600	17.442	.000
HILO	792.100	1	792.100	119.512	.000
AGE * HILO	152.100	1	152.100	22.949	.000
Error	238.600	36	6.628		
Total	6316.000	40			
Corrected Total	1298.400	39			

a. R Squared = .816 (Adjusted R Squared = .801)

Here we see that we have a significant effect due to age, with younger subjects outperforming older subjects, and a significant effect due to the level of processing, with better recall of material processed at a higher level. Most importantly, we have a significant interaction, reflecting the fact that there is no important difference between younger and older subjects for the task with low levels of processing, but there is a big difference when the task calls for a high level of processing—younger subjects seem to benefit more from that processing (or do more of it).

17.19 η^2 and ω^2 for the data in Section 17.7:

Summary table from Section 17.7:

Source	***df***	***SS***	***MS***	***F***
Education	1	67.69	67.69	6.39
Group	2	122.79	61.40	5.80
E × G	2	20.38	10.19	0.96
Error	42	444.62	10.59	
Total	47	655.48		

$$\eta^2_{Educ} = \frac{SS_{Educ}}{SS_{total}} = \frac{67.69}{655.48} = .10$$

$$\omega^2_{Educ} = \frac{SS_{Educ} - (e-1)MS_{error}}{SS_{total} + MS_{error}} = \frac{67.69 - (2-1)10.59}{655.48 + 10.59} = .086$$

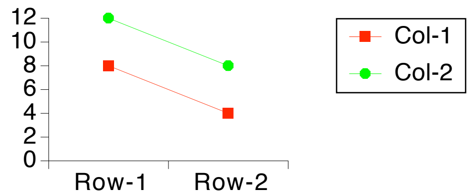
$$\eta^2_{Group} = \frac{SS_{Group}}{SS_{total}} = \frac{122.79}{655.48} = .19$$

$$\omega^2_{Group} = \frac{SS_{Group} - (g-1)MS_{error}}{SS_{total} + MS_{error}} = \frac{122.79 - (3-1)10.59}{655.48 + 10.59} = .15$$

$$\eta^2_{EG} = \frac{SS_{EG}}{SS_{total}} = \frac{20.38}{655.48} = .03$$

$$\omega^2_{EG} = \frac{SS_{EG} - (e-1)(e-1)MS_{error}}{SS_{total} + MS_{error}} = \frac{20.38 - (2-1)(3-1)10.59}{655.48 + 10.59} = .00$$

17.21 Set of data for a 2 × 2 design with a two main effects but no interaction:



Cell means:

8	12
6	8

17.23 The interaction was of primary interest in an experiment by Nisbett in which he showed that obese people varied the amount of food they consumed depending on whether a lot or a little food was visible, while normal-weight subjects ate approximately the same amount under the two conditions.

17.25 Calculation of Effect Size for main effects in Exercise 17.1

$$\hat{d}_{\text{Time}} = \frac{\bar{X}_{5\text{min}} - \bar{X}_{2\text{day}}}{\sqrt{MS_{error}}} = \frac{16.949 - 6.256}{\sqrt{19.466}} = \frac{10.693}{4.412} = 2.42$$

This represents a huge effect size for the Time variable.

It would be logical to compare condition in which participants select their own keyword to the rote learning condition.

$$\hat{d}_{\text{Condition}} = \frac{\bar{X}_{generate} - \bar{X}_{\text{rote}}}{\sqrt{MS_{error}}} = \frac{9.462 - 14.077}{\sqrt{19.466}} = \frac{-4.615}{4.412} = 1.05$$

The two groups differ by about one standard deviation, which is still a large effect.

17.27 Effect size for data in Exercise 17.13

It does not make a lot of sense to look for effect size measures for tasks, because there is nothing particularly interesting in that. However, because I asked the question, I ought to supply an answer, so I will compare the Pattern Recognition and Recall tasks, because they are both relatively passive.

$$\hat{d}_{\text{Task}} = \frac{\bar{X}_{Pat\,\text{Re}\,c} - \bar{X}_{\text{Recall}}}{\sqrt{MS_{error}}} = \frac{9.644 - 38.778}{\sqrt{107.835}} = \frac{-29.134}{10.3844} = -2.8$$

The two tasks differ by 2.8 standard deviations in the number of errors.

It does make sense to compare the non-smoking and actively smoking participants, because those are meaningful groups.

$$\hat{d}_{\text{Smoke}} = \frac{\overline{X}_{\text{NonSmoke}} - \overline{X}_{\text{Active}}}{\sqrt{MS_{error}}} = \frac{16.067 - 19.933}{\sqrt{107.835}} = \frac{-3.866}{10.3844} = -0.37$$

The two conditions differ by about a third of a standard deviation, which is a reasonably large difference by Cohen's standards.

17.29 The important test for the Eysenck data is the interaction, because Eysenck thought that the two age groups would only differ in important ways when the tasks involved a greater amount of cognitive processing.

17.31 Liddle's study of sexual orientation.

Sexual Identification:

	Disclose	Not Disclose	Mean
Female	37.15	36.56	36.86
Male	33.00	33.00	33.00
Means	35.08	34.78	34.93

Average within-cell variance = 20.74

$$SS_{Gender} = nc\sum\left(X_g - \overline{X}..\right)^2 = 15 \times 2[(36.86 - 34.93)^2 + (33.00 - 34.94)^2]$$
$$= 30(1.93^2 + (-1.93)^2) = 223.49$$
$$SS_{Condition} = ng\sum\left(X_c - \overline{X}..\right)^2 = 15 \times 2[(35.08 - 34.93)^2 + (33.28 - 34.94)^2]$$
$$= 30(0.15^2 + (-0.15)^2) = 1.35$$
$$SS_{cells} = n\sum\left(X_{cell} - \overline{X}..\right)^2 = 15 \times [(37.15 - 34.93)^2 + \ldots + (33.00 - 34.93)^2] = 30(4.928) = 225.53$$
$$SS_{GC} = SS_{Cells} - SS_{Gender} - SS_{Condition} = 225.53 - 223.49 - 1.35 = 0.69$$

Source	*df*	*SS*	*MS*	*F*
Gender	1	223.49	223.49	10.78
Condition	1	1.35	1.35	<1
G × X	1	0.69	0.69	<1
Error	56	1161.44	20.74	

(Liddle (1997) argued persuasively that the effect of Condition was so very small that the failure to find a significant difference could not be attributed to small sample sizes. There clearly seems to be no real difference between these groups.)