
Preface

Statistical design is one of the fundamentals of our subject, being at the core of the growth of statistics during the previous century. Design played a key role in agricultural statistics and set down principles of good practice, principles that still apply today. Statistical design is about understanding where the variance comes from, and making sure that is where the replication is. Indeed, it is probably correct to say that these principles are even more important today. Fisher (1947) compared a dataset to a sample of gold ore. The finest analysis could only extract the proportion of gold contained in the ore. But a good design could produce a sample with more gold.

There are plenty of “Design of Experiments” books available, many of which do a fine job of describing not only how to design experiments, but also how to analyze them. So why bother with another book? There were two main reasons.

The first reason is the observation that many of our “standard” analyses have become driven by the default setting of one’s favorite computer package and, unfortunately, many times these default settings provide an incorrect analysis. More frightening is the fact that sometimes such default analyses have found their way into textbooks.

The second reason is that, although design books have gotten broader in coverage of designs and often have much to say about analysis, the basic theoretical underpinnings are not always covered, and if they are, they are not covered in sufficient detail to understand how to construct the correct analysis and to understand why the computer package default analysis may be incorrect. Without fully understanding what the correct analysis *should be*, it is impossible to design a good experiment.

So ... the goal is to describe the principles that drive good design, which are also the principles that drive good statistics. Moreover, this will be done with detail and attention paid to the theoretical background – only by having more than a passing familiarity with the fundamental theory can one truly understand statistical design.

This book is not an encyclopedia of designs. There is no attempt to cover all designs, and no attempt to teach data analysis at all. Although we will analyze many datasets, we will usually use common anova techniques and not preach about transformations, heterogeneity, missing data, and all the other good stuff that a good data analyst needs to know. We assume that the design has been run with enough success so that rescue techniques are not needed for the analysis. We also assume that the student has been exposed to such data analysis strategies.

We will cover the most popular designs in depth, both with theory and examples, and datasets from consulting sessions and research publications. We emphasize basic principles and careful modeling and, armed with such tools, the student should be able to apply these principles in any situation

This book grew out of a course on Statistical Design taught at Cornell during the 1990s, and at Florida in the 2000s. Most of the examples and datasets are from consulting sessions with graduate students writing theses or professors writing papers, and the subjects span everything from planting alfalfa in the field to harvesting brain stem cells. (Most of my career has been spent at Colleges of Agriculture – Rutgers, Cornell, and split between Arts and Sciences and Agriculture at Florida. This will show in the examples.) I have found that although the data and the lab techniques have changed, the statistical principles remain quite constant.

Now, for the more important details:

- The level of the text is for first or second year graduate students. The students should be familiar with standard statistical methodology (anova, blocking, multiple regression) that one would get from a typical one-year methods sequence (from books such as Ott and Longnecker 2000, Rawlings *et al.* 1998).
- The material in the text is about right for a one-semester course. There is probably a bit more than can be comfortably covered, so some picking and choosing will be necessary. However, marching through the text is a reasonable strategy.
- The chapters cover, for the most part, the standard material of a design book, with mostly real examples, and applications of design in real situations. Although we cover many microarray designs, we do not have a special section, instead treating them as the topics apply. The only unusual chapter is Chapter 4, where the concept of blocking is explored further, and the effect of a random factor is examined. To me, blocks are not about being samples from a larger population (which can be difficult to justify) but rather about the correlation structure that they induce.
- Most chapters have a section *Technical Notes*, which contains the underlying theory in detail. The level in these notes is “anything goes”; we use a lot of matrix algebra, some calculus, and also some statistical concepts such as likelihood and sufficiency. These are in-depth looks that will enhance the understanding of the advanced student, but skipping these

sections will not hamper the beginner. To fully appreciate these sections it would be good if the student has had a course from a book like *Statistical Inference* by Casella and Berger (2001).

- The exercises are divided into “Essential” and “Accompaniment”. Everyone should do the essentials. The accompaniments tend to be more of a theoretical nature, going into the details of the procedures. I strongly suggest that, if the students are able, they should do these exercises too. I have always found that, for me, true understanding only comes from slogging through the details.
- The datasets are on the book web page found at

www.stat.ufl.edu/~casella

and, for most datasets in the examples there will be an accompanying R program. These will not be sophisticated analyses, but rather will serve as a starting point. I make no guarantees about the R programs. Smaller datasets may only appear in text.

My best advice is that if you really want to understand statistical design, read (or even better, reread) Fisher. His ideas, especially about blocking, have greatly influenced my thinking. In fact, what I perceived as mishandling of the randomized complete block design was one of the driving forces behind the text. Fisher, of course, got it right.

There are many gray areas in design – when to pool, how to replicate, etc. – some of which cannot be fully answered with statistical fact. This is where we enter the realm of opinions, where judgments are made more on anecdotal evidence and experience rather than formal calculations. After doing this stuff for over 20 years opinions form about how to do things – I have taken the liberty of sharing those thoughts.

Finally, thanks to all of my mentors. First there was Leon Gleser, my PhD advisor, who taught me to work hard and try to learn as much as possible. My design mentors were many, starting with Virgil Anderson at Purdue, and Walt Federer at Cornell, who not only made me really understand split-plot designs, but also taught me to say “a model” and *never* “the model”. And Carl Lowe, the Plant Breeding Professor at Cornell who taught the course “Field Plot Techniques”, and knew more about field plot layouts than anyone on the planet. When he retired he gave me all of his notes and examples, many of which appear as examples and exercises (and in the Appendix). I also thank the students and colleagues who suffered through my learning process, who listened as I thought through all of this, read through notes, and solved problems: Mihai Giurcanu, Jamie Jarabek, David Lansky, Michael Meredith, Deborah Reichert, and Andy Scherrer.

My most special mentor was Myra Samuels. A true sadness in my life is that she died in 1992. She taught me to always ask questions until you really understand. Thanks, Myra.

X Preface

And to all of my other mentors – you have all helped me more than you know – Bob Bechhoffer, Jim Berger, Larry Brown, Shanti Gupta, Jean-Pierre Habicht, George McCabe, Doug Robson, Bill Strawderman. And all of the others that I have learned from. Thanks.

George Casella
Gainesville, Florida

June 6, 2008

Completely Randomized Designs

And so it was ... borne in upon me that very often, when the most elaborate statistical refinements possible could increase the precision by only a few percent, yet a different design involving little or no additional experimental labour might increase the precision two-fold, or five-fold or even more

R. A. Fisher (1962)

If the idea looked lousy, I said it looked lousy. If it looked good, I said it looked good. Simple proposition.

Richard P. Feynman

Surely You're Joking, Mr. Feynman

2.1 Introduction

The terminology *Completely Randomized Design (CRD)* refers to the experiment design, not the data layout. In a CRD the treatments can be either crossed or nested,

but the key feature is that the randomization must be carried out throughout the data layout. One important implication of this is that all factors in a CRD must be fixed factors – as a random factor

CRDs have only fixed factors

is necessarily a restriction on randomization, there are no random factors in a CRD.

A theoretical consequence of the fact that all factors in a CRD must be fixed factors is that we can study the theory for *all* CRD simply by looking at the oneway CRD. This follows because of the simple error structure of the CRD, and the fact that any effect can be built up through contrasts. However, this fact is only useful as a theoretical tool, say when we are trying to develop some distributional properties, as the data layout and the treatment structure

always play an important practical role. But the oneway CRD is the place to start.

Most importantly, the randomization structure of the CRD implies that there is only one error term, the within error, and all effects are tested against it.

2.1.1 A Oneway CRD Model

Example 2.1. ONEWAY CRD An experiment was done to assess *in vitro digestibility* (IVD) of dried forage samples for alfalfa grown at different temperatures. The object is to determine if growing temperatures affected the feeding quality of alfalfa. There are four observations for each temperature, with data layout:

		Growing Temperatures (Celsius)			
		17°	22°	27°	32°
IVD Values	x	x	x	x	x
	x	x	x	x	x
	x	x	x	x	x
	x	x	x	x	x

The data are in dataset IVD. This experiment was done as a oneway CRD, which means that treatments were assigned at random to the different units. In particular, the four IVD values measured at 17° were *not* taken from four plots exposed together to the same 17° temperature. They were four separate trials, with independent exposure to the 17° temperature. That is, a treatment is selected at random, and the experimental unit is then subjected to the treatment. If 22° is selected twice in a row, the temperature must be reset! ||

The test on treatments is

$$F = \frac{MS(\text{Treatments})}{MS(\text{Within})},$$

which, under H_0 has an F -distribution. The validity of this test can be justified with Cochran's Theorem (see the discussion following Theorem 2.20).

2.1.2 CRD and the Two-sample t -test

The oneway anova is a generalization of the two sample t -test, and by building on the theory of contrasts, gives us an effective way of comparing many means. If we had just two growing conditions, say 17° and 22°, we could then do a two sample t -test to check if the responses are significantly different. If we denote the means, variances, and cell sizes by

	Growing Temperatures (Celsius)			
	17°	22°	27°	32°
IVD	\bar{y}_1	\bar{y}_2	\bar{y}_3	\bar{y}_4
Values	s_1^2	s_2^2	s_3^2	s_4^2
	n_1	n_2	n_3	n_4

the t -test is

$$t_{\text{calc}} = \frac{\bar{y}_1 - \bar{y}_2}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = 1.399, \quad p = .211,$$

where $s^2 = [(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2] / [n_1 + n_2 - 2]$ is the pooled variance. We see that there is no significant difference in response between the 17°C and 22°C growing temperature. We compare this t -test to a oneway anova using only treatments 17°C and 22°C.

Source	df	Sum Sq	Mean Sq	F	p
Trt	1	0.211	0.211	1.957	0.211
“Residuals”	6	0.647	0.107		

The anova table is taken from R output. Like all computer packages, R calls the last line of the anova “residuals”. However, for us it is very important to actually keep track of what this is: within? interaction? technical replication? Here, in the oneway CRD, we have “within” error.

Within error, sometimes called “pure” error, is very different from a “residual”. As the name implies, a residual is something that is left over. In statistics, the residual is left over from the model fit.

A digression on “residual” error

For example, if a higher-order interaction term is used as an error term, it is left over from a model fit, and its validity as an error estimate depends on the validity of the model. In contrast, the validity of a within error as an error estimate is not dependent on the fit of the model. Regardless of the model, each experimental unit within the cell receives exactly the same treatment, and hence any difference between the observations is not dependent on the model, and only reflects the inability of the experiment to replicate itself.

Example 2.2. IVD ONEWAY CRD CONTINUED

Returning to the IVD anova, we know that the oneway anova on two treatments is identical to the two sample t -test. We can continue, and if we (for example) use the simple contrasts of Exercise 1.10, we could run three t -tests and find the results in Table 2.1.

Although in this example the results are, qualitatively, about the same, it is important to realize that the anova uses the pooled MSE for its estimate of error, so the contrast tests have 12 degrees of freedom rather than 6. In the anova we get to use an error estimate based on pooling the within error

Table 2.1. Two-sample t tests and anova contrast t tests corresponding to IVD simple contrasts

Test	Two-Sample t			Anova contrasts		
	df	t_{calc}	p	df	t_{calc}	p
17°C vs. 22°C	6	1.399	0.211	12	1.181	.260
27°C vs. 32°C	6	-.160	.878	12	-.181	.858
17°C + 22°C vs. 27°C + 32°C	6	-5.851	.001	12	-6.102	< .0001

from all four treatments, even if those treatments are not being tested in the contrast.

Orthogonal contrasts do not guarantee independent tests

It is also important to remember that even though we are using orthogonal contrasts, which are independent in this case, the tests are *not* independent because they share the same denominator.

If we apply the formula for contrast sums of squares (Definition 1.14) to Table 2.1, we see that the contrast sums of squares will add to the anova treatment sum of squares. Moreover, the contrast t -statistics are a breakdown of the anova F -test in that (referring to Table 2.1)

$$\frac{(1.181)^2 + (-0.181)^2 + (-6.102)^2}{3} = 12.89$$

which is the F -statistic from the complete anova (Exercise 2.3). ||

2.1.3 CRD Anova

The oneway anova F -test is summarized in Table 2.2, which reflects the partitioning of the sums of squares according to (1.8)

The CRD design, of course, does not stop at the oneway. As we will see, the treatment design in a CRD can take many forms, the defining feature being that the randomization is unrestricted throughout the table, and there is no correlation between any two observations.

Example 2.3. RED CLOVER TOWAY CRD To investigate the effect of sulfur and nitrogen on the growth of red clover, a plant scientist conducted a greenhouse experiment using a CRD with the treatments in a crossed layout. The sulfur levels were applied at rates of 0, 3, 6, and 9 pounds/acre, and the rate of nitrogen application was either 0 or 20 pounds/acre. Greenhouse pots were prepared with uniform soil, allowing for $r = 3$ pots per treatment combination. The data are given in Table 2.3 ||

Table 2.2. Oneway anova table.

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F-Statistic
Treatments	$t - 1$	$SS(\text{Trt}) = \sum_i r(\bar{y}_{i.} - \bar{\bar{y}})^2$	$MS(\text{Trt}) = \frac{SS(\text{Trt})}{t-1}$	$F = \frac{MS(\text{Trt})}{MS(\text{Within})}$
Within	$t(r - 1)$	$SS(\text{Within}) = \sum_i \sum_j (y_{ij} - \bar{y}_{i.})^2$	$MS(\text{Within}) = \frac{SS(\text{Within})}{t(r-1)}$	
Total	$rt - 1$	$SS(\text{Total}) = \sum \sum (y_{ij} - \bar{\bar{y}})^2$		

Table 2.3. Dry matter yields, in grams/pot, of red clover.

		Sulfur			
		0	3	6	9
Nitrogen	0	4.48	4.70	5.21	5.88
		4.52	4.65	5.23	5.98
	20	4.63	4.57	5.28	5.88
		5.76	7.01	5.88	6.26
		5.72	7.11	5.82	6.26
		5.78	7.02	5.73	6.37

The partitioning of the sums of squares for the twoway anova is a straightforward generalization of the oneway partition, giving

$$SS(\text{Total}) = SS(\text{Treatment A}) + SS(\text{Treatment B}) + SS(\text{A} \times \text{B}) + SS(\text{Within}),$$

where, here, the interaction term measures the nonadditivity of the treatments, and the within serves as the error for all estimation and tests. The details of the partitioning are

$$\begin{aligned}
 \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r (y_{ijk} - \bar{\bar{y}})^2 &= \sum_{i=1}^t rb(\bar{y}_{i..} - \bar{\bar{y}})^2 + \sum_{j=1}^g ra(\bar{y}_{.j.} - \bar{\bar{y}})^2 \\
 (2.1) \qquad &+ \sum_{i=1}^t \sum_{j=1}^g (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{\bar{y}})^2 \\
 &+ \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r (y_{ijk} - \bar{y}_{ij.})^2,
 \end{aligned}$$

which results in the anova table given in Table 2.4.

Table 2.4. Twoway anova table

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square	F -statistic
T	$t - 1$	$SS(T) = \sum_i r g (\bar{y}_{i..} - \bar{y})^2$	$MS(T) = \frac{SS(T)}{t-1}$	$F = \frac{MS(T)}{MS(Within)}$
G	$g - 1$	$SS(G) = \sum_j r g (\bar{y}_{.j.} - \bar{y})^2$	$MS(G) = \frac{SS(G)}{g-1}$	$F = \frac{MS(G)}{MS(Within)}$
T × G	$(t - 1)(g - 1)$	$SS(T \times G) = \sum_{ij} r (\bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y})^2$	$MS(T \times G) = \frac{SS(T \times G)}{(t-1)(g-1)}$	$F = \frac{MS(T \times G)}{MS(Within)}$
Within	$tg(r - 1)$	$SS(Within) = \sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2$	$MS(Within) = \frac{SS(Within)}{tg(r-1)}$	
Total	$tgr - 1$	$SS(Total) = \sum_{ijk} (y_{ijk} - \bar{y})^2$		

Remember that the treatment design, which is a twoway factorial, is all that is needed to determine all of the anova table except the last column. The F -ratios can only be properly formed with knowledge of the experiment design, and those in Table 2.4 reflect the fact that we have a CRD.

The treatment design gives the anova table

The error term in the CRD anova is a true “within” error. It comes from true replication of the experimental unit within the smallest cell of the experiment. All of the units in the cell are subjected to exactly (we hope!) the same treatment combination, so any differences in the response is model-independent true error. In the CRD all tests are done against this term.

There is often discussion about how to perform tests on main effects, T and G , in the presence of interaction, and how to interpret these effects. We will address this topic later in Section 2.4, when we look at expected mean squares.

Example 2.4. TWOWAY CRD CONTINUED The anova for the data of Example 2.3 is given in Table 2.5. The data can be found in dataset `RedClover`. All tests are done against the within MSE, and all terms are wildly significant, showing that there are effects due to Sulphur and Nitrogen, and they also interact. To understand what is going on we need some further analysis. We will return to this example in Section 2.5. ||

2.2 Model and Distribution Assumptions _____

The IVD experiment (Example 2.1) is a oneway CRD with model (1.1), that is

Table 2.5. Twoway CRD anova for the red clover data.

Source	df	Sum Sq	Mean Sq	F-statistic	p
Sulfur	3	3.06	1.02	285.53	< .00001
Nitrogen	1	7.83	7.83	2185.63	< .00001
Sulphur \times Nitrogen	3	3.76	41.25	349.78	< .00001
Within	16	0.06	0.0036		

$$(2.2) \quad Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \dots, t; \quad j = 1, \dots, r,$$

where Y_{ij} is the observed response, μ is an overall mean, τ_i is the treatment effect, and ε_{ij} is the error. For identifiability we can assume that $\bar{\tau} = 0$; otherwise we can consider the treatment effect to be $\tau_i - \bar{\tau}$ and the overall mean to be $\mu + \bar{\tau}$ (see the discussion in Section 1.1). We also assume

- (i) The random variables $\varepsilon_{ij} \sim N(0, \sigma^2)$ for $i = 1, \dots, t$, and $j = 1, \dots, r$ (normal errors with equal variances).
- (ii) $\text{Corr}(\varepsilon_{ij}, \varepsilon_{i'j'}) = 0$.

It is somewhat more common to assume that the ε_{ij} are *iid*, from which (ii) would follow. Under normality, this is equivalent to the assumption of zero correlation. We think it is most important to emphasize correlation structure, however, and thus present the assumptions in this way.

The mean and variance of Y_{ij} s are

$$(2.3) \quad E(Y_{ij}) = \mu + \tau_i, \quad \text{Var}(Y_{ij}) = \sigma^2.$$

Building on the oneway CRD, more complicated CRD experiments will tend to have a factorial structure for the treatment design. We will mainly emphasize twoway factorials in our development, for most of the theoretical development is straightforward for higher-order factorials (see Exercise 2.23 and Section 2.5.2).

A model for the twoway CRD is

$$(2.4) \quad Y_{ijk} = \mu + \tau_i + \gamma_j + (\tau\gamma)_{ij} + \varepsilon_{ijk}, \\ i = 1, \dots, t; \quad j = 1, \dots, g, \quad k = 1, \dots, r,$$

where Y_{ijk} is the observed response, μ is an overall mean, τ_i is one treatment effect, and γ_j is the other treatment effect. The term $(\tau\gamma)_{ij}$ represents the interaction of the two factors, a deviation from an additive response. Finally, ε_{ijk} is the error.

We further assume

- (i) The random variables $\varepsilon_{ijk} \sim N(0, \sigma^2)$ for $i = 1, \dots, t$, and $j = 1, \dots, g, k = 1, \dots, r$ (normal errors with equal variances).
- (ii) $\text{Corr}(\varepsilon_{ijk}, \varepsilon_{i'j'k'}) = 0$.

Note that without loss of generality we can assume

$$(2.5) \quad \bar{\tau} = \bar{\gamma} = (\bar{\tau\gamma}) = 0,$$

for this is equivalent to redefining the mean level μ as $\mu + \bar{\tau} + \bar{\gamma} + (\bar{\tau\gamma})$ and has no effect on interpretation of the parameters. However, for identifiability, it is necessary to go further, and it is typical to also assume that

$$(2.6) \quad (\bar{\tau\gamma})_{i.} = (\bar{\tau\gamma})_{.j} = 0 \text{ for all } i, j$$

Restrictions
for
identifiability

This is because the data from such an experiment will be in an $g \times t$ table, with the parameter estimates coming from the cell means. Losing 1 degree of freedom for the overall mean, that leaves $tg - 1$ degrees of freedom to estimate the parameters. Since we have $g + t + tg$ effect parameters in model (2.4), the restrictions (2.6) ensure identifiability.

To estimate the parameters in the CRD, we use least squares. We give some details for the twoway CRD, and leave other cases for exercises. Under model (2.4), least squares is quite straightforward in the CRD design. The least squares estimates satisfy

$$(2.7) \quad \min_{\mu, \tau_i, \gamma_j, (\tau\gamma)_{ij}} \sum_{i=1}^t \sum_{j=1}^g \sum_{k=1}^r (y_{ijk} - \mu - \tau_i - \gamma_j - (\tau\gamma)_{ij})^2.$$

Under (2.5), the solution (see Exercise 2.7) is given by

$$(2.8) \quad \begin{aligned} \bar{y} &= \hat{\mu} \\ \bar{y}_{ij.} - \bar{y} &= \hat{\tau}_i + \hat{\gamma}_j + (\hat{\tau\gamma})_{ij} \\ \bar{y}_{i..} - \bar{y} &= \hat{\tau}_i + (\hat{\tau\gamma})_{i.} \\ \bar{y}_{.j.} - \bar{y} &= \hat{\gamma}_j + (\hat{\tau\gamma})_{.j} \end{aligned}$$

where we see that, due to the identifiability constraint, the parameter estimate contains two pieces - for example - $\tau_i + (\tau\gamma)_{i.}$, where τ_i is often interpreted as a “main effect”, an effect that is constant across the levels of γ , while $(\tau\gamma)_{i.}$ is an average effect. That is, the effect may be different in the levels of γ , and we are just estimating the average.

It is important to understand that we *cannot* separate these effects, main and average, without an additional *assumption*, such as $(\bar{\tau\gamma})_{i.} = 0$.

Note: It is very important to understand the difference between the identifiability constraint (2.6) and the *assumption* that $(\bar{\tau\gamma})_{i.} = 0$.

The identifiability constraint gets us a set of parameters to estimate. It does *not* imply that $(\bar{\tau\gamma})_{i.} = 0$, but rather has the effect of redefining the parameters. That is, the effect of (2.6) is equivalent to redefining the parameters as

$$\begin{aligned}\tau'_i &= \tau_i + (\bar{\tau\gamma})_{i.}, \\ \gamma'_j &= \gamma_j + (\bar{\tau\gamma})_{.j}, \\ (\tau\gamma)'_{ij} &= (\tau\gamma)_{ij} - (\bar{\tau\gamma})_{i.} - (\bar{\tau\gamma})_{.j}.\end{aligned}$$

Thus, the average interaction effect does not go away, it just relocates.

In our calculations we will assume that $\bar{\tau} = \bar{\gamma} = (\bar{\tau\gamma}) = 0$ or, equivalently, the overall mean is $\mu + \bar{\tau} + \bar{\gamma} + (\bar{\tau\gamma})$. Also, to keep notation manageable, we will also assume that $(\bar{\tau\gamma})_{i.} = (\bar{\tau\gamma})_{.j} = 0$. Thus, a treatment effect will be written as τ_i , but this should always be interpreted as $\tau_i + (\bar{\tau\gamma})_{i.}$.

2.3 Expected Squares and F -tests

The calculation of expected values of mean squares (EMS) is an important part of any design, as it indicates the correct denominators for F -tests in the anova and the correct error terms for testing and estimating contrasts. Moreover, it helps us to see which replication helps control different sources of variation, and thus helps us in setting up a better design.

Although the direct calculation of EMS can be painful, it is important to carry it out carefully, and not rely on so-called “EMS algorithms”, which can sometimes give incorrect results unless used very carefully. Moreover, after doing a few of the calculations, the procedure becomes fairly transparent and the actual calculations can go quite smoothly.

We will do the calculations for the twoway CRD of model (2.4), leaving the other CRDs for exercises (see Exercises 2.11 and 2.23). Calculation of the EMS is the first step in justifying the F -tests in Table 2.4.

Starting with the first treatment sum of squares, we have

$$\begin{aligned}\text{E}(\text{SS(T)}) &= \text{E} \sum_i rg (\bar{Y}_{i..} - \bar{Y})^2 \\ &= \text{E} \sum_i rg ([\mu + \tau_i + (\bar{\tau\gamma})_{i.} + \varepsilon_{i..}] - [\mu + \varepsilon_{...}])^2 \\ &= \sum_i rg \tau_i^2 + \text{E} \sum_i rg (\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2,\end{aligned}$$

where the cross term is zero in the last line. From Lemma 2.16 it follows that

$$(2.9) \quad \text{E} \sum_i rg (\bar{\varepsilon}_{i..} - \bar{\varepsilon}_{...})^2 = rg \left(1 - \frac{1}{t}\right) \sum_i \text{Var}(\bar{\varepsilon}_{i..}) = (t-1)\sigma^2$$

and thus

$$E(SS(T)) = (t - 1)\sigma^2 + rg \sum_i \tau_i^2.$$

Continuing in this fashion, we can produce Table 2.6 (see Exercises 2.22 and 2.20).

Table 2.6. Expected Mean Squares for twoway CRD anova.

Source	df	EMS
Treatment T	$t - 1$	$\sigma^2 + \frac{rg}{t-1} \sum_i \tau_i^2$
Treatment G	$g - 1$	$\sigma^2 + \frac{rt}{g-1} \sum_j \gamma_j^2$
T \times G	$(t - 1)(g - 1)$	$\sigma^2 + \frac{r}{(t-1)(g-1)} \sum_{ij} (\tau\gamma)_{ij}^2$
Within	$tg(r - 1)$	σ^2

The null hypothesis of no effect of treatment T is

$$(2.10) \quad H_0 : \tau_i = 0 \text{ for all } i$$

and is tested by

$$F_{t-1, tg(r-1)} = \frac{MS(\text{Treatment T})}{MS(\text{Within})}.$$

The other tests are formed similarly. Note that all tests are against the within error – a very simple situation. Justification of the tests comes through Cochran’s Theorem, which we relegate to Technical Note 2.8.3.

There is sometimes discussion about the interpretation of “main effects” in the presence of interaction. That is, if there is interaction, which means that one treatment acts differently depending on the levels of the other treatment, then there is sometimes concern about the interpretation of the “treatment effect”.

However, upon close examination of the parameterization of the model, there is really no complication here. Without restricting the parameters, the interaction test is of the hypothesis

$$H_0 : (\tau\gamma)_{ij} - (\bar{\tau\gamma})_{i.} - (\bar{\tau\gamma})_{.j} = 0 \text{ for all } i, j,$$

Interpreting
treatment
effects

and whether this is true has no bearing on the sizes of $(\bar{\tau\gamma})_{i.}$ and $(\bar{\tau\gamma})_{.j}$. Furthermore, as we cannot separate τ_i and $(\bar{\tau\gamma})_{i.}$ without further assumptions (that are unverifiable), the treatment effect is “always” an average effect. The treatment hypotheses, such as (2.10), are concerned with the sizes of the average effects.

2.4 Estimating Contrasts

Under model (2.4) we have already derived the least squares estimates of the parameters. Since contrasts in these parameters are the typical focus of inference, we now develop those inferences.

Point Estimates

As least squares estimates are unbiased, it follows immediately that, for example,

$$(2.11) \quad E(\bar{Y}_{i..} - \bar{Y}) = \tau_i,$$

and a contrast $\sum_i a_i (\bar{Y}_{i..} - \bar{Y})$ is an unbiased estimator of $\sum_i a_i \tau_i$ with variance

$$(2.12) \quad \text{Var} \left(\sum_i a_i (\bar{Y}_{i..} - \bar{Y}) \right) = \text{Var} \left(\sum_i a_i (\bar{\varepsilon}_{i..} - \bar{\varepsilon}) \right) = \frac{\sigma^2}{rg} \sum_i a_i^2,$$

with analogous formulas for the other effects (see Exercise 2.8).

For the oneway CRD it follows immediately that

$$(2.13) \quad E(\bar{Y}_{i.} - \bar{Y}) = \hat{\tau}_i, \quad \text{Var} \left(\sum_i a_i \hat{\tau}_i \right) = \frac{\sigma^2}{r} \sum_i a_i^2,$$

see Exercise 2.6.

Note that in the estimation of contrasts, the term involving \bar{Y} cancelled. Thus, the variance calculation only involved independent terms. If we had estimated τ_i alone, with $\bar{Y}_{i.} - \bar{Y}$, we would have to deal with a covariance term (see Exercise 2.8).

Variance Estimates

The residuals from the least squares fit of the model (2.4) are

$$y_{ijk} - \hat{\mu} - \hat{\tau}_i - \hat{\gamma}_j - (\hat{\tau\gamma})_{ij} = y_{ijk} - \bar{y}_{ij.} = \varepsilon_{ijk} - \bar{\varepsilon}_{ij.},$$

which represents the within variance. The within sum of squares has expected value

$$E[\text{SS(Within)}] = E \left(\sum_{ijk} (\varepsilon_{ijk} - \bar{\varepsilon}_{ij.})^2 \right) = rtg \left(1 - \frac{1}{r} \right) \sigma^2,$$

making $\text{MS(Within)} = \frac{1}{tgr(r-1)} \sum_{ijk} (y_{ijk} - \bar{y}_{ij.})^2$ an unbiased estimate of σ^2 , and, for example, we can estimate a contrast variance

$$(2.14) \quad \hat{\text{Var}} \left(\sum_i a_i \hat{\tau}_i \right) = \frac{\text{MS(Within)}}{rg} \sum_i a_i^2.$$

Inference in the CRD

The main objects of interest in an anova are treatment contrasts, which are usually tested following an anova. Inference for contrasts can be based on multiple comparison procedures (Miscellanea 2.9.1), or on individual tests and intervals. In either case, the inferential distribution is built up from the model (2.4)

Under model (2.4)

$$Y_{ijk} \sim N(\mu + \tau_i + \gamma_j + (\tau\gamma)_{ij}, \sigma^2), \quad \text{Cov}(Y_{ijk}, Y_{i'j'k'}) = 0,$$

and using results from Sections 1.8 and 2.4 we can build up the distribution of any contrast. For example,

$$(2.15) \quad \sum_i a_i Y_{i..} \sim N\left(\sum_i a_i \tau_i, \frac{\sigma^2}{rg} \sum_i a_i^2\right)$$

and hence

$$\frac{\sum_i a_i Y_{i..} - \sum_i a_i \tau_i}{\sqrt{\frac{\sigma^2}{rg} \sum_i a_i^2}} \sim N(0, 1).$$

We now can apply Theorem 2.23 to replace σ^2 with $\hat{\sigma}^2 = \text{MS}(\text{Within})/(tg(r-1))$ to get

$$\frac{\sum_i a_i Y_{i..} - \sum_i a_i \tau_i}{\sqrt{\frac{\hat{\sigma}^2}{rg} \sum_i a_i^2}} \sim t_{tg(r-1)}.$$

2.5 Deeper into Factorials

In this section we look a little closer at the interpretations of interaction in factorial experiments. There is often the urge to get as much as possible out of an experiment and, in doing so, to include many different treatments. However, interpretations of treatment effects, in the face of many interactions, can be difficult. We look at a number of examples of types of interactions that can be expected, and try to better understand them through contrasts.

2.5.1 Investigating Interactions

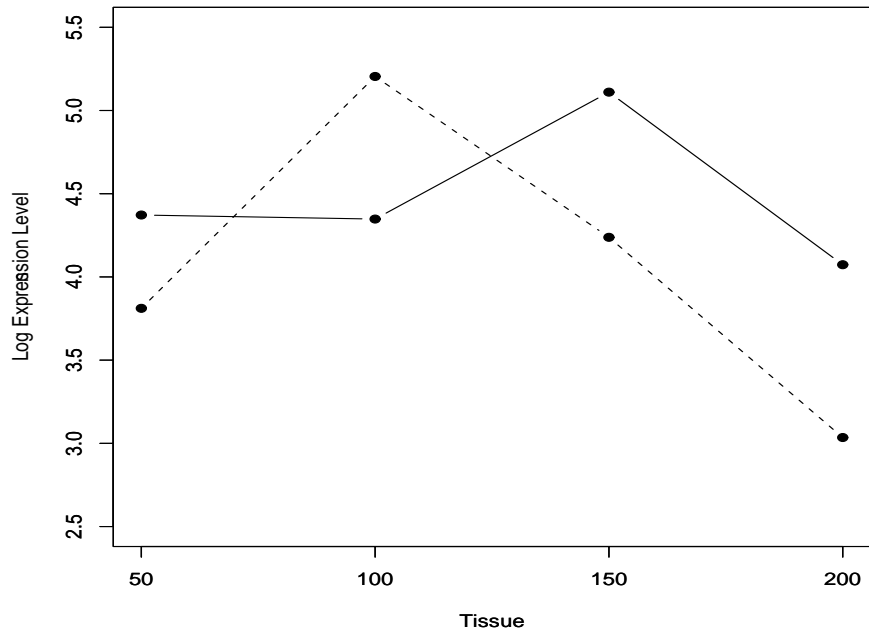
The existence of an interaction means that the effect of one treatment is dependent on the levels of another. This makes interpretation more difficult, and also can cause problems if the experimenter is looking to control future responses by setting treatment levels.

One overall distinction in interactions is between *qualitative* and *quantitative* interactions. In the first case, although the treatment response differs according to the levels of another factor, the response is only changed in

quantity. In the second case there is a change in quality, that is, the “better” treatment changes according to the level of another treatment.¹

Example 2.5. FISH MICROARRAYS REVISITED As an example, we revisit Example 1.16, where the experimenter measured gene expression level in fish tissue as a function of two treatments. Figure 2.1 shows a plot of the cell means, often called an “interaction plot”. There is can be seen that there is a qualitative interaction - the lines cross - and the effect of hCG is different at the different levels of tissue. ||

Fig. 2.1. Interaction plot for log expression-level data from the fish tissue experiment. The solid line corresponds to the absence of hCG, and the dashed line to the presence of hCG.



Example 2.6. RED CLOVER REVISITED We next revisit another experiment seen in Examples 2.3 and 2.4. The interaction plot for the two treatments is shown in Figure 2.2, where there is a strong quantitative interaction.

¹ Procedures that test for qualitative interactions have been developed, especially in the biostatistics literature. See, for example, Gail and Simon (1985) or Piantidosi and Gail (1993).

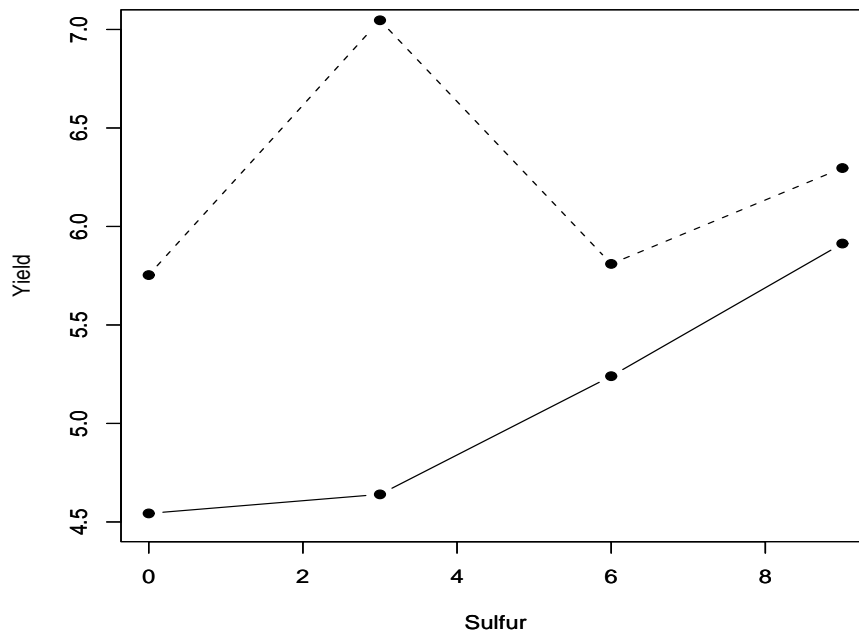
Although the lines do not cross, there seems to be a strong elevation of yield at one combination of Sulfur and Nitrogen. The anova, summarized in Table 2.7 shows that all effects are significant.

The fact that all the polynomial trends are significant does not really tell us much - in terms of using this information we see that Nitrogen = 20 is preferred. Because everything is so significant we suspect that the peak at Sulfur = 3 is real, which seems to point to an optimal combination. ||

Table 2.7. Breakdown of interaction effect for the red clover data

Source	df	Sum Sq	Mean Sq	F -statistic	p
Sulfur \times Nitrogen	3	3.76	1.25	349.78	< .00001
Linear Sulfur \times Nitrogen	1	1.40	1.40	388.2	< .00001
Quadratic Sulfur \times Nitrogen	1	0.72	0.72	199.33	< .00001
Cubic Sulfur \times Nitrogen	1	1.64	1.64	456.95	< .00001
Within	16	0.057	0.0036		

Fig. 2.2. Interaction plot for yield from the red clover experiment. The solid line corresponds to Nitrogen=0, and the dashed line to Nitrogen=20.



Before leaving the subject of interpretation of interactions, we look at one more example that arose in a QTL experiment², where the experimenter was investigating the interaction of two QTL, known in the genetics literature as *epistasis*.

Example 2.7. QTL EPISTASIS Data were collected in an experiment to examine the main effects and interactions of two QTL. The data look like

		QTL1		
		AA	Aa	aa
QTL2	BB	x x	x x	x x
	Bb	x x	x x	x x
	bb	x x	x x	x x

In QTL experiments there are a number of interaction patterns that are expected, and one researcher at the University of Florida wondered if, simply due to the interaction, could it happen that the mere presence of QTL2 in the experiment could increase the power to detect an effect due to QTL1. That is, for typical interaction patterns such as those shown in Figure 2.3, does increasing the difference in the means of QTL1 result in an increase in the means of QTL2?

To answer this question we examine the EMS of the effects, which are given in Table 2.6 for the twoway analysis. The EMS are important because they are the noncentrality parameters corresponding to the treatment hypothesis tests (Technical Note 2.8.4).

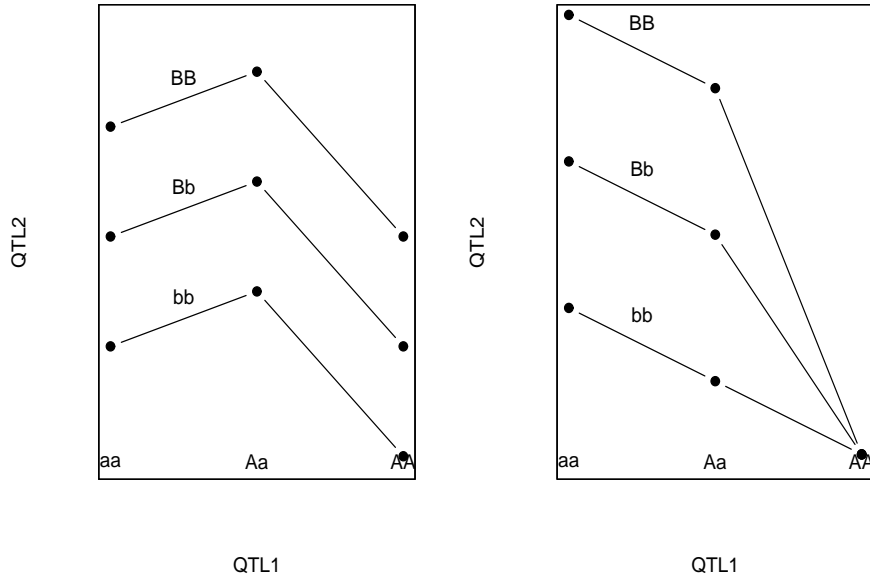
If $H_0 : \tau_i = 0$ for all i is true, then the usual ratio of mean squares has a central F distribution and, under the alternative, the distribution is noncentral F with noncentrality parameter $\sum_i \tau_i^2$. The F -statistic is stochastically increasing in its noncentrality parameter, which implies that the ratio of mean squares will tend to be large if this parameter is larger. Thus, the larger this value is, the more power we have to reject H_0 and detect QTL differences. Now we look more closely at the pattern of means in Figure 2.3, and show that by increasing the marginal mean difference in QTL1, we automatically increase the marginal mean difference in QTL2.

First, we show that when there is no interaction the values of one factor cannot influence the other, but when there is any interaction whatsoever, the levels of one factor can cause higher values in the other.

If there is no interaction, then $(\tau\gamma)_{ij} = 0$, so the margins control the cells, and when computing any marginal mean the effect of the other parameter disappears because of the restrictions (2.5). So in this case the effects are independent of one another, and no matter what happens in the means of one parameter, it can have no effect on the marginal means of the other parameter. This case is shown in the left panel of Figure 2.3.

² *Quantitative Trait Loci (QTL)* are regions on the genome that can be linked to quantitative traits. For example, one may find a certain region on the corn genome to be linked with yield.

Fig. 2.3. The left panel shows no interaction between two QTL, while the right panel shows a typical interaction pattern between two QTL.



We now illustrate a pattern of cell means that look like the right panel. The cell entries are the cell mean parameters and the marginal entries are the marginal mean parameters. We use two free parameters, a and b , to arrive at the pattern in Table 2.8.

Table 2.8. Cell means that correspond to the pattern displayed in the right panel of Figure 2.3.

		QTL1			Average
		$a + \frac{4}{3}b$	$\frac{4}{3}b$	$-a - \frac{2}{3}b$	$\frac{2}{3}b$
QTL2	BB	$a + \frac{4}{3}b$	$\frac{4}{3}b$	$-a - \frac{2}{3}b$	$\frac{2}{3}b$
	Bb	$a + \frac{1}{3}b$	$\frac{1}{3}b$	$-a - \frac{2}{3}b$	0
	bb	$a - \frac{2}{3}b$	$-\frac{2}{3}b$	$-a - \frac{2}{3}b$	$-\frac{2}{3}b$
Average		$a + \frac{1}{3}b$	$\frac{1}{3}b$	$-a - \frac{2}{3}b$	0

Since Table 2.8 adds to zero for any choice of a and b , any values represent a legitimate set of QTL parameters.

This is really
an interesting
occurrence

Now the main point is that if the main effect of QTL2 increases, that is, if b increases, because of the interaction this has the effect of increasing the main effect of QTL1. That is, as b increases, the marginal means of QTL1 separate more, giving us more power to detect a difference in QTL1, solely due to the interaction with QTL2.

||

Thus, the presence of interaction can make one marginal factor seem significant simply due to the presence of the other factor!

2.5.2 Higher-Order Factorials

Optimizing resources in experimentation would suggest that putting more treatments in one experiment is a good thing – we can argue that:

- (1) With many treatments in one experiment, we get simultaneous information. This allows better comparisons.
- (2) With more treatments to explain variation, the error will be reduced.
- (3) Running one experiment saves resources and time.

These are all excellent arguments, but in practice big factorials, except in certain situations, are not a good thing to do. What we saw in the previous section is that in the presence of interaction, effects can get muddled together and precise inferences become problematic. And this problem worsens when we move to higher-order factorial; threeway, fourway, ten-way! (Just take a look at the EMS from the threeway in Exercise 2.23, and the formal conclusions from the hypothesis tests.)

Having said this, we need to understand how to deal with higher-order factorials for at least two reasons: (i) Experimenters do them, and (ii) They are good for exploration.

Although many-way designs can be run with treatments having many levels, in such designs we run into the problem of interpretation of factors with many degrees of freedom. A significant threeway interaction with 18 degrees of freedom is difficult to make sense of - there are 18 contrasts to think about!

Higher order exploratory factorials are best run with treatments having two levels – the levels are typically chosen to span the range of possible treatment levels, representing a “high” and “low” setting. This gives the greatest chance of finding an effect. (See Section 6.3.) Also, since the treatment has only two levels and 1 degree of freedom, it is represented by one contrast, making interpretation easy. Moreover, all interactions between two-level treatments have 1 degree of freedom, making their interpretation easy

Another circumstance that sometimes arises in higher-order factorials is the lack of within error. Although sometimes the experimental conditions are replicated to get a within error, sometimes they are not. In such cases it

is typical to use the higher-order interactions (threeway and higher) as error terms. The thinking is that these higher-order factors do not carry much main effect information and are mostly error anyway. Whether this is true, to use these are errors means that we assume that their contribution to the EMS is only through σ^2 , and there is no contribution from the interaction. This leads to conservative tests, meaning that it is harder to reject H_0 , so if we do reject we have some confidence that a true effect has been found.

Example 2.8. **2 × 2 FACTORIAL** A 2 × 2 factorial can be represented by the four cell means

	Factor <i>A</i>	
Factor <i>B</i>	θ_{11}	θ_{12}
	θ_{21}	θ_{22}

and each effect can be represented by a contrast

<i>A</i> effect	<i>B</i> effect	<i>A</i> × <i>B</i>
1 -1	1 1	1 -1
1 -1	-1 -1	-1 1

where the contrasts merely account for the mean differences. Note, in particular, that the interaction contrast is the “difference of the differences”. ||

Example 2.9. **INTERACTION CONTRASTS** If factor *A* has four levels and factor *B* has three levels, then there are 6 interaction contrasts for *A* × *B*, each with 12 coefficients. These can be obtained as all products of three orthogonal contrasts in *A* and two orthogonal contrasts in *B*:

	<i>A</i>					
<i>B</i>					1	0
					-1/2	1
					-1/2	-1
	1	-1/3	-1/3	-1/3		
	0	1	-1/2	-1/2		
	0	0	1	-1		

The six interaction contrasts are now obtained as the products of the coefficients, and the interpretation is a bit more straightforward. For example, the first *B* contrast compares the first level of *B* with the average of the other two, and the first *A* contrast compares the first level of *A* with the average of the other three. The product of these contrasts is an interaction contrast.

	<i>A</i>			
<i>B</i>	1	-1/3	-1/3	-1/3
	-1/2	1/6	1/6	1/6
	-1/2	1/6	1/6	1/6

This interaction contrast compares the B contrast applied to level 1 of A with the average B contrast in the other levels of A . In other words, it is as if we collapsed the above table into a 2×2 table

		A	
		Level 1	Average of Levels 2-4
B	Level 1		
	Average of Levels 2-3		

and did an interaction contrast here.

So we see that it is possible to interpret interactions using contrasts, but the task is not simple and the interpretations may not be meaningful. And we have only done twoway interactions – adding a third factor complicates things even more (see Exercise 2.15). ||

Example 2.10. A SOCIAL SCIENCE EXPERIMENT To examine the effect of different factors on a citizenship test performance, a factorial experiment was set up with the following factors:

- (1) Education: Three levels (less than high school, high school, greater than high school),
- (2) Home Environment: Four levels,
- (3) Country of Origin: Seven countries.

The response variable was $Y =$ score on a citizenship aptitude test.

The twoway interaction of Education \times Home Environment would be of interest, but the other twoway interactions are of less interest. Moreover, the threeway interaction, with 36 degrees of freedom, would be very difficult to interpret. ||

Finally, we look at the classical higher-order factorial design (see Section 6.3 for another treatment of these designs), the 2^n design, where we have n factors, each at two levels. Things are actually quite simple here because everything has two levels – all effects are contrasts and all contrasts are easily obtainable as products.

It is common to specify the two levels of each factor, low and high, in the following way. The high level of A is denoted by a , and the low level is denoted by the absence of a . Thus the treatment combination ab has A and B at their high level and C at the low level, while ac has A and C high and B low. The treatment combination with all factors at their low level is denoted by (1).

Example 2.11. **A 2^4 FACTORIAL** A 2^4 factorial design would have $2^4 = 16$ treatment combinations, and the full anova would specify 15 factors: 4 main effects, $\binom{4}{2} = 6$ twoway interactions, $\binom{4}{3} = 4$ threeway interactions, and 1 fourway interaction. Each effect has one degree of freedom, and all contrasts are obtainable as products. This makes for easy interpretation. A is the difference, AB is the difference of the differences, ABC is the difference of the AB differences, etc.

The following table shows some of the effects in the 2^4 design. The low level is -1 in the contrast and the high level is 1. The interaction contrasts are all products; for example the AD contrast is obtained by multiplying the A contrast by the D contrast, and the ABD contrast comes from $AB \times D$ (or $AD \times B$ or $A \times B \times D$).

Treatment	A	B	D	AD	ABD
(1)	-1	-1	-1	1	-1
d	-1	-1	1	-1	1
c	-1	-1	1	-1	1
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
abc	1	1	-1	-1	-1
abcd	1	1	1	1	1

Unless there was express interest in the threeway and higher interactions, it would be typical to pool these terms for an error estimate (Exercise 2.16) ||

2.6 Adjusting for Covariates

The analysis of covariance (ancova) in some sense has no business being in a design book, as it is more of an analysis tool than a design tool. However, its function is to reduce variance of the treatment means, so from that view it is a design strategy.

Here we look at a few examples of ancova, and do some variance calculations to understand how ancova works, and how it may benefit (on not!) a design.

Example 2.12. **SOME ANCOVA EXAMPLES**

- (1) In agriculture, an experiment is done to examine the effect of treatments on yield. However, it is known that the density/plot of the plants could affect yield, and this variable is used as a covariate.
- (2) In a nutrition experiment where the growth of laboratory rats is tracked, the initial weight of the rats influences their growth rate, and could be used as a covariate.
- (3) In a microarray experiment with a spotted array, the florescence of a spot is related to the size of the spot, which could vary from gene to gene. Thus, spot size is a possible covariate.

||

In each of the above examples the candidate covariate satisfies two conditions:

- (1) The covariate is related to the response, and can account for variation in the response.
- (2) The covariate is *not* related to the treatment.

Note: We will see in (2.23) that the second condition is extremely important, and if it is not satisfied it can erase all the benefits of using a covariate.

A covariate functions somewhat like a block, in that it removes variation. However, it may not be a planned part of the experiment (it may be noticed after the design is set up) and it is typically continuous, so it really cannot be blocked over. If the covariate is recognized during the design phase, we could stratify over it, effectively using it to balance the observations by having a range of the covariate in each treatment.

Example 2.13. **CORN YIELD COVARIANCE** When trying to compare yields of different varieties of corn, the response is confounded due to the fact that the number of plants in a plot may vary. Since the number of plants clearly influences yield, and should not be related to any treatment, it is an ideal covariate. The following data are the yield (Y) (dry weight) of four varieties of corn, along with a covariate (X) measuring the number of plants per plot.

		Varieties							
		Cornell		Robson		Ohio		Ohio	
		M-4		360		K-24		M-15	
Obs.	X	Y	X	Y	X	Y	X	Y	
1	20	12.8	20	12.2	20	14.1	13	8.6	
2	17	11.0	20	10.0	20	13.1	18	10.2	
3	20	10.9	16	9.8	20	12.8	17	8.7	
4	15	9.1	20	9.8	20	11.8	14	7.3	
5	20	9.6	19	9.8	20	10.8	15	9.3	
6	15	9.3	20	12.1	13	7.8	11	8.2	

Examining the data suggests that, regardless of the variety, the yield is increased if the number of plants per plot is increased (Figure 2.4). ||

Here we will concentrate of the design effect of the covariate, in particular looking at the effect of the covariate on the variance of a treatment contrast, and attempting to document when the use of a covariate will be beneficial.

We will only look at models with one covariate, and stay with CRDs. The design effect in more complicated situations should be clear, and for details of more complex ancova models there are data analysis books that treat this. (A more advanced treatment can be found in Mead 1988; see also Federer and Meredith 1992.)

Recall the oneway anova model (2.2)

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad i = 1, \dots, t, \quad j = 1, \dots, r.$$

If there is a covariate, x_{ij} , we can use an ancova model

$$(2.16) \quad Y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij}, \quad i = 1, \dots, t, \quad j = 1, \dots, r,$$

where $\varepsilon_{ij} \sim N(0, \sigma^2)$, independent, and we add to the anova model a “regression” piece that adjusts for the covariate. Although it is not immediately clear how to estimate the treatment means in this model, we can appeal to least squares to obtain

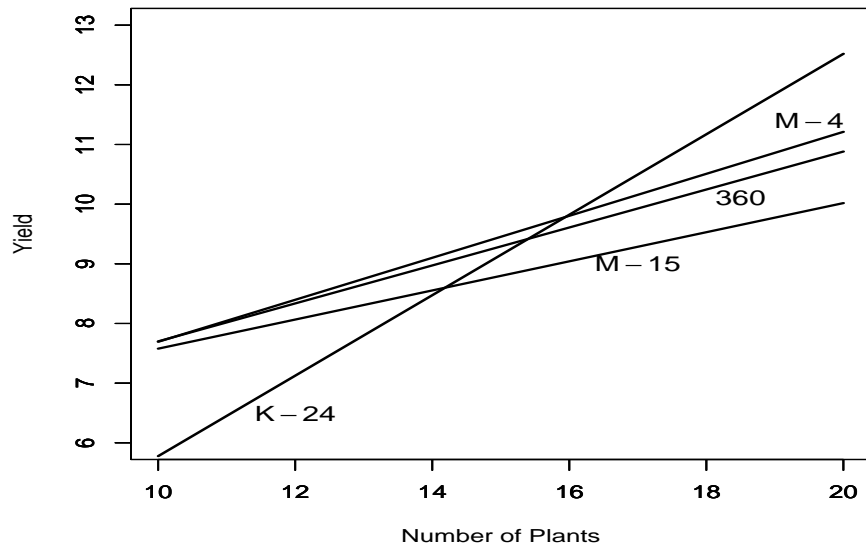
$$(2.17) \quad \widehat{\mu + \tau_i} = \bar{y}_i - \hat{\beta}_1(\bar{x}_i - \bar{x}),$$

where

$$\hat{\beta}_1 = \frac{\sum_{ij} (x_{ij} - \bar{x}_{i\cdot})(y_{ij} - \bar{y}_{i\cdot})}{\sum_{ij} (x_{ij} - \bar{x}_{i\cdot})^2}.$$

If we restrict $\sum \tau_i = 0$, then $\hat{\mu} = \bar{y}$ and we can estimate the effects τ_i . However, this breaks down if the cell sizes are unequal (Exercise 2.29).

Fig. 2.4. Least squares line fit to the four varieties of corn in Example 2.13. Regardless of the treatment, there is a positive relationship between yield and number of plants per plot.



The crucial covariance assumption

Note that the means are adjusted to by covariate mean with the assumption that, in each group, the slope is the same. This assumption is both crucial and bothersome. Without it, the mean comparisons could be very sensitive to where they are adjusted – look at Figure 2.4. There it seems that the equal slope assumption may be violated,

but if we fit different slopes then, depending on where we adjust the means, they may be far apart or close together. It seems best to try to live with the equal slope assumption if covariance analysis is to be performed, unless there is a good reason to adjust each mean by a specific covariate value.

Since the least squares estimate is unbiased, we have $E(\widehat{\mu + \tau_i}) = \mu + \tau_i$ with variance

$$\begin{aligned}
 \text{Var}(\widehat{\mu + \tau_i}) &= \text{Var}(\bar{Y}_i - \hat{\beta}_1(\bar{x}_i - \bar{x})) \\
 (2.18) \qquad &= \text{Var}(\bar{Y}_i) + (\bar{x}_i - \bar{x})^2 \text{Var}(\hat{\beta}_1) \\
 &= \sigma^2 \left(\frac{1}{r} + \frac{(\bar{x}_i - \bar{x})^2}{\sum_{ij}(x_{ij} - \bar{x}_i)^2} \right),
 \end{aligned}$$

where we use the fact that \bar{Y}_i and $\hat{\beta}_1$ are independent.

There are two ways we can *interpret* (2.16), neither of which is formally correct as an ancova model, but each of which lends some insight:

- (1) If we write (2.16) as

$$Y_{ij} - \mu - \tau_i = \beta(x_{ij} - \bar{x}) + \varepsilon_{ij},$$

then the left side of the equation are the anova residuals, so we can think of ancova as doing a regression on the anova residuals. Note that in this interpretation we see that the variation due to regression will come out of the anova residual, and thus should decrease experimental error.

- (2) If we write (2.16) as

$$Y_{ij} - \beta(x_{ij} - \bar{x}) = \mu + \tau_i + \varepsilon_{ij},$$

then we can think of ancova as an anova on the regression residuals. In this case the observations have been adjusted, and their variance has been decreased, so the anova error should be decreased.

Although both of these interpretations are useful, the ancova actually does things simultaneously, and fits the model

$$(2.19) \qquad Y_{ij} - \bar{Y}_i = \beta(x_{ij} - \bar{x}_i) + \varepsilon_{ij},$$

where we adjust the means in each treatment. Notice that the total sum of squares for this regression is, in fact, the within sum of squares from the oneway anova, that is,

$$\text{SS(Regression Total)} = \sum_{ij} (y_{ij} - \bar{y}_{i.})^2 = \text{SS(Within from anova)},$$

and thus the ancova adjustment does come out of the anova “residual”.

Partitioning the Sums of Squares

Model (2.19) actually leads to two different partitioning of the sums of squares. Since the regression comes out of the anova within, the ancova table can be written

Source	df	SS
Treatments	$t - 1$	$r \sum_i (\bar{y}_i - \bar{y})^2$
Within Treatments	$t(r - 1)$	$\sum_{ij} (y_{ij} - \bar{y}_{i.})^2$
Regression (Covariate after Trt)	1	$\hat{\beta}_1 \sum_{ij} (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})$
Residual	$t(r - 1) - 1$	$\sum_{ij} [(y_{ij} - \bar{y}_{i.}) - \hat{\beta}_1(x_{ij} - \bar{x}_{i.})]^2$

where $\hat{\beta}_1$ is given at (2.17). Here we have partitioned the sum of squares by first fitting the anova, and then pulling out the effect of the covariate from the regression, as in interpretation (1) above. This table allows us to do the usual anova test, and see if the regression is significant by forming the F -ratio $\text{MS(Regression)}/\text{MS(Residual)}$.

Although this test may be interesting, it does not get at the heart of the ancova rationale, in that we want to see if the *adjusted yields* are significantly different. To do this, we first partition the sums of squares by doing the regression, and then removing the treatment variability from the residuals, as in interpretation (2) above. The resulting ancova table is

Source	df	SS
Regression	1	$\hat{\beta}_0 \sum_{ij} (x_{ij} - \bar{x})(y_{ij} - \bar{y})$
Residual from Regression	$tr - 2$	$\sum_{ij} [(y_{ij} - \bar{y}) - \hat{\beta}_0(x_{ij} - \bar{x})]^2$
Treatment (after Covariate)	1	$\text{SS(Residual from Regression)} - \text{SS(Residual)}$
Residual	$t(r - 1) - 1$	$\sum_{ij} [(y_{ij} - \hat{\beta}_1 x_{ij}) - (\bar{y}_{i.} - \hat{\beta}_1 \bar{x}_{i.})]^2$

where

$$\hat{\beta}_0 = \frac{\sum_{ij} (x_{ij} - \bar{x})(y_{ij} - \bar{y})}{\sum_{ij} (x_{ij} - \bar{x})^2},$$

and it is easiest to get the adjusted treatment sum of squares by subtraction. The $\text{SS(Residual from Regression)}$ becomes the total sum of squares for the adjusted treatment anova, and SS(Residual) is actually a within sum of squares of the adjusted data $y_{ij} - \hat{\beta}_1 x_{ij}$; we have written it this way in the second table. However, realize that the bottom-line residual is the same in both tables.

Example 2.14. CORN YIELD COVARIANCE CONTINUED The two anova tables for the corn data are

Covariate After Treatment				Treatment After Covariate			
Source	df	SS	MS	Source	df	SS	MS
Varieties	3	27.955	9.318	Plants	1	43.916	43.916
Within	20	46.765	2.338	Residual (from Reg.)	22	30.804	1.400
Plants (after Var.)	1	21.729	21.729	Varieties (after Plants)	3	5.768	1.923
Residual	19	25.036	1.318	Residual	19	25.036	1.318

where the test statistic for varieties, after adjusting for the covariate, is

$$F_{3,19} = \frac{1.923}{1.318} = 1.459,$$

which is not significant. ||

Testing Treatments

The ancova test on treatments, done in the above example, is a test of the adjusted treatments. Formally, we test the hypotheses

$$H_0 : Y_{ij} = \mu + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij} \text{ vs. } H_1 : Y_{ij} = \mu + \tau_i + \beta(x_{ij} - \bar{x}) + \varepsilon_{ij},$$

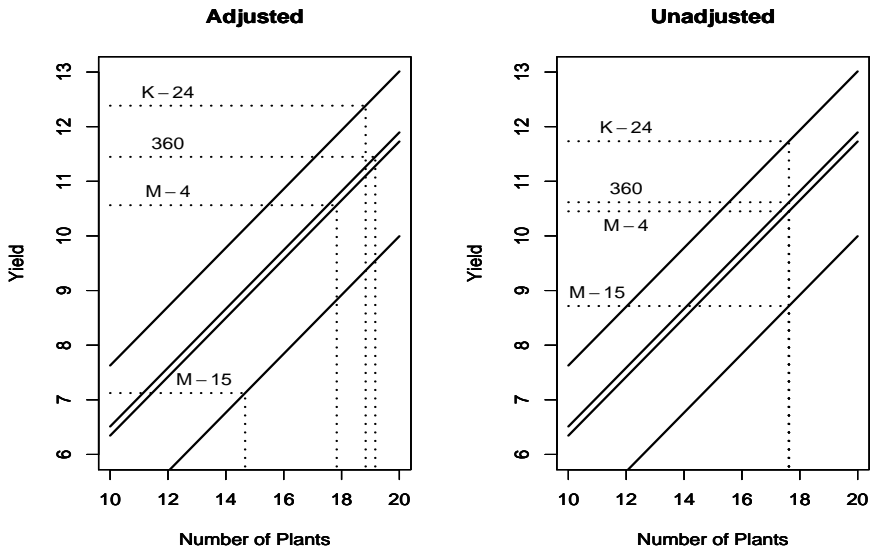
where the null hypothesis specifies only a regression relationship, and the alternative specifies a treatment effect in addition to the regression. To test these hypotheses note that the H_0 model is a special case of the H_1 model (we also say that the H_0 model is nested in the H_1 model), and thus the H_1 model will always provide a better fit (since it has more parameters). To measure if this fit is better, we can use the F -ratio

$$(2.20) \quad F = \frac{[\text{SS}(\text{Residual from } H_0) - \text{SS}(\text{Residual from } H_1)] / (\text{df}_{H_0} - \text{df}_{H_1})}{\text{SS}(\text{Residual from } H_1) / \text{df}_{H_1}},$$

where df_{H_0} and df_{H_1} are the residual degrees of freedom under the respective models. This is the test done in Example 2.14. Specifically, from the above ancova tables we have

$$F_{t-1, t(r-1)-1} = \frac{\left(\sum_{ij} [(y_{ij} - \bar{y}) - \hat{\beta}_0(x_{ij} - \bar{x})]^2 - \sum_{ij} [(y_{ij} - \bar{y}_{i.}) - \hat{\beta}_1(x_{ij} - \bar{x}_{i.})]^2 \right) / (t - 1)}{\left(\sum_{ij} [(y_{ij} - \bar{y}_{i.}) - \hat{\beta}_1(x_{ij} - \bar{x}_{i.})]^2 \right) / (t(r - 1) - 1)}.$$

Fig. 2.5. Adjusted and unadjusted treatment means for the data of Example 2.13. The fitted lines are from model (2.16). The ancova adjusts the means to their covariate mean (left panel), and the anova (unadjusted) is equivalent to adjusting the means to the overall covariate mean (right panel).



Note: The strategy to get the F -statistic in (2.20) is a general one for linear models.

If the H_0 model is nested in the H_1 model, then (2.20) provides a valid test of whether the richer model H_1 is a significant improvement.

Estimating Contrasts

Building on (2.18), the variance of a contrast is given by

$$\begin{aligned}
 \text{Var} \left(\sum_i a_i (\widehat{\mu} + \tau_i) \right) &= \text{Var} \left(\sum_i a_i \bar{Y}_i \right) + \text{Var}(\hat{\beta}_1) \left(\sum_i a_i (\bar{x}_i - \bar{x}) \right)^2 \\
 (2.21) \qquad \qquad \qquad &= \frac{\sigma^2}{r} \sum_i a_i^2 + \frac{\sigma^2}{\sum_{ij} (x_{ij} - \bar{x}_i)^2} \left(\sum_i a_i \bar{x}_i \right)^2,
 \end{aligned}$$

where we use the independence of \bar{Y}_i and $\hat{\beta}_1$ and the fact that $\sum_i a_i = 0$. We then estimate σ^2 with the MS(Residual) from the ancova.

Example 2.15. **CONTRASTS FOR CORN YIELD EXPERIMENT** Returning to Example 2.13, we calculate the estimated treatment means and their standard errors. Figure 2.5 illustrates the covariance adjustment of the means. There it can be seen that the adjustment increases the mean estimate as the number of plants increases, and results in a spreading out of the treatment means. The ancova also reduces the treatment variances, as the following table shows.

	Varieties			
	360	K-24	M-15	M-4
Unadjusted Mean	10.617	11.733	8.717	10.450
Std. Error	0.624	0.624	0.624	0.624
Adjusted Mean	11.447	12.384	7.124	10.562
Std. Error	0.496	0.486	0.563	0.469

Note, in particular, the adjustment for M-15, which had fewer plants per plot. The ancova tells us that the yield of M-15 is, in fact, lower than the others even after adjusting for the number of plants.

If we test the contrast $H_0 : \tau_{M-15} = \frac{1}{3}(\tau_{360} + \tau_{K-24} + \tau_{M-4})$, we find

$$\begin{aligned} \text{Unadjusted: } t_{20} &= -4.100, \\ \text{Adjusted: } t_{19} &= -6.659, \end{aligned}$$

both wildly significant. Thus, even though the ancova gave greater separation, this differences was large enough so that the original anova would find it. See Exercise 2.13 for more. ||

Variance of a Treatment Difference

We close this section with a deeper look at the variance of a treatment difference, which may add some insight as to when the ancova can be expected to improve things. From (2.21), the estimated variance of the difference of two adjusted means is

$$\text{Var}((\widehat{\mu + \tau_i}) - (\widehat{\mu + \tau_{i'}})) = \frac{2\hat{\sigma}^2}{r} + \frac{\hat{\sigma}^2}{\sum_{ij}(x_{ij} - \bar{x}_i)^2} (\bar{x}_i - \bar{x}_{i'})^2,$$

where $\hat{\sigma}^2$ is the MS(Residual) from the ancova and we recognize the sum in the denominator as the within sum of squares from an anova of X on Treatments; we denote it by $SS(\text{Within}_x)$. An idea that goes back to Finney (1946) is to replace this variance with a common one that can be used for all of the differences. To do this, we can replace the term $(\bar{x}_i - \bar{x}_{i'})^2$ with its average over all pairs.

First, it can be verified that for any numbers b_1, b_2, \dots, b_n we have

$$(2.22) \quad \sum_{i=1}^n \sum_{i'=1}^n (b_i - b_{i'})^2 = 2n \sum_{i=1}^n (b_i - \bar{b})^2.$$

There are $t(t-1)$ nonzero terms in the $\bar{x}_i - \bar{x}_{i'}$ pairs, and applying (2.22) we obtain

$$\begin{aligned} \text{Var}((\mu + \widehat{\tau}_i) - (\mu + \widehat{\tau}_{i'})) &\approx \frac{2\hat{\sigma}^2}{r} + \frac{\hat{\sigma}^2}{\sum_{ij}(x_{ij} - \bar{x}_{i\cdot})^2} \left[\frac{2}{t-1} \sum_i (\bar{x}_i - \bar{x})^2 \right] \\ (2.23) \qquad \qquad \qquad &= \frac{2\hat{\sigma}^2}{r} \left(1 + \frac{1}{t-1} \frac{\text{SS}(\text{Trt}_x)}{\text{SS}(\text{Within}_x)} \right), \end{aligned}$$

where $\text{SS}(\text{Trt}_x)$ is the treatment sum of squares from the anova of X on Treatments. From (2.23) we see two things very clearly

- (1) The variance of a treatment difference is reduced as the regression of Y on X in (2.19) improves, as $\hat{\sigma}^2$ will decrease.
- (2) The variance of a treatment difference is *increased* if the covariate X is related to the treatment, that is, if $\text{SS}(\text{Trt}_x)/\text{SS}(\text{Within}_x)$ increases. This is a clear message that it is never advantageous to use a covariate that is related to the treatment.

2.7 Exercises

Essential

2.1 Referring to Example 2.1

- (a) For this experiment to be a CRD, explain how the data would need to be collected.
- (b) Complete the anova table:

Source	df	SS	MS	F
Temp	-	5.86	-	-
Within	-	-	-	-
Total	-	7.67	-	-

- (c) The experimenter is trying to decide which of two different temperature contrasts to test (i) $(-3, 1, 1, 1)$ or (ii) $(-3, -1, 1, 3)$. Give the experimenter an interpretation of each of these contrasts.
- (d) For contrast (i) of part (c), find two others that are orthogonal.
- (e) Calculate the t -statistic for testing $H_0: -3\tau_1 - \tau_2 - \tau_3 + 3\tau_4 = 0$, and find the proportion of variation in Temperature that is *not* explained by the contrast $(-3, -1, 1, 3)$.

2.2 Referring to Section 1.2 of Chapter 1:

- (a) Write $\bar{Y}_i - \bar{Y}$ as a contrast in Y_{ij} .
- (b) Use the contrast formulas to find the mean and variance of $\bar{Y}_i - \bar{Y}$. Check that they agree with those in Section 1.2
- (c) Write $Y_{ij} - \bar{Y}_i$ as a contrast in Y_{ij} and then verify (1.6) and (1.6).

2.3 Referring to the IVD data of Example 2.1:

- (a) Describe, in words, the conclusions that you would draw from the results of Table 2.1.
- (b) Produce the complete anova table for the IVD data.

- (i) Break down the treatment sum of squares into the contrast sum of squares corresponding to Table 2.1.
- (ii) Reproduce the t -statistics for the contrasts.
- (iii) Verify that the average squared t -statistic is the anova F -ratio.
- (c) Use another set of orthogonal contrasts, those that correspond to the linear, quadratic and cubic trend, to break down the treatment sum of squares. Test the significance of each contrast and describe the conclusions that can be made.

2.4 Here we look more closely at parameter restrictions.

- (a) Referring to model (2.4), show that the restriction $\bar{\tau} = \bar{\gamma} = (\bar{\tau\gamma})_{i..} = (\bar{\tau\gamma})_{.j.} = 0$ results in exactly $tg - 1$ parameters to be estimated.
- (b) Extend the results of (a) to a threeway CRD with all interactions.
- (c) For a general factorial with T different treatments, each at t_i levels, $i = 1, \dots, T$, show that there are $\prod_i t_i - 1$ degrees of freedom to estimate effect parameters. Describe a restriction on the parameters that will result in identifiability, that is, that will reduce the parameters space to $\prod_i t_i - 1$ elements.

2.5 For the oneway model (2.2),

- (a) Show that under the assumption that $\bar{\tau} = 0$, which makes the τ_i estimable, the least squares estimates in the oneway CRD are given by

$$\hat{\tau}_i = \bar{y}_{i.} - \bar{y}, \quad \hat{\mu} = \bar{y}.$$

Derive the variances of $\hat{\tau}_i$ and $\hat{\tau}_i - \hat{\tau}_{i'}$.

- (b) For the following oneway CRD, provide the anova table and estimates of the parameter effects and their variances. To determine diet quality, male weanling rats were fed diets with various protein levels. Fifteen rats were randomly assigned to one of three diets, and their weight gain in grams was recorded. The data are (also in dataset **Protein**).

Diet Protein Level		
Low	Medium	High
3.89	8.54	20.39
3.87	9.32	24.22
3.26	8.76	30.91
2.70	9.30	22.78
3.82	10.45	26.33

2.6 Show the following for the oneway CRD (1.1):

- (a) $E(\bar{Y}_{i.} - \bar{Y}) = \hat{\tau}_i$.
- (b) A contrast $\sum_i a_i (\bar{Y}_{i.} - \bar{Y})$ is an unbiased estimator of $\sum_i a_i \tau_i$ with variance $\frac{\sigma^2}{r} \sum_i a_i^2$.
- (c) If there are r_i observations per treatment, then (a) and (b) hold with variance $\sigma^2 \sum_i a_i^2 / r_i$.

2.7 For the twoway CRD (2.4):

- (a) Verify the derivation of the least squares estimators and their variances. (*Hint*: First add $\pm y_{ij.}$ and show that the estimates will only depend on $y_{ij.}$. Then write $\tau_i + \gamma_j + (\tau\gamma)_{ij} = \gamma_{ij}$ and show that the least squares estimate of γ_{ij} is $y_{ij.}$.)

- (b) For the data of Example 2.4, found in dataset **RedClover**, estimate all of the treatment effects and their variances.

2.8 Calculation of variances in the twoway CRD.

- (a) Show that the estimator $\bar{Y}_{i..} - \bar{Y}$ has variance

$$\text{Var}(\bar{Y}_{i..} - \bar{Y}) = \text{Var}(\varepsilon_{i..} - \bar{\varepsilon}) = \left(1 - \frac{1}{t}\right) \frac{\sigma^2}{rg}.$$

Note that $\varepsilon_{i..}$ and $\bar{\varepsilon}$ are correlated, which has to be accounted for in the variance calculation. Justify the application of Lemma 2.16 with $W_i = \varepsilon_{i..}$.

- (b) Verify (2.12) by showing

$$\sum_i a_i(\bar{Y}_{i..} - \bar{Y}) = \sum_i a_i(\varepsilon_{i..} - \bar{\varepsilon}) = \sum_i a_i \varepsilon_{i..}$$

and then use the fact that $\varepsilon_{i..}$ are independent with variance σ^2/rg .

- (c) Find the expectation and variance for contrasts in the other effect estimates in (2.8).
 (d) Repeat (b) and (c) for the case of unequal cell sizes, where there are r_{ij} observations per cell. As a guide, first verify that

$$\text{Var}\left(\sum_i a_i(\bar{Y}_{i..} - \bar{Y})\right) = \sigma^2 \sum_i \frac{a_i^2}{\sum_j r_{ij}}$$

2.9 For the following experiment

- (i) specify the model equation
 (ii) set up the anova table (source, df and EMS)
 (iii) specify two hypotheses and how they would be tested

The cathode warm-up time in seconds was determined for three different tube types using eight observations on each tube type. The order of the experiment was completely randomized. The results were

	Tube Type		
	A	B	C
Warm-up	19 20	20 40	16 19
Time	23 20	20 24	15 17
(Seconds)	26 18	32 22	18 19
	18 35	27 18	26 18

2.10 (Finding an Optimal Region)³

If we are looking for maximum yield it would be wasteful to examine regions of low yield. Typically, the main features of such regions are *first-order*, that is, they are main effects rather than interactions. Thus, in a first search for such regions, we would be willing to use interaction terms as a denominator. Moreover, we consider this the first of a series of experiments; after we approximately locate a region of optimal response we will continue with a more precise experiment.

³ Methods for finding optimal regions are called *Response Surface* methods. Although we are not treating these methods in detail, they are an important topic. The textbooks of Mead (1988) and Dean and Voss (1999) contain chapter length treatments; a thorough introduction is the book by Khuri and Cornell (1996).

As a particular example, an experimenter is rearing beneficial insects and is trying to find the conditions that produce insects most rapidly. The researcher has four large environmental chambers that can be used to control temperature, photoperiod, and humidity. There are other factors, for example, diet, but these will not be considered at this time.

We consider four possible designs, where photoperiod is hours of light and dark in a 24-hour period; for example, 16:8 is 16 hours of light and 8 hours of dark. The designs are given below.

		Relative			
(1)		Chamber	Temperature (°C)	Humidity %	Photoperiod
One factor at a time	1	20	60	16:8	
	2	25	60	16:8	
	3	30	60	16:8	
	4	35	60	16:8	

		Relative			
(2)		Chamber	Temperature (°C)	Humidity %	Photoperiod
One factor at a time	1	20	60	16:8	
	2	20	60	16:8	
	3	25	60	16:8	
	4	25	60	16:8	

		Relative			
(3)		Chamber	Temperature (°C)	Humidity %	Photoperiod
Factorial, ignore humidity	1	20	60	16:8	
	2	20	60	14:10	
	3	25	60	16:8	
	4	25	60	14:10	

		Relative			
(4)		Chamber	Temperature (°C)	Humidity %	Photoperiod
Fraction	1	20	60	14:10	
	2	25	60	16:8	
	3	20	80	14:10	
	4	25	80	16:8	

(Experiment (4) is an example of a *fractional factorial*. See Section 6.3.)

- (a) Construct the anova table (just source and df) for each design.
- (b) Recall that the major goal of these experiments is to locate the condition for optimal growth. List at least one advantage and one disadvantage of each design.
- (c) Which design would you use for locating the region with the fastest insect growth?

2.11 Here we calculate expected mean squares to justify the *F*-test in Table 2.2.

- (a) Show that

$$E(SS(\text{Trts})) = rE \sum_{i=1}^t [\bar{y}_i - \bar{y}]^2 = rE \sum_{i=1}^t [(\tau_i - \bar{\tau}) + (\bar{\epsilon}_i - \bar{\epsilon})]^2$$

$$\begin{aligned}
 &= r \sum_{i=1}^t \tau_i^2 + r \sum_{i=1}^t E(\bar{\varepsilon}_{i\cdot} - \bar{\varepsilon})^2 \quad [\text{cross term is zero}] \\
 &= r \sum_{i=1}^t \tau_i^2 + (t-1)\sigma^2 \quad [\text{Lemma 2.16}].
 \end{aligned}$$

(b) Calculate the other EMS and produce the anova table

Source	df	EMS
Treatments	$t - 1$	$\sigma^2 + \frac{r}{t-1} \sum_i \tau_i^2$
Within	$r(t - 1)$	σ^2

(c) For model (2.2), suppose $j = 1, \dots, r_i$, so there are unequal numbers of experimental units per treatment. Produce the anova table analogous to that in part (b).

2.12 Kuehl (1994) reported data on weight gain (mg) of shrimp cultured in aquaria, subjected to different levels of temperature (T), density of shrimp populations (D), and water salinity (S). The factors were crossed, and the experiment was run as a threeway CRD. Here is a schematic of the data, given in dataset **Shrimp**.

	$T = 25^\circ\text{C}$		$T = 35^\circ\text{C}$	
	D		D	
	80	160	80	160
10%	x, x, x	x, x, x	x, x, x	x, x, x
S 25%	x, x, x	x, x, x	x, x, x	x, x, x
40%	x, x, x	x, x, x	x, x, x	x, x, x

- (a) Produce the anova table and do the appropriate F -tests.
- (b) You should find that, at the .05 level, the $T \times D$ and $S \times D$ interactions are not significant. Provide interpretations of this.
- (c) You should find that, at the .05 level, the $T \times S$ and $T \times S \times D$ interactions are significant. Provide interpretations of this.
- (d) Explore the interactions in part (c) further. Look at interaction with the linear and quadratic contrasts of S . Can you refine your conclusions from part (c)?
- (e) Suppose you were asked to design a second experiment to further explore the treatments that result in large weight gain. Based on what you have learned in parts (a)-(d), what would you suggest?

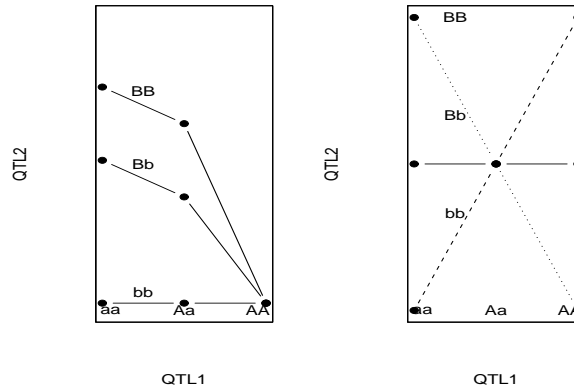
2.13 Referring to Example 2.5, look further into the interaction term to see if there is any significant effects.

- (a) Get the contrast coefficients for the three orthogonal (in the parameters) interaction contrasts for linear, quadratic and cubic tissue \times hCG.
- (b) Calculate the contrast sums of squares for each contrast, and perform the anova F -tests. Note that the estimated contrasts are correlated, and the sums of squares do not add to that of the 3 df interaction.
- (c) Perform the same linear, quadratic, and cubic breakdowns for the main effect of tissue.

(d) What can you conclude about the effects and interactions?

2.14 Referring to Example 2.7:

- (a) Produce a table like Table 2.7 for the left panel of Figure 2.3, where there is no interaction. Show that the marginal effect parameters act independently.
- (b) The following figure has two more possible QTL interactions patterns.



Show that, by constructing tables similar to Table 2.8, the left panel results in the marginal means of one QTL affecting the other, while for the second panel (where there is interaction) the marginal QTLs act independently (in fact, one has no effect).

2.15 Referring to Example 2.9:

- (a) Fill in the table with the six interaction contrasts for $A \times B$.
- (b) Make a similar table using polynomial contrasts.
- (c) Verify, using the cell means, that the contrast from the collapsed 2×2 table is exactly the same as the contrast with all twelve cells.
- (d) Suppose that a third factor C , with three levels, was crossed with A and B . Obtain the coefficients of the $A \times B \times C$ interaction contrast that compares the first level of B with the average of the other two, the first level of A with the average of the other three, and a linear trend in C . Also, give the collapsed table, and attempt to interpret the contrast in words.

2.16 Referring to Example 2.11:

- (a) Fill in the table with all of the effect contrasts.
- (b) Give the anova table, source, df, and EMS, where we pool the threeway and fourway interactions to get an error estimate. Specify the assumptions needed for this pooling.

2.17 Referring to Example 2.13:

- (a) Verify the analysis of covariance, the adjusted treatment means and their standard errors.
- (b) Test the significance of the pairwise differences, using both adjusted and unadjusted means. Do the conclusions change from the anova to the ancova?
- (c) Redo part (b) using the average standard error of the difference give in (2.23). How do the results compare to part (b)? Based on your assessment of $SS(\text{Trt}_x)$ was the ancova worthwhile?

- 2.18** An experiment described in Snedecor and Cochran (1989) (same Cochran⁴) concerned treatment of leprosy in the Philippines. On each patient, six sites were examined for leprosy bacilli, with the abundance of the bacilli being measured. The covariate, X , is a score representing the abundance before treatment, and the response Y is the same score after several months of treatment with antibiotics A and B or a control C . The data are in dataset **Antibiotic**.
- Run both an anova on Y and an ancova with covariate X . Explain the results with respect to the significance of the treatments.
 - The treatment design suggests testing the two orthogonal contrasts $A + B - 2C$ and $A - B$. Do this both for the adjusted and unadjusted means. Describe your conclusions.
 - Is the initial measurement X a good covariate? Explain. You may want to look at the anova of X on treatment.
 - Plot the individual regression lines for each drug, along with the overall regression (from the covariance analysis) and the unadjusted means. Explain the plot in terms of the results of the analysis.

Accompaniment

- 2.19** *A Useful Identity:*

- Show that for numbers x_{ij} ,

$$\sum_{i=1}^t \sum_{j=1}^t (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2 = \sum_{ij} x_{ij}^2 - b \sum_i \bar{x}_i^2 - t \sum_j \bar{x}_j^2 + bt\bar{x}^2.$$

- If X_{ij} are random variables with $EX_{ij} = 0$, show that

$$\begin{aligned} E \left[\sum_{i=1}^t \sum_{j=1}^t (X_{ij} - \bar{X}_i - \bar{X}_j + \bar{X})^2 \right] \\ = \sum_{ij} \text{Var} X_{ij} - b \sum_i \text{Var} \bar{X}_i - t \sum_j \text{Var} \bar{X}_j + bt \text{Var} \bar{X}. \end{aligned}$$

- 2.20** (a) Show how to apply Lemma 2.16 to establish (2.9).
 (b) Finish the calculations to fill in Table 2.6. To calculate the EMS for the interaction term you might first want to establish that

$$\bar{\varepsilon}_{ij.} - \bar{\varepsilon}_{i..} - \bar{\varepsilon}_{.j.} + \bar{\varepsilon}_{...} \sim N \left(0, \frac{(t-1)(g-1)}{r} \sigma^2 \right).$$

- 2.21** (Some anova theory: the t - F relationship)

- Referring to Example 2.2, there it was demonstrated numerically that the oneway CRD anova F -statistic is equal to the average squared t -statistics from uncorrelated contrasts. Show, analytically, that this is true for any set of orthogonal contrasts.

⁴ This is really a famous pair. You know some of what Cochran did. Among the accomplishments of Snedecor is the derivation of the anova F -statistic, which has sometimes been called “Snedecor’s F ”. Snedecor named it “ F ” in honor of Fisher.

- (b) Another way to derive the anova F -statistics is as a *maximum* of t -statistics. In a oneway CRD with t treatments and n_i observations per treatment, we test contrast hypotheses by

$$\text{reject } H_0^{(a)} : \sum_i a_i \mu_i = 0 \text{ if } |t^{(a)}| = \frac{|\sum_i a_i \bar{y}_i|}{\sqrt{\text{MSE} \sum_i a_i^2 / n_i}} > t_{\alpha/2}.$$

- (i) The *union-intersection* principle (see Casella and Berger 2001, Sections 8.3.3 and 11.2) tells us that $H_0 : \mu_1 = \mu_2 = \dots = \mu_t$ is true if and only if $H_0^{(a)}$ is true for *every* set of contrasts (a_1, a_2, \dots, a_k) . Prove this. (Thus, H_0 is the intersection of $H_0^{(a)}$.)
- (ii) Building on (i), show that H_0 will be rejected if at least one $H_0^{(a)}$ is rejected. (Thus, the rejection region of H_0 is a union of the individual rejection regions.)
- (iii) Finally, argue that H_0 will be rejected if and only if $\max_a |t^{(a)}|$ exceeds its critical point, and establish that

$$\max_{(a_1, a_2, \dots, a_k) : \sum_i a_i = 0} \frac{(\sum_i a_i \bar{y}_i)^2}{\sum_i a_i^2 / n_i} = \sum_i n_i (\bar{y}_i - \bar{\bar{y}})^2,$$

showing that the maximum of the squared t -statistics is the anova F . (This is a difficult maximization.)

- 2.22** Complete the calculations to produce the EMS in Table 2.6. In particular, apply the identity in Exercise 2.19 to show

$$E \sum_{ij} (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{\bar{Y}})^2 = bt [\text{Var} \varepsilon_{ij} - \text{Var} \bar{\varepsilon}_i - \text{Var} \bar{\varepsilon}_j + \text{Var} \bar{\bar{\varepsilon}}].$$

- 2.23** For a threeway CRD, all factors fixed, a model is

$$y_{ijkl} = \mu + \alpha_i + \delta_j + \gamma_k + (\alpha\delta)_{ij} + (\alpha\gamma)_{ik} + (\delta\gamma)_{jk} + (\alpha\delta\gamma)_{ijk} + \varepsilon_{ijkl},$$

where $i = 1, \dots, t, j = 1, \dots, b, k = 1, \dots, c, l = 1, \dots, r$, and $\varepsilon_{ijkl} \sim N(0, \sigma^2)$, independent, and we have the identifiability restrictions

$$\begin{aligned} \bar{\alpha} = \bar{\delta} = \bar{\gamma} = (\bar{\alpha}\delta) = (\bar{\alpha}\gamma) = (\bar{\delta}\gamma) = (\alpha\bar{\delta}\gamma) &= 0, \\ (\bar{\alpha}\delta)_i = (\bar{\alpha}\delta)_j = (\bar{\alpha}\gamma)_i = (\bar{\alpha}\gamma)_k = (\bar{\delta}\gamma)_j = (\bar{\delta}\gamma)_k &= 0, \\ (\alpha\bar{\delta}\gamma)_i = (\alpha\bar{\delta}\gamma)_j = (\alpha\bar{\delta}\gamma)_k = (\alpha\bar{\delta}\gamma)_{ij} = (\alpha\bar{\delta}\gamma)_{ik} = (\alpha\bar{\delta}\gamma)_{jk} &= 0. \end{aligned}$$

- (a) Verify that the EMS for factor A is given by

$$E(\text{MS}(A)) = \sigma^2 + \frac{1}{t-1} \sum_{i=1}^t \alpha_i^2.$$

- (b) Verify that the EMS for the $A \times B$ interaction is given by

$$E(\text{MS}(A \times B)) = \sigma^2 + \frac{1}{(t-1)(b-1)} \sum_{i=1}^t \sum_{j=1}^b (\bar{\alpha}\delta)_{ij}^2.$$

- (c) Based on what you learned in (a) and (b) fill in the complete threeway anova table (Source, df, EMS). (You do not need to do any more calculations – you should be able to deduce what the other EMS will be.)
- (d) An experimenter will typically state a hypothesis in words. For the two hypothesis

$$H_0 : \text{No B effect and } H_0 : \text{No BC effect}$$

state each hypothesis in terms of the appropriate parameter values.

2.24 Refer to Miscellanea 2.9.2. If the treatment is applied, we assume that either $\delta_{ij} \sim \text{iid } N(0, \sigma_\delta^2)$, or $\delta_i \sim \text{iid } N(0, \sigma_\delta^2)$, in both cases independent of $\varepsilon_{ij} \sim N(0, \sigma_\varepsilon^2)$.

- (a) Show that, if we have the case where $\delta_i \sim \text{iid}N(0, \sigma_\delta^2)$, then $\text{Cov}(Y_{ij}, Y_{ij'}) = \sigma_\delta^2$, and in the other cases the Y_{ij} are all independent.
- (b) Show that the following table summarizes the EMS for the case of modeling the error due to applying the treatment, and that if $\delta_{ij} = \delta_i$ there is no test on treatments.

EMS for Attributes				
Source	df	EMS $\delta_{ij} = 0$	EMS $\delta_i \sim \text{iid } N(0, \sigma_\delta^2)$	EMS $\delta_{ij} \sim \text{iid } N(0, \sigma_\delta^2)$
Treatments	$t - 1$	$\sigma_\varepsilon^2 + \frac{r}{t-1} \sum_i \tau_i^2$	$\sigma_\varepsilon^2 + r\sigma_\delta^2 + \frac{r}{t-1} \sum_i \tau_i^2$	$\sigma_\varepsilon^2 + \sigma_\delta^2 + \frac{r}{t-1} \sum_i \tau_i^2$
Within	$t(r - 1)$	σ_ε^2	σ_ε^2	$\sigma_\varepsilon^2 + \sigma_\delta^2$

(This table once again illustrates the importance of true replication. In the middle EMS column the treatment is not truly replicated, and the result is no test on treatments!)

2.25 Here we see that a single contrast has a chi-squared distribution.

- (a) Verify that the matrix \mathbf{aa}' in (2.24) is idempotent.
- (b) Show that (a) implies that for $Y_i \sim N(0, \sigma^2)$ and any contrast, we have

$$\frac{\left(\sum_i a_i Y_i\right)^2}{\sigma^2 \sum_i a_i^2} \sim \chi_1^2$$

2.26 Referring to Technical Note 2.8.3:

- (a) Show that A_1 is idempotent and $\text{SS}(\text{Within}) = \mathbf{Y}'A_1\mathbf{Y}$.
- (b) Show that A_2 is idempotent and $\text{SS}(\text{Within}) = \mathbf{Y}'A_2\mathbf{Y}$.
- (c) Show that $A_1A_2 = A_2A_1 = 0$, and hence $A_1 + A_2$ is idempotent and the assumptions of Cochran's Theorem are satisfied.
- (d) Prove Theorem 2.21.

2.27 Referring to Technical Note 2.8.3, here we will prove Theorem 2.22.

- (a) Establish (1)-(3) of Theorem 2.22, using arguments similar to those in Exercise 2.26.
- (b) To prove (4), we need to find the matrix A_3 that satisfies $\mathbf{Y}A_3\mathbf{Y} = \text{SS}(T \times G)$, and show that A_3 is idempotent and satisfies $A_1A_3 = A_3A_1 = 0$, where A_1 satisfies $\mathbf{Y}A_1\mathbf{Y} = \text{SS}(\text{Within})$. Cochran's Theorem can then be applied to get the F -test. (The matrix A_3 is constructed in a manner similar to the interaction matrix in Technical Note 3.8.2.)

2.28 Prove Theorem 2.23 by showing that, under model (2.4):

- (a) $\text{Cov}(\bar{Y}_i, Y_{i'jk} - \bar{Y}_{i'j}) = 0$ for all i, i', j, k .

- (b) $\text{Cov}(\bar{Y}_j, Y_{ij'k} - \bar{Y}_{ij'}) = 0$ for all i, j, j', k .
- (c) $\text{Cov}(\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j, Y_{ij'k} - \bar{Y}_{ij'}) = 0$ for all i, i', j, j', k .
- (d) Use the normality assumption to go from uncorrelated to independence in (a)-(c), and use the properties of the t -distribution (Section 2.8.2) to complete the proof.

2.29 Referring to Section 2.6:

- (a) Use model (2.16) to derive the least squares estimators (2.17) and show that the least squares estimator of $\mu + \tau_i$ is unbiased.
- (b) Show that, in (a), if $\sum_i \tau_i = 0$, then we can estimate τ_i . If the cell sizes are unequal (r_i instead of r), derive the least squares estimates. How can we now estimate τ_i ?
- (c) Derive the variance of the least squares estimate of $\mu + \tau_i$ in the case of unequal r_i .

2.30 For the variance of an anova contrast:

- (a) Verify the variance (2.21).
- (b) Prove the identity (2.22) and use it to verify the variance approximation (2.23). (Add $\pm \bar{b}$ on the left side and expand.)
- (c) Suppose that, in (2.22), each b_i is a mean based on r_i observations. Formulate and prove an analogous identity to (2.22), and then establish an analogous average variance to (2.23).

2.31 The following data are measurements on the strength index of three varieties of cotton, where the treatments are pounds of potassium oxide per acre (dataset **Imbalance**). Here we want to see the effect of imbalance on the anova.

Strength Index of Cotton			
Varieties			
Treatment	1	2	3
36	7.06	7.75, 8.22	7.95, 8.59
60	7.51, 7.5	8.08, 8.18	8.69, 8.39
84	7.27, 7.49	7.9	8.04
108	6.55, 6.47	7.26, 7.06	7.35
132	6.97, 7.14	7.52, 7.83	7.63, 7.2

- (a) Verify the anova tables in Miscellanea 2.9.3.
- (b) Produce one anova table that contains the partial sums of squares. (You can do this by combining the two tables from part (a) or by using the R command `drop1`.)
- (c) In terms of the analysis of these data, explain why the table in part (b) is most appropriate.
- (d) Calculate the contrast sum of squares for the linear effect of potassium oxide and test its significance.
- (e) Estimate the contrast and its standard error, and give a 95% confidence interval.

2.8 Technical Notes _____

2.8.1 Helpful Lemma I _____

The following lemma is helpful in calculating expected sums of squares in CRD designs.

Lemma 2.16. Suppose that W_1, W_2, \dots, W_n satisfy $EW = 0$, $\text{Var}W = \sigma^2$, and $\text{Cov}(W_i, W_{i'}) = 0$. Then

- (1) $E \sum_{i=1}^n (W_i - \bar{W})^2 = \sum_{i=1}^n \text{Var}W_i - n \text{Var}\bar{W} = (1 - \frac{1}{n}) \sum_{i=1}^n \text{Var}W_i$,
- (2) $\text{Cov}(W_i - \bar{W}, W_{i'} - \bar{W}) = -\frac{1}{n} \sigma^2$,
- (3) $\text{Var}(W_i - \bar{W}) = E(W_i - \bar{W})^2 = (1 - \frac{1}{n}) \sigma^2$.

If the W_i are normal, then

- (4) $W_i - \bar{W} \sim N(0, (1 - \frac{1}{n}) \sigma^2)$.

2.8.2 Anova Theory

In this section we outline the statistical theory behind anova, which leads us to t -tests and F -tests, as well as confidence intervals. Many of the details that we suppress here can be found in Casella and Berger (2001), especially in Section 5.3.

Theorem 2.17 (Independence of Linear Combinations). Let $Y_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, r$, be jointly normal, with $\text{Cov}(Y_i, Y_{i'}) = \phi_{ii'}$, and let $W_j = \sum_{i=1}^r a_{ij} Y_i$, $j = 1, 2, \dots, m$ be any collection of linear combinations. Then

- (1) The random variables W_j , $j = 1, 2, \dots, m$, are jointly normal, with

$$W_j \sim N \left(\sum_{i=1}^r a_{ij} \mu_i, \sum_{i=1}^r a_{ij}^2 \sigma_i^2 + 2 \sum_{i>i'} a_{ij} a_{i'j} \phi_{ii'} \right).$$

- (2) If $\text{Cov}(W_j, W_{j'}) = 0$, W_j and $W_{j'}$ are independent.
- (3) If $\text{Cov}(W_j, W_{j'}) = 0$ for all $j \neq j'$, then the W_j are all independent.

The proof follows from the properties of the normal distribution and some algebra. It uses the fact that, because of the special form of the normal distribution, if the covariances are zero the density function factors into products.

Theorem 2.18 (From normals to χ^2 random variables). Let χ_p^2 denote a chi squared random variable with p degrees of freedom.

- (1) If $Z \sim N(0, 1)$, then $Z^2 \sim \chi_1^2$.
- (2) If W_i , $i = 1, 2, \dots, r$ are independent, each distributed as $\chi_{p_i}^2$, then $\sum W_i \sim \chi_{\sum p_i}^2$.

So the degrees of freedom of independent χ^2 random variables sum. Of course, W_i could be the square of a standard normal or, more generally, if $Y_i \sim N(\mu_i, \sigma_i^2)$, $i = 1, 2, \dots, r$ are independent, then

$$\sum_{i=1}^r \left(\frac{Y_i - \mu_i}{\sqrt{\sigma_i^2}} \right)^2 \sim \chi_r^2.$$

Now, we take the final step to the t and F distributions.

Theorem 2.19 (t and F distributions).

(1) If $Z \sim N(0, 1)$ and $V \sim \chi_p^2$, independent of Z , then

$$\frac{Z}{\sqrt{V/p}} \sim t_p,$$

where t_p is Student's t distribution with p degrees of freedom.

(2) If $W \sim \chi_q^2$ and $V \sim \chi_p^2$, independent, then

$$\frac{W/q}{V/p} \sim F_{q,p},$$

where $F_{q,p}$ has an F distribution with q numerator degrees of freedom and p denominator degrees of freedom.

2.8.3 Cochran's Theorem for CRDs

Thus far, we have reached the t and F distributions when starting from independent normal random variables. However, we need more, as we will often be dealing with variables of the form $Y_i - \bar{Y}$, and these are not independent. More generally, in anova, we are typically interested in the distributions of sums of squares that are built up from linear combinations (actually contrasts) of normal variables. (For more details about anova theory see Graybill and Hultquist 1961, Albert 1976, Khuri 1982. Speed (1987) gives a very detailed derivation of this decomposition.)

For example, from Theorem 2.17, consider a linear combination $\sum_i a_i Y_i$ with $\sum_i a_i^2 = 1$. Write $\sum_i a_i Y_i = \mathbf{a}'\mathbf{Y}$, where $\mathbf{a} = (a_1, a_2, \dots, a_p)$ and $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$. Then

$$(2.24) \quad \left(\sum_i a_i Y_i \right)^2 = (\mathbf{a}'\mathbf{Y})^2 = \mathbf{Y}'(\mathbf{a}\mathbf{a}')\mathbf{Y},$$

where the matrix $(\mathbf{a}\mathbf{a}')$ satisfies $(\mathbf{a}\mathbf{a}')^2 = (\mathbf{a}\mathbf{a}')$. Recall that a matrix A that satisfies the condition $A^2 = A$ is called *idempotent*, and there is a strong connection between idempotent matrices and the chi squared distribution.

The major result that we need is known as Cochran's Theorem, which dates back to Cochran (1934)⁵.

Theorem 2.20 (Cochran's Theorem). Let $\mathbf{Y} \sim N(0, \Sigma)$, and let A_k , $k = 1, 2, \dots, m$ satisfy $\sum_{k=1}^m A_k = A$, where $A\Sigma$ is idempotent. If

$$A_k \Sigma \text{ is idempotent for every } k \text{ and } A_k \Sigma A_{k'} = 0, \quad k \neq k',$$

then

- (1) $\mathbf{Y}' A_k \mathbf{Y} \sim \chi_{\text{tr}(A_k \Sigma)}^2$ for every k ,
- (2) $\mathbf{Y}' A_k \mathbf{Y}$ and $\mathbf{Y}' A_{k'} \mathbf{Y}$ are independent for $k \neq k'$,
- (3) $\mathbf{Y}' A \mathbf{Y} \sim \chi_{\text{tr}(A \Sigma)}^2$.

⁵ A complete treatment of Cochran's Theorem can also be found in Stuart and Ord 1987, Chapter 15 or Scheffé 1959 Appendix VI.

If $\Sigma = \sigma^2 I$, a scalar multiple of the identity matrix, it follows from the previous discussion that $\mathbf{Y}'\mathbf{Y}/\sigma^2 = \sum Y_i^2/\sigma^2 \sim \chi_p^2$. Using Cochran's theorem, if A is an idempotent matrix then we can write

$$\mathbf{Y}'\mathbf{Y} = \mathbf{Y}'A\mathbf{Y} + \mathbf{Y}'(I - A)\mathbf{Y},$$

and, when divided by σ^2 , both quadratic forms on the right side are independent χ^2 random variables, with degrees of freedom equal to the trace of the matrix in the quadratic form. That is, for $Y \sim N(0, \sigma^2 I)$ and idempotent A we have

$$\mathbf{Y}'A\mathbf{Y}/\sigma^2 \sim \chi_{tr(A)}^2 \text{ and } \mathbf{Y}'(I - A)\mathbf{Y}/\sigma^2 \sim \chi_{tr(I-A)}^2,$$

where $tr(A)$ is the sum of the diagonal elements. Since $A(I - A) = 0$, we have partitioned $\chi_p^2 = \chi_{tr(A)}^2 + \chi_{tr(I-A)}^2$.

As we have seen, many of the sums of squares that we deal with in anova are built up from contrasts in normal random variables. We now look at the oneway CRD anova

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2), \text{ independent, } \quad i = 1, \dots, t, \quad j = 1, \dots, r,$$

and work our way to the F distribution for the test on treatments.

Within cell i , we have observations $Y_{i1}, Y_{i2}, \dots, Y_{ir}$ and we can write the deviations from the mean as

$$\begin{pmatrix} Y_{i1} - \bar{Y}_i \\ Y_{i2} - \bar{Y}_i \\ \vdots \\ Y_{ir} - \bar{Y}_i \end{pmatrix} = \left(I - \frac{1}{r}J\right) \begin{pmatrix} Y_{i1} \\ Y_{i2} \\ \vdots \\ Y_{ir} \end{pmatrix},$$

where I is the identity matrix and J is a matrix of ones. Now we note that

$$\left(I - \frac{1}{r}J\right) \left(I - \frac{1}{r}J\right) = \left(I - \frac{1}{r}J\right)$$

is an idempotent matrix. The entire set of deviations can be written as

$$\begin{pmatrix} Y_{11} - \bar{Y}_1 \\ Y_{12} - \bar{Y}_1 \\ \vdots \\ Y_{tr} - \bar{Y}_t \end{pmatrix} = \begin{pmatrix} \left(I - \frac{1}{r}J\right) & 0 & \cdots & 0 \\ 0 & \left(I - \frac{1}{r}J\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \left(I - \frac{1}{r}J\right) & \end{pmatrix}_{tr \times tr} \begin{pmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ Y_{tr} \end{pmatrix} \\ \stackrel{\text{def}}{=} A_1 \mathbf{Y}.$$

It can be shown that A_1 is idempotent and $SS(\text{Within}) = \mathbf{Y}'A_1\mathbf{Y}$. Similarly, we can write the deviations for the treatment sum of squares as

$$\begin{pmatrix} \bar{Y}_1 - \bar{\bar{Y}} \\ \bar{Y}_2 - \bar{\bar{Y}} \\ \vdots \\ \bar{Y}_t - \bar{\bar{Y}} \end{pmatrix} = \left(I - \frac{1}{t}J\right) \begin{pmatrix} \bar{Y}_1 \\ \bar{Y}_2 \\ \vdots \\ \bar{Y}_t \end{pmatrix} = \left(I - \frac{1}{t}J\right) \frac{1}{r} \begin{pmatrix} \mathbf{1}_{r \times 1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{r \times 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \mathbf{1}_{r \times 1} & \end{pmatrix}_{t \times tr} \mathbf{Y},$$

however, the matrix multiplying \mathbf{Y} is not square and cannot be idempotent. To remedy this we take t copies of it (this is why there is a t multiplying the treatment sum of squares) and define

$$A_2 = \begin{pmatrix} \left(I - \frac{1}{t} J \right) \\ \left(I - \frac{1}{t} J \right) \\ \vdots \\ \left(I - \frac{1}{t} J \right) \end{pmatrix}_{tr \times t} \frac{1}{r} \begin{pmatrix} \mathbf{1}_{r \times 1} & 0 & \cdots & 0 \\ 0 & \mathbf{1}_{r \times 1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \mathbf{1}_{r \times 1} & \end{pmatrix}_{t \times tr}$$

and then A_2 is idempotent and $\text{SS}(\text{Treatments}) = \mathbf{Y}' A_2 \mathbf{Y}$. We are now in the position to apply Cochran's Theorem to get the following, whose proof is left to Exercise 2.26.

Theorem 2.21 (Oneway CRD F -distribution). *For the oneway CRD*

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2), \quad \text{independent}, \quad i = 1, \dots, t, \quad j = 1, \dots, r,$$

- (1) $\text{SS}(\text{Within})/\sigma^2 \sim \chi_{t(r-1)}^2$.
- (2) Under $H_0 : \tau_i = 0$ for all i , $\text{SS}(\text{Trt})/\sigma^2 \sim \chi_{t-1}^2$, independent of $\text{SS}(\text{Within})$.
- (3) Under $H_0 : \tau_i = 0$ for all i , $F = \frac{\text{SS}(\text{Trt})/(t-1)}{\text{SS}(\text{Within})/(t(r-1))} \sim F_{t-1, t(r-1)}$.

A similar theorem holds for the twoway (and higher) CRDs, with the complication that we have to deal with the matrices for the interaction terms. We leave that pleasant task to Exercise 2.27, and simply state the theorem here.

Theorem 2.22 (Twoway CRD F -distributions). *For the twoway CRD of model (2.4),*

- (1) $\text{SS}(\text{Within})/\sigma^2 \sim \chi_{tg(r-1)}^2$.
- (2) Under $H_0 : \tau_i = 0$ for all i , $\text{SS}(T)/\sigma^2 \sim \chi_{t-1}^2$, independent of $\text{SS}(\text{Within})$ and $F = \frac{\text{SS}(T)/(t-1)}{\text{SS}(\text{Within})/(tg(r-1))} \sim F_{t-1, tg(r-1)}$.
- (3) Under $H_0 : \gamma_j = 0$ for all j , $\text{SS}(G)/\sigma^2 \sim \chi_{g-1}^2$, independent of $\text{SS}(\text{Within})$ and $F = \frac{\text{SS}(G)/(g-1)}{\text{SS}(\text{Within})/(tg(r-1))} \sim F_{g-1, tg(r-1)}$.
- (4) Under $H_0 : (\tau\gamma)_{ij} = 0$ for all i, j , $\text{SS}(T \times G)/\sigma^2 \sim \chi_{(t-1)(g-1)}^2$, independent of $\text{SS}(\text{Within})$ and $F = \frac{\text{SS}(T \times G)/((t-1)(g-1))}{\text{SS}(\text{Within})/(tg(r-1))} \sim F_{(t-1)(g-1), tg(r-1)}$.

Finally, the distribution of contrasts in the CRD is also straightforward. We state the theorem here and leave the proof to Exercise 2.28.

Theorem 2.23 (Contrasts in the twoway CRD). *For the twoway CRD of model (2.4), let $\hat{\sigma}^2 = \text{MS}(\text{Within})/(tg(r-1))$. Then for any contrast*

- (1) $\sqrt{gr} \sum_i a_i (\bar{Y}_i - \tau_i) / \hat{\sigma} \sim t_{tg(r-1)}$.
- (2) $\sqrt{tr} \sum_j a_j (\bar{Y}_j - \gamma_{ij}) / \hat{\sigma} \sim t_{tg(r-1)}$.
- (3) $\sqrt{r} \sum_{ij} a_{ij} (\bar{Y}_{ij} - \bar{Y}_i - \bar{Y}_j + \bar{Y} - (\tau\gamma)_{ij}) / \hat{\sigma} \sim t_{tg(r-1)}$.

There has been much research in answering the question "When is a sum of squares an analysis of variance", the title of Albert (1976). What is meant is, under what conditions can we partition the total sum of squares into orthogonal

pieces, and get out the usual F -tests. This was first addressed by Graybill and Hultquist (1961), who investigated the connections between such a decomposition and minimal sufficiency, which implies that the inference would then be based on the likelihood function. The most general treatment of this problem can be found in Speed (1987). On a related note, Bradley (1973) gives conditions for the equivalence of likelihood and least squares estimates in linear models (see Berger and Casella 1992 for a more elementary treatment of this equivalence).

2.8.4 Noncentral Distributions

In the anova, if we set up the ratio of mean squares correctly, then under the appropriate H_0 their ratio will have an F distribution. More precisely, it is a *central* F distribution as described in Theorem 2.19. If H_0 is false, and the appropriate parameter is nonzero, then the ratio of mean squares has a *noncentral* F distribution. This noncentral distribution, and its properties, is related to the power of the anova tests, and a good understanding of the behavior of the F -statistic under the noncentral distribution will be helpful in better evaluating anova strategies. Here we discuss the oneway and twoway CRD; extensions to other designs should be clear.

The central F distribution (we usually omit the adjective unless we need to be very clear) is obtained from the ratio of two independent central χ^2 random variables. The noncentral F comes from the noncentral χ^2 distribution.

Definition 2.24. If X_1, X_2, \dots, X_p are independent normal random variables, $X_i \sim N(\theta_i, \sigma^2)$, then

$$W = \frac{\sum_i X_i^2}{\sigma^2} \sim \chi_p^2(\lambda),$$

a *noncentral* χ^2 random variable with p degrees of freedom and noncentrality parameter $\lambda = \sum_i \theta_i^2$. If $V \sim \chi_q^2$ (central), independent of W , then

$$\frac{W/p}{V/q} \sim F_{p,q}(\lambda),$$

a *noncentral* F with degrees of freedom p and q and noncentrality parameter $\lambda = \sum_i \theta_i^2 / \sigma^2$.

The density function of the noncentral χ^2 is quite imposing, being an infinite sum of weighted central χ^2 densities. We will not display it here, but rather share a more useful characterization of the noncentral χ^2 : If

$$K \sim \text{Poisson}(\lambda) \text{ and } W|K \sim \chi_{p+2K}^2, \text{ then } W \sim \chi_p^2(\lambda),$$

which also tells us that the weights in that infinite sum are Poisson probabilities. The important property about these noncentral distributions is that they are *stochastically increasing* in the noncentrality parameter. That is, for a fixed value a

$$\begin{aligned} W \sim \chi_p^2(\lambda) &\Rightarrow P_\lambda(W > a) \text{ is an increasing function of } \lambda \\ T \sim F_{p,q}(\lambda) &\Rightarrow P_\lambda(T > a) \text{ is an increasing function of } \lambda. \end{aligned}$$

Thus, if W or T are test statistics, and we reject H_0 for large values of the statistic, then the power of the test increases with λ , since the probability of the statistics exceeding the cutoff point increases in λ .

Oneway CRD

In the oneway CRD of model (2.2), with EMS given in Exercise 2.11, we test $H_0 : \sum_i \tau_i^2 = 0$ with $F = \text{MS}(\text{Treatments})/\text{MS}(\text{Within})$, calculated as in Table 2.2. If H_0 is violated, then $E(\bar{Y}_i - \bar{Y}) = \tau_i$, the numerator sum of squares is noncentral χ^2 with noncentrality parameter $\sum_i \tau_i^2/\sigma^2$, and the F -statistic has the corresponding noncentral F distribution. The power of the F -test increases with $\sum_i \tau_i^2$. Thus the power increases no matter which τ_i increases, and the power is constant on spheres where $\sum_i \tau_i^2$ is constant. Note also that the denominator of the F -statistic remains an independent central χ^2 .

Twoway CRD

Similar conclusions hold for the twoway CRD of model (2.4) with respect to the tests of the main effects and the interactions, with denominator mean square $\text{MS}(\text{Within})$. That is, the power of the respective tests is increasing in $\sum_i \tau_i^2$, $\sum_i \gamma_i^2$, and $\sum_i (\tau\gamma)_i^2$ (see Table 2.6). There is an interesting occurrence, however, if we look at the practice of pooling error terms.

In Definition 2.24, if V is an independent *noncentral* χ^2 , say $V \sim \chi_q^2(\delta)$, then $T = (W/p)/V/q$ has a *doubly noncentral* F distribution, $F_{p,q}(\lambda, \delta)$. Moreover, it should be clear that $P_{\lambda,\delta}(T > a)$ is increasing in λ for fixed δ , and decreasing in δ for fixed λ (see Scheffé1959, Section 4.8).

Now, in the twoway CRD, suppose that we, for example, test $H_0 : \sum_i \tau_i^2 = 0$ using a denominator that is obtained by pooling the interaction and within sum of squares. (Possibly based on first accepting the null hypothesis that there is no interaction.) If, in fact there is an interaction, so $\sum_i (\tau\gamma)_i^2 > 0$, then we have increased the noncentrality parameter in the denominator of the F -ratio, which has a doubly noncentral F distribution. Thus, we have decreased the probability of the test statistic exceeding the cutoff or, in other words, we have decreased the power of the test. This is what is called a *conservative* test, where it is more difficult to reject H_0 because the distribution of the denominator of the statistic has been inflated. This procedure is more prone to making the Type I error, and a rejection is usually greeted with enthusiasm.

2.9 Miscellanea

2.9.1 Multiple Comparisons and Error Rates

Post-anova analysis will often involve many inferences, typically on contrasts or pairwise differences. Of course, we know that the simultaneous inference from many α level tests is not necessarily at level α , so the question is how do we control this in a meaningful way.

This is actually a difficult and important question, especially in light of microarray experiments where we might be faced with thousands of hypothesis tests. In such cases, we must intelligently balance power and false detection rates.

Classical Error Rate Control

First, there are many definitions of Type I Error for multiple comparisons, so exactly what is meant by “ α level” is not always clear. Some of the types of

error rates considered are *experimentwise error rate* or *comparisonwise error rate*. Miller (1981) and Hsu (1996) are good references for this topic.

Experimentwise error rate: Controls the error rate of the entire experiment, so a 5% experimentwise α level means that if all nulls are true, there will be a false rejection in 5% of experiments, no matter how many tests are done. This is a very conservative criterion and is virtually useless for exploratory studies (no power)

Comparisonwise error rate: Controls the error rate of comparisons, so a 5% comparisonwise α level means that if all nulls are true, there will be a false rejection in 5% of comparisons. So if you do 1000 comparisons you can expect 50 false rejections. This is a very liberal criterion with high power, and is good for exploratory studies. It will produce a lot of false rejections, however.

There is also a *familywise error rate*, which is the experimentwise error applied to a smaller group of comparisons, a “family”.

Note that, if we are combining results of *independent* tests, then things are simple in the following sense: If we do m tests, each at level α , and if all of the null hypotheses are true, then the overall Type I error rate (experimentwise) error is $1 - (1 - \alpha)^m$, so 20 independent .05 tests have an overall Type I error rate of .64 and 20 independent .01 tests have an overall Type I error rate of .18. Simultaneous coverage of the 20 intervals would be .36 and .82, respectively. However, this is not a realistic situation, as we will seldom find independent tests in any experiment. For example, if we have a set of uncorrelated contrasts under normality, and we *know* the variance, or we use an independent variance estimate for each contrast, then we have independent tests. In almost any other situation, such as correlated contrasts or the use of pooled variances, the tests are dependent.

As a more realistic example, if we want to make a simultaneous $1 - \alpha$ statement about the coverage of m confidence sets, then, from the Bonferroni Inequality, we can construct each confidence set to be of level $1 - \frac{\alpha}{m}$. If we do each test at level α/m then we are controlling the experimentwise error at level α . If we do each test at level α , then we are controlling the comparisonwise error at level α . In an anova with k treatments, simultaneous inference on all $k(k - 1)/2$ pairwise differences can be made with confidence $1 - \alpha$ if each t interval has confidence $1 - 2\alpha/[k(k - 1)]$.

An alternative to Bonferroni, which also controls the experimentwise error, is the Scheffé procedure, which provides simultaneous confidence intervals on *all* contrasts. The procedure states that simultaneously for all contrasts (a_1, \dots, a_k) ,

$$\sum_{i=1}^k a_i \bar{Y}_i - \mathbf{M} \sqrt{\hat{\sigma}^2 \sum_{i=1}^k a_i^2 / n_i} \leq \sum_{i=1}^k a_i \theta_i \leq \sum_{i=1}^k a_i \bar{Y}_i + \mathbf{M} \sqrt{\hat{\sigma}^2 \sum_{i=1}^k a_i^2 / n_i},$$

where $\mathbf{M} = \sqrt{(k - 1)F_{k-1, n-k, \alpha}}$ and $\hat{\sigma}^2$ has $n - k$ degrees of freedom.

One of the real strengths of the Scheffé procedure is that it allows legitimate “data snooping”, and we can test hypotheses that have been suggested by

the data. Suggested hypotheses can bias the results and, hence, invalidate the inference, but intervals or tests that are valid for *all* contrasts, whether they have been suggested by the data, have already been taken care of by the Scheffé procedure. So the bias is eliminated. Of course, we pay for this data snooping privilege because the confidence intervals are very wide, and the tests are not very powerful.

There is a plethora of other simultaneous inference procedures, most concerned with pairwise comparisons. A method due to Tukey gives simultaneous confidence intervals on all pairwise differences, not all contrasts.

Other types of multiple comparison procedures, which deal with pairwise differences, are more powerful than the Scheffé method or the Tukey method. Some procedures are the LSD (Least Significant Difference) Procedure, Protected LSD, Duncan’s Procedure, and Student–Neumann–Keuls’ Procedure.

These last two are *multiple range* procedures, where the cutoff point to which comparisons are made changes between comparisons. (These should be avoided, as they can lead to contradictory inferences.)

Avoid multiple range procedures

False Discovery Rate Control

There are many other error rate definitions, but we will only mention one more here, one that has become very popular in recent years, and has seen much use in microarray analysis. Table 2.9 is adapted from Benjamini and Hochberg (1995).

Table 2.9. Number of errors committed when testing m hypotheses, where m_0 nulls are true

	Declared Significant	Declared Nonsignificant	Total
True Null Hypotheses	V	$m_0 - V$	m_0
False Null Hypotheses	S	$m - m_0 - S$	$m - m_0$
	R	$m - R$	m

Note that the only numbers that we know in Table 2.9 are m and R , but error rates can be obtained as probabilities and expected values.

$$E\left(\frac{V}{m}\right) = \text{Comparisonwise Error Rate}$$

$$P(V \geq 1) = \text{Familywise Error Rate}$$

and testing each hypothesis at α/m guarantees that $P(V \geq 1) \leq \alpha$, while testing each hypothesis at α guarantees that $E(V/m) \leq \alpha$.

Benjamini and Hochberg (1995) suggest that we should control another error rate

$$E\left(\frac{V}{R}\right) = \text{False Discovery Rate,}$$

which is the proportion of discoveries (rejections of H_0) that are false (H_0 is true). They note that if all the nulls are true, this is equivalent to the familywise error rate, but is a smaller number once $m_0 < m$.

A procedure for controlling the false discovery rate (FDR) at a set value q^* is the following:

- (1) For testing m hypotheses, obtain the p -values and order them $p_{(1)} < p_{(2)} < \dots < p_{(m)}$. (Note that $p_{(i)}$ is not the p -value for the i^{th} hypothesis.)
- (2) Let k be the largest i for which $p_{(i)} < (i/m)q^*$.
- (3) Reject all nulls with a p -value less than $p_{(k)}$.

Controlling the FDR has almost become a standard in microarray analysis, partially due to the popularity of a procedure known as SAM (Statistical Analysis of Microarrays, Tusher *et al.* 2001). The SAM procedure is essentially multiple t -tests processed using the FDR control.

There has been much other work done in estimating the FDR (Storey and Tibshirani 2003) and on other variations of FDR. A large literature on FDR is developing, and good entries are Storey (2002, 2003), Genovese and Wasserman (2002, 2004), and Genovese *et al.* (2006).

2.9.2 Application or Attribute?

The question has sometimes come up about whether a treatment that is *applied* is different from a treatment that is an *attribute*. For example, a treatment could be a certain level of nitrogen in a fertilizer that is applied to a plot of land, or a treatment could be an age group, which is an attribute. Statistically, is there a difference?

This discussion could be about the difference between designed experiments and observational studies, but it is not. We are here dealing with designed studies, but could include, as treatments, factors that are attributable to the experimental units rather than applied to the experimental units. A possible way to understand the difference is with the following model, a variation on the oneway CRD (2.2):

$$Y_{ij} = \mu + \tau_i + \delta_{ij} + \varepsilon_{ij},$$

where we add δ_{ij} as the error in applying treatment i to the j^{th} unit in that treatment group. The error δ_{ij} can have one of three forms

$$\delta_{ij} = \begin{cases} 0 & \text{if the treatment is an attribute} \\ \delta_i & \text{if the treatment is applied once to the group} \\ \delta_{ij} & \text{if the treatment is applied independently to each unit.} \end{cases}$$

This model says that there is no error in the application of the treatment if the treatment is an attribute. So, for example, if τ_1 is the application of a fertilizer with 10% nitrogen, and the batch is mixed with 12%, that error is reflected in δ . (This is in contrast to ε , which would reflect the different responses in yield to the same treatment of 10% nitrogen.)

If we assume that the errors are independent, then there is no discernable difference between the first and third cases, that is, whether the treatment is an attribute, or applied independently to each unit, results in the same analysis. The only difference arises is if $\delta_{ij} = \delta_i$, for then a correlation is introduced and the analysis cannot proceed as a oneway CRD. In fact, in a formal sense the analysis cannot proceed at all (Exercise 2.24).

2.9.3 Imbalance

In this book we are dealing almost exclusively with balanced designs. When looking at the experiment from the design viewpoint, it would be silly to design an unbalanced experiment, as too many things are given up. Dealing with imbalance in data is a topic that is better treated in a book devoted to analysis rather than design (see, for example, Rawlings *et al.* 1998).

A design is *balanced* if there are the same number of observations in each cell. Imbalance only becomes a serious problem in twoway and higher designs, as the oneway CRD can handle imbalance pretty well (although it wreaks havoc with contrasts, as we have seen in Chapter 1; see Exercise 1.18).

Imbalance is a fact of life. Rats or plants can die, microarray wells may not be correctly read, data may be lost. Sometimes we have the luxury of having extra observations in each cell to plan for this— which might be one of the only valid reasons for have within observations in an RCB. However, discarding data for any reason is typically a bad idea.⁶

For example, if every cell but one in a twoway design has 5 observations, and one has four observations, balance can be achieved by discarding, at random, one observation from each of the cells with 5 observations. Never, never do this. Never throw away data unless there is a really good reason to do so. Each data point has information, and discarding information is never good.

Data are sacred!

There is a fundamental difference between unbalanced data and missing data. We take the approach that our analysis should always be based on the likelihood for the data. In the balanced case this agrees with the anova-based analysis, but in the unbalanced case things differ.

- In the unbalanced (but not missing) case, we can write down a model and a likelihood, and estimate and test based on the likelihood. In fact, this is regression analysis! There are problems, however, with identifiability constraints and estimability, but these can be handled.
- In the case where there is missing data, so we have empty cells, things get a bit more complicated. If the parameter for the cell is in the model then, formally, the likelihood function is calculated by averaging over all possible values that could have been in that cell. This is not a simple calculation, but can be handled by modern missing data techniques such as the EM algorithm (see, for example, Little and Rubin (2002)). Alternatively, we can do an *observed case* analysis using the likelihood for the observed data, but this can sometimes be a difficult calculation, and may sometimes require discarding data.

The main problem caused by imbalance is that, under the usual models, effects are no longer orthogonal. For example, consider model (2.4) with unequal cell sizes where, for simplicity, we assume no interaction:

$$Y_{ijk} = \mu + \tau_i + \gamma_j + \varepsilon_{ijk}, \quad i = 1, \dots, t; \quad j = 1, \dots, g, \quad k = 1, \dots, r_{ij},$$

and the r_{ij} are not necessarily equal. If we denote the total number of observations by $n = \sum_{ij} r_{ij}$, and the marginal totals by $r_{i.} = \sum_j r_{ij}$ and $r_{.j} = \sum_i r_{ij}$,

⁶ If there is a nonstatistical reason to believe that a data point is an outlier, an analysis with and without it can show if it is unduly affecting the inference.

then

$$E(\bar{Y}_i - \bar{Y}) = \tau_i - \frac{1}{n} \sum_i r_i \tau_i + \frac{1}{r_i} \sum_j r_{ij} \gamma_j - \frac{1}{n} \sum r_{.j} \gamma_j.$$

The identifiability constraints $\sum_i \tau_i = \sum_j \gamma_j = 0$ do not do much here, and thus there are “pieces” of extraneous parameters in this expected value no matter what we do. This means that the sums of squares will not be orthogonal – they will not add to the total – and we must be more careful if we do an anova. Specifically, we must decide between *sequential* or *partial* sums of squares.

The data in Exercise 2.31 is an unbalanced twoway. We can produce two different anova tables depending on which factor we fit first:

Treatments Before Varieties			Varieties Before Treatments		
Source	df	SS	Source	df	SS
Treatments	4	4.253	Varieties	2	3.563
Varieties	2	3.004	Treatments	4	3.694
Residual	19	1.023	Residual	19	1.023

Since the sums of squares are not orthogonal, fitting one will account for some variation in the other, so the order matters. In each of these tables we see the sequential sums of squares, which add to the total. Note that in the second table, where varieties are fit first, the sum of squares for treatments is smaller.

The second line of each table gives the partial sum of squares for that factor and this is the sum of squares that should be used for evaluating the effect. For example, if we want to evaluate the effect of treatment, it should be done after the variation due to varieties is removed, resulting in $SS(\text{Treatments After Varieties}) = 3.694$. These partial sums of squares can be presented in one anova table, where the sums of squares will not add to the total sum of squares. Also note that in each table, since the residual is fit last, its sum of squares is the same.

So, in general, what to do? The nonorthogonality problem, which leads to problems with the identifiability constraints, does not go away even if we calculate least squares means. However, it should be clear that this problem grows worse if the imbalance grows worse. Thus, if the imbalance is minimal, we can forge ahead with partial sums of squares and, for the most part, ignore the imbalance in our effect estimates. If the imbalance is great, or if there are cells without data, then it is probably best to turn to a cell means model, which we first saw in (1.2), and here would be

$$Y_{ijk} = \theta_{ij} + \varepsilon_{ijk}, \quad i = 1, \dots, t; \quad j = 1, \dots, g, \quad k = 1, \dots, r_{ij},$$

where θ_{ij} is the mean of cell (i, j) , and $E\bar{Y}_{ij} = \theta_{ij}$. We can then estimate treatment effects with contrasts in the \bar{Y}_{ij} .