2^k Factorial Experiments

Basic set-up:

- Each factor takes two-levels $\{+, -\}$;
- There are k different factors: $A_1, A_2, \ldots A_k$;
- There are 2^k treatment combinations in total;
- Simplest set-up there is one replication for each combination, hence 2^k observations in total.
- Note that there are also 2^k subsets of these factors (including the empty set).

Contrasts

Associated with a given subset of the factors, say $S \subseteq \{A_1, A_2, \ldots A_k\}$, we associate a contrast:

$$C_S = \sum_{(A_1,\ldots,A_k):\prod_{i\in S} A_i=+} y_{(A_1,\ldots,A_k)} \quad - \sum_{(A_1,\ldots,A_k):\prod_{i\in S} A_i=-} y_{(A_1,\ldots,A_k)}$$

e.g. if $k = 3$, so we have three factors A, B, C then

$$\begin{aligned} C_{AB} \quad = \quad & (y_{A+B+C+} + y_{A+B+C-} + y_{A-B-C+} + y_{A-B-C-}) \\ & -(y_{A-B+C+} + y_{A-B+C-} + y_{A+B-C+} + y_{A+B-C-}) \end{aligned}$$

(In the general case where there are n replicates of each treatment combination, each term in this expression simply represents the sum of the responses in the particular treatment group.)
We can alternately express the associated terms algebraically in the following way:

- Let a_1 correspond to the treatment combination:
  $A_1 = +, A_2 = -, \ldots, A_k = -$

- Let a_j correspond to the treatment combination:
  $A_1 = -, \ldots, A_{j-1} = -, A_j = +, A_{j+1} = -, \ldots A_k = -$
- Similarly let $a_1 a_2$ correspond to the treatment combination:
  $A_1 = +, A_2 = +, A_3 = -, \ldots, A_k = -$
- and so on....
- Until $a_1 a_2 \cdots a_k$ corresponds to the combination $A_1 = +, A_2 = +, \ldots, A_k = +$ in which every factor is $+$.

Under this scheme we can work out the contrast associated with a given set S by simply expanding:

$$\prod_{i \in S}(a_i - 1) \prod_{j \notin S}(a_j + 1)$$

For example, returning to the previous example, we obtain:

$$(a-1)(b-1)(c+1) = (ab-a-b+1)(c+1) = (abc+ab+c+1)-(ac+bc+a+b)$$

Effects

The *effect* associated with the contrast C_S then is obtained by dividing by the contrast by 2^{k-1}. This can be seen as being motivated by the fact that there are this many pairs of differences of means in the contrast. Hence we have:

$$\text{Effect}_S = \frac{1}{n2^{k-1}} C_S$$

Note that the *effect* here is also a contrast.

This also leads to the sum of squares associated with a given effect. Recall that we associated a sum of squares with a contrast in the following way:

$$\begin{aligned} SS_S &= \frac{n\left(\sum_i c_i \bar{y}_{i\cdot}\right)^2}{\sum_i c_i^2} \\ &= \frac{\left(\sum_i c_i y_{i\cdot}\right)^2}{n\sum_i c_i^2} = \frac{\hat{C}_S^2}{n2^k} = n2^{k-2}(\text{Effect}_S)^2 \end{aligned}$$

Note that as expressed here, all of the c_i's take value ± 1, hence the term $\sum_i c_i^2$ evaluates to 2^k.

Visualizing contrasts in the 2^3 case

Notice that any subset S assigns $+$ or $-$ signs to each treatment combination depending upon whether that combination assigns $+$ or $-$ to the product of the factors in S.

This we may visualize the effects in the case of a 2^3 factorial experiment, by constructing a cube containing the eight treatment combinations as vertices, and then looking at the set of points containing the $+$ vertices or the $-$ vertices. Specifically we obtain the following:

- Main effects: $+$ assigned to one face of the cube, $-$ assigned to the other parallel face of the cube;
- Two-way interactions: $+$ assigned to one plane passing through four corners of the cube. Two of the corners are joined by an edge of the cube (corresponding to the factor which does not occur in the two-way interaction), the other corners are diametrically opposite across a face of the cube.
- Three-way interactions: $+$ assigned to four corners which are not joined by an edge; $-$ assigned to the remaining four corners, again, not joined by an edge.
- Also note that there is 1 grand mean, 3 main effects; 3 two-way interactions, and then 1 three-way interaction.

Principles in setting up a 2^k factorial design:

- It is important to always try to make sure that the two levels for each factor are quite far apart; otherwise it is possible that a given factor might have an effect, but simply might not produce a large effect over the two values which are being used in the experiment, in which case we might fail to detect the effect. ("spread out the factor levels aggressively".)

- Often the function of a 2^k design is to *screen* a large set of variables, so as to identify those which are significant.
- The desire to examine many effects often means that it is not possible to do multiple replications at each treatment level. If only one observation is made for each treatment combination then the experiment is referred to as an 'unreplicated' factorial experiment. This is referred to as *Design Projection*.
- If one or more factors turns out to be irrelevant, then we can 'finesse' these extra observations in order to obtain 'hidden' replications.

Analyzing an unreplicated factorial

Since we have only one data-point for each treatment level, we cannot initially obtain an estimate of the error-variance: there simply is no data with which to estimate this.
If some of the high-order effects are zero we could combine their mean-squares to estimate the error. This approach is thus based on the heuristic, which Montgomery calls the *sparsity of effects principle*:

> ... most systems are dominated by some of the main effects and low-order interactions, and most high-order interactions are negligible

This leaves open the question as to *which* effects should be pooled in order to obtain our estimate of the error variance (σ^2). One approach to this problem is to construct a qq-plot for the estimated interactions. Because, under the null hypothesis that the corresponding (true) effect is zero, the associated mean-square estimates σ^2, if all effects were zero we would expect to see approximately a normal distribution centered on σ^2. Conversely, if the plot shows some outliers, then these may be viewed as likely candidates for non-zero interactions.
(see Spring example).

Unless we are happy with the full model, we will need to undertake some form of model selection. Usually this proceeds by starting from the full model, eliminating the terms which appear unnecessary (based on the qq-plot and analysis of contribution to the total sum of squares). Subsequent model reduction may proceed by examining F-statistics, and computing p-values.
Having arrived at a final model, as usual we must check residuals to examine the goodness-of-fit.

Randomized complete block designs

Recall our first design: **CRD:** to assess effects of a single factor, say F1 , on response, randomly allocate levels of F1 to experimental units.
Typically, one hopes the experimental units are *homogeneous* or nearly so:

- scientifically: If the units are nearly homogeneous, then any observed variability in response can be attributed to variability in factor levels.
- statistically: If the units are nearly homogeneous, then MSE will be small, confidence intervals will be precise and hypothesis tests powerful.

But what if the experimental units are not homogeneous?

Example: Let F_2 be a factor that is a large source of variation/heterogeneity, but is **not recorded** (e.g. age of animals in experiment, gender, field plot conditions, soil conditions).

We then have the following ANOVA table:

Source	SS	MS	F-ratio
F1	SS1	$MS1 = SS1/(t1-1)$	$MS1/MSE$
$(F2 + Error)$	$SS2 + SSE$	$(SS2 + SSE)/(N - t1)$	

If $SS2$ is *large*, then the F-statistic for F1 may be *small*.
If a factor

1. affects response

2. varies across experimental units

then it will increase the variance in response and also the experimental error variance/MSE if unaccounted for.

Blocking:

Blocking is the stratification of experimental units into groups that are more homogeneous than the whole.

Objective: To have less variation among units within blocks than between blocks.

Typical blocking criteria:

- location
- physical characteristics
- time

Example: Nitrogen fertilizer timing: How does the timing of nitrogen additive effect affect nitrogen uptake?

- Treatment: Sixdifferent timing schedules $1, \ldots, 6$: Level 4is "standard"
- Response: Nitrogen uptake($ppm \times 10^{?2}$)
- Experimental material: One irrigated field

Soil moisture is thought to be a source of variation in the response.

Design:

1. Field is divided into a 4×6 grid.
2. Within each row or *block* each of the six treatments are randomly allocated.

- The experimental units are *blocked* into presumably more homogeneous groups. i.e. it is supposed that the soil moisture is more similar within a row than between rows.

- The blocks are *complete* in that each treatment appears in each block.
- The blocks are *balanced* in that there are
  - $t_1 = 6$ observations for each level of block
  - $t_2 = 4$ observations for each level of treatment
  - $r = 1$ observations for each treatment $\times$ block combination

Thus this design is called a *randomized complete block design.*

Analysis of the RCB design with one rep:

Analysis proceeds just as in the two-factor ANOVA:

$$\begin{aligned} y_{ij} - \bar{y}_{\cdot\cdot} &= (\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}) + (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}) + (y_{ij} - \bar{y}_{i\cdot} - \bar{y}_{\cdot j} + \bar{y}_{\cdot\cdot}) \\ SSTotal &= SSTrt + SSB + SSE \end{aligned}$$

ANOVA table

Source	SS	df	MS	F-ratio
Trt	SST	$t_1 - 1$	$SST/(t_1 - 1)$	MST/MSE
Block	SSB	$t_2 - 1$	$SSB/(t_2 - 1)$	(MSB/MSE)
Error	SSE	$(t_1 - 1)(t_2 - 1)$	$SSE/(t_1 - 1)(t_2 - 1)$	
Total	SSTotal	$t_1 t_2 - 1$		

ANOVA for nitrogen example:

Source	SS	df	MS	F-ratio	p-value
Trt	201.32	5	40.26	5.59	0.004
Block	197.00	3	65.67	9.12	
Error	108.01	15	7.20		
Total	506.33	23			

Consider the following three ANOVA tables:

```
#######
> anova(lm(c(y)~as.factor(c(trt)) ))
```

```
Df Sum Sq Mean Sq F value Pr(>F)
as.factor(c(trt)) 5 201.316 40.263 2.3761 0.08024 .
Residuals 18 305.012 16.945
#######
#######
> anova(lm(c(y)~as.factor(c(trt)) + as.factor(c(rw)) ))
Df Sum Sq Mean Sq F value Pr(>F)
as.factor(c(trt)) 5 201.316 40.263 5.5917 0.004191 **
as.factor(c(rw)) 3 197.004 65.668 9.1198 0.001116 **
Residuals 15 108.008 7.201
#######
#######
> anova(lm(c(y)~as.factor(c(trt)):as.factor(c(rw)) ))
Df Sum Sq Mean Sq F value Pr(>F)
as.factor(c(trt)):as.factor(c(rw)) 23 506.33 22.01
Residuals 0 0.00
#######
```

Can we test for interaction? Do we care about interaction in this case, or just main effects? Suppose it were true that "in row 2, timing 6 is significantly better than timing 4, but in row 3, treatment 3 is better." Is this relevant in for recommending a timing treatment for other fields?

Did blocking help? Consider CRD as an alternative:

block	1	2	4	3	2	1	4
block	2	5	5	3	4	1	4
block	3	6	3	4	2	6	5
block	4	1	2	6	2	5	6

- Advantages:
  - more possible treatment assignments, so power is increased in a randomization test.

  - If we don't estimate block effects, well have more dof for error.

- Disadvantages
  - Its possible, (but unlikely) that some treatment level will get assigned many times to a "good" row, leading to post-experimental bias.
  - If "row" is a big source of variation, then ignoring it may lead to an overly large MSE.

Consider comparing the F-statistic from a Completely Randomized Design (CRD) with that from a Randomized Complete Block design (RCB). According to Cochran and Cox (1957):

$$\begin{aligned} MSE_{\text{crd}} &= \frac{SSB + r(t-1)MSE_{\text{rcb}}}{rt-1} \\ &= MSB\left(\frac{r-1}{rt-1}\right) + MSE_{\text{rcb}}\left(\frac{r(t-1)}{rt-1}\right) \end{aligned}$$

In general the effectiveness of blocking is a function of $MSE_{\text{crd}}/MSE_{\text{rcb}}$. If this is large, it is worthwhile to block. For the nitrogen example this is about 2.