Stat 471/571: Additional notes on factorial treatment designs.

Interactions: sometimes useful to distinguish quantitative from qualitative interaction

 **Quantitative interactions**: simple effects not similar, but same sign

 Transformation of Y can change, perhaps eliminate, the interaction



 **Qualitative** **interactions**: simple effects have different signs

 Transformation will not eliminate interaction

 Main effects very hard to interpret: main effect of factor B (the lines) = 0

 simple effects of B at each level of A are definitely not zero.



**Why use a factorial design?**

 Short answer: more precise than alternatives and can answer more questions.

 Context: Food preference study.

 Want to know about difference between sexes and differences between types

 Can study total of 144 people.

 Factorial design and two alternatives:

 1) two “one factor” experiments:

 72 folks: 36 men 36 women, both tasting the solid form: se diff = 0.27 70 df for error

 72 folks: 24 in each type, all women: se diff = 0.33 69 df for error

 2) “L” design, “one factor at a time”.

 4 treatments, C, L, S with women, S with men, 36 per group se diff = 0.27 140 df for error

 3) factorial: 6 treatments, 24 per treatment sex se diff = 0.19 138 df for error

 Focusing on main effects type se diff = 0.24 “

 More precise because of hidden replication, effectively 72 of each sex and 48 of each type.

 Also learn about interactions and all simple effects (can’t do with one-factor or L designs)

History:

 L = one factor at a time design very common through 1920’s.

 Although the Rothansted fertilizer studies used factorial designs starting in the 1840’s and 1850’s

 RA Fisher strenuously advocated for factorial designs because of their improved precision.

"No aphorism is more frequently repeated in connection with field trials, than that we must ask Nature few questions, or, ideally, one question, at a time. The writer is convinced that this view is wholly mistaken." (Fisher, RA 1926, The arrangement of field experiments. J. Ministry of Ag of Great Britain, 33:503-513). The middle of p 504 is especially interesting because that’s where Fisher suggests the 5% standard for p-values. This paper was followed up by his Design of Experiments book published in 1935.

**3 Factors**:

Soybean yield when grown at 2 levels of ozone, 2 levels of SO2 and either well watered or stressed.

3 replicates of all 8 treatment combinations.

Effects model has 3 main effects, 3 two way interactions and 1 three way interaction.

ANOVA table:

 Df Sum Sq Mean Sq F value Pr(>F)

stress.f 1 4079 4079 0.0633 0.80

so2.f 1 557371 557371 8.6439 0.0096

o3.f 1 19145174 19145174 296.9094 < 0.0001

stress.f:so2.f 1 11507 11507 0.1785 0.68

stress.f:o3.f 1 130833 130833 2.0290 0.17

so2.f:o3.f 1 156178 156178 2.4221 0.14

stress.f:so2.f:o3.f 1 1194945 1194945 18.5316 0.0005

Residuals 16 1031705 64482

All effects are averages over all factors “not in” that term

 Stress: difference between well watered and stress, averaged over SO2 level, O3 level, and replicates

 O3: difference between 0.01 and 0.1 O3, averaged over SO2 level, stress, and replicates

 O3\*SO2: interaction between O3 and SO2 averaged over stress level and replicates.

 Is the effect of SO2, averaged over stress levels, the same in low and high O3 plants.

 Stress\*SO2\*O3: is the two way interaction of O3\*SO2 the same in well watered and stressed plants?



Could pick any other variable to split: is the stress\*O3 interaction the same for all SO2 levels?

Qualitative interaction, so can’t transform it away.

 Three options:

 1) split the data, e.g. by stress level. I would probably do this.

 2) report appropriate simple effects

 3) decide the 3 way interaction is practically small and ignore

Degrees of freedom: Na levels of factor A, Nb  levels of factor B, Nc  levels of factor C

 A Na -1 B Nb -1 C Nc -1

 A\*B (Na -1)( Nb -1) A\*C (Na -1)( Nc -1) B\*C (Nb -1)( Nc -1)

 A\*B\*C (Na -1)( Nb -1)(Nc -1)

If there are missing cells, they reduce the df for the three way interaction.

Missing cells are bad news – marginal means are not estimable if you fit all interactions.

Discussed in detail below

Quick check for missing cells: check that the highest interaction has the correct df

**Four factors**: Same ideas, just more terms.

 4 main effects, 6 two way interactions, 4 three way interactions, 1 four way interaction.

 How to interpret the 4 way interaction?

 Are the three way interactions the same across all levels of the fourth factor?

 Not easy to make sense of

 Usual “boots on the ground” interpretation:

 the patterns of treatment effects are really complex

**Other practical issues** (all of these are opinions, intermingled with facts)

Should you remove the interaction when not significant? Usual context: 2 way factorial

 My general advice: no. Here’s a simple version of some deep statistical issues

 1) Dropping the interaction means the analysis loses the connection to contrasts among cell means

 2) remember that interaction contrasts have the largest se’s, so low power

 Dropping the interaction implies that you are certain there is no interaction.

 Data only tell you no evidence of one, and that can be very uncertain and inconclusive

 3) when the design is unbalanced, dropping the interaction changes the main effect hypothesis

 When no interaction, the main effect hypothesis now involves the sample sizes,

 many view as a weird thing to do.

 This does make sense when you are absolutely sure there is no interaction.

 Then you have multiple estimates of the effect of A: each simple effect in levels of B.

 Some are more precise than others, because some have larger sample sizes than others do.

 When you’re sure they are estimates of the same quantity (i.e., assume no interaction),

 it makes sense to emphasize the more precise estimates

 but this means the main effect hypothesis is no long an average of simple effects

 4) Dropping the interaction pools the interaction SS and df with the within-group error SS and df

 A data based decision to pool (e.g. interaction is small) biases the error variance

 Systematically too small.

 Numerous studies have shown that “pre-testing” the interaction, and dropping it when p > 0.05

 biases tests of interest (e.g. of main effects). They do not have the stated type I error (e.g. 5%).

What about 3 and 4 way interactions?

 Rarely answer questions of interest

 Common to assume they don’t exist and drop them

 Especially for observational data

 Experimental studies:

 Probably shouldn’t drop interactions so you maintain the cell means connection

 One reasonable strategy to maintain the connection to cell means:

 Fit main effects and two-way interactions, then “all other interactions”

 A+B+C+D+AB + AC + AD + BC + BD + CD + ABCD

 ABCD includes all the 3 way interactions: answers “Is there anything else?”

 When there are missing cells (most common in observational studies),

 analysis much simpler when drop the high order interactions

**Role of missing cells**:

 Consider a 2 way factorial, e.g. food preference study with men and women

 But men only given two types of product:

 Sex C L S marginal mean

 Men data ---- data ???

 Women data data data estimate

 Data ??? data

 If you include the interaction in the model (different mean for each cell)

 No obvious estimate of the marginal means for men or liquid.

 SAS, JMP and R/emmeans reports not-estimable.

 Any number depends on the choice of constraint.

 Even when main effect test is reported, it is meaningless (SAS is a bad offender here)

 Test of interaction still interpretable, but limited by data availability

 Is Men-Women similar in C and S. No evaluation of L because no M-W effect there.

 When you drop the interaction, you assume men-women difference is the same for C, L, and S.

 The mean difference between men and women can be estimated from the data in C and S

 The mean difference between C and L can be estimated from just the Women’s data

 So all main effects can be estimated and tested, assuming no interaction.

 In observational data with many factors, common to have missing cells,

 One reason I avoid models with 3 and 4 way interactions here

**What about studies with “almost” a complete factorial?**

 Best illustrated by example:

 2 factors: amount of X (3 levels, 0, 10, 20) and time of application of X (2 levels, Spring, Fall)

 5 treatments: 0, 10 in Spring, 10 in Fall, 20 in Spring, 20 in Fall

 Note that time is irrelevant when nothing added

 I call these augmented designs: 2 x 2 factorial augmented with a control

 But augmented design is more commonly used for different design issue

 Is this:

|  |  |
| --- | --- |
|  | Amount |
| Time | 0 | 10 | 20 |
| Never | data | --- | --- |
| Spring | --- | data | data |
| Fall | --- | data | data |

 Serious missing cells issues!

 Or, split the 0 group into two artificial times?

|  |  |
| --- | --- |
|  | Amount |
| Time | 0 | 10 | 20 |
| Spring | data | data | data |
| Fall | data | data | data |

 If time has an effect, notice that this artificially creates an interaction:

 No difference in 0 amount, but a difference in 10 and 20.

Could drop the 0 amount => usual 2 way factorial, but lose comparisons to the control

Best solution: write contrasts among 5 treatments. Most likely set are:

 Control – average of all the rest

 The usual 2 x 2 contrasts within the four 10/20, Spring/Fall treatments.

 Or, do all differences from Control to something else

 Dunnett’s multiple comparisons adjustment (one to many)

**Explosion of treatments**

 The number of treatments in a factorial treatment design can get very large

 3 levels A x 4 levels of B x 2 levels of C = 24 treatments

 Add 5 levels of D = 120 treatments

 120 experimental units if you have no replication!

 Strategies to reduce the number of eu’s:

 Mostly used in industrial experiments where error variance small.

 Often many possible factors, usually consider 2 or 3 levels of each

 1) Pick a “central” treatment (middle of 3 levels) and only replicate that.

 2) Fractional factorial designs: only use a carefully chosen subset of combinations

 Assume some or all interactions are small / zero,

 focus on estimating main effects or main effects + 2 way interactions

**Fractional factorial treatment design**, simple example: 3 factors, each 2 levels (+/-)

 Full factorial has 8 treatments, want 2 replicates, no more than 8 runs, so can only use 4 treatments.

|  |  |  |
| --- | --- | --- |
| A | B | C |
| - | - | + |
| - | + | - |
| + | - | - |
| + | + | + |

|  |  |  |
| --- | --- | --- |
| A | B | C |
| - | - | - |
| + | + | - |
| + | - | + |
| - | + | + |

 Fractional factorial design:

 Notice each trt occurs twice Omit these 4

 More precise estimates of treatments:

 main effects

 “half fraction of 2x2x2

 Design”

 Key assumption in this small example is no 2 way or 3 way interactions

 Other, larger, fractional factorial designs allow 2 way interactions but assume no higher interactions.

 Huge area of statistical research. Lots of published designs and software to produce them.

 My experience with fractional factorial designs: not good.

 Caveat: my experience is in ecological studies

 Grad student wanted to study 4 factors, many levels of each. Too many treatments

 I asked the student about 3 or 4 way interactions. Student said no, won’t ever happen.

 So I provided a fractional factorial design with 1/3 the number of treatments: practical.

 Reality was otherwise. Strong 3 and 4 way interaction, for good ecological reasons.

 Stats MS degree to figure out how to rescue the study

**Final thoughts:**

Treatment structure + experimental design.

 Can combine the two in all sorts of ways

 e.g. 3 way factorial in blocks

 Variation on a fractional factorial:

 Use all treatment combinations but confound blocks and high order interaction.

 e.g. one block has the first sets of 4 treatments above; the 2nd block has the other set