

# Factorial Treatment Structure

Gary W. Oehlert

School of Statistics  
University of Minnesota

September 22, 2014

## Definition and notation

As you recall a factor is an aspect of our treatment.

For example, my children used to take apple slices in their lunches. We didn't want the apples to brown, and various suggestions were made about keeping the slices cold and/or lowering the surface pH.

I could experiment to test browning with four treatments: room temperature, no lemon juice; room temperature, lemon juice rinse; cool, no lemon juice; cool, lemon juice rinse.

I have four treatments, but these treatments are formed by combining two factors (temperature and lemon juice) each at two levels.

Factorial treatment structure is simply the case where treatments are created by combining factors.

These could be Nisin and Vitamin E factors in potential antimicrobials; water/rice ratio and cooking time in steamed rice sensory properties; or intake temperature, intake pressure, injection pressure, and injection timing in fuel efficiency of diesel engines.

In each case, the treatments are the combinations of factor levels.

We will often refer to the factors generically as A, B, C, and so on. Factor A has  $a$  levels; factor B has  $b$  levels; and so on.

With two factors, there are  $g = ab$  treatments; with four factors, it is  $g = abcd$  treatments.<sup>1</sup>

We are still using a completely randomized design with  $N$  units applied to  $g$  treatments, and if you want to, you can ignore the factorial nature when you analyze.

I will argue that you should never ignore the factorial nature when you design.

---

<sup>1</sup>I once helped on an experiment with 24 factors—think about that for a moment.

Consider a two-factor design. We have data

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}$$

where  $i = 1, \dots, a$ ;  $j = 1, \dots, b$ ; and  $k = 1, \dots, n$ .

Here  $i, j$  together index the treatment by factor levels, and  $k$  indexes the replication within each treatment.

Notice that I wrote  $n$  and not  $n_{ij}$ . We begin with the case of balanced designs where every treatment has the same number of replications. Life is much easier under this assumption; we'll get to hard realities later.

For concreteness, consider a two-factor design with  $a=4$ ,  $b=3$ , and  $g=12$ . You can visualize this as a table of treatments:

11	12	13
21	22	23
31	32	33
41	42	43

We can do ANOVA on these 12 treatment groups, we can do pairwise comparisons, we can do contrasts (with coefficients  $w_{ij}$  summing to 0), and so on. Everything we've done to date just carries through.


However, the factorial treatment structure lets us organize things in another way.

For example, ignore B, and just do an ANOVA on the four groups determined by A. This will have 3 df, and let  $SS_A$  denote the between levels of A sum of squares.

This looks at variation between levels of A that is consistent (constant) across levels of B.

Denote the row means by  $\mu + \alpha_i$  with  $\sum \alpha_i = 0$ .<sup>2</sup>  $\alpha_i$  is called the main effect of level i of factor A.

---

<sup>2</sup>Other restrictions can be used, but I like this one. 

A somewhat less obvious way of exploring the same variation is to look at contrasts in the g groups of the form:

$w_{11}$	$w_{11}$	$w_{11}$
$w_{21}$	$w_{21}$	$w_{21}$
$w_{31}$	$w_{31}$	$w_{31}$
$w_{41}$	$w_{41}$	$w_{41}$

The coefficients all add to zero (as is normal for a contrast), but they are constant across levels of B. If we choose three orthogonal contrasts (e.g.,  $(-3,1,1,1)$ ,  $(0,-2,1,1)$ ,  $(0,0,-1,1)$ ), then the SS for those contrasts will add up to  $SS_A$ .



Now do the same thing for B, ignoring the levels of A. We will have 3 groups determined by B, with 2 df between them, and  $SS_B$  as the sum of squares.

This is variation between levels of B that is consistent (constant) across levels of A.

Denote the column means by  $\mu + \beta_j$  with  $\sum \beta_j = 0$ .  $\beta_j$  is called the main effect of level j of factor B.

As with A, look at contrasts in the g groups of the form:

$w_{11}$	$w_{12}$	$w_{13}$
$w_{11}$	$w_{12}$	$w_{13}$
$w_{11}$	$w_{12}$	$w_{13}$
$w_{11}$	$w_{12}$	$w_{13}$

The coefficients all add to zero (as is normal for a contrast), but they are constant across levels of A. If we choose three orthogonal contrasts (e.g.,  $(-1,0,1)$ ,  $(1,-2,1)$ ), then the SS for those contrasts will add up to  $SS_B$ .

There are 11 df between the 12 groups. We have 3 for factor A and 2 for factor B. Looking at the contrasts shows that they are orthogonal, so they are partitioning independent variation.

However,  $3 + 2 = 5 < 11$ ; there are 6 more df out there somewhere.

This variation is called the interaction variation, and it can only be seen if we vary both the levels of A and B simultaneously.

Interaction describes how the effect of changing levels of factor A varies by levels of factor B. Equivalently, it describes how the effect of changing levels of factor B varies with levels of factor A.

In general, factor A has  $a-1$  df; factor B has  $b-1$  df; the AB interaction has  $(a-1)(b-1)$  df. Moreover, there are contrasts

Coefs	Term	df
$w_i$	A	$a-1$
$w_j$	B	$b-1$
$w_{ij}$	AB	$(a-1)(b-1)$

The interaction contrasts  $w_{ij}$  satisfy  $\sum_i w_{ij} = \sum_j w_{ij} = 0$ .

There are  $(a-1)(b-1)$  contrasts (6 in our example) in a set of orthogonal contrast coefficients that describe the interaction variation. A main effect contrasts, B main effect contrasts, and AB interaction contrasts are all orthogonal to each other.

Recapping, factorial analysis with main effects and interactions is an option for analyzing factorial designs.

Factorial analysis is not required, but it often provides insights, particularly in cases where main effects are large and interactions are small.

A model that includes only main effects but not interaction is called additive.

An example where factorial analysis does more harm than good is looking at a reaction where three separate catalysts must be present for the reaction to go forward and produce the product.

We can do a three factor design with eight treatments (combinations of each catalyst present or absent). Given the above, we only get product when all three catalysts are present. We learn that fact in the factorial experiment.

However, the main effects and interaction decomposition will just disguise that simple result.

There are only two situations in which you should use factorial design, i.e., factorial treatment structure:

- 1 When interaction is present.
- 2 When interaction is absent.

If interaction is present, a factorial will allow you to study, estimate, and test it.

When interaction is absent, a factorial is more efficient than two designs that study A and B separately. (In the factorial, each data point tells you about A and about B.)

## Remember

Our definition of interaction is based on statistical modeling, not on scientific principles or considerations. In particular,

The need (or not) for interaction in modeling a data set depends on the scale of the response.

This means that a data set could look interactive on the natural scale, but look additive after transformation, or vice versa.

Interaction is always in the context of scale. Interaction as we use it is a modeling concept, not a scientific concept.



## Interaction plots

ANOVA will allow us to determine whether interaction terms in our model are statistically significant, but it won't help us understand the interaction.

A tool I like for understanding interaction is the interaction plot.

This is a plot where we put points at  $(i, \bar{y}_{ij\bullet})$  for all  $i, j$  combinations. Then we “connect-the-dots” between adjacent  $(i, \bar{y}_{ij\bullet})$  that share the same  $j$ .

Alternatively, you can reverse the roles of  $i$  and  $j$  and plot the pairs  $(j, \bar{y}_{ij\bullet})$  and then connect the dots between adjacent points that share the same  $i$ .

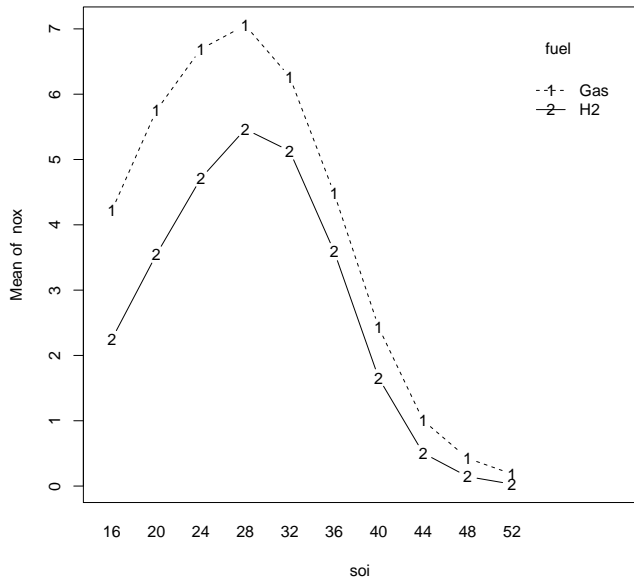
The two directions of plotting often give very different impressions.

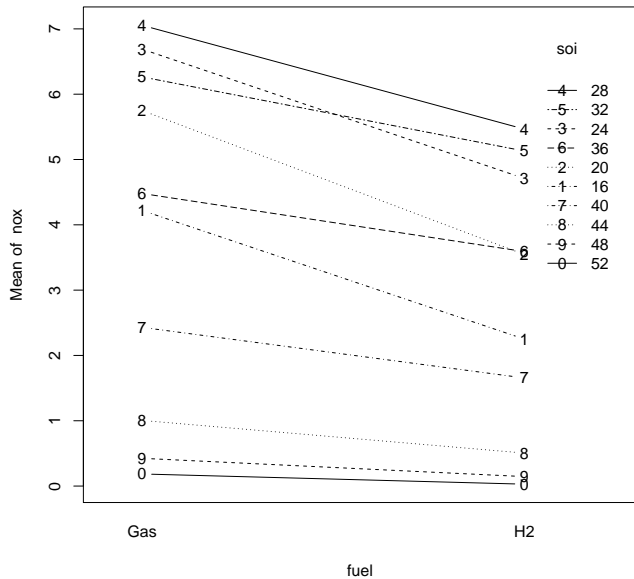
NOX emissions from dual fuel engine (diesel and something, either gasoline or hydrogen) with the start of injection (SOI) varied between 16 and 52 degrees before top of dead center.

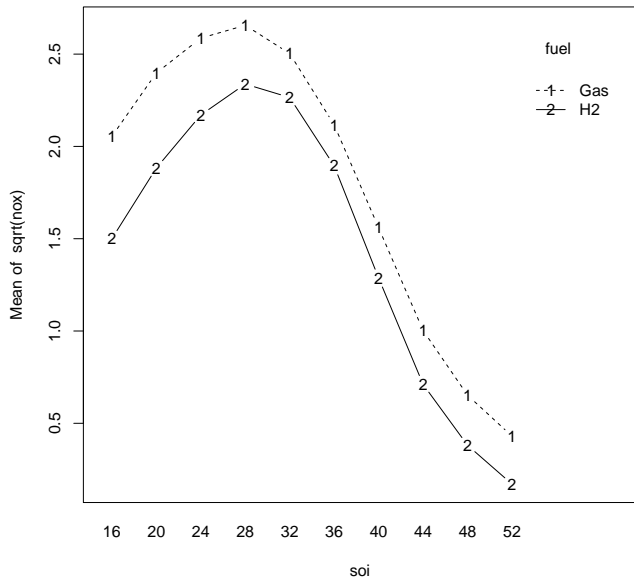
Both plots show us a much greater effect of fuel at low SOI, although I like the first one better.

The third plot uses the square root of NOX, which is what is needed to stabilize the residual variance.

Variability in these data is tiny relative to mean differences; everything is highly significant.







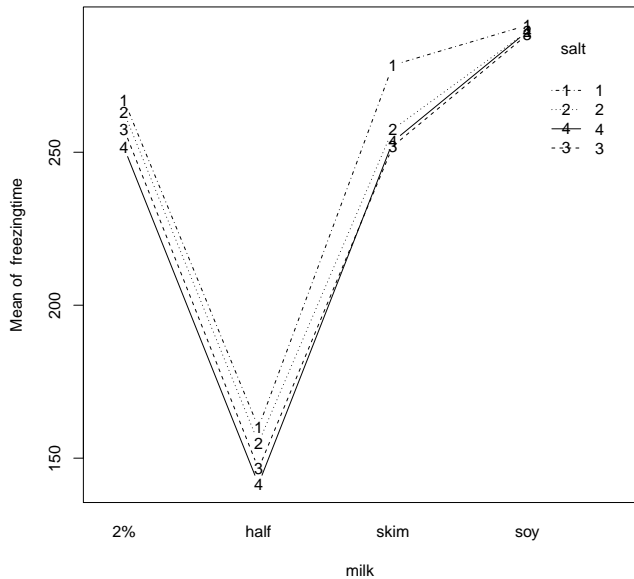
Freezing time of small samples of ice cream mix using different kinds of milk and different amounts of salt in the ice slurry.

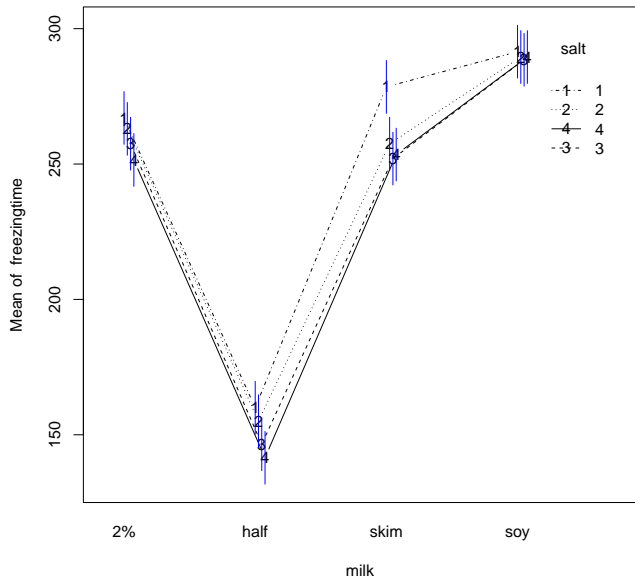
Both main effects are significant, but interaction is not.

Response: freezingtime

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
milk	3	89966	29988.6	710.3151	< 2.2e-16	***
salt	3	1104	368.1	8.7192	0.001169	**
milk:salt	9	501	55.6	1.3173	0.301901	
Residuals	16	675	42.2			

(This is a student project I can really appreciate!)







# Models and parameters

We begin with a two-factor model and then generalize.

When we just thought about  $g$  treatments and used an overall mean plus treatment effects, we had an extra parameter. We solved that via a constraint (e.g.,  $\alpha_1 = 0$  or  $\sum_i \alpha_i = 0$ ).

The standard factorial model has an overall mean, main effects for each factor and interaction effects. We will have a lot of extra parameters and need quite a few constraints to settle things down.

Overall mean, main effects, interaction effects

$$y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$$

$$i = 1, 2, \dots, a; \quad j = 1, 2, \dots, b; \quad k = 1, 2, \dots, n$$

$$\sum_{i=1}^a \alpha_i = \sum_{j=1}^b \beta_j = \sum_{i=1}^a \alpha\beta_{ij} = \sum_{j=1}^b \alpha\beta_{ij} = 0$$

$$g = ab; \quad N = nab = Ng$$

$$df_A = a - 1; \quad df_B = b - 1; \quad df_{AB} = (a - 1)(b - 1)$$

$$df_E = ab(n - 1) = N - ab = N - g$$

Note that  $(a - 1) + (b - 1) + (a - 1)(b - 1) = ab - 1 = g - 1$ .

$\mu$	$\beta_1$	$\beta_2$	$\beta_3$
$\alpha_1$	$\alpha\beta_{11}$	$\alpha\beta_{12}$	$\alpha\beta_{13}$
$\alpha_2$	$\alpha\beta_{21}$	$\alpha\beta_{22}$	$\alpha\beta_{23}$
$\alpha_3$	$\alpha\beta_{31}$	$\alpha\beta_{32}$	$\alpha\beta_{33}$
$\alpha_4$	$\alpha\beta_{41}$	$\alpha\beta_{42}$	$\alpha\beta_{43}$

Any particular treatment mean in the table is the interaction effect, plus the effect for the row, plus the effect for the column, plus the overall effect.

Note:  $\alpha\beta_{ij}$  does not mean  $\alpha_i \times \beta_j$ .

## Decomposing the data

$$\begin{aligned} y_{ijk} &= \bar{y}_{\bullet\bullet\bullet} + && \hat{\mu} \\ &(\bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet\bullet\bullet}) + && \hat{\alpha}_i \\ &(\bar{y}_{\bullet j\bullet} - \bar{y}_{\bullet\bullet\bullet}) + && \hat{\beta}_j \\ &(\bar{y}_{ij\bullet} - \bar{y}_{i\bullet\bullet} - \bar{y}_{\bullet j\bullet} + \bar{y}_{\bullet\bullet\bullet}) + && \hat{\alpha}\hat{\beta}_{ij} \\ &(y_{ijk} - \bar{y}_{ij\bullet}) && r_{ijk} \end{aligned}$$

$$\begin{aligned} y_{ijk} &= \bar{y}_{\bullet\bullet\bullet} + && \hat{\mu} \\ &(\bar{y}_{i\bullet\bullet} - \hat{\mu}) + && \hat{\alpha}_i \\ &(\bar{y}_{\bullet j\bullet} - \hat{\mu}) + && \hat{\beta}_j \\ &(\bar{y}_{ij\bullet} - [\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j]) + && \hat{\alpha}\hat{\beta}_{ij} \\ &(y_{ijk} - [\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\hat{\beta}_{ij}]) && r_{ijk} \end{aligned}$$

For balanced data, the SS decomposition is also easy.

$$\begin{aligned}\sum_{ijk} y_{ijk}^2 &= \sum_{ijk} (\hat{\mu} + \hat{\alpha}_i + \hat{\beta}_j + \hat{\alpha}\hat{\beta}_{ij} + r_{ijk})^2 \\ &= \sum_{ijk} \hat{\mu}^2 + \sum_{ijk} \hat{\alpha}_i^2 + \sum_{ijk} \hat{\beta}_j^2 + \sum_{ijk} \hat{\alpha}\hat{\beta}_{ij}^2 + \sum_{ijk} r_{ijk}^2 \\ &= N\hat{\mu}^2 + \sum_i nb \hat{\alpha}_i^2 + \sum_j na \hat{\beta}_j^2 + \sum_{ij} n \hat{\alpha}\hat{\beta}_{ij}^2 + \sum_{ijk} r_{ijk}^2 \\ &= SS_{Const} + SS_A + SS_B + SS_{AB} + SS_E\end{aligned}$$

Balance lets us go from the first line of the decomposition to the second, because all the cross products add to 0 in the balanced case. Without balance, life is much harder.

$SS_{Const}$  is usually ignored.

# ANOVA

Source	df	SS	MS	F
A	a-1	$\sum_i nb \widehat{\alpha}_i^2$	$SS_A/(a-1)$	$MS_A/MS_E$
B	b-1	$\sum_j na \widehat{\beta}_j^2$	$SS_B/(b-1)$	$MS_B/MS_E$
AB	(a-1)(b-1)	$\sum_{ij} n \widehat{\alpha}\widehat{\beta}_{ij}^2$	$SS_{AB}/[(a-1)(b-1)]$	$MS_{AB}/MS_E$
Error	N-ab	$\sum_{ijk} r_{ijk}^2$	$SS_E/(N-ab)$	

If  $H_0 : \alpha_j \equiv 0$  is true,  $MS_A/MS_E$  is F with (a-1) and N-ab df.

If  $H_0 : \beta_j \equiv 0$  is true,  $MS_B/MS_E$  is F with (b-1) and N-ab df.

If  $H_0 : \alpha\beta_{ij} \equiv 0$  is true,  $MS_{AB}/MS_E$  is F with (a-1)(b-1) and N-ab df.

Reject for big F, but still need to check assumptions.

The SS for various terms can also be considered as “improvement” SS for model fit. With the model  $A + B + AB$ :

- The SS for A is the improvement in adding A main effects to a constant mean model.
- The SS for B is the improvement in going from a model with just A to an additive model with A and B.
- The SS for AB is the improvement in going from an additive model to the full model with  $g$  treatment means.

# General factorials

Factorials with more than two factors are just like factorials with two factors, only more so:

- More factors, so more subscripts.
- More factors usually means more data.
- More terms; each additional factor doubles the number of terms in the model.
- More zero sum (or other) constraints on coefficients.
- More confusion with higher order interactions.

and no doubt more where those came from.

But the ideas are just like for two-factor factorials.



## Example: four factor model

$$\begin{aligned} Y_{ijklm} = & \mu + \\ & \alpha_i + \beta_j + \gamma_k + \delta_l + \\ & \alpha\beta_{ij} + \alpha\gamma_{ik} + \alpha\delta_{il} + \beta\gamma_{jk} + \beta\delta_{jl} + \gamma\delta_{kl} + \\ & \alpha\beta\gamma_{ijk} + \alpha\beta\delta_{ijl} + \alpha\gamma\delta_{ikl} + \beta\gamma\delta_{jkl} + \\ & \alpha\beta\gamma\delta_{ijkl} + \\ & \epsilon_{ijklm} \end{aligned}$$

$\beta\gamma_{jk}$ : how B effects change across levels of C (or vice versa).

$\beta\gamma\delta_{jkl}$ : how the BC interaction changes across levels of D (or some other version of two in and one out).

$\alpha\beta\gamma\delta_{ijkl}$ : how the BCD interaction changes across levels of A (or some other version of three in and one out).

These terms add to zero across any subscript (32 total zero sum constraints).

These terms are estimated by taking the mean in the data for the corresponding subscripts and subtracting out estimates of any “lower order” terms. So for  $\widehat{\alpha\beta\gamma}_{ijk}$  you would use

$$\widehat{\alpha\beta\gamma}_{ijk} = \bar{y}_{ijk\bullet\bullet} - [\widehat{\mu} + \widehat{\alpha}_i + \widehat{\beta}_j + \widehat{\gamma}_k + \widehat{\alpha\beta}_{ij} + \widehat{\alpha\gamma}_{ik} + \widehat{\beta\gamma}_{jk}]$$

SS are estimated effect squared, times number of units receiving the effect, added over levels.  $SS_{BCD} = \sum_{jkl} na \widehat{\beta\gamma\delta}_{jkl}^2$ .

DF are the product of the levels of factors appearing in the term, each reduced by 1. BCD has  $(b-1)(c-1)(d-1)$  df.

ANOVA has usual columns with MS as SS over DF and F tests for each term as the MS for the term over MS for error. (And you need to be careful in your naming once you get to five factors!)

To test the null hypothesis that all parameters of a given term are zero, compute the p-value for the F statistic from the F distribution with corresponding df.

Check assumptions as usual. Remember that interaction depends on scale.

Be glad you are children of a younger generation with easy access to computers and statistical software.

## Single replications

If  $n=1$ , then we have no df for estimating (pure) error. We need some surrogate estimate of error to do inference.

The usual approach is to leave one or more high-order interaction terms out of the model. Their SS and df then show up as “residual” variation with a corresponding MS. If the effects for the neglected terms are small, then this surrogate estimate of error will work pretty well.

The expected value for a mean square is  $\sigma^2$  plus something depending on the effects for the term. This makes our “MSE” potentially too big, leading to conservative tests (reject too rarely).

Choose the terms to leave out before looking at the data. If you cherry pick just the small MS to put accumulate into “error,” then your surrogate MSE will tend to be too small, leading to too many rejections.

A half-way version of dropping interaction terms out is to a parametric model of interaction with fewer df than the full  $(a-1)(b-1)(c-1)$  or whatever is in the term you would remove. This lets you keep the “big” part of the interaction in the model but allow the random/residual part of the interaction to serve as error. More on this much later.

A less common approach is to use some external estimate of error, perhaps from previous experiments or from the literature. This is a risky, risky approach, as you are taking on faith that what you are doing, how you are doing it, and what you are doing it with are just like these other guys. Don't forget to buy a lottery ticket.

# Hierarchy

A hierarchical model (in our sense) is one where the presence of a term implies the presence of all included terms. For example, presence of ABC in the model would imply the presence of the overall mean, A, B, C, AB, AC, and BC.

I always use hierarchical models unless I am very, very sure that it makes sense to use a non-hierarchical model.

The issue is parameterizations.

Remember that the parameterization of the means in a factorial is not unique. Here is a set of data with two different, equally valid, decompositions into constant effect, row effect, column effect, and interaction effect.

10	20	21.33	-1.33	1.33	23.33	0.0	0.0
20	24	-6.33	-3.67	3.67	-8.33	-5.0	5.0
30	24	0.67	-.67	.67	-1.33	-2.0	2.0
		5.67	4.33	-4.33	3.67	3.0	-3.0

Do these data have zero column effects or not?

Tests to remove a lower-order term (i.e., to consider a non-hierarchical model) are tests of parameters and depend delicately on the parameterization you choose.

If you only consider and compare hierarchical models, then you do not have to worry about the parameterization chosen.

The standard model and ANOVA tests assume that the correct parameterization is the one that derives from looking at simple averages wherever we take averages (rather than weighted averages, which lead to other parameterizations).

If that is really and truly the correct weighting, then making non-hierarchical tests is valid. If things are otherwise, or if you are not sure, stick to hierarchical models.

In the second breakdown above, rows were weighted 1, 2, 3.



# Pooling

Pooling (into error) refers to the practice of removing non-significant terms from the model, thus “pooling” them into the error or residual.

I discourage this practice in general and believe that there are only a couple of situations where pooling is advantageous.

The principal problem with pooling is that you risk biasing your  $MS_E$  upward.

The first situation where pooling makes sense is when you have very few df for error. In this case, the gain in error df often makes up for the risk in biasing the  $MSE$ .

Consider pooling a term if

- 1 You have few df for error, say less than 10 or 12.
- 2 The term has a low F-ratio,  $F < 2$ .
- 3 You are still maintaining hierarchy.

We have been talking about balanced designs. The second situation when you should use pooling is when you have an unbalanced design and you wish to examine or use the coefficients. You should

- 1 Select terms to retain in the model using  $MS_E$  from the full model.
- 2 Pool unselected terms into error before examining coefficients of selected terms.

The issue is that coefficients for terms in unbalanced models can depend on what other terms are in the model.

# Unbalanced factorials

In every life, a little rain must fall. — Longfellow

The first time you deal with unbalanced factorials, they seem like a downpour.

When factorial data are not balanced (i.e., when we do not have  $n_{ijk} \equiv n$ ), then

- 1 Row and column contrasts are not orthogonal.
- 2 The distinction between choosing a hierarchical model and testing hypotheses about parameters becomes more than philosophical.
- 3 Because of 2, there are multiple ways of building ANOVA tables and test statistics, each appropriate for their own purposes.

ANOVA still provides a comparison between a smaller model and an including, larger model.

Usually these models differ by a single term, and that is how we compute the SS for the term. Our two models are a base model and the base model plus the term of interest. This gives us an SS for our term of interest, adjusted for the terms in the base model.

So the question boils down to this: in order to test a term  $X$ , what terms should be in the base model?

Most software can provide a “sequential” ANOVA; this is often the default. In SAS, this is called Type I.

Suppose you have the following model:

$$y \sim 1 + A + B + C + AC + BC + AB$$

Term	Terms in base model
A	1
B	1, A
C	1, A, B
AC	1, A, B, C
BC	1, A, B, C, AC
AB	1, A, B, C, AC, BC

In sequential, each term is adjusted for those that precede it in the model.

Sequential is simplest from a programming perspective, and that is likely why it is the default.

Sequential has the advantage that the SS for the individual terms will add up to the SS for the model as a whole.

However, with unbalanced data you will get a different decomposition for terms in a different order. For example, B adjusted for A is likely not the same as B adjusted for A and C.

In the context of unbalanced factorials, building a model means selecting the main effects and interactions required to describe the mean structure adequately.

Building models does not depend on the parameterization.

Building models is done by comparing hierarchical models.

For any term  $X$ , choose as base model the largest hierarchical model that does not include  $X$ .



This choice of base model is called Type II in SAS. It has also been called “the method of Yates’ fitting constants,” but Type II is certainly quicker.

Suppose you have the following model:

$$y \sim 1 + A + B + C + AC + BC + AB + ABC$$

Then the Type II base models will be:

Term	Terms in base model
A	1, B, C, BC
B	1, A, C, AC
C	1, A, B, AB
AC	1, A, B, C, AB, BC
BC	1, A, B, C, AB, AC
AB	1, A, B, C, AC, BC
ABC	1, A, B, C, AB, AC, BC

It is possible to figure out what hypotheses about treatment means are being tested in type II (and it is hypotheses and contrasts between treatment means, not something about parameters).

However, these are technical and tedious, and it is easier to think of this as “Does adding this term to my base model really improve the fit?”

Seriously? You really want to see the hypotheses?

Consider a two-factor design. We have

weighted row means  $\mu_{i\star} = \sum_j n_{ij}\mu_{ij} / \sum_j n_{ij}$

weighted column means  $\mu_{\star j} = \sum_i n_{ij}\mu_{ij} / \sum_i n_{ij}$

Now form weighted averages for columns, but use the row weighted averages for cell means instead of the real cell means:

$$(\mu_{i\star})_{\star j} = \sum_i n_{ij}\mu_{i\star} / \sum_i n_{ij}$$

The Type II hypothesis for B adjusted for A is

$$\mu_{\star j} = (\mu_{i\star})_{\star j} \quad \text{for all } j$$

That is, the only reason column averages differ is because they are different weighted averages of row weighted averages.

We will look at one more approach called “standard parametric.” SAS calls this (wait for it . . .) Type III.

The tests in Type III have great allure, for example,  $H_0 : \alpha_j \equiv 0$ .

Testing in this fashion depends on the parameterization.

Specifically, this works for the basic equally-weighted-effects-add-to-zero parameterization. If you really believe that parameterization, then Type III is a possibility.

For standard parametric, the base model for term X is all terms except term X. This base model will usually be a non-hierarchical model.

You can construct Type II SS from Type I SS for several models in several orders. In R, you cannot construct Type III from various Type I analyses.

In R, the term  $A:B$  is going to force the modeling of the means for all A and B combinations. If you already have A and B in the model, then this will be the AB interaction. If A and/or B are not in the model (ahead of  $A:B$ ), then whatever was left out will get added in with the pure interaction.

This in itself is a not-so-subtle commentary on non-hierarchical models.

Sums of squares for contrasts are always Type III.

If your contrast is purpose-built for all  $g$  treatments, that is fine.

If your contrast is for row effects, or column effects, or some other set of parameters, bear in mind that this implies equal weighting across the factors not occurring in the contrast. This might, or might not, be appropriate.

For balanced data, Types I, II, and III are all the same.

For balanced data, contrasts coming from different standard terms in the model are orthogonal, so order does not matter.

The lack of balance breaks orthogonality, which is the root of the problem.

# Empty cells

Things really go pear shaped when one or more of the treatment counts  $n_{ijk}$  is 0.

Even if you have chosen a parameterization and weighting scheme, empty cells mean that your parameters are not well-defined.

Type I and Type II analyses are well-defined, but the typical Type III analysis is hopeless.

Your best bet is to choose meaningful contrasts among the cells where you do have data and work with them.