Stat 471/571: Factorial ANOVA, part 2

Reminder: more than 1 “factor”. Food preference: sex and type of product
6 treatments (all combinations of 3 types, 2 sexes).

	Type			
Sex	old	liquid	solid	average
F	0.24	1.12	1.04	0.80
M	0.20	1.24	1.08	0.84
average	0.22	1.18	1.06	0.82

Goal: the “usual” Two-way factorial ANOVA table:

Source	d.f.	SS	MS	F	p-value
Sex	1	0.06	0.06	0.04	0.83
Type	2	27.36	13.68	10.12	< 0.0001
Sex*type	2	0.16	0.08	0.06	0.94
Error	144	194.56	1.35		
c.Total	149				

Cell means model:

$$Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$$

Means, μ_i, for each treatment.
Answer questions via contrasts
Not the “usual” model for a factorial ANOVA

Effects model for 2 way anova:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk}$$

Focus on the parameters defining the means for each group: μ, α_i, β_j, and γ_{ij}.
μ: “something” common to all observations
1 parameter
α_i: “something” common to all observations in row i
2 parameters for 2 rows
β_j: “something” common to all observations in column j
3 parameters for 3 columns
γ_{ij}: “something” common to all observations in cell i, j
6 parameters for 6 cells

Serious problem:
12 parameters to be estimated from 6 cell means.

Same problem in a one-way ANOVA

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$$

7 parameters (1 μ, 6 α's) For 6 group means

Same problem in a 2 way ANOVA, but way more extreme

too many parameters!

Solution:

impose constraints on the parameters (e.g. force some parameters to be 0 or to sum to 0).

to get a unique solution, need # constraints = # parameters - # means

With 7 parameters (1 way ANOVA) and 6 means, need 1 constraint

With 12 parameters, 6 means, need 6 constraints

can estimate remaining 6 parameters.

Choice of constraint is arbitrary.

Choice changes the estimates of the remaining parameters!

Illustration with made up data with three sets of constraints:

μ	α_M	α_W	β_C	β_S	β_L	γ_{MC}	γ_{ML}	γ_{MS}	γ_{WC}	γ_{WS}	γ_{WL}
1	2	**0**	-1	2	**0**	1	2	**0**	**0**	**0**	**0**
2	**0**	-2	**0**	3	1	1	2	**0**	**0**	**0**	**0**
2.33	**1**	**-1**	**-1.33**	**1.66**	**-0.33**	1	2	**0**	**0**	**0**	**0**

All three sets of parameters fit the data equally well!

My reaction when I first saw this:

What! One data set, many different answers.

That's too confusing! Which one is right?

They're all right, but you're not interested in specific values for these parameters.

Interested in things like treatment means, marginal means, simple effects, and main effects.

Statistical theory:

the things you're really interested in are the same values

no matter which set of constraints you use.

Estimable function: a quantity that does not depend on the arbitrary choice of constraint.

Examples:

$\mu_{ML} = \mu + \alpha_M + \beta_L + \gamma_{ML}$

$\mu_{ML} - \mu_{WL} = (\mu + \alpha_M + \beta_L + \gamma_{ML}) - (\mu + \alpha_W + \beta_L + \gamma_{WL}) = (\alpha_M - \alpha_W) + (\gamma_{ML} - \gamma_{WL})$

$\mu_{M.} - \mu_{W.} = \frac{1}{3}\Sigma_{C,L,S}(\mu + \alpha_M + \beta_i + \gamma_{Mi}) - \frac{1}{3}\Sigma_{C,L,S}(\mu + \alpha_W + \beta_i + \gamma_{Wi}) = (\alpha_M - \alpha_W) + \frac{1}{3}\Sigma_{C,L,S}(\gamma_{Mi} - \gamma_{Wi})$

Illustration with made up data with three sets of constraints:

μ	α_M	α_W	β_C	β_S	β_L	γ_{MC}	γ_{ML}	γ_{MS}	γ_{WC}	γ_{WS}	γ_{WL}	μ_{ML}	$\mu_{ML} - \mu_{WL}$	$\mu_{M.} - \mu_{W.}$
1	2	0	-1	2	0	1	2	0	0	0	0	5	4	3
2	0	-2	0	3	1	1	2	0	0	0	0	5	4	3
2.33	1	-1	-1.33	1.66	-0.33	1	2	0	0	0	0	5	4	3

If your software complains about "non-est" or "not estimable",
you have asked for a quantity that depends on the choice of constraints.
Your software is telling you that what you asked for has multiple possible answers.
The most likely reason is that you asked for the wrong thing.

SS by Model comparison:

Reminder: 1 way ANOVA, SS can be computed by comparing two models:
Full: $Y_{ij} = \mu_i + e_{ij}$
Reduced: $Y_{ij} = \mu + e_{ij}$
The difference in error SS = the SS for "groups"

Model comparison the hard way:
Fit the reduced model:

```
lm(y ~ 1, data=food)
proc glm data=food; model Y = ; run;
```

SS_{error} = 222.14, with 149 d.f. (= 150 - 1)

Fit the full model:

```
lm(y ~ group, data=food)
proc glm data=food; class group; model Y = group; run;
```

SS_{error} = 194.56, with 144 d.f. (= 150 - 6)

SS for groups by subtraction: SS_{groups} = 222.14 - 194.56 = 27.58 with 149 - 144 = 5 d.f.
Exactly the same SS_{groups} as by contrasts or formulae!

Model comparison for a 2 way factorial:

Idea: Compute SS for Sex by comparing model with sex effect ($\alpha's$) to one without
Fit the reduced model:

```
lm(y ~ 1, data=food)
proc glm data=food; class sex; model Y = ; run;
```

SS_{error} = 222.14, with 149 d.f. (= 150 - 1)

Fit the full model:

```
lm(y ~ sex, data=food)
proc glm data=food; class sex; model Y = sex; run;
```

SS_{error} = 222.08, with 148 d.f. (= 150 - 2)

SS for sex by subtraction: SS_{sex} = 222.14 - 222.08 = 0.06 with 149 - 148 = 1 d.f.
Exactly the same SS_{sex} as by contrasts or formulae!

But, which pair of models?

Effect		model	error SS	SS for effect
Sex	Full	$\mu + \alpha_i$	222.08	
	Red.	μ	222.14	0.06
Sex	Full	$\mu + \alpha_i + \beta_j$	194.72	
	Red.	$\mu \quad + \beta_j$	194.78	0.06
Sex	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	194.56	
	Red.	$\mu \quad + \beta_j + \gamma_{ij}$	194.62	0.06
Type	Full	$\mu + \beta_j$	194.78	
	Red.	μ	222.14	27.36
Type	Full	$\mu + \alpha_i + \beta_j$	194.72	
	Red.	$\mu + \alpha_i$	222.08	27.36
Type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	194.56	
	Red.	$\mu + \alpha_i \quad + \gamma_{ij}$	221.92	27.36
Sex*type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	194.56	
	Red.	$\mu + \alpha_i + \beta_j$	194.72	0.16

Very nice consequence of equal sample sizes (also called balanced data):
When sample sizes are equal (balanced), choice of model pair doesn't matter.
consequence of orthogonality

When sample sizes are not equal, choice does matter.

Types of SS, illustrated using unbalanced data.
Food palatability study with sample sizes ranging from 21 to 25 people per group.

Type I SS, also called sequential SS: each term compared to model with only 'earlier' terms

Model	Effect		model	error SS	SS for effect
Sex Type Sex*type	Sex	Full	$\mu + \alpha_i$	209.87	
		Red.	μ	210.00	0.13
	Type	Full	$\mu + \alpha_i + \beta_j$	186.46	
		Red.	$\mu + \alpha_i$	209.87	23.41
	Sex*type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu + \alpha_i + \beta_j$	186.46	0.34
Type Sex Sex*Type	Type	Full	$\mu + \beta_j$	186.60	
		Red.	μ	210.00	23.40
	Sex	Full	$\mu + \alpha_i + \beta_j$	186.46	
		Red.	$\mu \quad + \beta_j$	186.60	0.14
	Sex*type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu + \alpha_i + \beta_j$	186.46	0.34

Type III SS, also called partial SS: Each term compared to all other terms except term of interest

SAS model	Effect		model	error SS	SS for effect
sex type sex*type	Sex	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu \quad + \beta_j + \gamma_{ij}$	186.26	0.14
	Type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu + \alpha_i \quad + \gamma_{ij}$	209.44	23.32
	Sex*type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu + \alpha_i + \beta_j$	186.46	0.34
type sex sex*type	Sex	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu \quad + \beta_j + \gamma_{ij}$	186.26	0.14
	Type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu + \alpha_i \quad + \gamma_{ij}$	209.44	23.32
	Sex*type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	186.12	
		Red.	$\mu + \alpha_i + \beta_j$	186.46	0.34

Type I SS depend on order of terms in the model, when data are unbalanced.
Differences are small here; can be huge.
Type III SS (and F tests) are the same for any orders.

Big advantage to type III tests.
Same as SS derived using contrasts among cell means.

Type II SS (and F tests) similar to type III, but assume no interaction.

R implementation

anova() applied to an lm or lmer object gives you sequential (Type I) SS and tests.
When design is balanced, order of terms doesn't matter. anova() results are reasonable.

When the design is unbalanced, even accidentally, anova() results are not reasonable. To quote John Fox, "The standard R anova function calculates sequential ("type-I") tests. These rarely test interesting hypotheses in unbalanced designs."

Can get partial (Type III) SS and tests in various ways:
`joint_tests()` in the emmeans library is the most reliable and does not require care setting up the model.
`anova()` in the **lmerTest** library gives type III tests, but only for models fit with lmer().
Confusing because anova() has an lm definition and an lmer definition, but only when lmerTest activated

`drop1()` in the base stats library provides tests when dropping one term at a time from the model (the definition of type III tests). The current implementation only drops the interaction in a factorial model. It doesn't provide tests for main effects.

`Anova()` in the car library provides type II and type III tests, but the implementation requires setting some options before fitting the lm() model. To quote Terry Thernau, a statistican at Mayo Clinic: "The standard methods for computing type 3 that I see in the help lists are flawed, giving seriously incorrect answers unless sum-to-zero constraints were used for the fit (contr.sum). This includes both the use of drop.terms and the Anova function in the car package."

JMP implementation Use Analyze / Fit model. The Effect Tests box has type III tests. Will give sequential tests upon request (red triangle in top left, Estimates / Sequential Tests)

SAS implementation SAS reports both type I and type III SS by default. I concentrate on type III SS and tests to answer the standard questions about means.

When the design is complex, comparing type I and type III SS provides a quick check that the data are balanced. If the Type I and III SS are the same, the data is balanced. If not, the data are not.

Missing cells

No observations for one or more combinations of levels. E.g. Men only tasted old and solid products.

	Type			
Sex	old	liquid	solid	average
F	0.24	1.12	1.04	0.80
M	0.20		1.08	??
average	0.22	??	1.06	??

1 way ANOVA has 4 d.f. (only 5 groups). Divide into usual 2 way ANOVA quantities:

Source	d.f.	
Sex	1	
Type	2	
Sex*Type	**1**	should be 2. lose the d.f. here, in the interaction!

If you have missing cells, SAS (and some other programs) will report type I and type III SS.
Main effect tests (sex, type) are **meaningless!**,
because those tests correspond to meaningless comparisons among models.
Can interpret interaction (but limited to a subset of the data).

Often the first clue: LSMEANS are non-est.
Remember, row average is average of three cell means.
Can't estimate Men, liquid, so can't calculate men or liquid marginal means.

Solutions to missing cells:
Best: Write your own contrasts:
what questions can you answer here using the means you have?
OK: Test interaction, if very ns. e.g. $p > 0.20$, drop interaction from model
Main effect tests do make sense when no interaction in model.
Make sure you know what you're doing!

Estimates of marginal means

Balanced data easy, unbalanced data requires some care
Type III approach with all interactions in the model usually makes most sense

Estimates of sex effect (difference between men and women)
Same ideas for type, but have to deal with 3 levels

Balanced data: (25 per cell) All roads lead to Rome

Model	M	W	diff.	s.e.
sex only, ignore type	0.840	0.800	-0.40	0.20
"type II", ignore interaction	0.840	0.800	-0.40	0.120
type III	0.840	0.800	-0.40	0.120
type III without interaction in model	0.840	0.800	-0.40	0.19

Unbalanced data: (21-25 obs per cell) Choice makes a difference

Model	M	W	diff.	s.e.
sex only, ignore type	0.863	0.803	-0.060	0.203
"type II", ignore interaction:	0.859	0.795	-0.064	0.194
type III:	0.858	0.795	-0.062	0.194
type III without interaction in model:	0.859	0.795	-0.064	0.192

Differences are small here. Can be large.
My view: Type III with interaction makes most "sense"
Main effects are equally weighted averages of simple effects.

But, assumes population of interest has equal amounts of each group.
Usually reasonable in a designed experiment.
May be a problem in an observational study.

Sometimes this isn't (may not be) appropriate:
see me for more details if interested

Beware:
Dropping interaction from model gives you type II estimates (even though labelled type III).
Moral: always include all interactions in your model
(unless you have good reasons to do otherwise, and you know what you're doing!

Putting the pieces together; Doing the analysis:

Start with the ANOVA table (1 way or 2 way, depending on question(s).

Use F tests based on type III SS to answer "standard" factorial questions.
F tests are the start, not the end of the analysis
What are the means, differences/contrasts that answer important questions?
How precise are means, differences, contrasts?
Most useful number in the ANOVA table: often the MSE
Plot the means in a way that communicates the key results

Some useful things to check:
Check residuals to make sure model reasonable
Especially if important effects have large s.e.'s
Look for equal variances, additive effects
If not reasonable, correcting often increases power   If you believe you have balanced data:
Check whether Type I SS = Type III SS. Should be same when balanced.
Often find they're not! especially useful when many factors or levels
Student forgot about the missing observation(s).
One or more lines of data accidently left out.
R/JMP/SAS didn't read data correctly.
Check d.f. for highest order interaction.
Should be product of main effect d.f.
If not, you're missing one or more cells. Stop and think hard!

Study design: Choosing a sample size:

Easy using t-statistics.
Generalize the approach from 1way ANOVA.
Choose the quantity of greatest interest or least precisely estimated

Specify the difference of interest and error s.d.

Calculate the s.e. of the quantity of interest. Plug into power calculation. Can often use software by using contrasts and 1-way ANOVA approach. If not, will probably have to do by hand.

Example: want 80% power to detect a difference of 0.5 in palatibility, s.d. =1.16.
Evaluate main effect of sex, main effect of food (liquid-solid), one simple effect, and interaction (M l-s - W l-s)

Effect	s.e.	n per group
Sex	$\sqrt{2/3n}$	29
Type: L - S	$\sqrt{2/2n}$	43
Simple effect: L-S in M	$\sqrt{2/n}$	85
Interaction	$\sqrt{4/n}$	170

SS in ANOVA table by averaging observations and using formulae:

Notation: I rows (here I=2, sex), J columns (here J=3, type), n obs per cell

Y_{ijk}: observation k for sex i, type j.
$Y_{ij.}$: average of observations from sex i, type j. (n=25)
$Y_{i..}$: average of observations from sex i (nJ=75)
$Y_{.j.}$: average of observations from type j (nI=50)
$Y_{...}$: average of all observations (nIJ=150)

SS as variability between averages (works when data are balanced)

Source	d.f.	here	Sum of Squares	here
Treatment	$IJ-1$	5	$n\sum_{ij}\left(\overline{Y}_{ij.}-\overline{Y}_{...}\right)^2$	27.58
Error	$IJ(n-1)$	144	$\sum_{ijk}\left(Y_{ijk}-\overline{Y}_{ij.}\right)^2$	194.56
c.total	$IJn-1$	149	$\sum_{ijk}\left(Y_{ijk}-\overline{Y}_{...}\right)^2$	222.14

Source	d.f.	here	Sum of Squares	here
Sex	$I-1$	1	$nJ\sum_i\left(\overline{Y}_{i..}-\overline{Y}_{...}\right)^2$	0.06
Type	$J-1$	2	$nI\sum_j\left(\overline{Y}_{.j.}-\overline{Y}_{...}\right)^2$	27.36
Sex*Type	$(I-1)(J-1)$	2	$n\sum_i\left(\overline{Y}_{ij.}-\overline{Y}_{i..}-\overline{Y}_{.j.}+\overline{Y}_{...}\right)^2$	0.16
Error	$IJ(n-1)$	144	$\sum_{ijk}\left(Y_{ijk}-\overline{Y}_{ij.}\right)^2$	194.56
c.total	$IJn-1$	149	$\sum_{ijk}\left(Y_{ijk}-\overline{Y}_{...}\right)^2$	222.14

Notice:
Sex SS is variability between averages for each sex, 0 when all sexes have same average

Type SS is variability between averages for each type, 0 when all types have same average
Will come back to Sex*Type
Error same in both ANOVA's: pooled variability between obs in the same treatment.
df for Sex, Type and Sex*type add up to df for trt in 1 way ANOVA
quick algebra, always so
SS for Sex, Type and Sex*type add up to SS for trt in 1 way ANOVA
tedious algebra, always so when balanced

Approach completely falls apart if sample sizes are unequal.