Stat 402A: Introduction to Factorial ANOVA

Data from a study of food preference:
- 3 types of protein supplement, control, new liquid, new solid
- randomly assigned to 75 men and 75 women, each tasted one. 25 per type.
- response is a measure of preference (-3 = absolutely dislike, 3 = wonderful)

6 treatments (3 types, 2 sexes).
Could ask about differences between the 6 treatments (1 way ANOVA)

Treatments have a natural structure, which suggests specific questions:
- Averaged over types, is there a difference between sexes?
- Averaged over sexes, is there a difference between types?
  - Is there a difference between new liquid and new solid?
  - Is there a difference between control (old) and average of the new types?
- Do men and women react similarly to the types? i.e.
  - Is the difference between liquid and solid the same in men and women?
  - Is the difference between control and new the same in men and women?

	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

When each cell has the same number of observations (25), 3 ways to compute SS
1) averaging observations (e.g. all 75 men, all 50 control)
2) comparing models using error SS
3) computing contrast SS

Looking ahead:
- When unequal numbers per treatment:
  - (1) fails miserably, (2) works when done right, (3) works
- When missing cells (e.g. control not given to men)
  - (1) and (2) fail miserably, (3) works if done carefully

Take home messages:
- Use Type 3 SS to construct F tests, unless there are missing cells
- Estimate marginal means using averages of cell means
- Think about what you're doing, especially if there is an interaction
- Missing cells are a big problem

Next few lectures will provide the “why” underlying these points.

Interactions:

Simple effect: difference (or contrast) between levels of one factor at one specific level of the other. e.g., difference between men and women in liquid, or difference between control and average of liquid and solid in men.

Interaction exists when simple effects are not the same.
Equivalent to non-parallel lines in a plot of means.

Sometimes (GxE study in plant/animal breeding): interactions are the goal of the study. Test of interactions is a key results

Usually, focus on main effects or simple effects. Interaction test is produced automatically.

When interaction not significant:
Nice, easy interpretation of main effects. Interpret main effects as estimates of each simple effect.
Here, report that palatability of the liquid product is 0.96 units larger than the control.

When interaction is significant:

1. Dogma (common in texts, journal reviewers): split data e.g. separately consider men (N=75) and women (N=75). analyze each group separately, report simple effects and tests within each group.

   If more interested in men-women, would split into three groups (control, liquid, solid)

2. My approach (1): Remember that the marginal mean is an average and the difference in marginal means is an average of simple effects.

   Do these averages "make sense"? If so, then report the marginal means and their differences.

   Don't forget that the simple effects could be quite different

   Examples (assuming a significant interaction):
   Makes sense: Will sell product to a population of 50% men, 50% women. Marginal mean is the average palatability in this population.

   Doesn't: Fertilizer x Corn variety. Can only plant a field with one corn variety. Want to know which fertilizer is best for that variety. Average doesn't make sense, so need simple effects.

3. My approach (2): Look at how large the interaction effects are, not just whether they are significant. Esp. useful if the error is small or n is large. Sometimes will decide to report main effects and ignore the interaction, even if statistically significant

   Example: fertilizer x corn variety: diff between fertilizers for variety A: 5.2 bu/ac (se = 0.01), diff between fertilizers for variety B: 5.3 bu/ac (se = 0.01). Test is significant, but may decide to ignore.

Precision of marginal and cell means and their contrasts or differences:

Need estimate of within-group variability = $s = \sqrt{\text{MSE}} = \sqrt{1.35} = 1.16$

s.e. of a row marginal mean = $s\sqrt{1/nJ} = s\sqrt{1/\#\text{ men in study}} = 1.16/\sqrt{75} = 0.13$.
s.e. of a col marginal mean = $s\sqrt{1/nI} = s\sqrt{1/\#\text{ liquid values in study}} = 1.16/\sqrt{50} = 0.16$.
s.e. of a cell mean = $s\sqrt{1/n} = s\sqrt{1/\#\text{ men,liquid in study}} = 1.16/\sqrt{25} = 0.23$.

s.e. of a diff. between XXX = $\sqrt{2}$s.e. of XXX
substitute row mean, col mean or cell mean for XXX, as appropriate.
s.e. diff. in row means = 0.19, s.e. diff in col means = 0.23, s.e. diff in cell means = 0.33

s.e. of an interaction, e.g. $(\mu_{11} - \mu_{12}) - (\mu_{21} - \mu_{22}) = s\sqrt{4/n} = 0.46$

Some marginal means estimated more precisely,
all marginal means estimated more precisely than any cell mean.

Interaction effects are the least precisely estimated!

Difference between Liquid and Control in men:
two possible estimates: marginal means: 1.18 -0.22 = 0.96, s.e. = 0.23
simple effect: 1.24 - 0.20 = 1.04, s.e. = 0.33
which to use?

My approach: plot the cell means to show the pattern, use interaction to decide whether to report marginal or simple effects

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

SS by Contrasts: $\gamma = \sum_{ij} c_{ij}\, \mu_{ij}$, $\quad g = \sum_{ij} c_{ij}\, \overline{Y}_{ij.}$, $\quad$ s.e. $g = s_p \sqrt{\sum_{ij} (c_{ij})^2 / n_{ij}}$

	μ_{FC}	μ_{FL}	μ_{FS}	μ_{MC}	μ_{ML}	μ_{MS}
M - F, av. over type	-1/3	-1/3	-1/3	1/3	1/3	1/3
L - S, av over sex	0	1/2	-1/2	0	1/2	-1/2
C - (L+S)/2, av over sex	1/2	-1/4	-1/4	1/2	-1/4	-1/4
L-S, same in M and F	0	1	-1	0	-1	1
C-(L,S), same in M and F	1	-1/2	-1/2	-1	1/2	1/2

Contrasts have SS, $\text{SS}_{\text{contrast}} = \dfrac{g^2}{\sum_{ij} (c_{ij})^2 / n_{ij}}$

Contrast	g	s.e.	$\sum_{ij} (c_{ij})^2 / n_{ij}$	SS
M - F, av. over type	0.04	0.19	2/75	0.06
L - S, av over sex	0.12	0.23	1/25	0.36
C - (L+S)/2, av over sex	-0.9	0.20	3/100	27.00
L-S, same M and F	-0.08	0.46	4/25	0.04
C-(L,S), same M and F	0.12	0.40	3/25	0.12

When two contrasts are orthogonal, they represent statistically unrelated pieces of information

$\sum_i l_i \mu_i$ and $\sum_i k_i \mu_i$ are orthogonal when $\sum_i \dfrac{l_i\, k_i}{n_i} = 0$

Are L-S and C-(L+S)/2 orthogonal? (remember all cells have 25 people, $n_i = 25$)

L - S, av over sex (l_i)	0	1/2	-1/2	0	1/2	-1/2
C - (L+S)/2, av over sex (k_i)	1/2	-1/4	-1/4	1/2	-1/4	-1/4
product	0	-1/8	1/8	0	1/8	-1/8

$\sum_{ij} l_{ij}\, k_{ij} / n_{ij} = (0 + -1/8 + 1/8 + 0 + -1/8 + 1/8)/25 = 0/25 = 0$, so yes.

When contrasts are orthogonal, can add SS to simultaneously test both components

No differences between types: L-S = 0 and C-(L+S)/2 = 0

SS for L-S = 0.36, SS for C-(L+S)/2 = 27, sum = 27.36

SS for 'no differences between types' using contrasts = SS using averages (above)

Orthogonal contrasts are not unique: L-C: SS = 23.04, S - (L,C): SS = 4.32, sum=27.36

But sum of non-orthogonal contrasts is wrong: L-C: SS = 23.04, S-C: SS = 17.64, sum=40.68

or, L-C: SS = 23.04, L-S: SS = 0.36, sum = 23.40

What about more than 2 contrasts?

A set of 3, 4, 5, $\cdots$ contrasts are orthogonal if all pairs are orthogonal

All pairs above are orthogonal. What do the SS add up to?

0.06 + 0.36 + 27.0 + 0.04 + 0.12 = 27.58 = trt SS from 1 way ANOVA

The traditional lines in the ANOVA table for a 2 way ANOVA represent one specific subdivision of the SS between all IJ groups.

Sometimes useful to think backwards with contrasts: answers to new questions.

Food palatability study: there are 6 trts, so 5 orthogonal contrasts.
You might have 2 important questions, e.g. diff between types: L-S, C-(L+S/2). Can compute SS for those contrasts.
If other questions are very much less important, may only want to ask 'is anything else different?'. I.e. partition trt variability into (between types = 2 df) + (everything else = 3 d.f.) = (diff btwn trt = 5 d.f.).

Could figure out 3 orthog contrasts for (everything else), but don't need to. Logic: There are 5 orthog. contrasts. L-S and C-(L+S)/2 are two of them. Can get the other 3 by subtraction.

Sometimes called a lack of fit test or 'leftover SS' test.

Source	d.f.	SS	MS	F	p
Treatments	5	27.58			
L-S	1	0.36			
C-(L+S)/2	1	27.00			
rest	3	0.22	0.07	0.054	>0.5
		= 27.58-27.36			
Error	144	194.56	1.35		

Interpretation: No evidence of any difference in palatability, other than the differences between types.

Model for 2 way anova (effects form):

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

Ignore variance, σ^2. Focus on the parameters defining the means for each group: μ, α_i, β_j, and γ_{ij}.

Same problem as with 1-way ANOVA: too many parameters. In total, 12 parameters to be estimated from 6 cell means. Here, way too many!

Solution: impose constraints on the parameters (e.g. force = 0). With 6 constraints, can estimate remaining 6 parameters. Choice of constraint is arbitrary.

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 SAS complains about 'non-est' you have asked for a quantity that depends on the choice of constraints. SAS is telling you that what you asked for is non-estimable.

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 using SAS, the hard way:
Fit the reduced model:

```
proc glm; class group; model Y = ; run;
```

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

Fit the full model:

```
proc glm; 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:

```
proc glm; class sex; model Y = ; run;
```

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

Fit the full model:

```
proc glm; 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?

SAS name	Type I	Effect		model	error SS	SS for effect
	a	Sex	Full	$\mu + \alpha_i$	222.08	
			Red.	μ	222.14	0.06
Type II	b	Sex	Full	$\mu + \alpha_i + \beta_j$	194.72	
			Red.	$\mu \quad + \beta_j$	194.78	0.06
Type III		Sex	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	194.56	
			Red.	$\mu \quad + \beta_j + \gamma_{ij}$	194.62	0.06
	b	Type	Full	$\mu + \beta_j$	194.78	
			Red.	μ	222.14	27.36
Type II	a	Type	Full	$\mu + \alpha_i + \beta_j$	194.72	
			Red.	$\mu + \alpha_i$	222.08	27.36
Type III		Type	Full	$\mu + \alpha_i + \beta_j + \gamma_{ij}$	194.56	
			Red.	$\mu + \alpha_i \quad + \gamma_{ij}$	221.92	27.36
Type II, III	a,b	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.

Type I SS are sequential: each term compared to model with only 'earlier' terms
a) model Y = sex type sex*type
b) model Y = type sex sex*type

Type I SS depend on order of terms in the model, when data are unbalanced

Putting the pieces together; Doing the analysis:

Start with the ANOVA table.

Source	d.f.	SS	MS	F = MS/MS_{error}	p-value
Type	2	27.36	13.68	10.12	<0.0001
Sex	1	0.06	0.06	0.04	0.83
Type*Sex	2	0.16	0.08	0.06	0.94
Error	144	194.56	1.35		

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?
Plot the means in a way that communicates the key results

Conclusions: (without p-values or s.e's, format for these depends on journal)
- no evidence of any difference between sexes, average over types, estimate = 0.04
- strong evidence of at least one difference between types, averaged over sexes
  - contrasts: No evidence of difference between L and S, estimate = 0.12
  - Large difference between old and the average of the new types, estimate = 0.90.
- No evidence of an interaction between sex and type

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