Experimental Design 2: Replicated Factorial Designs

Outline:

1. The idea of a factorial designs and orthogonal contrasts
2. Two-factor designs
   A. Analysis using contrasts ("Hard" way, for illustration)
   B. Analysis using factorial effects ("Easy" way, standard)
   C. Power for a two-factor design
3. Three-factor designs
4. Four-factor designs

Examples:

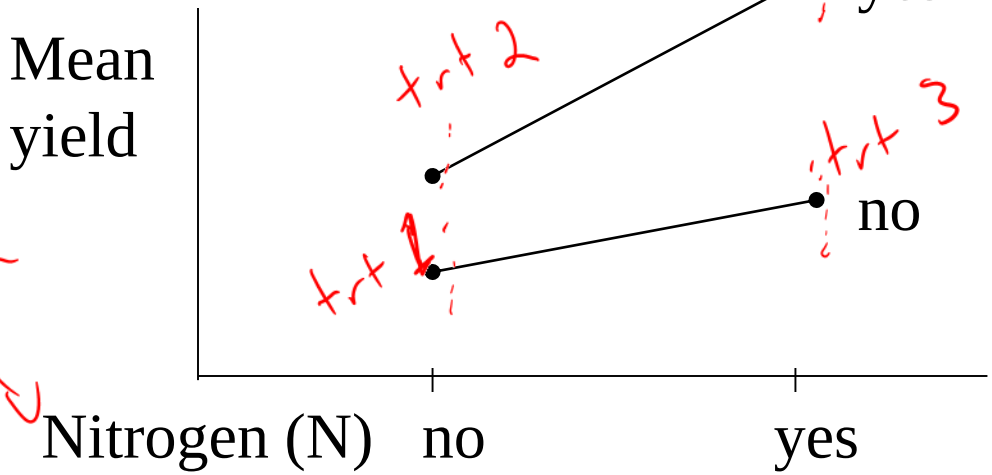
1. Two-way with contrasts (hard way)
2. Two-way as a factorial (easy way)
3. Anova vs anova in R
4. Battery two-way example
5. Power for two-way using Lenth
6. Fish three-way example
7. Poplar four-way example

1. The idea of a factorial design

Example: Four treatments in an RCB or CRD design

	P no	**P yes**
N no	Trt 1	Trt 2
N yes	Trt 3	Trt 4

A Profile, Interaction or Line Plot:



Although this was described a one-way design with four treatments, these treatments have a special structure called a "**factorial structure**".

"Factors"	"Levels" of the factors
1. Nitrogen	(yes, no)
2. Phosphorus	(yes, no)

It is called a 2 x 2 (or 2^2) factorial design

Plan:

ANOVA

1. Study factorial designs using a one-way layout with contrasts ("hard" way, for illustration).
2. Then we do a factorial ANOVA ("easy" way, standard)

Linear Contrasts

A linear contrast (usually just called a contrast) is a comparison of trt means. A contrast is written as:

$l = a_1\mu_1 + a_2\mu_2 + \cdots + a_t\mu_t = \sum_{i=1}^{t} a_i\mu_i$

The contrast is described by a list of coefficients $(a_1, a_2, \ldots a_t)$.

A linear contrast has the property that the sum of the contrast coefficients (the sum of the a_i's) is zero: $\sum_{i=1}^{t} a_i = 0$.

To estimate, just substitute sample means as estimates of the μ_i's:

$\hat{l} = a_1\overline{y_1} + a_2\overline{y_2} + \cdots + a_t\overline{y_t} = \sum_{i=1}^{t} a_i\overline{y_i}$

Note: Pairwise comparisons (Ex H_0: $\mu_1 - \mu_2 = 0$) are the simplest and most common contrasts.

$a_1 = 1$

$a_2 = -1$

Orthogonal Contrasts

Suppose we have two contrasts l_1 and l_2:

$$l_1 = \sum a_i \mu_i \qquad l_2 = \sum b_i \mu_i$$

l_1 and l_2 are orthogonal if:

$$\sum a_i b_i = a_1 b_1 + a_2 b_2 + \cdots + a_t b_t = 0$$

A set of contrasts is said to be mutually orthogonal if all pairs of contrasts in the set are orthogonal.

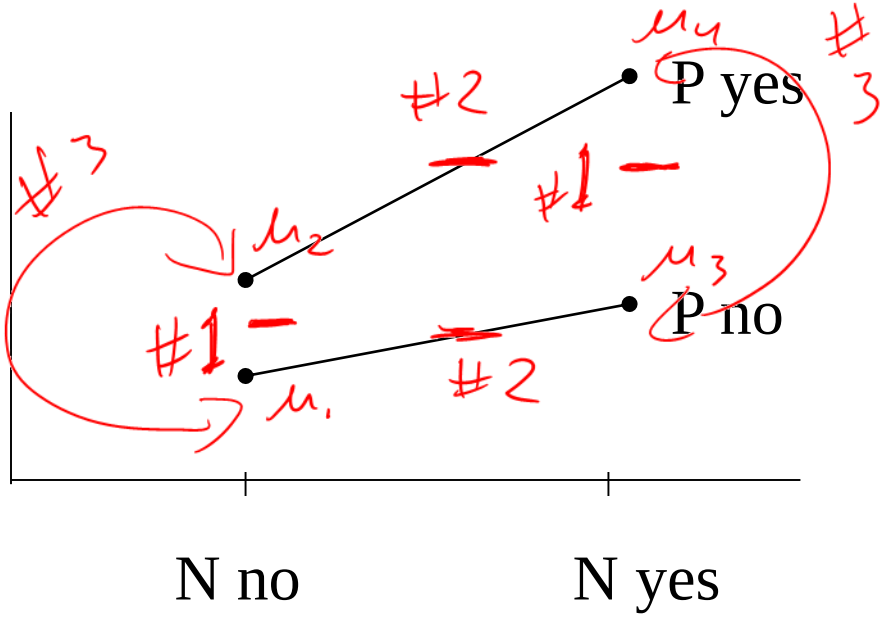
Properties of Orthogonal Contrasts:

1. Independence
2. We can partition the SSTrt using dfTrt = t-1 orthogonal contrasts. (Historically used as an error check).

Read more in O&L Section 9.2

Return to our **Example:**

	P no	**P yes**
N no	μ_1	μ_2
N yes	μ_3	μ_4



Now lets set up contrasts.

1.Main Effect of N

In words: Difference N no vs N yes (averaging over P)?
Using graph: Average of the left two points vs the average of the right two points.

H_0: $(\mu_1 + \mu_2)/2 - (\mu_3 + \mu_4)/2 = 0$
Coefficients (a's): (0.5, 0.5, -0.5, -0.5)
Check Sum to Zero: +0.5 + 0.5 -0.5 -0.5 = 0

2. Main Effect of P

In words: Difference P no vs P yes (averaging over N)?
Using graph: Average of the bottom line vs the average of the top line.

H_0: $(\mu_1 + \mu_3)/2 - (\mu_2 + \mu_4)/2 = 0$
Coefficients (a's): (0.5, -0.5, 0.5, -0.5)
Check Sum to Zero: +0.5 - 0.5 +0.5 -0.5 = 0

3. N by P interaction

In words: Does the effect of P depend on the level of N?
Compare the effect of P with N vs effect of P without N
Using graph: Are the lines parallel?

H_0: $(\mu_4 - \mu_3) - (\mu_2 - \mu_1) = 0$
Coefficients (a's): (1, -1, -1, 1)
Check Sum to Zero: +1 -1 -1 +1 = 0

Summary of Three contrasts of interest

Treatment number	1	2	3	4
Level of P	no	yes	no	yes
Level of N	no	no	yes	yes
1. Main effect of N	1	1	-1	-1
2. Main effect of P	1	-1	1	-1
3. N*P Interaction:	1	-1	-1	1

Notes:

1. Multiplying through by a constant (ex: 0.5) does not change the test (or the orthogonality).
2. N*P = pairwise multiplication of N and P.

3. This set of contrasts is mutually orthogonal because all pairs of contrasts in the set are orthogonal.

Advantages of Factorial Designs:

1. If there is an interaction, you can study it.
2. If there is no interaction, you can more accurately estimate the effect of each factor by averaging over the levels of the other factor. Each estimate uses all the data:

   Effect of N (2n obs compared to 2n obs)

   Effect of P (2n obs compared to 2n obs)

Note: we are using n to represent the number of reps per NxP combination.

The "**Factorial Principle**" says: If there is an interaction, study it. If there is no interaction, make conclusions about one factor, averaging over levels of the other factor. (But also consider your research questions!)

2A. Two-factor design with contrasts

Two-way example: A 3 x 2 factorial treatment structure in a CRD. **Three tillage methods** (1, 2, 3) tested on each of **two varieties** (1, 2) with n=4 obs per treatment combination. So a total of 3 x 2 x 4 = 24 plots.

	Till 1	**Till 2**	**Till 3**
Var 1	Trt 1	Trt 2	Trt 3
Var 2	Trt 4	Trt 5	Trt 6

For illustration, we will start by fitting the one-way model:

$$y_{ij} = \mu + \alpha_i + e_{ij} \qquad i = 1,..,6 \qquad j = 1,...,4$$

In R: `lm(Y ~ trt)`

Then test comparisons of interest using contrasts.

Specially chosen contrasts (mutually orthogonal)

1. Variety 2 vs Variety 1 (averaging over Tillage)

   H_0: $(\mu_4+\mu_5+\mu_6)/3 - (\mu_1+\mu_2+\mu_3)/3 = 0$

   Coefficients (a's): (-1/3, -1/3 -1/3, +1/3, +1/3, +1/3)

   Multiply by 3: (-1, -1, -1, +1, +1, +1)

2. Till1 vs average of Tills 2,3 (averaging over Variety)

   H_0: $(\mu_1+\mu_4)/2 - (\mu_2+\mu_3+\mu_5+\mu_6)/4 = 0$

   Coefficients (a's): (1/2, -1/4, -1/4, 1/2, -1/4, -1/4)

   Multiply by 4: (2, -1, -1, 2, -1, -1)

3. Till 3 vs Till2 (averaging over Variety):

   H_0: $(\mu_3+\mu_6)/2 - (\mu_2+\mu_5)/2 = 0$

   Coefficients (a's): (0, -1/2, +1/2, 0, -1/2, +1/2)

   Multiply by 4: (0, -1, +1, 0, -1, +1)

"Specially chosen" contrasts continued

interactions

4. (Till1 vs average of Tills 2, 3 for Variety 2) vs
   (Till1 vs average of Tills 2, 3 for Variety 1)
   H_0: $(\mu_4 - (\mu_5+\mu_6)/2) - (\mu_1 - (\mu_2+\mu_3)/2) = 0$
   Coefficients (a's): (-1, +0.5, +0.5, 1, -0.5, -0.5)
   Multiply by 2: (-2, +1, +1, +2, -1, -1)

5. (Till2 vs Till3 for Variety 1) vs (Till2 vs Till3 for Variety 2)
   H_0: $(\mu_2-\mu_3) - (\mu_5-\mu_6) = 0$
   Coefficients (a's): (0, +1, -1, 0, -1, +1)

Summary of Contrasts of Interest:

Treatment number	1	2	3	4	5	6
Level of Variety	1	1	1	2	2	2
Level of Tillage	1	2	3	1	2	3
1. Variety main effect (averaging over tillage)	-1	-1	-1	1	1	1
Tillage main effect (averaging over variety)						
2. Till 1 vs avg Tills 2, 3	2	-1	-1	2	-1	-1
3. Till 3 vs Till 2	0	-1	1	0	-1	1
Till x Variety Interaction						
4. Till1 vs avg Tills 2, 3 by Var	-2	1	1	2	-1	-1
5. Till2 vs Till 3 by Var	0	1	-1	0	-1	1

Comments about Contrasts:

1. The five contrasts are orthogonal. ← not fun
2. The # of contrasts is equal to dfTrt = 6 - 1 = 5.
3. Multiplying through by a constant (ex: 1/3 or ½) does not change the test (or the orthogonality).
4. We will see that R uses different contrasts, but the type 3 test results are the same.
5. The interaction contrasts (4 and 5) can be obtained by multiplying the coefficients for the main effects contrasts (1 x 2, 1 x 3).

Return to the **Two-way Example using Contrasts (Hard way):**

1. We start by running the analysis as a one-way ANOVA with 6 trts (2 varieties x 3 tillage methods) and running all pairwise comparisons (using emmeans). None of the comparisons are significant at the alpha = 0.05 level.

2. We then use the "no intercept" or "means" model and test orthogonal contrasts using `lht()` from the car package.

Contrast	SS	F Test Stat	p-value
1. Cvar	1.0004	0.2285	**0.6384**
2. Ctill1	40.1500	9.1705	0.0072
3. Ctill2	0.2756	0.0630	0.8047
4. Cint1	3.5752	0.8166	0.3781
5. Cint2	0.0006	0.0000	0.9906

Sim Test #1
F = 4.6167
p = 0.02407

Sim Test #2
F = 0.4084
p = 0.6707

Conclusions about the **Two-Way Example using Contrasts**:

- Recall the "**Factorial Principle**", which says: If there is an interaction, study it. If there is no interaction, focus on main effects. (But also consider your research questions!)

- With this in mind, we start by looking at the test of **interaction** (Sim Test #2, $p = 0.6707$). Since $p > 0.05$, there is no evidence of interaction. In other words, no evidence that the effect of tillage depends on variety.

- Since there is no significant interaction, we focus on main effects of Tillage and Variety. In other words, compare the tillage methods averaging over variety and the varieties averaging over tillage.

Conclusions about the **Two-Way Example using Contrasts** continued:

- We look at the **Variety main effect** (Cvar, p = 0.6384). Since p > 0.05, there is no evidence of Variety main effect. In other words, no evidence of a difference between varieties (averaging over Tillage methods).

- We look at the **Tillage main effect** (Sim Test #2, p = 0.0241). Since p < 0.05, we conclude that there are some differences between the three tillage methods (averaging over varieties). In particular, Till 1 differs from the average of Tills 2, 3 (Ctill 1, p = 0.0072). But no evidence of a difference between Tills 2, 3 (Ctill 2, p = 0.8047)

Notes about Contrast Analysis:

1. Notice that the sum of SS for the 5 contrasts, adds up to SS Trt (from Model1, one-way ANOVA)

1.00 + 40.15 + 0.28 + 3.58 + 0.00 = 45.00 = SSTrt

This is because we used orthogonal contrasts.

2. Constructing factorial contrasts works for small numbers of factors and levels. There are more efficient ways to construct the contrasts (which we study next).

2B. Two-factor design using factorial effects

Let y_{ijk} = the response for the

i^{th} level of factor A, $i = 1,\ldots,a = 2$ (Variety)

j^{th} level of factor B, $j = 1,\ldots,b = 3$ (Tillage)

k^{th} replication (or block, if blocked), $k = 1,\ldots n = 4$

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + \varepsilon_{ijk}$$

α_i = the effect of the i^{th} level of factor A (variety)

β_j = the effect of the j^{th} level of factor B (tillage)

$(\alpha\beta)_{ij}$ = the interaction effect when the i^{th} level of factor A and the j^{th} level of factor B are used together.

γ_k = the k^{th} block effect (omit if no blocks)

Two-way Analysis using R:

```
options(contrasts = c("contr.sum",
"contr.poly"))
Model <- lm(resp ~ till*var, data = InData)
Anova(Model, type = 3)
emmeans(Model, pairwise ~ till*var)
```

Notes:

1. The summary() output does not directly answer common research questions. Use `Anova()` from car and `emmeans()` from emmeans package instead.
2. The contrasts option statement is critical when using Type 3 tests! With the default contrasts, the Type 3 tests are meaningless.

3. `lm(resp ~ till*var)` is equivalent to using `lm(resp ~ till + var + till:var)`.

4. When the data is **balanced (no missing values):**
   - Type 1 tests from `anova()` are the same ast Type 2 and 3 tests from `Anova()`.
   - emmeans and simple means are the same.

5. If the data is **not balanced**, Type 3 tests from Anova() are recommended. emmeans should still be used. See note #2!

Two-way Example as Factorial: ANOVA table and interpretation

	Sum Sq	Df	F value	Pr(>F)	
till	40.43	2	4.6167	**0.02407**	*
var	1.00	1	0.2285	**0.63839**	
till:var	3.58	2	0.4084	**0.67074**	
Residuals	78.81	18			

see slide 60

- Note that these results exactly match the "contrast" (hard way) approach!
- We recall the "Factorial principle" and start by looking at the **interaction** (till:var, p = 0.6707). Since p > 0.05, there is no evidence of interaction. In other words, no evidence that the effect of tillage depends on variety.
- Since there is no significant interaction, we focus on main effects of Tillage and Variety.

Two-way Example as Factorial continued:

- We look at the **Variety main effect** (var, p = 0.6384). Since p > 0.05, there is no evidence of Variety main effect. In other words, no evidence of a difference between varieties (averaging over Tillage methods).

- We look at the **Tillage main effect** (till, p = 0.0241). Since p < 0.05, we conclude that there are some differences between the three tillage methods (averaging over varieties).

- Then use emmeans to get **pairwise comparisons for Tillage** methods (averaging over Varieties). We find Till 1 has higher average yield than Till 3 (estimate = +2.875, p = 0.0337). The difference between Till 1 and Till 2 is borderline significant (estimate = +2.6125, p = 0.0557).

The factorial ANOVA table (balanced case):

start here

Thus

SSTotal = total sum of squares (corrected) $= \sum_{i,j,k} (y_{ijk} - \bar{y}_{...})^2$

SSA(var) = sum of squares for factor A $= bn\sum_{i=1}^{a} (\bar{y}_{i..} - \bar{y}_{...})^2$

SSB(till) = sum of squares for factor B $= an\sum_{i=1}^{b} (\bar{y}_{.j.} - \bar{y}_{...})^2$

SSAB = sum of squares for A*B $= n\sum_{i,j} (\bar{y}_{ij.} - \bar{y}_{...})^2 - \text{SSA} - \text{SSB}$

SSblk = sum of squares for blocks $= ab\sum_{j=1}^{n} (\bar{y}_{..k} - \bar{y}_{...})^2$

$\text{SSResid} = \text{SSTotal} - \text{SSA} - \text{SSB} - \text{SSAB} - \text{SSblk}$

Omit SSblk if completely randomized (no blocks)

The formulas are summarized in the ANOVA table (no blocking):

Source	df	MS = SS/df	F	p-value
A	a-1	MSA	MSA/MSResid	p_a
B	b-1	MSB	MSB/MSResid	p_b
AB	(a-1)(b-1)	MSAB	MSAB/MSResid	p_{ab}
Resid	ab(n-1)	MSResid		

Note dfResid formula works when data is balanced (no missing values) and no blocking.
Another way to calculate is by subtraction:
dfResid = Total # observations – 1 – dfA – dfB - dfAB

Parameter Interpretation:

As we have seen, we can usually do a factorial analysis without working with the parameters directly. (In other words, we don't need to directly use the information from `summary()`).

1. A factorial analysis is just a regression on indicator variables.
2. The model is over parameterized, some indicator variables are omitted. The number of indicator variables corresponding to each "effect" (main effect or interaction) is the df for that effect.
3. The parameter estimates can be interpreted in terms of the mean responses.
4. We can change the contrasts used by R (which will change the parameter estimates and interpretation). This is important for the Type 3 tests!

emmeans: Based on the table of cell parameter means. emmeans are estimates of the means as if each cell had an equal sample size.

	Till 1	Till 2	Till 3
Var 1	Trt 1	Trt 2	Trt 3
Var 2	Trt 4	Trt 5	Trt 6

marginal
Var 1 avg

$$LSmean\ \ Var1 = \frac{1}{3}\sum_{j=1}^{3}(\hat{\mu} + \widehat{\alpha_1} + \widehat{\beta_j} + \widehat{\alpha\beta_{1j}})$$

$$LSmean\ Till1 = \frac{1}{2}\sum_{i=1}^{2}(\hat{\mu} + \widehat{\alpha_i} + \widehat{\beta_1} + \widehat{\alpha\beta_{i1}})$$

This data is balanced (equal number of observations for each trt combination), so emmeans are the same as simple means.

Describing Designed Experiments

It is common to describe designed experiments using two properties:

1. "**Blocking Structure**" or "**Design Structure**":
   The method of controlling variation. So far we have considered: completely randomized (CRD), randomized complete block (RCB), Latin Square (LS) and Graeco Latin Square (GLS). Later we will talk about split plot and repeated measures.

2. "**Treatment Structure**": The way in which the treatments are organized, one factor, two factors, three factors, etc.

A blocking structure can be combined with several possible treatment structures: two factor treatment structure in an RCB.

Sometimes it is not clear whether an effect is a treatment factor or a type of variation: e.g., gender in a weight loss program, or soil moisture in a varietal trial.

Battery Example: a two-way with interaction

Battery life, studied in a two-factor (treatment structure) completely randomized design (blocking structure), with n=4 reps per treatment combination.
Response: Y=life (hrs).
Factors: Temp (1,2,3) and Material (1,2,3). This is a 3x3 or 3^2 factorial.

Battery life (at a fixed output) is known to decrease with temperature. The objective is to compare the performance of the three materials.

Note: Temperature is a numerical variable. We could analyze these data using an ANCOVA model, but with only 3 temperatures, we question whether the response curves would be useful.

Battery Results:

```
Anova Table (Type III tests)

             Sum Sq Df  F value     Pr(>F)
mat           10684  2   7.9114  0.001976 **
temp          39119  2  28.9677 1.909e-07 ***
mat:temp       9614  4   3.5595  0.018611 *
Residuals     18231 27
```

drives analysis

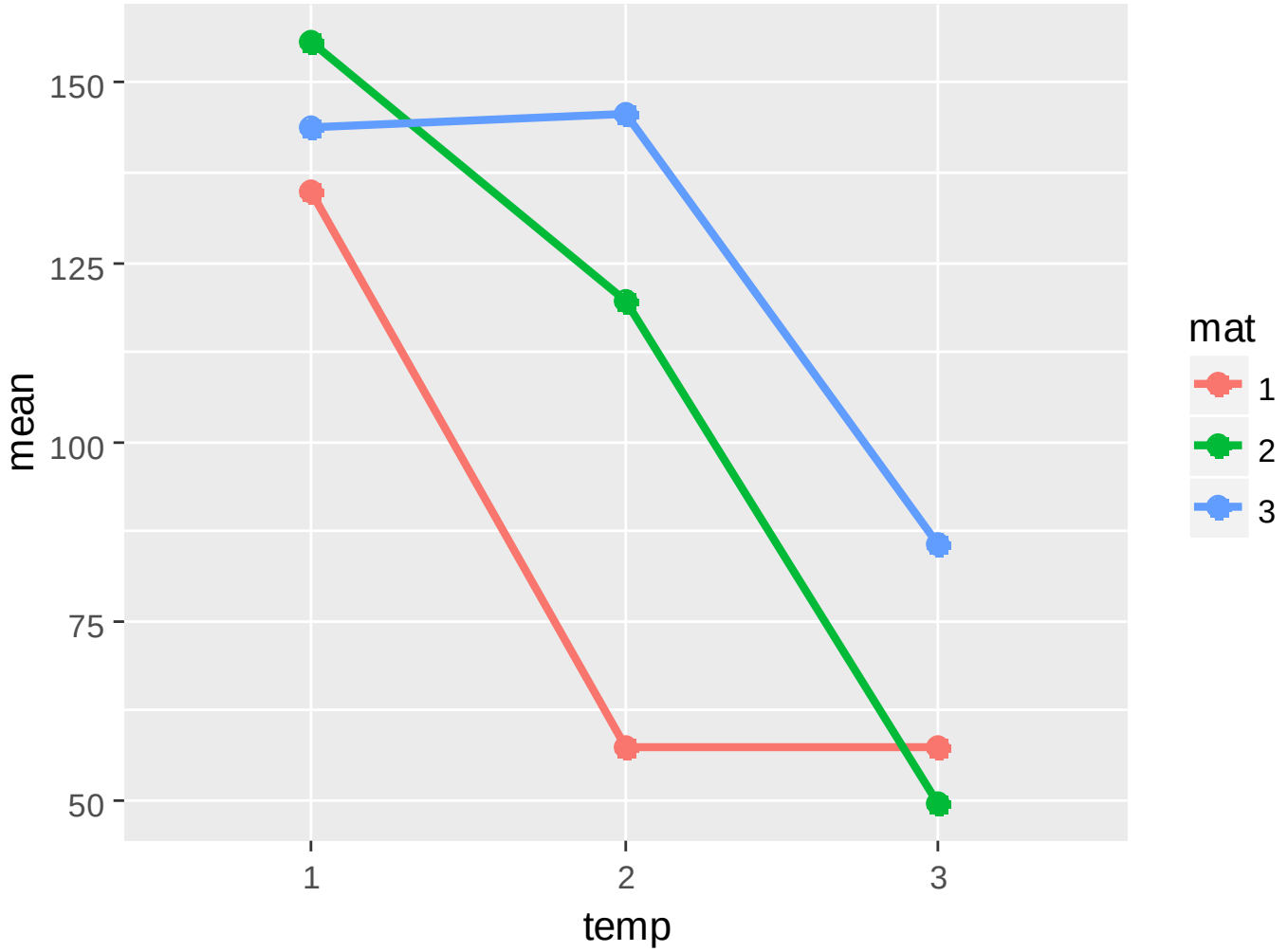
- Test of main effect of mat: Are there differences between materials (averaging over temperatures)?
- Test of main effect of temp: Are there differences between temperatures (averaging over materials)?
- Test of mat:temp interaction: Does the effect of material depend on the level of temperature? (Or does the effect of temperature depend on the level of material?)

We recall the “Factorial principle” and start by looking at the **interaction** (mat:temp, p = 0.0186). Since we find a significant interaction, we want to study it.

In other words, the significant interaction tells us that the effect of material depends on the level of temperature. Hence it does not make sense to compare materials averaging over temperature.

But what strategy should we use to understand the interaction?
Start by looking at the interaction graph!
Also think about research questions.

Battery Example: Interaction Plot



1. **Compare all 9 trts to each other using Unadjusted (LSD) or Tukey (HSD).**

This can be done with emmeans:

```
emmeans(Model, pairwise ~ mat*temp)
```

Problems with this approach:

- The Tukey (or HSD) method (default with emmeans) makes you pay a penalty in terms of power for making so many comparisons. It is not worth paying the penalty, since many of the comparisons are not of interest (e.g. mat 1 at temp 1 versus mat 2 at temp 2).
- The unadjusted (or LSD) method has high type I error rates when making so many comparisons. Can be done using `adjust = "none"` options from emmeans.
- This approach doesn't make use of the factorial structure.

2. **Compare levels of one factor, holding the level of the other factor fixed.** This is often a useful way to make sense of the interaction. Since we have a 2 factors, we have 2 options:
   (a) compare materials for each temperature
   (b) compare temperature for each material.

Since battery life dropping with temperature is not controversial, the experimenter probably would prefer (a). In practice, you think about relevant research questions.

This can be done with emmeans:

```
emmeans(Model, pairwise ~ mat|temp)
```

Family-wise

This will give us Tukey comparisons of materials, separately within each level of temp. This approach will control the MEER separately for each group of comparisons (not overall MEER). Usually reasonable.

Experimentwise Error Rate

Calculation of LSD and HSD values

Recall that the LSD (unadjusted) and HSD (Tukey adjusted) values just give the Margin of Error (ME) for a pairwise comparison.
So they follow the standard form: ME = TV x SE

In the formulas below, **t = #trts being compared**, **n = #obs contributing to the mean**. t table values can be found in Table 2 and q table values can be found in Table 10.

$$LSD_{0.05} = t_{\alpha/2, df\ \mathrm{Re}sid}\sqrt{\frac{2MS\ \mathrm{Re}sid}{n}}$$

$$HSD_{0.05} = q_{0.05}(t, df\ \mathrm{Re}sid)\sqrt{\frac{MS\ \mathrm{Re}sid}{n}}$$

Battery Example #1: Calculate LSD and HSD for pairwise comparisons of Materials at a fixed level of Temperature.

In R, we accomplish this using

```
emmeans(Model, pairwise ~ mat|temp)
```

$$LSD_{0.05} = t_{\alpha/2, df\ \mathrm{Re}sid}\sqrt{\frac{2MS\ \mathrm{Re}sid}{n}} = 2.052\sqrt{\frac{2*(18231/27)}{4}}$$

$$= 2.052*18.374 = 37.70$$

$$HSD_{0.05} = q_{0.05}(t, df\ \mathrm{Re}sid)\sqrt{\frac{MS\ \mathrm{Re}sid}{n}}$$

$$= q_{0.05}(3,27)\sqrt{\frac{(18231/27)}{4}} = 3.51\sqrt{\frac{675.2}{4}} = 45.60$$

```
emmeans(Model, pairwise ~ mat|temp)

temp = 1:
 contrast estimate       SE df t.ratio p.value
 1 - 2       -21.00 18.37407 27  -1.143  0.4967
 1 - 3        -9.25 18.37407 27  -0.503  0.8703
 2 - 3        11.75 18.37407 27   0.639  0.7998
```

A short cut is to use the SE information from emmeans to calculate the LSD/HSD values. See previous slide for table values (2.052, 3.51).

LSD = 2.052 * 18.374 = 37.70
HSD = 3.51* 18.374/sqrt(2) = 45.60

Note: Watch out for the sqrt(2)!!!!

Battery Example #2: Calculate LSD and HSD for pairwise comparisons of Materials averaging over Temperature.

In R, we accomplish this using

```
emmeans(Model, pairwise ~ mat)
```

NOTE: Due to significant interaction, this comparison probably NOT of interest!

$$LSD_{0.05} = t_{\alpha/2, df\ \mathrm{Re}sid}\sqrt{\frac{2MS\ \mathrm{Re}sid}{n}} = 2.052\sqrt{\frac{2*(18231/27)}{12}}$$

$$= 2.052*10.608 = 21.768$$

power ?

$$HSD_{0.05} = q_{0.05}(t, df\ \mathrm{Re}sid)\sqrt{\frac{MS\ \mathrm{Re}sid}{n}}$$

$$= q_{0.05}(3,27)\sqrt{\frac{(18231/27)}{12}} = 3.51\sqrt{\frac{675.2}{12}} = 26.33$$

```
emmeans(Model, pairwise ~ mat)

contrast  estimate       SE df t.ratio p.value
 1 - 2    -25.16667 10.60827 27  -2.372  0.0628
 1 - 3    -41.91667 10.60827 27  -3.951  0.0014
 2 - 3    -16.75000 10.60827 27  -1.579  0.2718
```

A short cut is to use the SE information from emmeans to calculate the LSD/HSD values. See previous slide for table values (2.052, 3.51).

LSD = 2.052 * 10.608 = 21.77
HSD = 3.51* 10.608/sqrt(2) = 26.33

Note: Watch out for the sqrt(2)!!!!

2C. Power in a two-factor design

To compute power for the two-factor we need the usual info:

A. Conjectured means

B. Conjectured standard deviation (σ)

Example: Suppose we are interested in calculating power for a 2x4 factorial (factor A with 2 levels, factor B with 4 levels) with conjectured σ =3 and conjectured means:

	B1	B2	B3	B4
A1	1	2	3	4
A2	5	6	7	8

μ's

See **"Power for two-way using Lenth"**

3. Three-factor designs (3way ANOVA)

Let y_{ijkl} = the response for the

i^{th} level of factor A, $i = 1,\ldots,a$

j^{th} level of factor B, $j = 1,\ldots,b$

k^{th} level of factor C, $k = 1,\ldots,c$

l^{th} replication (or block, if blocked), $l = 1,\ldots n$

$$y_{ijkl} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} + (\alpha\beta\gamma)_{ijk} + \tau_l + e_{ijkl}$$

SS formulas and an ANOVA table for the balanced three-factor design are given in O&L. It is not important to know these formulas, because we will depend on R to compute them. The idea of the formulas is the same as for all factorial designs.

SS formula idea: terminology

1. SSA, SSB and SSC, (called "main effects") are calculated using the formula used in the one-way, ignoring the other treatments. (e.g. SSA would be from a one-way with $t=a$ trts, and $b*c*n$ reps)

2. Two-way interaction SS's: SSA*B ,SSA*C and SSB*C, are calculated using the formulas used in a two-factor design, treating the third factor as if it were additional reps.

3. SSA*B*C is calculated using the formula for a one-way ANOVA treatment SS and the three-way means with $t=a*b*c$ and n reps, minus all of the other SS's.

4. The SSResid is the same as the SSResid in the one-way ANOVA with $t=a*b*c$ treatment and n reps.

5. The df for each main effect is: (# of levels -1). For each interaction df is the product of the df for the individual terms. dfResid is *a*b*c*(n-1)* (the same as a one-way design with *a*b*c* treatments an n reps.)

6. All of the above sums of squares could be constructed from the one-way ANOVA with *a*b*c* treatments, using appropriate contrasts. (But that would be tedious.) In practice, we will use `lm()` with `Anova( , type = 3)`.

Fish Example: three-way factorial

The growth of a certain species of fish was studied in a 3 x 2 x 2 factorial design. 36 tanks were randomly assigned in a CRD (n=3) as follows:

Factors	Levels
(1) Water temp	(cold, lukewarm, warm)
(2) Water movement	(still, flowing)
(3) Light level	(high, low)

The ANOVA table is given in the output. We check the highest order interaction first:
Temp:Move:Light p=0.0311

Since the three-way interaction is significant, we want to study the three-way interaction graph.

Studying the three-way interaction:
There is the appearance of an interaction between Move and Temp when Light is low, but no evidence of an interaction between Move and Temp when Light is high. The fact that the size of the interaction differs between the two graphs is established by the significance of the three-way interaction. We look for a logical way to handle this three way interaction.

Options:

1. **Use emmeans() to make comparisons of interest.** This is a very flexible function and can address many questions of interest. See example.
2. **Split the analysis into high and low light parts and make separate conclusions.** This is a reasonable idea if making explicit light comparisons is not as important as describing how the relationship between Move and Temp differs between lighting conditions.

Conclusions from split analysis (Separate 2way at each Light)

High light: Since there is no significant interaction (p=0.4436), we make conclusions about Move, averaging over Temp (there is more weight gain in still water, p=0.0001). We make conclusions about Temp, averaging over Move (temperatures are different, p=0.0009) We follow this with comparisons of individual temperatures, averaging over Movement. We conclude that warmer temperatures yield greater weight gains.

Low light: There is a strongly significant interaction between Temp and Move, in low light (p=0.0022). We use the emmeans function to compare movement at each of the temperatures. For "cold" temp, still is significantly higher than flowing (p=0.0050). For "luke", there is no significant difference (p=0.4384). For "warm", still is significantly lower than flowing (p=0.0099).

Comments on "splitting" the analysis:

1. Some considerations when deciding whether to split analysis:
   - desire to simplify
   - specific comparisons of interest
   - significant higher order interaction
   - unequal variance (more on this later)
2. For the Fish analysis, there is a significant 3 way interaction. From the residual diagnostic plot, there is no evidence of unequal variance. See next slide for formal test.
3. The trade off is reduced dfResid (and hence reduced power) for the split analysis. For the Fish analysis, the 3way analysis (with both light groups) has dfResid = 24. For the 2way analyses, dfResid = 12.

Formal test of variances for the Fish analysis:

When the analysis is split (by light level), we get an estimate of error variance for high light (0.0178) and for low light (0.0283). (These are the MSResids's from the separate ANOVAs, with 12 df each). If it is justifiable, it is preferable to use the MSResid from the combined ANOVA (0.023), which is the pooled value with 24 df. Using the methods of Chapter 7, we can test the hypothesis that the two true error variances being estimated by the two MSResids's are the same.

$$F = \frac{\text{MSResid}_{\text{larger}}}{\text{MSResid}_{\text{smaller}}} = \frac{0.0283}{0.0178} = 1.59 \qquad F_{0.25} = 1.49$$

$0.10 < p - value < 0.25$ From the F-table with 12 and 12 d.f.

Since this F is not significant we might prefer to use the pooled MSResid for our tests.

Summary: When there is a significant three-way interaction, we inspect the three-way interaction graph and interpret the three-way means.

Main effects (averaged over two factors) and two-way interactions (averaged over one factor) are generally not of interest.

We search for a strategy to make sense of the three-way interaction by focusing on two-way interactions at each level of a third factor.

4. Four-factor designs (4way ANOVA)

Poplar Example:

Growth of hybrid poplar trees in concrete "soil frames" was studied. There were four factors, each at 2 levels ($2^4 = 16$ treatments) run in a RCB with 3 blocks.

Y = dry weight of new growth

Factors	Levels
1) Lime (L)	(no, yes)
2) Nitrogen (N)	(no, yes)
3) Phosphorus (P)	(no, yes)
4) Potassium (K)	(no, yes)

Poplar Results:

1. No main effect or interaction involving P was significant. Therefore further conclusions will be drawn averaging over P. (Here is where we get some benefit from the factorial structure.)
2. A very large effect of N and large effect of L, with a relatively small, but significant interaction. We want to look at the L*N interaction graph. There is only an effect of L when N is present.
3. No significant main effect of K, but a borderline significant N*K interaction. We want to look at the N*K interactions graph. There is only an effect of K when N is present.

4. The diagnostic plot shows unequal variance. Summary statistics suggest that the variance is systematically higher when N=2. What can we do?
   - Try a transformation.
   - Split analysis by Nitrogen level.
   - Allow different variances for the different levels of Nitrogen.