Experimental Design 3: Unbalanced and Unreplicated Designs

1. Unbalanced factorial designs
2. Factorial design with missing cells
3. Unreplicated factorial designs
4. Balanced incomplete blocks (BIBD)

Examples:

1. Unbalanced Two Factor Example
2. Missing Cells Two Factor Example
3. Disruption Example: Unreplicated four-factor
4. Pillow Example: BIBD

1. Unbalanced Factorial Designs

The term unbalanced refers to when the numbers of observations per treatment are unequal.

This happens when:

1. There are missing data.
2. A sampling plan does not control the cell numbers. For example, observational studies.

3. A sampling plan is designed to have unequal cell numbers. Sometimes chosen when some experimental units are much more expensive or hard to get than others.

Example: <u>Unbalanced</u> Two Factor Analysis

In this example there are 3 Drugs each studied at 4 Doses. However, the number of observations for each trt combination is different.

	Dose 1	Dose 2	Dose 3	Dose 4
Drug 1	n=3	n=6	n=4	n=7
Drug 2	n=6	n=8	n=4	n=3
Drug 3	n=3	n=7	n=7	n=5

Comments on Unbalanced Two Factor Analysis

1. No changes are required to run the analysis with unbalanced data.

2. The Type3 tests are preferred. They are reasonable and interpretable.

3. The emmeans are equivalent to the simple means for interactions, but not for main effects. The SE are not the same for each trt combination due to imbalance.

4. Because the data is unbalanced, Type 1, Type 2 and Type 3 tests are not the same (not shown)

2. Factorial Designs with Missing Cells

Example: Two Factor Analysis with <u>Missing Cells</u>

In this example there are 3 Drugs each studied at 4 Doses. The data is still unbalanced, but now there are 2 trt combinations that are not represented by any observations!

	Dose 1	Dose 2	Dose 3	Dose 4
Drug 1	n=3	n=6	**n=0**	n=7
Drug 2	n=6	n=8	n=4	**n=0**
Drug 3	n=3	n=7	n=7	n=5

Comments for Two Factor Analysis with <u>Missing Cells</u>

1. `Anova( ,type = 3)` generates an error and does not return any results.

2. Use of the `emmeans()` function is OK, but not all pairwise comparisons are possible.

2. When there are missing cells, you can <u>treat the factorial design as if it were a one-way design</u>, and then compute contrasts that are of interest.

4. Another approach would be to split the analysis by one of the factors.

A related example:

Some times an experiment is run as a factorial design but with an “added” control treatment.

For example:

	A1	A2	
B1	Trt 1	Trt 2	
B2	Trt 3	Trt 4	
			Trt 5= Control

Two Possibilities:

1. Run the analysis as a one-way ANOVA on with 5 treatments. This makes sense if you want to compare active treatments to control.
2. Omit the control from the analysis and run as a factorial design. This makes sense if you are not interested in comparing to the control treatment.

3. **Unreplicated factorial designs:**

“Disruption” Example:

An unreplicated 3^2 x 2^2 factorial (36 obs) in a CRD. The effects of four factors on the delivery of brain cancer antibody were studied. Subjects were female rats of about the same age and weight. All subjects were inoculated with cancer cells at the beginning of the experiment. n=1 observations per treatment combination!

Factors and levels:
Dayspost: number of days (after inoculation) that treatment was started (8, 12, 16). This is a measure of tumor size.

Trt: whether the blood-brain barrier was disrupted (BD=disrupted by intra-arterial injection of mannitol to one side of the brain, NS=not disrupted, saline injection)

Antibody agent used (AIB, MTX, DEX70). These antibodies have different molecular weights.

Route: delivery method of the antibody agent (IA=intra-arterial, IV=intra-venous)

Response Variables:
BAT=concentration of antibody around the tumor
LH= concentration of antibody in the other (the lateral) half of the brain.

The measure of primary interest is the ratio:

$$Y=BAT/LH$$

This gives a measure of the antibody around the tumor, relative to the antibody in the unaffected half of the brain. We will consider whether the response is best modeled in the log scale.

If we fit the full factorial design with all main effects and interactions:

4 main effects
6 two-way interactions
4 three-way interactions
1 four-way interaction

dfResid will be **zero**!

Main Problem: How to test hypotheses and estimate standard errors without an estimate of error?

Solution: If any interactions are zero, then MS's for those interactions estimate σ^2. Those MS's can be **pooled** to form an estimate of error. We depend on the principle that higher-way interactions are likely to be small, so we usually start with the tentative assumption that the highest-way (four way) interaction is zero.

Disruption Analysis:

0. Omitting the four-way interaction from the model pools it with and error estimate, giving us a tentative ANOVA table, and some residuals to look at.

1. The residual plot in the **original scale** has mild evidence of a cornucopia shape (increasing variance and nonlinearity).

2. The residual plot in the **log scale** looks improved. I choose the log scale.

3. In the log scale ANOVA (with 4 df error), none of the 3way interactions are significant (smallest p-value is 0.2104). Trt and Route main effects are significant (p-values < 0.05).

5. Noting that 4 dfResid is very low and no 3way interactions show evidence of significance, we now drop the three-way interactions from the model. dfResid is now 16. We have increased dfResid, at the risk is having inflated the estimate of error by pooling interactions that are not really zero.

   - Very significant TRT effect (p=0.0001)
   - Very significant ROUTE effect (p=0.0001)
   - Borderline ROUTE*DAYSPOST (p=0.0463)
   - No other effects even close to significant (smallest p=0.2574).

6. Compute and compare the trt means using emmeans. Because TRT does not interact with any other factors, this is sufficient.

7. Look at the interaction graph for route*dayspost. Make specific statements about the interaction using the emmeans.

Unreplicated Analysis Notes:

1. When there is no replication, all methods of estimating error are risky to some degree. You have to weigh the cost savings against this risk.

2. Sometimes replications are run sequentially. If there appears to be enough information from one rep, we stop the experiment. If there appears to be a need for a second rep, we continue.

3. Usually, we select the highest level interactions for pooling into error. Since large effects tend to have important interactions, don't pool interactions among effects that appear large, without first checking.

4. Balanced Incomplete Block (BIBD)

Basic Problem: What to do if there are more treatments than the "natural" number of EU's in a block.

Example 1: Trts are A, B, C We want to use "lab day" as a block, but the lab can do only 2 runs per day. The "natural" block size is 2. Consider the following design:

Block	1	2	3
	A	B	B
	C	C	A

Notation:
t = number of treatments = 3
r = number of reps = 2
b = number of blocks = 3
k = natural block size (k < t) = 2
N = total obs = 6

This design is “balanced” in the sense that each treatment occurs in a block with each other treatment the same number of times (λ). In our example $\lambda=1$.

It is <u>not</u> “balanced” in the sense that each combination of trt and block occurs the same number of times. Use Type 3 SS!

When t=b the design is called “symmetric”.

NOTE: BIBD designs do not exist for all situations.

For a BIBD to exist we must have <u>both</u> conditions:

1. $N = (\#\text{trts})(\#\text{reps}) = tr$

   $= (\#\text{blocks})(\text{natural block size}) = bk$

   Therefore: tr = bk,

2. The number of times each treatment occurs with each other treatments in the same block (λ), must be an integer.

$$\lambda = \frac{r(k-1)}{t-1}$$

For Example 1: 3x2 = 2x3 (condition 1 met)

$$\lambda = \frac{2(2-1)}{3-1} = 1 \text{ (condition 2 met)}$$

Example 2: Treatments are A, B, C, D. We use 4 blocks of size 3.

A	A	B	A
C	B	C	B
D	C	D	D

So t = 4, r=3, b = 4, k = 3

Condition 1: $tr = bk$

$$(4)(3) = (4)(3)$$

Condition 2: $$\lambda = \frac{r(k-1)}{t-1}$$

$$\lambda = \frac{3(3-1)}{4-1} = 2$$

Analysis of a BIBD:

A "fixed block" analysis is the same as RCB with missing data:

$$y_{ij} = \text{response for trt } i, i = 1,\ldots,t$$

$$\text{block } j, j = 1,\ldots,b$$

$$y_{ij} = \mu + \alpha_i + \beta_j + e_{ij} \quad (\text{Same as RCB})$$

$$(\text{Note}: tb > N)$$

In R:

```
Model <- lm(Y ~ trt + block)

Anova(Model ,type =3)

Emmeans(Model, pairwise ~ trt)
```

Example3: Pillow data (from O&L)

Eight experimental pillows and a standard control pillow (A, B, C, D, E, F, G, H, I) are tested by potential customers (t=9). Because it was believed that more accurate results would be obtained if individuals were asked to test only three pillows at a time, pillows were tested in groups of three (k=3). Each individual tested all twelve (b=12) groups of pillows. The assignment of pillows to groups was a BIBD. The response is total firmness score for all testers.

A	B	C
D	E	F
G	H	I
A	D	G
B	E	H
C	F	I

A	E	I
B	F	G
C	D	H
A	F	H
B	D	I
C	E	G

Pillow Results and Comments:

1. Significant differences between means for Pillows (based on ANOVA table and pairwise comparisons).
2. The SE is the same for every pair of pillows. This is due to balance and a key feature of BIBD. Example hand calculation:

$$SE(diff\ emmeans) = \sqrt{\frac{2kMS\,\mathrm{Re}\,sid}{t\lambda}} = \sqrt{\frac{2*3*31.74}{9*1}} = 4.6$$

4. To get LSD (unadjusted) or HSD (Tukey) value, multiply the above SE by appropriate table value: $t_{\alpha/2}$ for LSD, $q_{\alpha}/\sqrt{2}$ for HSD, with df=dfResid.
5. In this analysis, we are treating blocks as fixed. We will return to this example and treat blocks as random in the Random2 notes. The results will be different!