34 Lecture 34: April 19

Last time

- hypothesis test and confidence intervals for one-way random-effects model
- review of one-way random effects ANOVA model

Today

- HW3 deadline extended to Tuesday 04/18 midnight.
- nested design
- Two-factor designs

Additional reference

Course notes by Dr. Jason Osborne.
Lecture notes from Lukas Meier on ANOVA using R

Nested design

Factor B is nested in factor A if there is a new set of levels of factor B for every different level of factor A.

To illustrate the concept of nested design, we consider the "Pastes" data set in "lme4" package in R. The strength of a chemical paste product was measured for a total of 60 samples coming from 10 randomly selected delivery batches each containing 3 randomly selected casks. Hence, two samples were taken from each cask. We want to check what part of the variability of strength is due to batch and cask.

Let Y_{ijk} be the strength of the kth sample of cask j in batch i. We can use the model

$$Y_{ijk} = \mu + A_i + B_{j(i)} + \epsilon_{ijk}$$

where A_i is the random effect of batch and $B_{j(i)}$ is the random effect of cask **within** batch. Note the special notation $B_{j(i)}$ emphasizes that cask is nested in batch.

Two-factor designs with factors that are fixed/random and nested/crossed

There are in total six types of two-factor models with fixed/random effects factors that are either crossed or nested.

1. $Y_{ijk} = \mu + A_i + B_j + (AB)_{ij} + \epsilon_{ijk}$	crossed/random
2. $Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$	nested/fixed
3. $Y_{ijk} = \mu + A_i + B_{j(i)} + \epsilon_{ijk}$	nested/random
4. $Y_{ijk} = \mu + \alpha_i + B_j + (\alpha B)_{ij} + \epsilon_{ijk}$	crossed/mixed
5. $Y_{ijk} = \mu + \alpha_i + B_{j(i)} + \epsilon_{ijk}$	nested/mixed
6. $Y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$	crossed/fixed

In the models above,

- GREEK symbols parameterized FIXED, unknown treatment means (except for the error term)
- CAPITAL letters represent RANDOM effects (including the error term)
- for Model 1, $A_i, B_j, (AB)_{ij}$ are all independent
- for Model 2, $\sum \alpha_i = \sum_j \beta_{j(i)} \equiv 0$
- for Model 3, $A_i, B_{j(i)}$ are all independent
- for Model 4, $\sum \alpha_i = 0$ and $B_j, (\alpha B)_{ij}$ are all independent
- for Model 5, $\sum \alpha_i = 0$
- for Model 6, $\sum \alpha_i = \sum \beta_j = \sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} \equiv 0$

Now let's consider the following 6 examples from Dr. Osborne's notes.

1. Entomologist records energy expended (y) by $N = 27$ honeybees
   - at three TEMPERATURES (20, 30, 40$^\circ C$)
   - consuming three levels of SUCROSE (20%, 40%, 60%)

Temp	Suc	Sample		
20	20	3.1	3.7	4.7
20	40	5.5	6.7	7.3
20	60	7.9	9.2	9.3
30	20	6	6.9	7.5
30	40	11.5	12.9	13.4
30	60	17.5	15.8	14.7
40	20	7.7	8.3	9.5
40	40	15.7	14.3	15.9
40	60	19.1	18.0	19.9

- First factor:
- Second factor:
- Fixed or random?
- Crossed or nested?
- Model: $Y_{ijk} = \mu + \qquad\qquad + \epsilon_{ijk}$

Answer:

2. Experiment to study effect of drug and method of administration on fasting blood sugar in a random sample of $N = 18$ diabetic patients.

- First factor is drug: brand I tablet, brand II tablet, insulin injection
- Second factor is type of administration (see table)

Drug (i)	Type of Administration (j)	Mean $\bar{y}_{j(i)}$	Variance $s^2_{j(i)}$
($i = 1$) Brand I tablet	($j = 1$) $30mg \times 1$	15.7	6.3
	($j = 2$) $15mg \times 1$	19.7	9.3
($i = 2$) Brand II tablet	($j = 1$) $20mg \times 1$	20	1
	($j = 2$) $10mg \times 1$	17.3	6.3
($i = 3$) Brand I tablet	($j = 1$) before breakfast	28	4
	($j = 2$) before supper	33	9

- First factor:

- Second factor:
- Fixed or random?
- Crossed or nested?
- Model: $Y_{ijk} = \mu + \qquad\qquad + \epsilon_{ijk}$

Answer:

3. An experiment is conducted to determine variability among laboratories (interlaboratory differences) in their assessment of bacterial concentration in milk after pasteurization. Milk w/ various degrees of contamination was tested by randomly drawing four samples of milk from a collection of cartons at various stages of spoilage. Y is colony-forming units/μl. Labs think they are receiving 8 independent samples.

	Sample			
Lab	1	2	3	4
1	2200	3000	210	270
	2200	2900	200	260
2	2600	3600	290	360
	2500	3500	240	380
3	1900	2500	160	230
	2100	2200	200	230
4	2600	2800	330	350
	4300	1800	340	290
5	4000	4800	370	500
	3900	4800	340	480

- First factor:
- Second factor:
- Fixed or random?
- Crossed or nested?
- Model: $Y_{ijk} = \mu + \qquad\qquad + \epsilon_{ijk}$

Answer:

4. An experiment measures Campylobacter counts in $N = 120$ chickens in a processing plant, at four locations, over three days. Means (std) for $n = 10$ chickens sampled at each location tabulated below:

	Sample			
	Before	After	After	After
Day	Washer	Washer	mic. rinse	chill tank
1	70070.00	48310.00	12020.00	11790.00
	(79034.49)	(34166.80)	(3807.24)	(7832.05)
2	75890.00	52020.00	8090.00	8690.00
	(74551.32)	(17686.27)	(4848.01)	(5526.19)
3	95260.00	33170.00	6200.00	8370.00
	(03176.00)	(22259.08)	(5028.81)	(5720.15)

- First factor:
- Second factor:
- Fixed or random?
- Crossed or nested?
- Model: $Y_{ijk} = \mu + \qquad\qquad + \epsilon_{ijk}$

Answer:

5. An experiment to assess the variability of a particular acid among plants and among leaves of plants:

Plant i	1			2			3			4		
Leaf j	1	2	3	1	2	3	1	2	3	1	2	3
$k = 1$	11.2	16.5	18.3	14.1	19.0	11.9	15.3	19.5	16.5	7.3	8.9	11.3
$k = 2$	11.6	16.8	18.7	13.8	18.5	12.4	15.9	20.1	17.2	7.8	9.4	10.9
$k = 3$	12.0	16.1	19.0	14.2	18.2	12.0	16.0	19.3	16.9	7.0	9.3	10.5

- First factor:
- Second factor:
- Fixed or random?

- Crossed or nested?
- Model: $Y_{ijk} = \mu +$ $\qquad\qquad$ $+\epsilon_{ijk}$

Answer:

6. Plant heights from 20 pots randomized to 10 treatment combinations

Treatment	Dark	Source	Intensity	Pot	Seedling 1	Seedling 2
DD	1	D	D	1	32.94	35.98
DD	1	D	D	2	34.76	32.40
AL	0	A	L	1	30.55	32.64
AL	0	A	L	2	32.37	32.04
AH	0	A	H	1	31.23	31.09
AH	0	A	H	2	30.62	30.42
BL	0	B	L	1	34.41	34.88
BL	0	B	L	2	34.07	33.87
BH	0	B	H	1	35.61	35.00
BH	0	B	H	2	35.65	32.91

- First factor:
- Second factor:
- Fixed or random?
- Crossed or nested?
- Model: $Y_{ijk} = \mu +$ $\qquad\qquad$ $+\epsilon_{ijk}$

Answer:

Tables of expected mean squares (EMS)

When factors A and B are CROSSED, and no sum-to-zero assumptions are made on random effects, expected means associated with sums of squares are given in the table below:

Source	*df*	A, B fixed	A, B random	A fixed B random
A	$a-1$	$\sigma^2 + nb\psi_A^2$	$\sigma^2 + nb\sigma_A^2 + n\sigma_{AB}^2$	$\sigma^2 + nb\psi_A^2 + n\sigma_{\alpha B}^2$
B	$b-1$	$\sigma^2 + na\psi_B^2$	$\sigma^2 + na\sigma_B^2 + n\sigma_{AB}^2$	$\sigma^2 + na\sigma_B^2 + n\sigma_{\alpha B}^2$
AB	$(a-1)(b-1)$	$\sigma^2 + n\psi_{AB}^2$	$\sigma^2 + n\sigma_{AB}^2$	$\sigma^2 + n\sigma_{\alpha B}^2$
Error	$ab(n-1)$	σ^2	σ^2	σ^2

When factor B is NESTED in factor A, expected means associated with sums of squares are given in the table below:

Source	*df*	A, B fixed	A, B random	A fixed B random
A	$a-1$	$\sigma^2 + nb\psi_A^2$	$\sigma^2 + nb\sigma_A^2 + n\sigma_{B(A)}^2$	$\sigma^2 + nb\psi_A^2 + n\sigma_{B(A)}^2$
$B(A)$	$a(b-1)$	$\sigma^2 + n\psi_{B(A)}^2$	$\sigma^2 + n\sigma_{B(A)}^2$	$\sigma^2 + n\psi_{B(A)}^2$
Error	$ab(n-1)$	σ^2	σ^2	σ^2

where ψ^2 and σ^2 values are defined below

$$\psi_A^2 = \frac{1}{a-1}\sum_1^a \alpha_i^2 \quad \text{effect size of factor } A$$

$$\psi_B^2 = \frac{1}{b-1}\sum_1^b \beta_i^2 \quad \text{effect size of factor } B$$

$$\psi_{AB}^2 = \frac{1}{(a-1)(b-1)}\sum_{i=1}^a\sum_{j=1}^b (\alpha\beta)_{ij}^2 \quad \text{effect size of interaction}$$

$$\psi_{B(A)}^2 = \frac{1}{a(b-1)}\sum_{i=1}^a\sum_{j=1}^b \beta_{j(i)}^2 \quad \text{effect size of factor } B$$

$$\begin{aligned}
\sigma_A^2 &= \text{Var}(A_i) \quad \text{variance component for factor } A\\
\sigma_B^2 &= \text{Var}(B_i) \quad \text{variance component for factor } B\\
\sigma_{AB}^2 &= \text{Var}((AB)_{ij}) \quad \text{variance component for interaction}\\
\sigma_{B(A)}^2 &= \text{Var}(B_{j(i)}) \quad \text{variance component for factor } B\\
\sigma^2 &= \text{Var}(E_{ijk}) \quad \text{error variance}
\end{aligned}$$

The term *effect size* is often used in power considerations and sometimes involves division by σ^2.

Using expected mean squares to analyze data in mixed-effects models

F-tests and estimating variance components.

1. To test for interaction effect, use $F_{AB} = \frac{MS[AB]}{MS[E]}$
2. To test for main effect of A, use $F_A = \frac{MS[A]}{MS[AB]}$
3. To test for main effect of B, use $F_B = \frac{MS[B]}{MS[AB]}$

Note the departure from fixed-effects analysis, where $MS[E]$ is always used in the denominator.

The estimated variance components satisfy the system of equations by equate (observed) mean squares to their expected values.

For example, for a 2 factor crossed, random-effects model

$$\begin{aligned} MS[E] &= \hat{\sigma}^2 \\ MS[AB] &= \hat{\sigma}^2 + n\hat{\sigma}^2_{AB} \\ MS[A] &= \hat{\sigma}^2 + nb\hat{\sigma}^2_A + n\hat{\sigma}^2_{AB} \\ MS[B] &= \hat{\sigma}^2 + na\hat{\sigma}^2_B + n\hat{\sigma}^2_{AB} \end{aligned}$$

Analysis of variance in nested designs

Consider a two-factor design in which factor B is nested in factor A. Let Y_{ijk} denote the k^{th} response at level j of factor B within level i of factor A. A model:

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{ijk}$$

for $i = 1, 2, \ldots, a$, $j = 1, 2, \ldots, b_i$, $k = 1, 2, \ldots, n$ $SS[Tot]$ can be broken down into components reflecting variability due to A, $B(A)$ and variability not due to either factor ($SS[E]$): $SS[Tot] = SS[A] + SS[B(A)] + SS[E]$

$$\begin{aligned} SS[Tot] &= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{+++})^2 \\ SS[A] &= \sum_i \sum_j \sum_k (\bar{y}_{i++} - \bar{y}_{+++})^2 \\ SS[B(A)] &= \sum_i \sum_j \sum_k (\bar{y}_{ij+} - \bar{y}_{i++})^2 \\ SS[E] &= \sum_i \sum_j \sum_k (y_{ijk} - \bar{y}_{ij+})^2 \end{aligned}$$

The ANOVA table looks like

Source	*df*	Sum of Squares	Mean Square	F
A	$a-1$	$SS[A]$	$MS[A] = \frac{SS[A]}{a-1}$	$F_A = \frac{MS[A]}{MS[E]}$
$B(A)$	$\sum_i (b_i - 1)$	$SS[B(A)]$	$MS[B(A)] = \frac{SS[B(A)]}{\sum_i (b_i-1)}$	$F_{B(A)} = \frac{MS[B(A)]}{MS[E]}$
Error	$N - \sum b_i$	$SS[E]$	$MS[E] = \frac{SS[E]}{N-\sum b_i}$	
Total	$N-1$	$SS[Tot]$		

And with random-effect, the test statistic becomes

Test for	$A, B(A)$ fixed	A fixed, $B(A)$ random	$A, B(A)$ random
Factor A	$MS[A]/MS[E]$	$MS[A]/MS[B(A)]$	$MS[A]/MS[B(A)]$
Factor $B(A)$	$MS[B(A)]/MS[E]$	$MS[B(A)]/MS[E]$	$MS[B(A)]/MS[E]$