Associated Reading

Section 17.1 - 17.8

Honeybees

Example 9.1

An entomologist records energy expended (y) by $N = 27$ honeybees at three temperatures (20, 30, 40 C). The honeybees consumed three levels of sucrose (20, 40, 60 %) The 27 honeybees are assigned at random to receive one of the 9 treatment combinations. (Compare to Example 3.6)

Temperature	Sucrose	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	6.0	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

Honeybees

Example 9.1

- Factors: Temperature and Sucrose
- Fixed or Random: Both factors are fixed
- Crossed or Nested: Temperature and Sucrose are crossed
- Name: Two-factor, fixed effects, factorial design
- Model: $y_{ijk} = \mu + \alpha_i + \beta_j + \left(\alpha\beta\right)_{ijk} + \epsilon_{ijk}$

  $\epsilon_{ijk} \overset{iid}{\sim} N\left(0, \sigma^2\right)$

Drug Administration

Example 9.2

Experiment to study effect of drug and method of administration on fasting blood sugar in a random sample of $N = 18$ diabetic patients. (Compare to Example 8.1)

Drug	Type of Administration
Brand I tablet	$30mg \times 1$
	$15mg \times 2$
Brand II tablet	$20mg \times 1$
	$10mg \times 2$
Insulin injection	before breakfast
	before supper

Drug Administration

Example 9.2

- Factors: Drug and Administration
- Fixed or Random: Both factors are fixed
- Crossed or Nested: Nested - Admin(Drug)
- Name: Two-factor, fixed effects, nested design
- Model: $y_{(i)jk} = \mu + \alpha_i + \beta_{(i)j} + \epsilon_{(i)jk}$

  $\epsilon_{(i)jk} \overset{iid}{\sim} N\left(0, \sigma^2\right)$

Milk Pasteurization

Example 9.3

An experiment is conducted to determine variability among laboratories (interlaboratory differences) in their assessment of bacterial concentration in milk after pasteurization. Milk with various degrees of contamination was tested by randomly drawing four samples of milk from a collection of cartons at various stages of spoilage. Each of the four samples was split into 10 parts and two were sent to each of the 5 laboratories. Y is colony-forming units/μl. Labs think they're receiving 8 independent samples.

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

Milk Pasteurization

Example 9.3

- Factors: Lab and Sample
- Fixed or Random: Both factors are random
- Crossed or Nested: Crossed
- Name: Two-factor, random effects, factorial design
- Model: $y_{ijk} = \mu + \alpha_i + \beta_{ij} + (\alpha\beta)_{ij} + \epsilon_{ijk}$

$$\alpha_i \overset{iid}{\sim} N\left(0, \sigma_A^2\right), \quad \beta_i \overset{iid}{\sim} N\left(0, \sigma_B^2\right), \quad (\alpha\beta)_{ij} \overset{iid}{\sim} N\left(0, \sigma_{AB}^2\right)$$

Chicken Processing Plant

Example 9.4

An experiment measures *Campylobacter* counts in $N = 120$ chickens in a processing plant, at four locations, over three days. Sample group means are given below. Students visited the plants on three random sampled winter days. On each day they sampled $n = 10$ chickens at each of four locations, or sites, along the washing line: (before first washer, after 3rd washer, after microbial rinse, after chill tank).

	Location			
Day	Before Washer	After Washer	After mic. rinse	After chill tank
1	70070.00	48310.00	12020.00	11790.00
2	75890.00	52020.00	8090.00	8690.00
3	95260.00	33170.00	6200.00	8370.00

Data courtesy of Michael Bashor, General Mills

Chicken Processing Plant

Example 9.4

- Factors: Location and Day
- Fixed or Random: Location is fixed and Day is random
- Crossed or Nested: Crossed
- Name: Two-factor, mixed effects, factorial design
- Model: $y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

$$\beta_i \overset{iid}{\sim} N\left(0, \sigma_B^2\right), \quad (\alpha\beta)_{ij} \overset{iid}{\sim} N\left(0, \sigma_{AB}^2\right)$$

Acid Concentration

Example 9.5

An experiment to assess the variability of a particular acid among plants was done. Many plants were planted and four were randomly selected. From each plant three leaves were selected to be measured.

	Plant (i)			
Leaf (j)	1	2	3	4
1	11.60	14.03	15.73	7.37
2	16.47	18.56	19.63	8.87
3	18.67	12.10	16.87	10.90

Data from Neter, et al (1996)

Acid Concentration

Example 9.5

- Factors: Plant and Leaf
- Fixed or Random: Both factors are random
- Crossed or Nested: Nested - Leaf(Plant)
- Name: Two factor, random effects, nested design
- Model: $y_{(i)jk} = \mu + \alpha_i + \beta_{(i)j} + \epsilon_{(i)jk}$

$$\alpha_i \stackrel{iid}{\sim} N\left(0, \sigma_A^2\right), \quad \beta_{(i)j} \stackrel{iid}{\sim} N\left(0, \sigma_{B(A)}^2\right)$$

Light Intensity

Example 9.6

Five treatments of light intensity were assigned randomly to ten pots of plants. Each pot had two seedlings per pot. For each seedling the plant height was measured for a total of 20 measurements.

Treatment	Pot	Seedling 1	Seedling 2
1	1	32.94	35.98
1	2	34.76	32.40
2	1	30.55	32.64
2	2	32.37	32.04
3	1	31.23	31.09
3	2	30.62	30.42
4	1	34.41	34.88
4	2	34.07	33.87
5	1	35.61	35.00
5	2	33.65	32.91

Light Intensity

Example 9.6

- Factors: Intensity Treatment and Pot
- Fixed or random: Treatment is fixed and Pot is random
- Crossed or Nested: Nested - Pot(Treatment)
- Name: Two-factor, mixed effects, nested design
- Model: $y_{(i)jk} = \mu + \alpha_i + \beta_{(i)j} + \epsilon_{(i)jk}$

$\beta_{(i)j} \overset{iid}{\sim} N\left(0, \sigma^2_{B(A)}\right)$

Recap

- For two-factor studies there are six possible designs we've seen so far:
  - Fixed and Crossed; Fixed and Nested; Random and Crossed; Random and Nested; Mixed and Crossed; Mixed and Nested
- While the models look *very* similar - they are actually quite different
  - The constraints and assumptions are different

    (though we always assume fixed effects sum to zero, we do not assume this for random effects)
  - The covariance matrix for the parameter estimates will be different
  - The covariance matrix for the responses will be different!
- Inference is different because
  - Expected Mean Squares (EMS) are different
  - Observations can now be correlated (e.g. the birth weight and sire example)

Equating Mean Squares: Method of Moments

- Can be used to estimate variance components
- Also allows us to derive correct F test by using EMS
- To make the notation simpler:
  - **Balanced design:** $n_{ij} = n$ for all (i, j)
  - **Constant number of nested levels:** $b_i = b$ for all i
  - **Fixed effects:** ψ_A^2, $\psi_{B(A)}^2$, etc. used to denote sum of squared fixed effects

EMS: Two-Factor, Factorial Design

When factors A and B are crossed we have three possible scenarios.

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

Note: fixed effects get a ψ^2 term; random effects get a σ^2 term.
More on this later...

Note: These EMS assume no “sum to zero” constraint for random effects

EMS: Two-Factor, Nested Designs

When factor B is nested in factor A, we have the same three possible scenarios.

Source	df	A, B fixed	A, B random	A fixed, B random
A	$a-1$	$\sigma^2 + nb\psi^2_A$	$\sigma^2 + nb\sigma^2_A + n\sigma^2_{B(A)}$	$\sigma^2 + nb\psi^2_A + n\sigma^2_{B(A)}$
$B(A)$	$a(b-1)$	$\sigma^2 + n\psi^2_{B(A)}$	$\sigma^2 + n\sigma^2_{B(A)}$	$\sigma^2 + n\sigma^2_{B(A)}$
Error	$ab(n-1)$	σ^2	σ^2	σ^2

Note: fixed effects get a ψ^2 term; random effects get a σ^2 term.
More on this next...

Note: These EMS assume no "sum to zero" constraint for random effects

Defining the Terms

$$\psi_A^2 = \frac{\sum_{i=1}^{a} \alpha_i^2}{a-1} \qquad \sigma_A^2 = \mathrm{V}\left[\alpha_i\right]$$

$$\psi_B^2 = \frac{\sum_{j=1}^{b} \beta_j^2}{b-1} \qquad \sigma_B^2 = \mathrm{V}\left[\beta_i\right]$$

$$\psi_{AB}^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} (\alpha\beta)_{ij}^2}{(a-1)(b-1)} \qquad \sigma_{AB}^2 = \mathrm{V}\left[(\alpha\beta)_{ij}\right]$$

$$\psi_{B(A)}^2 = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b} \beta_{(i)j}^2}{a(b-1)} \qquad \sigma_{B(A)}^2 = \mathrm{V}\left[\beta_{(i)j}\right]$$

$$\sigma^2 = \mathrm{V}\left[\epsilon_{ijk}\right] = \mathrm{V}\left[\epsilon_{(i)jk}\right]$$

- For a given factor:
  - **Effect sizes** (ψ^2 terms) measure the overall difference in means
  - **Variance components** (σ^2 terms) measure the overall variance

So many formulas!

How are EMS used to create a test?!

1. Identify the factor for which you want to construct a test
2. Write down the EMS for this factor
3. Write down your null hypothesis
4. Re-write your factor's EMS (from Step 2) ***assuming your null is true***
5. Locate an EMS in your table that matches your answer in Step 4*
6. Your F_{obs} statistic is the ***Observed*** Mean Square from Step 2 divided by the ***Observed*** Mean Square from Step 5
7. The DF for your F statistic are the DF associated with the Observed Mean Squares in Step 6

*This may not always be possible - if not you need to find a way to approximate the F test. (Or just use another method...)

Using EMS: An Example

Suppose we consider a case where A is fixed, B is random, and A and B are crossed.

1. Find the correct test for Factor A
2. $EMS = \sigma^2 + nb\psi_A^2 + n\sigma_{AB}^2$
3. $H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_a = 0$
4. H_0 true $\Rightarrow \psi_A^2 = 0 \Rightarrow EMS = \sigma^2 + n\sigma_{AB}^2$
5. Interaction has $EMS = \sigma^2 + n\sigma_{AB}^2$ exactly!
6. $F_{obs} = \frac{MSA}{MSAB}$
7. Use $F(a-1,\ (a-1)(b-1))$ to compute $p-value$

The denominator wasn't MSE!
Very common in many designs and it is exactly why understanding EMS is crucial!

Knowing EMS in Balanced Designs

- More than two factors?! We can't list all possible ANOVA tables!
- Remember - each random effect gets a σ^2 and fixed effects get a ψ^2
- Random effects dominate - if a fixed and random factor interact/nest then treat it as random

Suppose you have a three-factor, crossed design with B fixed and A and C random. We know the model is

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

1. **EMS for B:** $\sigma^2 + nac\psi^2_B + nc\sigma^2_{AB} + na\sigma^2_{BC} + n\sigma^2_{ABC}$
2. **EMS for ABC:** $\sigma^2 + n\sigma^2_{ABC}$
3. **EMS for A:** $\sigma^2 + nbc\sigma^2_A + nc\sigma^2_{AB} + nb\sigma^2_{AC} + n\sigma^2_{ABC}$
4. **Test Factor B?:** No matching EMS in ANOVA table - need to use $MSAB + MSBC - MSABC$