Associated Reading

Sections 17.1 - 17.2

Birth weights example

Example 7.1

- Response variable = birth weight
- $t = 5$ sires are selected and each mated to $n = 8$ dams
- $N = 40$ calves from a balanced, factorial, CRD (data below)

Sire #	Factor Level	Calf 1	Calf 2	$\cdots$	Calf 7	Calf 8	$\bar{y}_{i\cdot}$	$s_{i\cdot}$
177	1	61	100	$\cdots$	75	62	83.6	22.6
200	2	75	102	$\cdots$	98	94	97.5	11.2
201	3	58	60	$\cdots$	54	100	63.1	15.0
202	4	57	56	$\cdots$	101	101	77.5	25.9
203	5	59	46	$\cdots$	105	75	91.0	28.0

Identifying Effect Type

Criteria	Fixed	Random
Choosing Levels?	finite number of possible levels	select from "infinite" population of levels
Repeating your experiment?	must use exactly the same levels as previous experiment	new sample of levels from same population
Goal of current experiment?	Estimate population means for the current factor levels, not the variance components for the factor	Estimate variance components for the factor, not the population means of the current levels
Inferences will about...?	set of levels that ***were used*** in the experiment	population of all levels that ***could have been used*** in the experiment

Fixed vs. Random Examples

- Antibiotics in the Binding Fraction Example were **fixed effects**
  - The model we used was a **one-way, fixed effects, ANOVA model**
- Sires in the Birth Weights Example are **random effects**
  - The model we would use here is a **one-way, random effects, ANOVA model**

You wish to compare the effect of three different fertilizers on the yield (in bushels) of apples from trees in your orchard. You want to compare the fertilizers across apple varieties, so you select 5 of your commercial varieties.

- Is fertilizer a fixed or random effect?
- Variety is what type of effect?
- What type of model should be fit to this data?

Fixed vs. Random Effects Models

	Fixed Effects	Random Effects
Model?	$y_{ij} = \underbrace{\mu}_{\text{fixed}} + \underbrace{\tau_i}_{\text{fixed}} + \underbrace{\epsilon_{ij}}_{\text{random}}$	$y_{ij} = \underbrace{\mu}_{\text{fixed}} + \underbrace{\tau_i}_{\text{random}} + \underbrace{\epsilon_{ij}}_{\text{random}}$
Assumptions?	$\epsilon_{ij} \overset{iid}{\sim} N(0,\sigma^2)$	$\epsilon_{ij} \overset{iid}{\sim} N\left(0,\sigma^2\right)$ $\tau_i \overset{iid}{\sim} N\left(0,\sigma_T^2\right)$ all τ_i, ϵ_{ij} are independent
Effect?	τ_i = mean of the i^{th} group	τ_i = outcome of your random level selection
Expected Value?	$\mathrm{E}\left[y_{ij}\right] = \mu + \tau_i$	$\mathrm{E}\left[y_{ij}\right] = \mu$
Variance?	$\mathrm{V}\left[y_{ij}\right] = \sigma^2$	$\mathrm{V}\left[y_{ij}\right] = \sigma_T^2 + \sigma^2$

Variance Components!

- Goal of RE: Estimate **variance components** of the random effects
- One-way Random Effects Model:
  - Total Variance of any observation is $\sigma_y^2 = \sigma_T^2 + \sigma^2$
  - σ_T^2 Component: variance associated with randomly selecting t factor levels
  - σ^2 Component: variance associated with randomly selecting N experimental units in a CRD
- What does it mean if $\sigma^2 = 0$? What if $\sigma^2 > 0$?
- What does it mean if $\sigma_T^2 = 0$? What if $\sigma_T^2 > 0$?

Conceptual Differences and Similarities

- Fixed and Random Effects Models clearly have the same form
    - Parameters have different interpretations
    - Assumptions are not identical
- What about the analysis?
    - ANOVA table looks exactly the same!
    - Still use F test to make inference about the factor!
    - Rationale for the test, hypotheses, and conclusion are all different
- How can we use the same F test but for a different set of hypotheses and rationale?!

Expected Mean Squares: aka Why do F tests work?

- Independence, normality, homoscedasticity assumptions $\Rightarrow$ Mean Squares follow prescribed distributions
- CRD + Fixed Effects Assumptions + ST 511 knowledge $\Rightarrow$
  - $\mathrm{E}\left[MSE\right] = \sigma^2$
  - $\mathrm{E}\left[MSTrt\right] = \sigma^2 + \frac{n}{t-1}\sum_{i=1}^{t}\tau_i^2$
  - What happens to $F = \frac{MSTrt}{MSE}$ under H_0? Under H_1?
- CRD + Random Effects Assumptions + ST 511 Knowledge $\Rightarrow$
  - $\mathrm{E}\left[MSE\right] = \sigma^2$
  - $\mathrm{E}\left[MSTrt\right] = \sigma^2 + n\sigma_T^2$

ANOVA Table and Overall F test

Source	DF	SS	MS	F
Treatments	t - 1	SSTrt	$MSTrt = \frac{SSTrt}{t-1}$	$F = \frac{MSTrt}{MSE}$
Error	n - t	SSE	$MSE = \frac{SSE}{n-t}$	
Total	n - 1	SST		

- Overall F test hypothesis: $H_0 : \sigma_T^2 = 0$ vs $H_1 : \sigma_T^2 > 0$
- What happens to F under H_0? Under H_1?

Estimating the model parameters

- Estimating the overall mean is unchanged: $\hat{\mu} = \overline{y}_{..}$
- What about the variance components? There's more than one approach!
- Approach #1: Equating Mean Squares (formal name is **<u>Method of Moments</u>**)
  - $\sigma^2 = \mathrm{E}[MSE] \Rightarrow \hat{\sigma}^2 = MSE$
  - $\sigma^2 + n\sigma_T^2 = \mathrm{E}[MSTrt] \Rightarrow \hat{\sigma}^2 + n\hat{\sigma}_T^2 = MSTrt \Rightarrow \hat{\sigma}_T^2 = \frac{MSTrt - MSE}{n}$
- Pros and Cons? → Easy to use and easy to solve, but poor mathematical properties.
- Approach #2: **<u>Maximum Likelihood</u>** is preferred, but formulas/justification too complicated for ST 512.

Back to our example: Hypothesis Testing

Example 7.2

Source	DF	SS	MS	F	p
Sire	4	5591.15	1397.78750	3.01	0.0309
Error	35	16232.75	463.79286		
Total	39	21823.90			

1. Does the population of sires vary in terms of the birth weight of offspring? (Equivalently, does sire contribute significantly to the variability in birth weight of calves?)
2. What aspects of this analysis would be different for had we treated sire as a fixed effect?
3. Estimate the variance in birth weights by equating mean squares (e.g. method of moments).
4. Estimate the proportion of this variance due to the effect of randomly selecting sires.
5. Estimate the proportion of this variance due to the effect of randomly assigning sires to dams.

Answering the questions: Part 1!

Example 7.2

1. Does the population of sires vary in terms of the birth weight of offspring? (Equivalently, does sire contribute significantly to the variability in birth weight of calves?)
   - $y_{ij} = \mu + \tau_i + \epsilon_{ij}$
   - μ is the mean birth weight for all calves, τ_i is the random effect from the i^{th} sire, and ϵ_{ij} is the residual for the j^{th} calf from sire i.
   - $H_0 : \sigma^2_{sire} = 0$ vs. $H_1 : \sigma^2_{sire} > 0$
   - $\tau_i \overset{iid}{\sim} N\left(0, \sigma^2_{sire}\right)$, $\epsilon_{ij} \overset{iid}{\sim} N\left(0, \sigma^2\right)$, and each τ_i and ϵ_{ij} are independent
   - $F = \frac{MS\ Sire}{MSE} = \frac{1397.78750}{463.79286} = 3.01$
   - $p - value = 0.0309$
   - $p - value = 0.0309 < 0.05 = \alpha$, so we reject the null hypothesis
   - There is sufficient sample evidence to suggest a portion of the variability in birth weight is due to different sires.

Answering the questions: Part 2!

Example 7.2

2. What aspects of this analysis would be different for had we treated sire as a fixed effect?
   Research question, hypotheses, assumptions, conclusion. (Mechanics such as F and p values would remain unchanged.)

3. Estimate the variance in birth weights.

$$\hat{\sigma}^2 = MSE = 463.79286$$

$$\hat{\sigma}^2_{sire} = \frac{MS\ Sire - MSE}{n}$$

$$= \frac{1397.78750 - 463.79286}{8} = \boxed{116.74933}$$

$$\hat{\sigma}^2_y = \hat{\sigma}^2 + \hat{\sigma}^2_{sire}$$

$$= 463.79286 + 116.74933 = \boxed{580.54219}$$

Answering the questions: Part 3!

Example 7.2

4. Estimate the proportion of this variance due to the effect of randomly selecting sires.

$$\frac{\hat{\sigma}^2_{sire}}{\hat{\sigma}^2 + \hat{\sigma}^2_{sire}} = \frac{116.74933}{580.54219} = 0.2011$$

5. Estimate the proportion of this variance due to the effect of randomly assigning sires to dams.

$$\frac{\hat{\sigma}^2}{\hat{\sigma}^2 + \hat{\sigma}^2_{sire}} = \frac{463.79286}{580.54219} = 0.7989$$