

Lesson 23: Fixed, Random and Mixed Effects

MA376 Applied Statistics, AY27-1

Cadet: _____

Section: _____

Objectives

1. **Frame.** A factor is fixed when its levels are the only ones you care about and random when they are a sample of many.
2. **Apply.** Fit a random-intercept model with `lme4` and recover the ICC from its variance components.
3. **Mean.** A random intercept shrinks a small group toward the overall mean. The ICC is the correlation between two observations in the same group.
4. **Communicate.** Explain what a company effect means when the companies are a sample.

1. The AFT data

Fall Army Fitness Test (AFT) scores, 0 to 500. _____ cadets in _____ companies. Fewest cadets in a company: _____ Most: _____

The commander's question: which company is the most fit?

1. Look at the plot of company averages. Which companies sit at the top and the bottom? How many cadets did they test?

2. Model 1: no company effect

$$y_{ij} = \mu + \varepsilon_{ij} \quad \text{lm(score ~ 1)} \quad \hat{\mu} = \underline{\hspace{2cm}}$$

Predict a cadet in	D3	H2	J5 (new company)
Model 1			

2. What does Model 1 tell the commander? Does the plot agree?

3. Model 2: fixed company effect

$$y_{ij} = \beta_0 + \beta_j + \varepsilon_{ij} \quad \text{lm(score ~ company)} \quad \hat{\beta}_0 \text{ (A1's average)} = \underline{\hspace{2cm}}$$

	D3	H2	J5 (new company)
Cadets tested n_j			0
$\hat{\beta}_j$			
Prediction $\hat{\beta}_0 + \hat{\beta}_j$			

3. Model 2 says D3 is the most fit. Why should the commander be careful with that answer? What happens with J5?

4. Model 3: random company effect

$$y_{ij} = \mu + u_j + \varepsilon_{ij}, \quad u_j \sim N(0, \sigma_u^2), \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad \text{lmer(score} \sim 1 + (1 \mid \text{company}))$$

From `summary(fit3)`:

$$\hat{\mu} \text{ (Intercept)} = \underline{\hspace{2cm}} \quad \hat{\sigma}_u^2 \text{ (company)} = \underline{\hspace{2cm}} \quad \hat{\sigma}^2 \text{ (Residual)} = \underline{\hspace{2cm}}$$

The Model 3 formula.

$$\hat{y}_{ij} = \hat{\mu} + \hat{u}_j = \hat{\mu} + w_j (\bar{y}_j - \hat{\mu})$$

The overall average plus a share w_j of the company's gap: w_j runs from 0 (Model 1) to 1 (Model 2).

$$w_j = \frac{n_j}{n_j + \lambda} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}^2/n_j}, \quad \lambda = \frac{\hat{\sigma}^2}{\hat{\sigma}_u^2} = \underline{\hspace{2cm}}$$

	D3	H2	J5 (new company)
n_j			0
$\bar{y}_j$			none
$w_j = n_j/(n_j + \lambda)$			
$\hat{u}_j = w_j (\bar{y}_j - \hat{\mu})$			
$\hat{y} = \hat{\mu} + \hat{u}_j$			

A company needs about _____ cadets before its own average gets half the weight.

Why shrink? D3's average is off by about $\sqrt{\hat{\sigma}^2/n_j} = \underline{\hspace{2cm}}$ points from luck alone. Companies really differ by about $\hat{\sigma}_u = \underline{\hspace{2cm}}$.

4. What makes λ big? What makes it small? Why did D3 move so much more than H2?

5. Shrinkage

Rank	Model 2: company (cadets)	Estimate	Model 3: company (cadets)	Estimate
1				
2				
3				
4				
5				

5. Which companies moved the most between the two rankings? What do they have in common?

6. Side by side

	Parameters	D3	H2	J5 (new company)
Model 1: no company effect				
Model 2: fixed company effect				
Model 3: random company effect				

7. How much does company matter? The ICC

$$\text{ICC} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + \hat{\sigma}^2} = \underline{\hspace{2cm}}$$

_____ of the variation in a cadet's score comes from the company. Two cadets in the same company have scores correlated at _____.

The ICC sets the pull. $\lambda = \frac{1 - \text{ICC}}{\text{ICC}} = \underline{\hspace{2cm}}$ Same λ as Part 4? _____

6. If the ICC were 0.50 instead, would small companies get pulled more or less? Why?

8. Tell the commander

7. In two or three sentences: how much does company matter, and which company is the most fit?

9. Fixed or random?

	Fixed	Random
The levels		
The question		
Estimates		
Small groups		
Pick it when		

Mixed model: some effects fixed, some random. Model 3 is one: _____ fixed, _____ random.

8. Fixed or random? Circle one.

- | | | |
|---|-------|--------|
| 1. Three fertilizer brands. You will buy one. | fixed | random |
| 2. 12 recruiting stations sampled from the hundreds in USAREC. | fixed | random |
| 3. Day 1 and Day 2 sections of MA376. | fixed | random |
| 4. 30 patients in a drug trial, each measured every week. | fixed | random |
| 5. The four AFT test sites at West Point. The Master of the Sword wants to know which one scores low. | fixed | random |
| 6. 20 Army units chosen for a survey on retention. | fixed | random |

10. Practice: radon in Minnesota

L23_radon_counties.csv: _____ homes in _____ counties. log_radon is radon on the log scale. Fewest homes in a county: _____ Most: _____

The state's question: which county has the worst radon?

P1. Highest county average: _____ on _____ homes. Lowest: _____ on _____ homes.

P2. Model 2, $\text{lm}(\log_radon \sim \text{county})$, calls the worst county _____, at _____.

P3. Model 3, $\text{lmer}(\log_radon \sim 1 + (1 \mid \text{county}))$:

$\hat{\mu} =$ _____ $\hat{\sigma}_u^2 =$ _____ $\hat{\sigma}^2 =$ _____ $\lambda =$ _____

P4 and P5. Fill in both forms of w and the prediction.

	Lac Qui Parle	St. Louis	a county never tested
n_j			0
$\bar{y}_j$			none
$w_j = n_j / (n_j + \lambda)$			
$w_j = \hat{\sigma}_u^2 / (\hat{\sigma}_u^2 + \hat{\sigma}^2 / n_j)$			
$\hat{y} = \hat{\mu} + w_j (\bar{y}_j - \hat{\mu})$			

P6. ICC = _____. _____ of the radon variation is county to county.

P7. Top five counties.

Rank	Model 2: county (homes)	Model 3: county (homes)
1		
2		
3		
4		
5		

Who drops out, and why?

P8. Tell the state in two or three sentences.

P9. An inspector asks, "Is Lac Qui Parle worse than Blue Earth?" Fixed or random? Why?

Key Takeaway. Grouped data gives three choices: ignore the group, a dummy for each group, or a random intercept. Dummies take every group's average at face value, so small groups land at the extremes by luck. A random intercept blends each group's own average with the overall average, and pulls a small group the most. Fixed: you care about these levels. Random: the levels are a sample, and you care about how much they vary.