DESIGNS WITH RANDOM EFFECTS

Chapter 17

LEARNING OBJECTIVES

- Argue whether a factor should be fixed or random
- Write statistical models and their new assumptions when including random effects
- Write statistical model for CRD that includes analysis of subsampling and random block models
- Perform analysis and interpret output in R (finals week Lab)

FIXED AND RANDOM EFFECTS

- Each treatment effect is said to be a fixed effect
  - Effect is repeatedly observable across different EUs
- Responses also affected by uncontrollable variation

$$Y_{ij} = \mu + \tau_i + E_{ij}$$

- Error terms also affect response but in a much less predictable way
- E_{ij}'s are your first introduction to random effects
  - Observed effects generated from normal distribution (not repeatable by experimenter)

INFERENCE FOR RANDOM FACTORS

- **Fixed factor:** variable whose levels are pre-selected
  - Effect on the response is repeatable, call it a **fixed effect**
- **Random factor:** variable whose levels are a random sample of all possible levels
  - Effect on response not repeatable without knowledge of level, call these **random effects**
  - Determine whether the **variability of the levels** produces **variability in the response**

- With CRD model, cared only about the **variability of the EUs**, not estimating the individual EU effects
  - If we did the CRD again with new EUs we would expect new random effects, but their variance should be the same

FIXED OR RANDOM?

- Difficult to decide whether a factor is fixed or random
- Fixed factor tips:
  - Levels specifically chosen and only care about inference between these chosen levels
  - Level choices and their effects are repeatable
- Random factor tips:
  - Levels represent random sample from larger population of possible levels
  - Have no control over the choice of levels
  - Observed a reasonable number of such levels***
  - Care only about estimating variability between effects

BREAD EXAMPLE

- Remember the bread example from early on
  - 3 treatments: rise time = 35, 40, 45
  - 4 loaves of bread per rise time, all from larger batch
- Rise time is a fixed factor because this is set by the experimenter, even though there are a large number of possible rise times
  - In terms of this specific data collection procedure, the levels were not randomly generated, only randomly assigned
- Even if we wanted to consider it as a random factor, we haven't observed many of these levels
  - Our estimate for the rise time variance would be bad

INTERLABORATORY TESTING

- Apolipoproteins (Apo A-I) known to clear cholesterol from arteries
- Compare measured Apo A-I concentrations across laboratory on same reference material
- All of their measurements should give the same value

- Want to determine:
  - Measurement variability between labs (interlaboratory)
  - Measurement variability within lab (intralaboratory)

- Had a total of 28 labs each with 7-10 measurements

INTERLABORATORY TESTING

- Lab is now a factor with 28 levels (random because it is a random sample of all possible labs)
- If we cared about inference for only the 28 labs we would treat it as a fixed factor
- Statistical model has two random effects

$$Y_{ij} = \mu + T_i + E_{ij}$$

$$\left.\begin{aligned} T_i &\sim^{iid} N(0, \sigma_T^2) \\ E_{ij} &\sim^{iid} N(0, \sigma_E^2) \end{aligned}\right\} \text{Mutually indpt}$$

- Notation: use capital Roman letters for random factors, Greek letters for fixed factors

INTERLABORATORY TESTING

- This is not a randomized, comparative experiment because we are not applying any treatment
- Goal: Determine whether $\sigma_T^2 > 0$
- Analysis challenge: measurements with the same lab will have the exact same $T_i = t_i$ level
- Result: repeated measurements from same level of random factor are correlated with each other
- How can we estimate the two variances σ_T^2 , σ_E^2

MIXED MODELS

- Focus our time on introducing random factors into designs we have already seen
- Introduces models with both fixed and random effects, call these mixed models
- First introduction to this will be an analysis of a CRD with duplicate observations per EU
- Before we just averaged over the duplicates but this is throwing away valuable information!

CRD WITH SUBSAMPLING
PAPER AIRPLANES

- Replicate our classroom experiment to allow each person to throw paper airplane multiple times.

- EU: Paper which is constructed into airplane
- OU: Each throw of airplane.

- Analyzing each throw as a replicated treatment would be pseudoreplication.

GOLF EXAMPLE

- Recall golfer example when we talked about GRCBDs
  - 10 golfers, 3 tee heights, 5 replicates per tee height
- Not interested in making comparisons between specific golfers so we could consider this to be a random effect
- Golfer effect: $B_h \sim^{iid} N(0, \sigma_B^2)$
- Treatment effect: τ_i
- Golfer/treatment interaction: $B\tau_{hi} \sim^{iid} N(0, \sigma_{Bt}^2)$
- Interaction effect random because it involves random factor

GRCBD STATISTICAL MODEL

- Model looks similar as before, but now with random effects instead

$$Y_{hij} = \mu + B_h + \tau_i + B\tau_{hi} + E_{hij}$$

- The result is the following

$$E(Y_{hij}) = \mu + \tau_i \qquad Var(Y_{hij}) = \sigma^2_B + \sigma^2_{Bt} + \sigma^2_E$$

LEARNING OBJECTIVES REVIEW

- Argue whether a factor should be fixed or random
- Write statistical models and their new assumptions when including random effects
- Write statistical model for CRD that includes analysis of subsampling and random block models
- Perform analysis and interpret output in JMP