Applied Statistical Methods II

Chapter #25

Random and Mixed Effects Models

Looking forward:

- Random Effects Models
- Start with one random effect model.

An Example

- The gene Per3 is a gene that helps regulate a person's circadian pattern.
- We want to know if the level of expression of this gene is the same among all shift workers.
- The expression level is measured by:
  1. Taking a blood sample.
  2. Extracting the mRNA from this sample.
  3. Using a microarray to measure the amount of mRNA from Per3.

- The processing, extracting, and measuring of mRNA makes recorded variables indirect measures that have random noise.
- Several samples are taken from each subject which are assumed to be noisy measures of a subject's mean expression level.
- In our study, we have 5 subjects and take 4 samples from each subject.

Random Effect Model

- We assume that the distribution of measured expression from samples from subject i is $N(\mu_i, \sigma^2)$.
  - Same idea as the fixed effects one-way ANOVA.
- We assume that the mean subject expression levels μ_i among shift workers are random variables
  - μ_i iid $\sim N(\mu_., \sigma_\mu^2)$
- Y_{ij} is the expression for the j^{th} sample from subject i.
- The random effects model is:
  - $Y_{ij} = \mu_i + \epsilon_{ij}$
  - $\mu_i \sim$ iid $N(\mu_., \sigma_\mu^2)$
  - $\epsilon_{ij} \sim$ iid $N(0, \sigma^2)$
  - μ_i is independent of ϵ_{jk}
- Also known as a repeated measures model.

Philosophy of Random Effects

- Our primary interest is in the variability of mean expression among shift-workers.
- Might have a small interest in the difference in mean expressions between two subjects, but not main goal.
- Philosophical differences:
  - Fixed Effects - each subject has his own mean rating.
  - Random Effects - mean rating of subject is a random variable.
- Computational differences between fixed and random effects models:
  - With balanced one-factor study fit through sums-of-squares - nothing.
  - With unbalanced data fit through sums-of-squares - a lot.
  - Through MLE - can speak about within group variance only in random-effects models.

Properties of Random Effect Model

$$EY_{ij} = E(\mu_i) + E(\epsilon_{ij}) = \mu.$$

$$\begin{aligned} Var(Y_{ij}) &= Var(\mu_i) + Var(\epsilon_{ij}) \\ &= \sigma_\mu^2 + \sigma^2 \end{aligned}$$

$$\begin{aligned} Cov(Y_{ij}, Y_{kl}) &= 0 \\ Cov(Y_{ij}, Y_{ik}) &= Var(\mu_i) = \sigma_\mu^2 \end{aligned}$$

$$\begin{aligned} Y_{ij} &= \mu. + \epsilon_{ij}^* \\ \epsilon_{ij}^* &\sim N(0, \sigma^2 + \sigma_\mu^2) \\ Cov(\epsilon_{ij}^*, \epsilon_{k\ell}^*) &= \delta_{ik}\sigma_\mu^2 + \delta_{ik}\delta_{j\ell}\sigma^2 \end{aligned}$$

Intraclass Correlation

- The intraclass correlation is the amount of variance accounted for by the random effect.
  - % of total variability attributed to there being a random factor
- Also the correlation between any two responses from the same subject.

$$
\begin{aligned}
Corr(Y_{ij}, Y_{ik}) &= \frac{Cov(Y_{ij}, Y_{ik})}{\sqrt{Var(Y_{ij})Var(Y_{ik})}} \\
&= \frac{\sigma_\mu^2}{\sigma^2 + \sigma_\mu^2} \\
&= \frac{Var(\mu_i)}{Var(Y_{ij})}
\end{aligned}
$$

Testing Random Effect

If there is not much variability between subjects:

- It would be hard to distinguish samples from one subject from another.
- σ_μ^2 should be small.
- To test if there is a significant difference in expression among different subjects:
  - $H_0 : \sigma_\mu^2 = 0$
  - vs $H_a : \sigma_\mu^2 > 0$

Approaches to Testing H_0

Sums-of-Squares/OLS approach:

- Can use various sums-of-squares to estimate variance components (σ^2 and σ_μ^2).
- Only nice if you have balanced data. Can make adjustment for unbalanced data.
- Will focus on sum-of-squares approach for balanced data today.
- Can be fit by Proc GLM or Proc MIXED

Maximum Likelihood Approach

- Write out the likelihood function for the model.
- Estimate variance components by maximizing the likelihood.
- Non-linear and must use an interactive algorithm.
- Fit by Proc Mixed.
- Valid for unbalanced data.

Approach to Estimation

- For now: ASSUME BALANCED DATA.
  - Y_{ij} where $i = 1, \ldots, r$ and $j = 1, \ldots, n$
- Fit the fixed effects one-way ANOVA to get MSTR and MSE.
- We will show that
  - $E(MSE) = \sigma^2$
  - $E(MSTR) = \sigma^2 + n\sigma_\mu^2$
  - Show that F-test from the one-factor fixed effect ANOVA can test $H_0 : \sigma_\mu^2 = 0$.

Some Notation

Since we have balanced data:

$$\begin{aligned}
\overline{Y}_{i\cdot} &= n^{-1} \sum_j Y_{ij} \\
\bar{\epsilon}_{i\cdot} &= n^{-1} \sum_j \epsilon_{ij} \\
\overline{Y}_{\cdot\cdot} &= r^{-1} \sum_i \overline{Y}_{i\cdot} \\
\bar{\epsilon}_{\cdot\cdot} &= r^{-1} \sum_i \bar{\epsilon}_{i\cdot} \\
\bar{\mu}_{\cdot} &= r^{-1} \sum_i \mu_i
\end{aligned}$$

MSE

- $SSE = \sum_{ij} \left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2$
- $MSE = SSE/(r(n-1))$
- Conditional on μ_i, Y_{ij} are independent $N(\mu_i, \sigma^2)$
- Properties of the sample variance
  $E_{\mu_i}\left\{E\left[(n-1)^{-1}\sum_j \left(Y_{ij} - \overline{Y}_{i\cdot}\right)^2 \mid \mu_i\right]\right\} = E_{\mu_i}\sigma^2 = \sigma^2$
- Average over the r groups/subjects:
  $EMSE = \sigma^2$

MSTR

$$\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot} = (\mu_i - \overline{\mu}_{\cdot}) + (\overline{\epsilon}_{i\cdot} - \overline{\epsilon}_{\cdot\cdot})$$

$$SSTR = \sum_{i,j} (\mu_i - \overline{\mu}_{\cdot})^2 + \sum_{i,j} (\overline{\epsilon}_{i\cdot} - \overline{\epsilon}_{\cdot\cdot})^2 + 2\sum_{i,j} (\mu_i - \overline{\mu}_{\cdot})(\overline{\epsilon}_{i\cdot} - \overline{\epsilon}_{\cdot\cdot})$$

$$E(\mu_i - \overline{\mu}_{\cdot}) = 0, \text{ so cross term is zero.}$$

Conditional on group, use the sample covariance properties:

$$E\left[(r-1)^{-1} \sum (\mu_i - \overline{\mu}_{\cdot})^2\right] = \sigma_\mu^2$$

- $\bar{\epsilon}_{i\cdot} \sim N(0, \sigma^2/n)$
- Sample variance properties tell us that

$$E\left[(r-1)^{-1}\sum(\bar{\epsilon}_{i\cdot}-\bar{\epsilon}_{\cdot\cdot})^2\right]=\sigma^2/n$$

- Putting it all together:

$$EMSTR=\sigma^2+n\sigma_\mu^2$$

Testing $H_0 : \sigma^2_\mu = 0$

- $MSTR/MSE \geq 1$.
- $E(MSTR) = E(MSE)$ only when $\sigma^2_\mu = 0$.
- Can show that $F^* = MSTR/MSE \sim F_{r-1,r(n-1)}$ when $\sigma^2_\mu = 0$.
  - Stack to get $Y \sim N(\mu.1, V)$
  - $Z = Y/\sqrt{n\sigma^2_\mu + \sigma^2}$
  - Note the form of V!
  - $(r-1)MSTR/E(MSTR) = \frac{(r-1)MSTR}{n\sigma^2_\mu+\sigma^2} = Z'AZ$
  - AV is idempotent, $rk(A) = r - 1$, $1'A1 = 0$.
  - so $\frac{(r-1)MSTR}{n\sigma^2_\mu+\sigma^2} \sim \chi^2_{r-1}$.
  - Similarly, $\frac{r(n-1)MSE}{\sigma^2} \sim \chi^2_{r(n-1)}$.
  - How do you show that these are independent?

- Can use the fixed effects ANOVA F-test to test if mean expression differs by subject.
- Adjustment can be made for unequal sample sizes.
    - Unbalanced data become a problem with more than one random effect.

Mean Estimate

- The random effects model can also estimate the overall mean score.
- $E\overline{Y}_{..} = \mu_{.}$ so the overall mean is an unbiased estimate.

$$
\begin{aligned}
Var(\overline{Y}_{..}) &= Var(\overline{\mu}_{.}) + Var(\overline{\epsilon}_{..}) \\
&= \sigma_{\mu}^2/r + \sigma^2/(nr) \\
&= (\sigma^2 + n\sigma_{\mu}^2)/(rn)
\end{aligned}
$$

- Consistent as both the number of subjects and the number of samples go to infinity.
- What is the asymptotic variance if we only have a fixed number of subjects but can take as many samples as we want?

Inference on $\mu.$

- $E(MSTR/(rn)) = \text{Var}(\overline{Y}_{..})$
- After some work:
  - $MSTR$ is independent of $\overline{Y}_{..}$
  - $MSTR \sim \frac{\sigma^2+n\sigma_\mu^2}{r-1}\chi^2_{r-1}$
- $\frac{\overline{Y}_{..}-\mu_.}{s(\overline{Y}_{..})} \sim t_{r-1}$
  - $s(\overline{Y}_{..}) = \sqrt{MSTR/(rn)}$
- Can preform inference and make confidence intervals:
  - i.e: $\overline{Y}_{..} \pm t_{r-1}(1-\alpha/2)s(\overline{Y}_{..})$
- Let's explore in SAS.

Estimation of ICC

- Remember the intraclass correlation coefficient
  - $ICC = \frac{\sigma_\mu^2}{\sigma^2+\sigma_\mu^2}$
  - Correlation between two observations from the same subject.
- Some properties:
  - SSTR is independent of SSE
  - $SSTR \sim (\sigma^2 + n\sigma_\mu^2)\chi^2_{r-1}$
  - $SSE \sim \sigma^2\chi^2_{r(n-1)}$
- $\frac{MSTR}{n\sigma_\mu^2+\sigma^2}\frac{\sigma^2}{MSE} = \frac{MSTR}{MSE}\frac{\sigma^2}{\sigma^2+n\sigma_\mu^2} \sim F_{r-1,r(n-1)}$

- “Inverting” we can find that a $(1-\alpha)\%$ confidence interval for σ_μ^2/σ^2 is $[L, U]$.
  - $L = \frac{1}{n}\left[\frac{MSTR}{MSE}\left(\frac{1}{F_{r-1,r(n-1)}(1-\alpha/2)}\right)-1\right]$
  - $U = \frac{1}{n}\left[\frac{MSTR}{MSE}\left(\frac{1}{F_{r-1,r(n-1)}(\alpha/2)}\right)-1\right]$
- And $(1-\alpha)\%$ confidence interval for $ICC = \sigma_\mu^2/(\sigma^2+\sigma_\mu^2)$ is $[L^*, U^*]$.
  - $L^* = \frac{L}{1+L}$
  - $U^* = \frac{U}{1+U}$
- Note the two-side nature of the test despite it begins with an F-test.
- How to estimate ICC?
- Can program yourself.

Estimation of σ^2 and σ_μ^2

- What if you wanted to estimate and get confidence intervals for σ^2 and σ_μ^2?
- σ^2 is easy.
  - MSE is an unbiased estimator and we know its distribution.
- σ_μ^2 is harder.
  - Do not have a mean sums of squares estimator of σ_μ^2.
  - Must take the linear combination $\frac{MSTR - MSE}{n}$.
  - Do not know the exact distribution of linear combinations of sums-of-squares.

Estimating σ^2

- Already showed that *MSE* is an unbiased estimator.
- $\frac{r(n-1)}{\sigma^2} MSE \sim \chi^2_{r(n-1)}$
- Invert the statistic to get $(1-\alpha)\%$ confidence interval:

$$\frac{r(n-1)MSE}{\chi^2_{r(n-1)}(1-\alpha/2)} \leq \sigma^2 \leq \frac{r(n-1)MSE}{\chi^2_{r(n-1)}(\alpha/2)}$$

Inference for σ_μ^2

- Note that $\sigma_\mu^2 = \frac{E(MSTR)}{n} - \frac{E(MSE)}{n}$.
- We do not know the distribution of linear combinations of sums-of-squares.
- There are several procedures for the approximation of this distribution:
  - We will focus on the Satterthwaite procedure.
  - Has some poor asymptotics with certain weights.
  - But you need a lot of subjects in general to get good variance components estimates.
- Consider $L = \sum_j c_j E(MS_j)$
  - MS_j is some mean square with degrees of freedom df_j.
  - Unbiased estimator is $\hat{L} = \sum_j c_j MS_j$.

Satterthwaite

- Idea is to approximate $\frac{df\,\hat{L}}{L} \sim \chi^2_{df}$.
  - $df = \frac{\left(\sum_j c_j MS_j\right)^2}{\sum_j (c_j MS_j)^2/df_j}$
- CI: $\frac{df\,\hat{L}}{\chi^2_{df}(1-\alpha/2)} \leq L \leq \frac{df\,\hat{L}}{\chi^2_{df}(\alpha/2)}$
- To get a CI for σ^2_μ:
  - $MS_1 = MSTR$, $MS_2 = MSE$
  - $df_1 = r-1$, $df_2 = r(n-1)$
  - $c_1 = n^{-1}$, $c_2 = -n^{-1}$
  - $\hat{L} = (MSTR - MSE)/n$
  - $df = \frac{n^2\hat{L}^2}{MSTR^2/(r-1) + MSE^2/(r(n-1))}$
  - CI for σ^2_μ:
    $[\frac{df\,\hat{L}}{\chi^2_{df}(1-\alpha/2)}, \frac{df\,\hat{L}}{\chi^2_{df}(\alpha/2)}]$

Looking forward:

- Continue with Ch 25.
- Talk about two-way mixed effects ANOVA.
  - One factor is random and the other is fixed.
- Talk about two-way random effects ANOVA.
  - Both factors are random.

Two-Factor Mixed Effects Models

- Consider an example.
- We want to look at how growth promotion methods and variety of seeds affect the growth of turf grass.
- We consider three growth promotion methods and 5 variety of seeds.
- Six pots of each variety and method combination are planted.
- The pots are placed in a chamber that controls the environment.
- After several weeks, we measure the amount of growth as a continuous variable.

Two-Factor Mixed Effects Model

- We are specifically interested in the mean growth of these three growth promotion methods.
  - Method is a fixed effect.
- We think that the five types of seed are representative of a large population of turf seeds.
  - Main interest is not in the mean growth of each variety.
  - Interest is in characterizing mean growth variability between variety of seeds.
  - Variety is a random effect.
- There might be an interaction between variety and method.
  - Since variety is random, interactions are random effects.

Two-Way Random Effects ANOVA

- What if we are not directly interested in these three types of methods?
- What if these methods are representative of a much larger sample?
- We can model method as a random variable as well.
- When should we consider a factor as fixed rather than random?
  - When we do not have a large number of factor levels.
  - When we are specifically interested in inference at the factor levels represented in the study.
- First, we will focus on two-way mixed effects model.
- Then, we will consider the two-way random effects model.

The Model from the Text

- $Y_{ijk} = \mu. + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
- $\sum_i \alpha_i = 0$
- $\beta_j \sim$ iid $N(0, \sigma^2_\beta)$
- $(\alpha\beta)_{ij} \sim N(0, \frac{a-1}{a}\sigma^2_{\alpha\beta})$
- $\sum_i (\alpha\beta)_{ij} = 0$ for all j.
- $Cov((\alpha\beta)_{ij}, (\alpha\beta)_{kj}) = -\frac{1}{a}\sigma^2_{\alpha\beta}$ for $i \neq k$
- $\epsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
- $Var(Y_{ijk}) = \sigma^2_\beta + \frac{a-1}{a}\sigma^2_{\alpha\beta} + \sigma^2$
- $Cov(Y_{ijk}, Y_{ijl}) = \sigma^2_\beta + \frac{a-1}{a}\sigma^2_{\alpha\beta}$ for $k \neq l$.
- $Cov(Y_{ijk}, Y_{ljm}) = \sigma^2_\beta - \frac{1}{a}\sigma^2_{\alpha\beta}$ for $i \neq l$.
- $Cov(Y_{ijk}, Y_{lmp}) = 0$ for $j \neq m$

Some Comments

- To make this full rank, we are placing the assumptions that $\sum \alpha_i = 0$ and $\sum_i (\alpha\beta)_{ij} = 0$ for all j.
- SAS does not assume these constraints.
  - Constraint is somewhat old-fashioned.
- We will look at both this model and the model without constraints.
- In practice, you will use the model without the interactions summing to zero.

Mixed Effects Model - Book's Formulation

Assume that factor A is fixed and factor B is random.

- $E(MSA) = \sigma^2 + nb\frac{\sum \alpha_i^2}{a-1} + n\sigma^2_{\alpha\beta}$
  - df = $a-1$
- $E(MSB) = \sigma^2 + na\sigma^2_{\beta}$
  - df = $b-1$
- $E(MSAB) = \sigma^2 + n\sigma^2_{\alpha\beta}$
  - df = $(a-1)(b-1)$
- $E(MSE) = \sigma^2$
  - df =$ab(n-1)$
- Can see the test statistics:
  - $H_0 : \alpha_i = 0$ for all $i = 1, \ldots, a$: $MSA/MSAB \sim F_{a-1,(a-1)(b-1)}$
  - $H_0 : \sigma^2_{\beta} = 0$: $MSB/MSE \sim F_{b-1,ab(n-1)}$
  - $H_0 : \sigma^2_{\alpha\beta} = 0$: $MSAB/MSE \sim F_{(a-1)(b-1),ab(n-1)}$

SAS's Formulation

SAS with no sum-to-zero constraints.

- $Y_{ijk} = \mu_{\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
- $\beta_j \sim$ iid $N(0, \sigma^2_\beta)$
- $(\alpha\beta)_{ij} \sim$ iid $N(0, \sigma^2_{\alpha\beta})$
- $\epsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
- Pairwise independence between β_j, $(\alpha\beta)_{ij}$, and ϵ_{ijk}.
- $i = 1, \ldots, a$, $j = 1, \ldots, b$, $k = 1, \ldots, n$.
- $VarY_{ijk} = \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2$
- $Cov(Y_{ijk}, Y_{ijl}) = \sigma^2_\beta + \sigma^2_{\alpha\beta}$ for $k \neq l$.
- $Cov(Y_{ijk}, Y_{ljm}) = \sigma^2_\beta$ for $i \neq l$.
- $Cov(Y_{ijk}, Y_{lmp}) = 0$ for $j \neq m$

Mixed Effects Model

Assume factor A is fixed and factor B is random, and can show

- $E(MSA) = \sigma^2 + nb\frac{\sum(\alpha_i - \overline{\alpha})^2}{a-1} + n\sigma^2_{\alpha\beta}$
  - df = $a - 1$
- $E(MSB) = \sigma^2 + na\sigma^2_{\beta} + n\sigma^2_{\alpha\beta}$
  - df = $b - 1$
  - This is what has changed by using this parameterization.
- $E(MSAB) = \sigma^2 + n\sigma^2_{\alpha\beta}$
  - df = $(a - 1)(b - 1)$
- $E(MSE) = \sigma^2$
  - df = $ab(n - 1)$
- Can see the test statistics:
  - $H_0 : \alpha_i - \alpha_{i+1} = 0$ for all $i = 1, \dots, a$: $MSA/MSAB \sim F_{a-1,(a-1)(b-1)}$
  - $H_0 : \sigma^2_{\beta} = 0$: $MSB/MSAB \sim F_{b-1,(a-1)(b-1)}$
  - $H_0 : \sigma^2_{\alpha\beta} = 0$: $MSAB/MSE \sim F_{(a-1)(b-1),ab(n-1)}$

What can we estimate?

- How are the estimable functions under this two-way mixed-effects model different than the estimable functions of the two-way fixed-effects model?
- It is often of interest to compare $\alpha_i - \alpha_j$.
- The difference in group means.
- It makes sense even when there are interaction.
- Let's look at the turf example in SAS.

The ANOVA Table For Interaction Model

- $E(MSE) = E(SSE)/[ab(n-1)] = \sigma^2$
- Expected values of other mean squares are $\sigma^2 +$ "other terms"
- Ratios of MS's have F-distributions for testing effects.
- What if we think that the interactions are zero?
  - The fixed treatment effect acts additively with the random effect (maybe block effect).
  - Similar idea to using the additive model in fixed effects ANOVA.

Additive Two-Way Mixed Effect ANOVA Model

- $Y_{ijk} = \mu_{\cdot} + \alpha_i + \beta_j + \epsilon_{ijk}$
- $\beta_j \sim$ iid $N(0, \sigma_\beta^2)$
- $\epsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
- β_j is independent of ϵ_{ijk}.
- $E(Y_{ijk}) = \mu_{\cdot} + \alpha_i$
- $Var(Y_{ijk}) = \sigma_\beta^2 + \sigma^2$
- $Cov(Y_{ijk}, Y_{i'jk'}) = \sigma_\beta^2$
- $Corr(Y_{ijk}, Y_{i'jk'}) = \sigma_\beta^2 / \left[\sigma^2 + \sigma_\beta^2\right]$
- $Cov(Y_{ijk}, Y_{i'j'k'}) = 0$ if $j \neq j'$

Mean Squares for Additive ANOVA

- SSA and SSB are defined as in the interaction model.
  - df for SSA is $a - 1$.
  - df for SSB is $b - 1$.
- $SSE_p = SSE + SSAB$ is the sums-of-squares error or the additive model.
  - $df = [(a-1)(b-1)] + ab(n-1) = nab - a - b + 1$

Testing In The Additive Model

- $E(MSA) = \sigma^2 + nb\frac{\sum(\alpha_i - \overline{\alpha})^2}{a-1}$
- $E(MSB) = \sigma^2 + na\sigma_\beta^2$
- $E(MSE_p) = \sigma^2$
- Testing a factor A effect:
  - $H_0 : \alpha_i - \alpha_{i+1} = 0$ for $i = 1, \ldots, (a-1)$
  - $MSA/MSE_p \sim F_{a-1, nab-a-b+1}$ under H_0
- Testing a factor B effect:
  - $H_0 : \sigma_\beta^2 = 0$
  - $MSB/MSE_p \sim F_{b-1, nab-a-b+1}$ under H_0

What the Book Covers

- The book only covers the additive model for randomized complete block designs.
  - The case where $n = 1$.
  - Application of this special case is obvious when you understand the general case.
- Should I eliminate an interaction?
  - Report the additive model if you think there is no interaction.
  - Makes interpretation much clearer.
  - Can base your view of the presence of interactions not just on p-values but also on science and plots.

Candy Example

- A new example:
  - We are working for a candy factory and want to know how sweetness affects the likability of our chocolate.
  - Likeness is rated on a scale from 1-100.
  - We are interested in two specific levels of sweetness that our machines can manufacture.
  - We are concerned that the moisture of the bar can also affect likability.
  - There are a lot of moisture levels but we choose 4 levels.

Two-Way Random Effects Models:

- Consider the turf grass example again.
- What if we are not concerned about these three specific promotion methods?
- We are concerned if growth varies with different methods and variety of seeds.
- We think that these methods are sampled from possible methods.
- Will have two random effects.
- No ambiguity about parameterizations if you do not have interactions between random and fixed effects.
  - Book agrees with SAS now.

Two-Way Random Effects ANOVA Model

- $Y_{ijk} = \mu_{\cdot} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$
- $\alpha_i \sim$ iid $N(0, \sigma^2_\alpha)$
- $\beta_j \sim$ iid $N(0, \sigma^2_\beta)$
- $(\alpha\beta)_{ij} \sim$ iid $N(0, \sigma^2_{\alpha\beta})$
- $\epsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
- α_i, β_j, $(\alpha\beta)_{ij}$ and ϵ_{ijk} are pairwise independent.
- $E(Y_{ijk}) = \mu_{\cdot}$
- $Var(Y_{ijk}) = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta} + \sigma^2$
- $Cov(Y_{ijk}, Y_{ijk'}) = \sigma^2_\alpha + \sigma^2_\beta + \sigma^2_{\alpha\beta}$ if $k \neq k'$
- $Cov(Y_{ijk}, Y_{i'jk'}) = \sigma^2_\beta$ for $i \neq i'$
- $Cov(Y_{ijk}, Y_{ij'k'}) = \sigma^2_\alpha$ if $j \neq j'$

Two-Way Random Effects Model

Assume that factor A and B are random. We will derive:

- $E(MSA) = \sigma^2 + nb\sigma^2_\alpha + n\sigma^2_{\alpha\beta}$
  - df = $a - 1$
- $E(MSB) = \sigma^2 + na\sigma^2_\beta + n\sigma^2_{\alpha\beta}$
  - df = $b - 1$
- $E(MSAB) = \sigma^2 + n\sigma^2_{\alpha\beta}$
  - df = $(a - 1)(b - 1)$
- $E(MSE) = \sigma^2$
  - df = $ab(n - 1)$
- Can see the test statistics:
  - $H_0 : \sigma^2_\alpha = 0$: $MSA/MSAB \sim F_{a-1,(a-1)(b-1)}$
  - $H_0 : \sigma^2_\beta = 0$: $MSB/MSAB \sim F_{b-1,(a-1)(b-1)}$
  - $H_0 : \sigma^2_{\alpha\beta} = 0$: $MSAB/MSE \sim F_{(a-1)(b-1),ab(n-1)}$

Inference On σ^2

- Easy.
- We know that $E(MSE) = \sigma^2$
- We know that $\frac{ab(n-1)MSE}{\sigma^2} \sim \chi^2_{ab(n-1)}$
- Invert the statistic to get $(1-\alpha)\%$ CI's
$$\left[\frac{ab(n-1)MSE}{\chi^2_{ab(n-1)}(1-\alpha/2)}, \frac{ab(n-1)MSE}{\chi^2_{ab(n-1)}(\alpha/2)}\right]$$

Inference On σ_α^2

- Harder than for σ^2.
- Must use a Satterthwaite approximation.
  - What type of an approximation is done via the Satterthwaite?
- $E(MSA) = \sigma^2 + nb\sigma_\alpha^2 + n\sigma_{\alpha\beta}^2$
- $E(MSAB) = \sigma^2 + n\sigma_{\alpha\beta}^2$
- $\sigma_\alpha^2 = \frac{1}{nb}E(MSA) + \frac{-1}{nb}E(MSAB)$
- $\hat{\sigma}_\alpha^2 = \frac{1}{nb}MSA + \frac{-1}{nb}MSAB$
- $\frac{df\,\hat{\sigma}_\alpha^2}{\sigma^2}$ is approximately χ_{df}^2
- $df = \frac{(\hat{\sigma}_\alpha^2)^2}{\frac{((nb)^{-1}MSA)^2}{a-1} + \frac{((nb)^{-1}MSAB)^2}{(a-1)(b-1)}}$
- Example in SAS.

Looking forward:

- Methods for unbalanced data.
  - Example using ML.
  - Talk about the REML.

Unbalanced Data

UNBALANCED DATA

- Thus far we have only considered random and mixed-effects models for balanced data.
- Recall the problem with unbalanced data in fixed-effects models.
  - SSA, SSB, and SSAB are not orthogonal.
- Non-orthogonality is a problem in mixed-effects models.
  - How do we now estimate σ_β^2?
- There are sums-of-squares approaches but they are complicated and don't work that well.
- Most popular solution is to use normality based maximum likelihood.

Mixed-Effects Model

Consider the mixed-effects model:

- $Y_{ijk} = \mu. + \alpha_i + \beta_j + \epsilon_{ijk}$
- $\sum \alpha_i = 0$
- $\beta_j \sim$ iid $N(0, \sigma^2_\beta)$
- $\epsilon_{ijk} \sim$ iid $N(0, \sigma^2)$
- β_j is independent of $\epsilon_{ij'k}$
- $i = 1, \ldots, a$; $j = 1, \ldots, b$; $k = 1, \ldots, n_{ij}$.
- Think of fixed effects as regression components
  - $\mu.$ and α_i
- Write out the likelihood function.
- Conditional on variance components:
  - weighted regression to estimate fixed effects.
  - Inference will be based on the marginal likelihood assuming variance components are known.
- Two approaches to estimating variance components:
  - Maximum likelihood (ML).
  - Restricted Maximum Likelihood (REML).

Marginal Likelihood

$$\mathbf{Y}_{ij} = \begin{bmatrix} Y_{ij1} \\ \vdots \\ Y_{ijn_{ij}} \end{bmatrix}$$

$$\begin{aligned} E\mathbf{Y}_{ij} &= \mu. + \alpha_i \\ &= A_{ij}\beta \end{aligned}$$

$$A_{ij} = \begin{bmatrix} 1 & 0 & \dots & 1 & 0 & \dots & 0 \\ \vdots & & & & & & \vdots \\ 1 & 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

- $Var(\mathbf{Y}_{ij}) = V_{ij} = \sigma^2 \mathbf{I} + \sigma_\beta^2 \mathbf{J}$
  - $\mathbf{I}$ is the identity and $\mathbf{J}$ is the matrix of ones.
- $\mathbf{Y}_j = [\mathbf{Y}'_{1j}, \ldots, \mathbf{Y}'_{aj}]'$.
- Then $E\mathbf{Y}_j = A_j\beta$, where $A_j = [A'_{1j}, \ldots, A'_{aj}]'$
  $Var(\mathbf{Y}_j) = V_j = \sigma^2 \mathbf{I} + \sigma_\beta^2 \mathbf{J}$
- $\mathbf{Y}_j$ is independent of $\mathbf{Y}_{j'}$
- The -2 log likelihood is then:

$$\sum_{j=1}^{b} \left\{ n_{.j} \log(2\pi) + \log(|V_j|) + (\mathbf{Y}_j - A_j\beta)' V_j^{-1} (\mathbf{Y}_j - A_j\beta) \right\}.$$

Estimation of Main Effects

- Remember that V_j depends on the variance components σ^2, σ^2_β.
- If we knew these variance components, we can differentiate to find that the minimizer is:
  - $\hat{\beta} = \left(\sum_{j=1}^{b} A_j' V_j^{-1} A_j\right)^{-1} \sum_{j=1}^{b} A_j V_j^{-1} \mathbf{Y}_j$
  - $Var(\hat{\beta}) = \left(\sum_{j=1}^{b} A_j' V_j^{-1} A_j\right)^{-1}$
- In practice we do not know σ^2, σ^2_β so we don't know V_j.
- Two ways to estimate σ^2 and σ^2_β:
  - Maximum likelihood.
  - REML (restricted maximum likelihood).

ML Estimation

- Maximize the likelihood jointly over β and σ^2, σ_β^2
- Equivalent to choosing $\hat{\sigma}^2$ and $\hat{\sigma}_\beta^2$ that minimizes

$$\sum_{j=1}^{b} \left\{ n_{\cdot j} \log(2\pi) + \log(|V_j|) + (\mathbf{Y}_j - A_j\hat{\beta})' V_j^{-1} (\mathbf{Y}_j - A_j\hat{\beta}) \right\}$$

- This function is non linear in σ^2, σ_β^2.
  - Both V_j and $\hat{\beta}$ depend on σ^2, σ_β^2.
- SAS uses some numerical method to minimize the function.
  - Default and preferred is a Newton-Raphson method.

Inference

- Confidence intervals for $\beta = (\mu, \alpha_1, \cdots, \alpha_a)^T$ can be achieved via Wald intervals with variance $\left(\sum_{j=1}^{b} A_j' \hat{V}_j^{-1} A_j\right)^{-1}$.
  - Common methods do not take into account the variability of $\hat{\sigma}^2$ and $\hat{\sigma}^2_\beta$.
- $\hat{\sigma}^2$ and $\hat{\sigma}^2_\beta$ are biased.
  - Remember for regression that the ML estimator of σ^2 is $n^{-1}(Y - Xb)'(Y - Xb)$ which has expected value $\sigma^2 \frac{n-1}{n}$.
  - Bias due to a loss of degrees of freedom.
  - Asymptotically okay.
  - Will talk about REML estimates later.

- Tests for σ_β^2
- LR tests can be used to test $H_0 : \sigma_\beta^2 = 0$
    - Compare -2 log likelihood (L) from full and reduced model.
    - Must be careful with the degrees of freedom.
    - Assume you have parameters Θ_0 that is a subset of Θ.
    - You want to test if Θ_0 holds.
    - Let $\hat{\Theta}$ and $\hat{\Theta}_0$ be the ML estimates under the full and reduced model, respectively.
    - If Θ_0 is an interior point of Θ then: $L(\hat{\Theta}_0) - L(\hat{\Theta}) \sim \chi_d^2$
    - d is the difference in dimension between Θ and Θ_0
    - In linear regression, if we want to test if $\beta_1 = \beta_2 = 0$, then d=2.

- Does not hold if Θ_0 is on the boundary of Θ
- Since $\sigma^2_\beta \in [0, \infty]$, $H_0 : \sigma^2_\beta = 0$ is on the topological boundary.
- To test $H_0 : \sigma^2_\beta = 0$ vs $H_a : \sigma^2_\beta > 0$
  $L(\hat{\Theta_0}) - L(\hat{\Theta}) \sim F$
- F is a 50/50 mixture of χ^2_1 and χ^2_0 where χ^2_0 is point mass at 0.

Example

- We are working for the Army.
- Soldiers who come back from combat often have a hard time sleeping.
- We enroll patients at one of 5 VA hospitals who have a PSQI score $\geq$ 18.
- We randomly give each patient one of two types of therapy then measure their PSQI after three months.
- Goal: do the different therapies have different effects on lowering the PSQI?
- We also would have to control for sex. No reason to believe that different treatments work better for different sexes.
- What model should we fit?

A Simpler Problem

- Consider the simple linear regression $\mathbf{Y} = \mathbf{X}\beta + \epsilon$
  - $\mathbf{Y}$ is a n vector of observations.
  - $\mathbf{X}$ is $n \times (p+1)$
  - $\epsilon \sim N(0, \sigma^2 I)$
- Let $SSE = \left(\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right)' \left(\mathbf{Y} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right)$.
- MLE for σ^2 is SSE/n
- We know that $E(MSE) = SSE/\left[n - (p+1)\right] = \sigma^2$
- MLE for variance is biased downward by $(n-p-1)/n$.
- Biased because estimate is based on an estimate of parameters (fixed effects) $\rightarrow$ loss in degrees of freedom.
- Let A be any $n \times (n-p-1)$ matrix whose columns are orthogonal to X.
- Estimate σ^2 as the MLE of $A'\mathbf{Y} \sim N(0, \sigma^2 A'A)$.

The REML

- Let us go back to the mixed effects model.
- Stack the $\mathbf{Y}_j$ to put all data in a single vector.
- Idea is to base the estimates of variance components on a contrast of the data that will not depend on the mean effects.
- Theory is involved, see Harville paper for details.
- Can show that the likelihood of the contrast does not depend on the matrix A.
- The -2 log-likelihood of any contrast orthogonal to the design matrix is:

$$(n_{\cdot\cdot} - a)\log(2\pi) - \log\left|\sum_{j=1}^{b} X_j' X_j\right| + \log\left|\sum_{j=1}^{b} X_j' V_j^{-1} X_j\right| +$$

$$\sum_{j=1}^{b}\left[\log|V_j| + (\mathbf{Y} - X_j\hat{\beta})' V_j^{-1}(\mathbf{Y} - X_j\hat{\beta})\right]$$

About the REML and ML Approaches

- ML and REML approaches are most popular for unbalanced data.
- A downside is inference for variance components.
- Likelihood ratio test for variance components.
  - Remember that the LRT follows a mixture of χ^2's.