ST518 - Mixed effects models

Mixed Effects Models

Jason A. Osborne

N. C. State Univ.

Outline

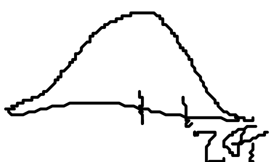
Topic: Mixed effects models

- One-way random effects model to study *variances*
- Mixed effects models
- Subsampling
- Expected mean squares for mixed models

One-way random effects model

Example:

- Genetics study w/ beef animals. Measure birthweight Y (*lbs*).
- $t = 5$ sires, each mated to a separate group of $n = 8$ dams.
- $N = 40$, completely randomized.



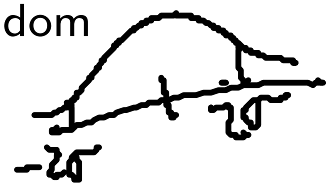
					Birthweights						
Sire #	Level				Sample					$\bar{y}_{i\cdot}$	s_i
177	1	61	100	56	113	99	103	75	62	83.6	22.6
200	2	75	102	95	103	98	115	98	94	97.5	11.2
201	3	58	60	60	57	57	59	54	100	63.1	15.0
202	4	57	56	67	59	58	121	101	101	77.5	25.9
203	5	59	46	120	115	115	93	105	75	91.0	28.0

Q: Statistical model for these data? $Y_{ij} = \mu + \tau_i \qquad + E_{ij}$

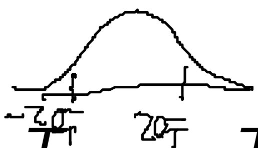
Random effects model

The one-way random effects model:

$$Y_{ij} = \underbrace{\mu}_{\text{fixed}} + \underbrace{T_i}_{\text{random}} + \underbrace{E_{ij}}_{\text{random}} \quad \text{for } i = 1, 2, \ldots, t \text{ and } j = 1, \ldots, n$$



with



- $T_1, T_2, \ldots, T_t \overset{iid}{\sim} N(0, \sigma_T^2)$
- $E_{11}, \ldots, E_{tn} \overset{iid}{\sim} N(0, \sigma^2)$
- $T_1, T_2, \ldots, T_t$ independent of $E_{11}, \ldots, E_{tn}$

Features

- $T_1, T_2, \ldots$ denote *random* effects, drawn from some population of interest. That is, $T_1, T_2, \ldots$ is a $\boxed{\text{random sample}}$!
- σ_T^2 and σ^2 are called $\boxed{\text{variance components}}$
- conceptually different from one-way fixed effects model

Beef animal genetic study, continued

With $t = 5$ and $n = 8$, the random effects $T_1, T_2, \ldots, T_5$ reflect sire-to-sire variability. (ie genetics)

No particular interest in $\tau_1, \tau_2, \ldots, \tau_5$ from the (misspecified) fixed effects model:

$$Y_{ij} = \underbrace{\mu}_{\text{fixed}} + \underbrace{\tau_i}_{\text{fixed}} + \underbrace{E_{ij}}_{\text{random}} \quad \text{for } i = 1, 2, \ldots, t \text{ and } j = 1, \ldots, n$$

fixed

with

- $\tau_1, \tau_2, \ldots, \tau_t$ unknown model parameters
- $E_{11}, \ldots, E_{tn} \overset{iid}{\sim} N(0, \sigma^2)$

We're not trying to *estimate* linear combos of fixed effects such as $\mu + \tau_1$. Instead, we care about the population from which T_1 was sampled, which is $N(0, \sigma_T^2)$.

One-way random effects model continued

Exercise: Using the random effects model, specify

$$E(Y_{ij}) \text{ and } \mathrm{Var}(Y_{ij})$$

$E(Y_{ij}) = E(\mu + T_i + E_{ij}) = E(\mu) + E(T_i) + E(E_{ij}) = \mu + 0 + 0 = \mu$

$V(Y_{ij}) = V(\mu + T_i + E_{ij}) = 0 + V(T_i) + V(E_{ij}) = \boxed{\sigma_T^2 + \sigma^2 = V(Y_{ij})}$

$V(\bar{Y}_{\cdot\cdot}) = V\left(\frac{1}{40}\sum\sum Y_{ij}\right) = V\left(\frac{1}{40}\sum_i\sum_j(\mu + T_i + E_{ij})\right) = V(\mu + \bar{T}_{\cdot} + \bar{E}_{\cdot\cdot}) = 0 + \frac{\sigma_T^2}{5} + \frac{\sigma^2}{40}$

- Two *components* to variability in data: σ^2, σ_T^2
- T_1, T_2, T_3, T_4, T_5 a random sample of sire effects
- Sire effects is a population in its own right.

Contrast this situation with the binding fractions. Why not model antibiotic effects as random? Why fixed?

$E(\bar{Y}_{\cdot\cdot}) = E(\mu + \bar{T}_{\cdot} + \bar{E}_{\cdot\cdot}) = \mu$

Model parameters: $\sigma^2, \sigma_T^2, \mu$

Sums of squares, mean squares - same as in fixed effects ANOVA:

$$SS[T] = \sum_i \sum_j (\overline{Y}_{i\cdot} - \overline{Y}_{\cdot\cdot})^2$$

$$SS[E] = \sum_i \sum_j (Y_{ij} - \overline{Y}_{i\cdot})^2$$

$$SS[Tot] = \sum \sum (y_{ij} - \overline{y}_{\cdot\cdot})^2$$

The ANOVA table is almost the same, it just has a different expected mean squares column:

Source	SS	df	MS	Expected MS
Treatment	$SS[T]$	$t-1$	$MS[Trt]$	$\sigma^2 + n\sigma_T^2$
Error	$SS[E]$	$N-t$	$MS[E]$	σ^2
Total	$SS[Tot]$	$N-1$		

BTW, if $H_0 : \sigma_T^2 = 0$, what is $E(MS(Trt))/E(MS(E))$ 1

Estimating parameters of one-way random effects model

(Solve a linear system of *estimating equations* obtained by equating statistics to their expected values and solving for unknown parameters:)

$$
\begin{aligned}
E(\bar{Y}_{\cdot\cdot}) = E(\hat{\mu}) &= \mu = \bar{Y}_{\cdot\cdot} \\
E(MS(T)) &= \sigma^2 + n\sigma_T^2 = MS(T) \\
E(MS(E)) &= \sigma^2 = MS(E)
\end{aligned}
$$

set & solve

leading to the solution

$$
\begin{aligned}
\hat{\mu} &= \bar{Y}_{\cdot\cdot} \\
\hat{\sigma}^2 &= MS(E) \\
\hat{\sigma}_T^2 &= \frac{MS(T) - MS(E)}{n}
\end{aligned}
$$

We've derived these estimators:

$$\begin{aligned}\widehat{\mu} &= \bar{y}_{..} \\ \widehat{\sigma}^2 &= MS[E] \\ \widehat{\sigma}_T^2 &= \frac{MS[T] - MS[E]}{n}\end{aligned}$$

For sires data, we observed $\bar{y}_{..} = 82.6$ and

Source	SS	df	MS	Expected MS
Sire	5591	4	1398	$\sigma^2 + 8\sigma_T^2$
Error	16233	35	464	σ^2
Total	21824	39		

leading to the observed estimates

$$\begin{aligned}\widehat{\mu} &= \underline{\quad 82.6 \quad} (lbs) \\ \widehat{\sigma}^2 &= \underline{\quad 464 \quad} (lbs^2) \\ \widehat{\sigma}_T^2 &= (1398 - 464)/8 \\ &= \underline{\quad 116.8 \quad} (lbs^2)\end{aligned}$$

Questions pertaining to this type of study:

Consider the birthweight of a randomly sampled calf

1. What is the estimated variance of such a calf? $\hat{V}(Y_{ij}) = \hat{\sigma}^2 + \hat{\sigma}_T^2 = 464 + 116.8 = 581$
2. Estimate how much of this variation is due to the sire effect.
3. Estimate how much of this variation is not due to the sire effect.

(2) $116.8/581 = 20\%$ | $464/581 = 80\%$ (3)

General questions:

1. Is it possible for an estimated variance component to be negative? yes
2. How? if $MS(T) < MS(E)$
3. What do you do in that case? Set $\hat{\sigma}_T^2 = 0$

not necessarily dropping the random effect from the model

$$
\begin{aligned}
\mathrm{Var}(Y_{ij}) &= \sigma^2 + \sigma_T^2 \\
\widehat{\mathrm{Var}}(Y_{ij}) &= MS(E) + \frac{MS(T) - MS(E)}{n} \\
\mathrm{Var}(T_i)/\mathrm{Var}(Y_{ij}) &= \sigma_T^2/(\sigma^2 + \sigma_T^2) \\
\mathrm{Var}(E_{ij})/\mathrm{Var}(Y_{ij}) &= \sigma^2/(\sigma^2 + \sigma_T^2)
\end{aligned}
$$

1. Yes, it is possible for $\widehat{\sigma}_T^2 < 0$ even though $\sigma_T^2 \geq 0$.
2. $\widehat{\sigma}_T^2 < 0 \Leftrightarrow$ ______________________________?
3. Inference concerning σ_T^2? ______________________________

Other parameters of interest in random effects models

Functions of

Coefficient of variation (CV):

$$CV(Y_{ij}) = \frac{\sqrt{\text{Var}(Y_{ij})}}{|E(Y_{ij}|)} =? \frac{\sqrt{Var(Y_{ij})}}{|E(Y_{ij})|} = \frac{\sqrt{\sigma^2 + \sigma_\tau^2}}{|\mu|}$$

$\times 100$

Note: this is *not* estimated by `Coeff Var` in PROC GLM output.

Intraclass correlation coefficient

$$\rho_I = \frac{\text{Cov}(Y_{ij}, Y_{ik})}{\sqrt{\text{Var}(Y_{ij})\text{Var}(Y_{ik})}} = \frac{cov(\mu + T_i + E_{ij}, \mu + T_i + E_{ik})}{\sqrt{(\sigma^2 + \sigma_\tau^2)(\sigma^2 + \sigma_\tau^2)}} = \frac{cov(T_i, T_i)}{\sigma^2 + \sigma_\tau^2} = \frac{\sigma_\tau^2}{\sigma^2 + \sigma_\tau^2}$$

$cov(T_i, E_{ik}) = 0$

- Interpretation: the correlation between two responses receiving the same level of the random factor.
- Bigger values of ρ_I correspond to (bigger/smaller?) random treatment effects.
- Answers questions like: How much of this variation is due to the sire effect?

For sires,

$$\widehat{CV} = \frac{\sqrt{\hat\sigma^2 + \hat\sigma_T^2}}{\hat\mu} = \sqrt{464 + 116.8}\,/\,82.6 = 0.29 \quad \text{or } 29\%$$

$$\hat\rho_I = \frac{\hat\sigma_T^2}{\hat\sigma^2 + \hat\sigma_T^2} = \frac{116.8}{464 + 116.8} = 0.20$$

Interpretations:

- The estimated standard deviation of a birthweight, 24.1 is 29% of the estimated mean birthweight, 82.6.
- The estimated correlation between any two calves with the same sire for a male parent, or the estimated *intrasire* correlation coefficient, is 0.20