unbiased

It can be shown that the weights on $\overline{y}_{11\cdot}$ and y_{121} are

$$\frac{\frac{1}{\text{Var}(\overline{y}_{11\cdot})}}{\frac{1}{\text{Var}(\overline{y}_{11\cdot})} + \frac{1}{\text{Var}(y_{121})}} \text{ and } \frac{\frac{1}{\text{Var}(y_{121})}}{\frac{1}{\text{Var}(\overline{y}_{11\cdot})} + \frac{1}{\text{Var}(y_{121})}}, \text{ respectively.}$$

$=$

This is a special case of a more general phenomenon: the BLUE is a weighted average of independent linear unbiased estimators with weights for the linear unbiased estimators proportional to the inverse variances of the linear unbiased estimators.

// end lecture 26

3-31-25

Of course, in this case and in many others,

$$\hat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}} = \begin{bmatrix} \frac{2\sigma_e^2+2\sigma_u^2}{3\sigma_e^2+4\sigma_u^2}\,\bar{y}_{11\cdot} & + & \frac{\sigma_e^2+2\sigma_u^2}{3\sigma_e^2+4\sigma_u^2}\,y_{121} \\ & y_{211} & \end{bmatrix}$$

is not an estimator because it is a function of unknown parameters.

Thus, we use $\hat{\boldsymbol{\beta}}_{\hat{\boldsymbol{\Sigma}}}$ as our estimator (i.e., we replace σ_e^2 and σ_u^2 by estimates in the expression above).

as n increases the approximation improves

- $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}}$ is an approximation to the BLUE.

- $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}}$ is not even a linear estimator in this case.

- Its exact distribution is unknown.

- When sample sizes are large, it is reasonable to assume that the distribution of $\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}}$ is approximately the same as the distribution of $\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}$.

$$\begin{aligned}
\mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\boldsymbol{\Sigma}}) &= \mathrm{Var}[(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{y}] \\
&= (\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\underline{\mathrm{Var}(\boldsymbol{y})}[(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}]^\top \\
&= (\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X}(\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{X})^{-1}
\end{aligned}$$

$$\mathrm{Var}(\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}}) = \mathrm{Var}[(\boldsymbol{X}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{y}] = ???? \approx (\boldsymbol{X}^\top\widehat{\boldsymbol{\Sigma}}^{-1}\boldsymbol{X})^{-1}$$

Summary of Main Points

- Many of the concepts we have seen by examining special cases hold in greater generality.
- For many of the linear mixed models commonly used in practice, balanced data are nice because...

When data are unbalanced we can still do inference but need to keep in mind that our results are approximations only!

1. It is relatively easy to determine degrees of freedom, sums of squares, and expected mean squares in an ANOVA table.

2. Ratios of appropriate mean squares can be used to obtain exact F-tests.

3. For estimable $\boldsymbol{C\beta}$, $\boldsymbol{C}\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}} = \boldsymbol{C}\widehat{\boldsymbol{\beta}}$. (OLS $=$ GLS).

4. When $\mathrm{Var}(\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}}) =$ constant $\times\ \mathrm{E}(MS)$, exact inferences about $\boldsymbol{c}^\top \boldsymbol{\beta}$ can be obtained by constructing t-tests or confidence intervals based on
$$t = \frac{\boldsymbol{c}^\top \widehat{\boldsymbol{\beta}} - \boldsymbol{c}^\top \boldsymbol{\beta}}{\sqrt{\text{constant} \times (MS)}} \sim t_{DF(MS)}.$$

5. Simple analysis based on experimental unit averages gives the same results as those obtained by linear mixed model analysis of the full data set.

When data are unbalanced, the analysis of linear mixed may be considerably more complicated.

1. Approximate F-tests can be obtained by forming linear combinations of Mean Squares to obtain denominators for test statistics.

2. The estimator $C\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}}$ may be a nonlinear estimator of $C\boldsymbol{\beta}$ whose exact distribution is unknown.

3. Approximate inference for $C\boldsymbol{\beta}$ is often obtained by using the distribution of $C\widehat{\boldsymbol{\beta}}_{\widehat{\boldsymbol{\Sigma}}}$, with unknowns in that distribution replaced by estimates.

Whether data are balanced or unbalanced, unbiased estimators of variance components can be obtained using linear combinations of mean squares from the ANOVA table.