EINDHOVEN UNIVERSITY OF TECHNOLOGY

2AMS10: Longitudinal Data Analysis 2021-2022 Q&A Session

October 26, 2021

Edwin van den Heuvel
e.r.v.d.heuvel@tue.nl
Professor in Statistics

Questions

- **Q:** What do we need to know/be able to derive for unbalanced data?
  - **A:** All we discussed for one-way ANOVA
  - **A:** All that was discussed in lecture notes for the two-way nested model

- **Q:** Must we know how to deal with situations where $\mathrm{COV}(a_i, a_{i'}) = \rho\sigma_a^2$ for $i \neq i'$?
  - **A:** Yes, you should know how to determine the covariance/correlation of y's for ANOVA and linear mixed models when the model is explicitly formulated

- **Q:** What is the dispersion matrix for y in the random model?
  A: For linear mixed models $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$, with $\boldsymbol{X}_i$ an observed matrix of covariates, $\boldsymbol{Z}_i$ a known matrix describing the structure of the random effects, $\boldsymbol{b}_i \sim N(\boldsymbol{0}, \boldsymbol{G}_i)$, $\boldsymbol{e}_i \sim N(\boldsymbol{0}, \boldsymbol{R}_i)$, and $\boldsymbol{b}_i$ and $\boldsymbol{e}_i$ independent. The variance of $\boldsymbol{y}_i$ is given by $\boldsymbol{V}_i = \mathrm{VAR}(\boldsymbol{y}_i) = \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{Z}_i^T + \boldsymbol{R}_i$
  A: It is not needed to obtain the inverse of $\boldsymbol{V}_i$
  A: You may need to be able to determine the inverse Fisher information matrix

Questions

- **Q:** Could you please derive the BLUP once more?
  - **A:** assume that $\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{b}_i + \boldsymbol{e}_i$, then we can study the joint distribution of $\boldsymbol{b}_i$ and $\boldsymbol{y}_i$, which is the following

    $$\begin{pmatrix} \boldsymbol{b}_i \\ \boldsymbol{y}_i \end{pmatrix} \sim N\left(\begin{pmatrix} \boldsymbol{0} \\ \boldsymbol{X}_i\boldsymbol{\beta} \end{pmatrix}, \begin{pmatrix} \boldsymbol{G}_i & \boldsymbol{Z}_i\boldsymbol{G}_i \\ \boldsymbol{Z}_i\boldsymbol{G}_i & \boldsymbol{V}_i \end{pmatrix}\right)$$

    The conditional distribution of the random effects $\boldsymbol{b}_i$ given the observations $\boldsymbol{y}_i$ is given by a normal distribution

    $$\boldsymbol{b}_i | \boldsymbol{y}_i \sim N(\boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{V}_i^{-1}(\boldsymbol{y}_i - \boldsymbol{X}_i\boldsymbol{\beta}), \boldsymbol{G}_i - \boldsymbol{Z}_i\boldsymbol{G}_i\boldsymbol{V}_i^{-1}\boldsymbol{G}_i^T\boldsymbol{Z}_i^T)$$

- **Q:** How can we derive that the SSA and the SSE are independent?
  - **A:** We have not derived the independence of sums of squares or mean squares for any of the ANOVA models during the course. Thus you just need to remember and possibly use that sums of squares or mean squares are independent.

TU/e

Questions

- Q: What does the F-statistic test for a random model? And is this also the test that SAS performs?
  - A: All F-tests in the ANOVA table test if the term it quantifies is different from zero.
  - A: Not all F-tests are exact. For unbalanced data they are not exact since the involved mean squares are not chi-square. For balanced data they may not be exact, since the denominator is not one mean square but a linear combination of mean square. In the three-way random effects model on Slide 88, the F-test for "type-of-class" ($H_0\colon \sigma_T^2 = 0$) is given by
    SAS: $F_{\text{SAS}} = MS_T/[MS_{ST} + MS_{C(T)} - MS_{SC(T)}]$ with df_T the numerator degrees of freedom and df_D the denominator degrees of freedom:
    $$df_D = \left[MS_{ST} + MS_{C(T)} - MS_{SC(T)}\right]^2/[MS_{ST}^2/df_{ST} + MS_{C(T)}^2/df_{C(T)} + MS_{SC(T)}^2/df_{SC(T)}]$$
    Alternative: $F_{\text{ALT}} = [MS_T + MS_{SC(T)}]/[MS_{ST} + MS_{C(T)}]$ with the following degrees of freedom
    $$df_N = \left[MS_T + MS_{SC(T)}\right]^2/[MS_T^2/df_T + MS_{SC(T)}^2/df_{SC(T)}]$$
    $$df_D = \left[MS_{ST} + MS_{C(T)}\right]^2/[MS_{ST}^2/df_{ST} + MS_{C(T)}^2/df_{C(T)}]$$

Questions

- **<u>Q:</u>** Could you discuss the significance of BLUPs in data analytics? How to interpret them? Also in regard to Fractional polynomials?
  - **<u>A:</u>** The BLUP's are predictions of the random effects. When you implement them you would be able to predict the results of an individual
  - **<u>A:</u>** Consider the model: $y_{ij} = b_{0i} + b_{1i}t_{ij}^{p_1} + b_{2i}t_{ij}^{p_2} + e_{ij}$, which describes the outcomes of individual i. The time profile of this individual is $b_{0i} + b_{1i}t_{ij}^{p_1} + b_{2i}t_{ij}^{p_2}$, but we do not know the coefficients b_{0i}, b_{1i}, and b_{2i}. The BLUP's predicts these coefficients $\hat{b}_{0i}$, $\hat{b}_{1i}$, and $\hat{b}_{2i}$. This gives us the opportunity to predict the time profile $\hat{y}_{ij} = \hat{b}_{0i} + \hat{b}_{1i}t_{ij}^{p_1} + \hat{b}_{2i}t_{ij}^{p_2}$ at time point t. The predictions may help explain how well our model predicts the data

- **<u>Q:</u>** Can we address covariates in the ANOVA model too?
  - **<u>A:</u>** Yes, but then it is called ANCOVA. The ANCOVA models are just simpler than the linear mixed effects models due to the covariance structure.

Questions

- <u>**Q:**</u> What to do if the normality of residuals is violated?
  - <u>**A:**</u> Use an alternative distributions instead of normality [outside scope of course]
  - <u>**A:**</u> Use a transformation of the outcome
  - <u>**A:**</u> Use other forms of the continues fixed effects (fractional polynomials for time, transformed covariates like log(BMI))

- <u>**Q:**</u> As Type 1 is suitable when orders of factors are known, could you give a few more examples to understand such order?
  - <u>**A:**</u> In a reliability study where multiple participants are being tested by multiple examiners, the order would be participants before examiners, since the participants data exists before the examiners can look at it.
  - <u>**A:**</u> in an investigation of the effect of BMI on blood pressure for men and women, the effect of women should be determined before the effect of BMI. Causally, sex will effect BMI and blood pressure. BMI will not effect sex.

Questions

- **Q:** What is the significance of degrees of Freedom in Solution for Fixed Effects Table?
  - **A:** The degrees of freedom is for all output the same. The degrees of freedom informs us of how much independent data is available to calculate effect sizes, confidence intervals and P-values.

- **Q:** I understood Interaction effects a*b, b*c, c*d is different from a*b*c*d. Is it correct? Is there a best practice when to chain multiple interaction effects?
  - **A:** The interaction effects a*b, b*c, and c*d are called two-factor interactions, while a*b*c*d are called four-factor interactions. Thus they are all interaction effects The single effect of a, b, c, or d are all called main effects (mostly in experimental designs)
  - **A:** The interpretation of three-factor or higher-factor interaction terms are more difficult to interpret. A two-factor interaction effect means that the effect of one factor depends on the level of another factor. A three-factor interaction means that the two-factor interaction depends on a third factor.