Additional note 10

Lecture 27

1. **Revision questions of L26**

   - Why consider nested factor?

     Answer: In repeated measurement, subjects are different, not homogenous as in a block factor. In this case, factor B other than treatment should be suppressed from ANOVA table and goes to the pooled interaction B(A) which contains all subject levels.

   - Where does the SS for factor B go?

     Answer: Absorb to the SS for interaction A:B in the 2-way ANOVA table.

   - How to estimate these parameters?

     Answer: $(\alpha\beta)_{j(i)}$ is estimated by $\bar{Y}_{ij\bullet} - \bar{Y}_{i1\bullet}$.

2. In *nested design*, we first need to distinguish *cross factor* and *nested factor*. For block say, a factor is a cross factor if it is independent of treatment factor but if the levels of block are similar but not identical for different levels of treatment, it may be considered as nested factor. A common example is when different animals are assigned to receive different treatments, each animal should not be considered as identical to the others.

3. In this case, animals are experimental unit (EU) and repeated measurements from each animal are observational unit (OU). The EU and OU may be considered as *biological replication* and *technical replication* respectively and RSS can be splitted between EU and OU.

4. In *split plot design*, the whole plot can be considered as EU and subplot as OU. RSS can be splitted according to the whole plot (EU) and subplot (OU), depending also on the way to assign the whole plot (EU) and subplot (OU). One possible case is when animals say are assigned as blocks/EU/whole plot (with variability across animals) and a 2-way ANOVA with interaction is performed. Then SS_B and $\text{SS}_{A:B}$ with df $a(b-1)$ (A as treatment) which measure the variability across blocks (animals, EU etc) can be treated as RSS across EU and the usual RSS with df $ab(r-1)$ as RSS across OU (repeats).

Lecture 28

1. **Revision questions of L27**

   - What is the Balanced nested design?

     Answer: When there are an equal number of levels of B within each level of A and an equal number of repeats.

   - In the calf feeding experiment, what is technical replication and biological replication?

     Answer:
     Technical replication: 4 calf within each pen and
     Biological replication: 2 pen within each feed.

   - In running a 2-way ANOVA model Y A/B which distinguish RSS between B and R (label of repeat), what is the EU (ie B) residual SS and OU (ie R) residual SS?

     Answer: First consider 4 models within 2 cases: 2-way or 1-way ANOVA both at splitting the SS at B and R (under B) levels:

     Case 1: `y ~ A/B + Error(B)` or `y ~ A*B + Error(B)` or `y ~ A+B + Error(B)`

```
summary(aov(y ~ A/B + Error(B)))

  Error: B
                  Df   Sum Sq  Mean Sq
  A               a-1    xxxxx    xxxxx
  A:B          a(b-1)    xxxxx    xxxxx

  Error: Within
                  Df   Sum Sq Mean Sq F value Pr(>F)
  Residuals   ab(r-1)    xxxxx   xxxxx    xxxx   xxxx
```

For `y ~ A*B + Error(B)` and `y ~ A*B + Error(B)`, `A:B` is changed to `B`. The model is saturated at Error: B level with no RSS (df=0) and so no F-test. But there is F-test at Error: Within (ie repeats R).

Case 2: `y ~ A + Error(B)`, ie 1-way ANOVA

```
summary(aov(y ~ A + Error(B)))

  Error: B
                  Df   Sum Sq  Mean Sq
  A               a-1    xxxxx    xxxxx
  Residual     a(b-1)    xxxxx    xxxxx

  Error: Within
                  Df   Sum Sq Mean Sq F value Pr(>F)
  Residuals   ab(r-1)    xxxxx   xxxxx    xxxx   xxxx
```

That is, replace A:B in case 1 with Residuals, then there will be F test for A and p-value under Error: B. In summary,
EU residual SS: $\text{SS}_B + \text{SS}_{A:B}$ with df $a(b-1)$.
OU residual SS: RSS with df $ab(r-1)$.

(For other way of defining EU, result may change, eg seed growth experiment.)

- In a split plot design, does whole plot/subplot associate with EU or OU?

  Answer: Whole plot with EU and subplot with OU.

2. Because the ANCOVA model can be written as

$$Y'_{ij} = Y_{ij} - \beta x_{ij} = \mu + \alpha_i + \epsilon_{ij},$$

the estimate of contrast $\sum_i c_i\alpha_i$ is $\sum_i c_i\bar{Y}'_{i\bullet} = \sum_i c_i(\bar{Y}_{i\bullet} - \hat{\beta}\bar{x}_{i\bullet})$ and the variance is

$$\text{Var}\left(\sum_{i=1}^{t} c_i\left(\bar{Y}_{i\bullet} - \hat{\beta}\bar{x}_{i\bullet}\right)\right) = \sigma^2\left(\sum_{i=1}^{t}\frac{c_i^2}{n_i} + \frac{(\sum_{i=1}^{t} c_i\bar{x}_{i\bullet})^2}{S_{0xx}}\right).$$

3. This contrast result and others can be obtained by setting up the design matrix $\boldsymbol{X}$ such that
   1. Parameter estimates are $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^\top\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{Y}$.
   2. Covariance $\widehat{\text{Cov}}(\boldsymbol{\beta}) = \hat{\sigma}^2(\boldsymbol{X}^\top\boldsymbol{X})^{-1}$, $\hat{\sigma}^2 = \frac{\text{RSS}}{n-t}$.
   3. Prediction value is $\widehat{Y}_0 = \boldsymbol{x}_0^\top\widehat{\boldsymbol{\beta}}$ and the variance is $\widehat{\text{Cov}}(\widehat{Y}_0) = \boldsymbol{x}_0^\top\widehat{\text{Cov}}(\widehat{\boldsymbol{\beta}})\boldsymbol{x}_0 + \hat{\sigma}^2$.
   4. The contrast is $\mathbf{c}^\top\boldsymbol{\beta}$ and its estimate is $\mathbf{c}^\top\widehat{\boldsymbol{\beta}}$ and its variance is $\hat{\sigma}^2\mathbf{c}^\top(\boldsymbol{X}^\top\boldsymbol{X})^{-1}\mathbf{c}$.

---

last adjustments: May 16, 2021 by JC