Chapter: Recovering the information from the Between Runs stratum

1. Intro

* + The problem – Why do I want to estimate the variance components (VCs) and approximate the effective degrees of freedom (EDF)? How does it help me in designing and analysing the experiments?
  + Describe how the VCs and EDF relevant to the problem.
  + Note that VC can be estimated by either linear combinations (LC) of the expected mean squares or restricted maximum likelihood methods (REML).
  + To decide which method to use, it depends on the EDF estimated, the higher the EDF estimated, the better the method.
  + The designs to be compared are completely randomised design (CRD), randomised block design (RBD) or balanced incomplete block design (BIBD). Hence, one question can be answer here is that which method is the best to use under which design.

2. An illustrative example

* + Using the example with v=2, r\_b = 3, n\_gamma = 4
  + Present the design and the theoretical ANOVA table

3. Theory for estimating the variance components

* + Briefly describe the estimation of VC using the LC method.
  + Describe the estimation of VC using the REML method
    - Construct the log-Likelihood function of the random variables base on chi-square distribution, where the random variables are the mean squares
    - Score function and Fisher’s information matrix
    - Fisher’s scoring algorithm for variance component estimation

4. Satterthwaite approximation for computing the EDF

* + Twice the square of the mean divided by the variance
  + The mean is computed from the associated mean squares by the VC estimated.
  + The variance is computed from the inverse of the score function which give the variance covariance matrix.

5. Illustrate EDF estimated from VC estimated using either LC or REML methods

* + Back to the example in Section 2.
  + EDF is in the range of 2 and 3.
  + EDF plots comparison
  + Power analysis which shows it is much better using the EDF estimated having the negative VC unadjusted.
  + This is because the negative VC will lead to higher EDF and lower estimated residual MS for conducting the tests for the treatment effects.

6. Illustrate the theory with the optimal designs found where the Phase 1 experiment is arranged in CRD

* Compare the EDF obtained from different methods, LC tend to be better with the low VC
* Compare between the 4-plex and 8-plex systems
  + v = 2, r\_b = 6: the simplest example where both 4- and 8-plex system has some DF associated with Between Animals stratum in the Between Runs stratum
  + v = 4, r\_b = 6: a case where EDF crossing over with different VC of between animals
  + v = 8, r\_b = 2: a case where EDF is better using 8-plex system
* Summary: The 4-plex is better in general because the theoretical ANOVA table from the optimal design with 4-plex system generally have 1 DF associated with Tag effect in the Between Animals Within Runs stratum, whereas the 8-plex system generally have 3 DF associated with tag effect in the Between Animals stratum. Thus, the experiments with 4 tags tend to have more than 2 DF to the experiment with 8 tags for estimating the residual MS in the Between Animals strata in Between and Within Runs strata.
* Based on the optimal designs founded, apart from the experiments comparing 2, 7 and 8 treatments, the crossing over from 8-plex becomes better when the experiment consists of more than 6 biological replicates and then the 4-plex experiment will become better as the VC of between animals increases.
* The cause of crossing over is that the 4-plex experiment will have more DF associated with animals confounded with runs than using the 8-plex experiment when the number of biological replicates is high.
* When there is some treatment information in the Between Runs stratum, the DF with pure animal effects in the Between Runs stratum will then decrease. This causes by the DF associated with the treatment effect. Thus, as the number of treatment increase, the DF with pure animal effects in the Between Runs stratum will then decrease.
* Based on the optimal design founded:
* When comparing between 2 treatments, 4-plex system is always better.
* When comparing between 6 or 7 treatments using 4 biological replicates, 8-plex is better.
* When comparing between 8 treatments, 8-plex system is always better, because the treatment does not confound with animals in the Between Runs stratum which maximised the DF with pure animal effects in the Within Runs stratum.

7. Illustrate the theory with the optimal designs found where the Phase 1 experiment is arranged in RBD

* Additional comparison with different initial designs where cages is confounded more with runs or tags.
* Compare between different initial designs as well as between the 4-plex and 8-plex systems
* Example with v = 4, r\_b = 4, n\_B = 4, the treatment average efficiency factor is always 1
  + 4-plex: initial design with cage confounded more with run, the EDF is always **8**. The initial design where cage is confounded more with tag, the EDF can be within **7 and 9** if VC of between animals increases.
  + 8-plex: initial design with cage confounded more with run, the EDF is always **7**. The initial design where cage is confounded more with tag, the EDF can be within **8 and 9** if VC of between animals increases.
  + Thus, using the **initial design where cage is confounded more with tag with 8-plex system** is better.
* I found that the VC of between cages does not affect the EDF. This is because the animals are nested from cages; thus, when the EMS consists of VC of between cages the given EMS should also consists of VCs of between animals.
* Recall from the previous section, the experiments with 4-plex system generally have 1 DF associated with Tag effect in the Between Animals Within Runs stratum, whereas the 8-plex system generally have 3 DF associated with tag effect in the Between Animals stratum.
* Only when the cage number is even, using the initial design where the cage is confounded more with tags is better, because it allows by having more DF associated with tag effects in the Between Cages stratum; thus it maximised the overall DF associated Between Animals Within Cages strata in both Between and Within Runs strata.
* More specifically, when using experiment consist of even cages, but apart from 4 or 8 cages, 4-plex system is better than 8-plex system. It is because there is only 1 DF associated with cages that can be confounded with the tag effects. If the 4-plex system is used, it is possible to confound that the cages with that1 DF associated the tag effects; hence, the animals is then orthogonal to tags. If the 8-plex system is used, there is still at least 2 DF associated with the tag effects in the Between Animals Within Cages Within Runs stratum. e.g. v = 2, r\_b = 6, n\_B = 2, n\_gamma = 4, EDF is between 7 and 9. n\_gamma = 8 EDF is between 6 and 7.
* The 8-plex system is only can be better if the experiment consists of 4 or 8 cages to allow the cages confounded all 3 DF of the tags effects and maximised the overall DF associated in the Between Animals Within Cages strata, e.g. the example presented.
* The remaining cases follows the examples where the Phase 1 experiment is arranged in CRD.

8. Illustrate the theory with the optimal designs found where the Phase 1 experiment is arranged in BIBD

* Follow the same rule as described in the previous two sections.
* For the experiment examples when comparing between 4 and 8 treatments, since these two examples used 4 and 8 cages, respectively, the 8-plex system is better 4-plex system.
* For the examples when comparing between 5 or 7 treatments, since these two examples does not use either 4 or 8 cages, the 4-plex system is better than the 8-plex system.

9. Summary and Conclusion

* + Discuss the implementation of the getVcEDF