MA50259: Statistical Design of Investigations

Dr. Sandipan Roy

Lecture 5: Blocking

Blocking

- **Objective:** Reduce the variance of the experimental error (σ^2) and increase the power for detecting treatment factor effects so that results **generalise to whole population**
- Choose the experimental units for a study to be as homogeneous as possible. Sometimes difficult!
- Heterogeneous experimental units are grouped into homogeneous subgroups before they are randomly assigned to treatment factor levels
- The act of grouping the experimental units together in homogeneous groups is called blocking.
- In a **randomized block design**, a group of heterogeneous experimental units is used so that the conclusions can be more general

Examples of blocking

- Plots of land in agricultural experiments are usually **blocked by proximity** because plots in close proximity normally have similar soil characteristics
- When experimental units are animals, the **grouping (blocking) of genetically similar animals**, such as littermates, often reduces variability within groups
- When experimental units are trials, or points in time where treatments will be applied, they are often **blocked by time** since many lurking variables may change over time and trials in close temporal proximity are more alike

Randomized complete block design (RCB) with one treatment factor

- Treatment factor has t levels
- b blocks (or subgroups of homogeneous experimental units)
- Each block contains exactly t experimental units for a total of $t \times b$ experimental units
- The t experimental units within each block are as similar as possible
- The groups of experimental units vary enough from block to block to allow general conclusions to be drawn
- The randomization of experimental units to treatment factor levels is performed within each block.

Comparison betwwen CRD and RCB $t = 3$, $b = 4$

```
levels<-c("level 1","level 2","level 3")
fac <- levels %>% rep(each = 4) %>% sample(12) %>% factor()
blocks <- factor( rep(c("block 1", "block 2", "block 3", "block 4"), each=3))
CRD <- data.frame( units=1:12,block=blocks,treatmentCRD=fac)
block1 <- sample(levels,3); block2 <- sample(levels,3)
block3 <- sample(levels,3); block4 <- sample(levels,3)
t<-c(block1,block2,block3,block4) %>% factor()
RCB<-data.frame(block = blocks, treatmentRCB = t)
cbind(CRD,RCB)
   units   block treatmentCRD   block treatmentRCB
1      1 block 1      level 2 block 1      level 3
2      2 block 1      level 3 block 1      level 2
3      3 block 1      level 3 block 1      level 1
4      4 block 2      level 1 block 2      level 3
5      5 block 2      level 2 block 2      level 2
6      6 block 2      level 2 block 2      level 1
7      7 block 3      level 1 block 3      level 2
8      8 block 3      level 2 block 3      level 1
9      9 block 3      level 1 block 3      level 3
10    10 block 4      level 3 block 4      level 2
11    11 block 4      level 3 block 4      level 1
12    12 block 4      level 1 block 4      level 3
```

Statistical model

$$y_{ij} = \mu + b_i + \tau_j + \epsilon_{ij}$$

where

- b_i is the effect of block $i \in \{1, \ldots, b\}$
- $\tau_1, \ldots, \tau_t$ are the treatment effects
- All $\{\epsilon_{ij}\}$ are $N(0, \sigma^2)$ and independent
- Note: There is no interaction between block and treatment
- Only $t \times b$ experimental units, there would be zero degrees of freedom for the error term *ssE* if a block by treatment interaction term were included!