

Assignment 1

Your Name

2025-08-25

Goal of this assignment

The goals of this assignment are to:

- Review a few basic concepts, and
- Make sure that you can knit an Rmd document with the basic features that will be needed throughout this course.

1.

Using your own words, describe a randomized complete block design (RCBD) in 3-5 sentences. Then write out a statistical model (using mathematical notation) that could be fitted to data generated by an RCBD.

An RCBD is a type of experiment design where there were groups of similar experimental units (i.e., the blocks) and all treatments fit once in each block. The capacity to run an experiment in an RCBD is independent of the treatment structure. To design an RCBD, we would randomize all treatments within each block. The “block effect” is typically considered a nuisance parameter, but they account for known variability in the data and may reduce the residual variance σ_ε^2 . We assume that there is no block \times treatment interaction.

A statistical model that could model data generated by an RCBD is

$$y_{ij} = \mu + T_i + b_j + \varepsilon_{ij}, \varepsilon_{ij} \sim N(0, \sigma^2),$$

where:

- y_{ij} is the observation of the i th treatment in the k th block,
- μ is the overall mean,
- T_i is the effect of the j th treatment,
- b_j is the effect of the k th block,
- ε_{ij} is the residual of the i th treatment in the j th block,
- σ^2 is the residual variance.

A more broad statistical model that could model data generated by an RCBD is

$$y_{ij} \sim P(\mu_{ij}, \phi), g(\mu_{ij}) = \eta_{ij} = \eta_0 + T_i + b_j,$$

where:

- y_{ij} is the observation of the i th treatment in the k th block, that arises from probability distribution P ,
- μ_{ij} is the mean of said observation and distribution,
- ψ is the dispersion parameter of the distribution P ,
- $g(\cdot)$ is the link function that connects the linear predictor to the mean, so that the value for the mean is always within its support,
- η_0 is the overall mean of the linear predictor,

- T_i is the effect of the j th treatment,
- b_j is the effect of the k th block.

2.

Edit the R code below:

- Silence the `messages` and `warnings`: your submitted pdf or html should include the code, but not the warnings/messages seen below.
- Print a summary of `model1`.
- Find and print the value for $\hat{\sigma}^2$ estimated in `model1`.

```
library(tidyverse)
library(agridat)
data("omer.sorghum")
df <- omer.sorghum %>% filter(env == "E3")
model1 <- lm(yield ~ 1 + gen + rep, data = df)
model1

##
## Call:
## lm(formula = yield ~ 1 + gen + rep, data = df)
##
## Coefficients:
## (Intercept)      genG02      genG03      genG04      genG05      genG06
##      672.03      -66.52      223.07      167.40       61.05     -267.21
##      genG07      genG08      genG09      genG10      genG11      genG12
##      355.20      159.53      231.30      156.66      146.29      -42.87
##      genG13      genG14      genG15      genG16      genG17      genG18
##      243.17        1.06       64.72       22.81     -296.66      448.28
##      repR2      repR3      repR4
##     -150.83     -85.66    -124.21

# different ways to get #sigma^2
sigma(model1)^2

## [1] 25627.36

summary(model1)$sigma^2

## [1] 25627.36

sum(model1$residuals^2) / model1$df.residual

## [1] 25627.36
```