Assignment 1

Test Student

2025-06-13

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 the same number of times in the same 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 normally assume that there is no block $\times$ treatment interaction.

A statistical model that could correspond to an RCBD is

$$y_{ijk} = \mu + T_j + b_k + \varepsilon_{ijk}, \varepsilon_{ijk} \sim N(0, \sigma^2),$$

where:

- μ is the overall mean,
- T_j is the effect of the jth treatment,
- b_k is the effect of the kth block,
- ε_{ijk} is the residual of the ith observation, jth treatment, and kth block,
- σ^2 is the residual variance.

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`.

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

```r
summary(model1)
```

```
##
## Call:
## lm(formula = yield ~ 1 + gen + rep, data = df)
##
## Residuals:
##      Min      1Q  Median      3Q     Max
```

```
## -349.66  -81.58   -1.08   78.47  318.59
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    672.03      86.46   7.773 3.29e-10 ***
## genG02         -66.52     113.20  -0.588 0.559393
## genG03         223.07     113.20   1.971 0.054201 .
## genG04         167.40     113.20   1.479 0.145336
## genG05          61.05     113.20   0.539 0.591996
## genG06        -267.20     113.20  -2.361 0.022111 *
## genG07         355.20     113.20   3.138 0.002826 **
## genG08         159.53     113.20   1.409 0.164820
## genG09         231.31     113.20   2.043 0.046198 *
## genG10         156.66     113.20   1.384 0.172393
## genG11         146.29     113.20   1.292 0.202071
## genG12         -42.87     113.20  -0.379 0.706453
## genG13         243.18     113.20   2.148 0.036467 *
## genG14           1.06     113.20   0.009 0.992565
## genG15          64.72     113.20   0.572 0.570007
## genG16          22.81     113.20   0.201 0.841122
## genG17        -296.66     113.20  -2.621 0.011534 *
## genG18         448.28     113.20   3.960 0.000233 ***
## repR2         -150.83      53.36  -2.827 0.006704 **
## repR3          -85.66      53.36  -1.605 0.114592
## repR4         -124.21      53.36  -2.328 0.023936 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 160.1 on 51 degrees of freedom
## Multiple R-squared:  0.6748, Adjusted R-squared:  0.5473
## F-statistic: 5.291 on 20 and 51 DF,  p-value: 7.64e-07
```

```r
# you only need to do one of these. this is just to show 3 possible responses:
sigma(model1)^2
```

```
## [1] 25627.36
```

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

```
## [1] 25627.36
```

```r
summary(model1)$sigma^2
```

```
## [1] 25627.36
```