

Purpose:

Simulate phenotypic values based on  $P = \mu + E_i + G_j + GE + \varepsilon_{ijk}$ ;

- a.  $\varepsilon_{ijk} \sim \text{i.i.d. } N(0, \sigma_\varepsilon)$ ,
- b.  $G_j \sim \text{i.i.d. } N(0, \sigma_g)$

Goal:

Build confidence in the use of simulation models

Keywords:

Simulation, modeling.

ALA: How to use simulation modeling.

Create a data set consisting of 400 observations consisting of 50 entries evaluated in two reps at four locations using the simulation model:

$$y_{ld} = \mu + E_i + G_j + GE_{ij} + \varepsilon_{ijk};$$

- $\mu = 150$
- $E_i = \{125, 135 \text{ for } i = 1, 2 \text{ respectively}\} \text{ and } \{165, 175 \text{ for } i = 3, 4 \text{ respectively}\}$
- $G_j \sim N(0, 11)$
- $GE_{ij}$ : multiply  $G_j$  by 7/11 for  $i = 1, 2$  and by 15/11 for  $i = 3, 4$
- $\varepsilon_{ijk} \sim \text{i.i.d. } N(0, 12)$

1. Conduct an EDA on "QG\_Mod13\_ALA13.1\_ds.csv".
2. Estimate the genotypic and residual variance for each environment.
3. Estimate the genotype x environment variance. What proportion is due to heterogeneity and what proportion is due to change of rank?
4. Are the relative contributions of heterogeneity and change of rank consistent with the values that you simulated?
5. Conduct an EDA on "QG\_Mod13\_ALA13.1\_dsb.csv" and perform similar analyses as in questions 2 to 4.
6. Describe the impact of missing data on the outcomes.
7. Create a new data set consisting of 400 observations consisting of 50 entries evaluated in two reps at four locations using the simulation model:

$$y_{ld} = \mu + E_i + G_j + GE_{ij} + \varepsilon_{ijk};$$

- $\mu = 150$
- $E_i = \{125, 135 \text{ for } i = 1, 2 \text{ respectively}\} \text{ and } \{165, 175 \text{ for } i = 3, 4 \text{ respectively}\}$
- $G_{j(1,2)} \sim \{\text{i.i.d. } N(0, 11)\}$  and  $\text{correlation}(G_{j(1,2)}, G_{j(3,4)}) = 0.65$

- $\varepsilon_{ijk} \sim \text{i.i.d. } N(0,12)$

Name the file "QG\_Mod13\_ALA13.1\_dsc.csv".

( See QG\_Mod13\_ALA13.1.xlsx)

8. Conduct an EDA on "QG\_Mod13\_ALA13.1\_dsc.csv" and perform similar analyses as in questions 2 to 4. (See QG\_Mod13\_ALA13\_1.R).

Randomly convert 15% of the yld values to missing values and name the file "QG\_Mod13\_ALA13.1\_dsd.csv"

9. Conduct an EDA on "QG\_Mod13\_ALA13.1\_dsd.csv", perform similar analyses as in questions 2 to 4 and describe the impact of missing data on the outcomes.