

**Purposes:**

1. Assess whether students know how to conduct appropriate data analyses on data from a Multi-Environment-Trial.
2. Assess whether students know how to interpret results from data analyses and communicate their interpretations.
3. Assess whether students understand how to make and communicate decisions based on justifiable reasons.

**Keywords:** Exploratory Data Analyses (EDA), Histogram, Boxplot, Multi-Environment Trial (MET)

**References:**

Plant Breeding Basics

**Useful R functions**

- `getwd()`
- `setwd()`
- `read.csv()`
- `rm()`
- `rm(list=())`
- `hist()`
- `attach()`
- `boxplot()`
- `str()`
- `as.factor()`
- `aov()`
- `summary()`

ALA:

Let's consider a data set from an early stage of field trials. At this phase the primary goal is to identify potential cultivars (lines, hybrids, synthetics) that yield more than a check cultivar by an average of 2% across a set of environments that are representative of the market place (also known as the target population of environments). You need to identify experimental lines that are 2% better than a check cultivar and make a decision as to whether these experimental lines should be selected for the breeding nursery (for the genetic improvement project) and for further cultivar development. The stage after this stage of cultivar development is very expensive because many traits need to be evaluated in many more environments. At the same time the sales force, consisting mainly of agronomists as well as production staff and the owners of the company are anxious to have a cultivar that will enable your organization to gain market share, i.e., your future depends on your decision.

After the preliminary field trial in which experimental lines were evaluated at a single location for maturity, height and other non-yield traits, a set of 49 experimental lines and a check cultivar were entered into a replicated multi-environment trial (MET). The experimental lines are designated 1 – 49 and the check is designated as line 50. Harvestable yield from two row plots were obtained from 10 environments representing the targeted market in a single growing season. The lines were assigned to the field plots using a randomized complete block design, where each block is represented as one of two replicates per location.

1. Download the data set (Review EDA with R `ds3.csv`) to a file folder in your personal computer.
2. Read the data set from your desktop into the R base-package
3. Provide a brief interpreted report of the following:
  - a. Evidence that your installation of R is reading the data correctly.
  - b. Evidence that you know how to conduct Exploratory Data Analyses (EDA: histograms and boxplots of yield data, estimates of means, medians, standard errors, CV's etc. by location and across locations).
  - c. Evidence that you understand how to make decisions based on the quality of data from the EDA.