Purposes:

1. Distinguish prediction from estimation.

Keywords: Multi-Environment Trial (MET), residual variance, i.i.d. $N(0,\sigma^2)$ Data transformation,

Useful R commands

- getwd()
- setwd()
- read.csv()
- rm()
- attach()
- as.factor()
- lm()
- aov()
- summary()

ALA:

Historically plant breeders have relied heavily on retrospective analyses such as the analyses that we have conducted to date. At this point, we have some reasonably precise estimates of yield that our experimental lines produced. As Plant Breeders we select those lines that met or exceeded some criteria. For example we found experimental line 17 exceeded the check and sub-consciously, we fully expect that the line will exceed the check by about the same amount in the future. This is not a prediction, it's an estimate of past performance. Most experienced plant breeders also worry that this may be a 'one year wonder'.

An alternative is to predict the yield of the experimental line. Animal breeders have been predicting values for at least 40 years while the concept of prediction has been implemented only recently in the most progressive and well-funded plant breeding programs. The primary challenge of implementing prediction methods in plant breeding is cultural. Most plant breeder's education has not included sufficient emphasis on regression and analyses of covariance. Rather considerable emphasis has been placed on experimental designs used by Agronomists, but seldom (if ever) used in genetic improvement programs.

In the data set "Review Statistical Inference Regression and Prediction ds4.csv", consider the variable labeled mscore. This is a molecular marker score that your colleagues in the molecular biology lab have developed. They claim that if you know the mscore, it is possible to predict the yield of the line.

1. Is this true?
2. Is it true for some of the environments, but not others? Hint: Model the relationship between mscore and yld as unique for types of environments.
3. What is the predicted value for the check based on the mscore? What are the predicted values for experimental lines with higher mscores? Are these predicted values significantly different than the check?