



The SolCAP Tomato Phenotypic Data: Estimating Heritability and Trait BLUPs



Dr. Heather L. Merk
The Ohio State University, OARDC



Before Moving Forward, You May Wish to...

- Download, install, and open R
 - R is available at www.r-project.org
- Download the script and data
 - Available at www.extension.org/pages/61006
 - Script - TBRT2011.txt
 - Data - TBRTQuality.csv

Objectives

- Locate and obtain SolCAP phenotypic data
- Estimate trait heritability on a line mean basis
- Estimate BLUPs for a trait on a per line basis

Experiment Design

- Field evaluations conducted in 2009, 2010
- Locations in CA, OH
- 2 reps in each year in each location (OH 2010 1 rep only)
- Randomized complete block design
- 143 processing breeding lines and varieties
 - Donated by public breeding and genetics programs or commercially available
- Multiple fruit collected for each plot

Data Collection

- Fruit cut & scanned for size, shape, & color analysis
- Images analyzed using TomatoAnalyzer
<http://www.extension.org/pages/32374>
- Quality analysis – pH, Brix, titratable acids
- Obtain standardized data that can be stored in databases



locate Data – SolCAP Website

www.solcap.msu.edu



Solanaceae Coordinated Agricultural Project



Home

Potato

Tomato

Education/Outreach

Tools

Publications

The SolCAP project links together people from public institutions, private institutions and industries who are dedicated to the Solanaceae crops: potato and tomato. Through research, education and training the SolCAP project will focus on genetic advances to US tomato and potato breeding that will lead to significantly improved varieties that benefit the processors, the consumer and the environment.

Germplasm


Phenotype Data

GenotypeData

http://www.solcap.msu.edu/tomato_phenotype_data.shtml

Home	Tomato Phenotype Data
Potato	The tomato cooperator's guide organizes standardized phenotypic data from a collection of elite US tomato germplasm that will be collected over two years across distinct environments (CA, OH, NC, FL). Trials have been organized by market class with fresh market material grown in CA, NC, and FL. Processing germplasm is evaluated in CA and OH. Plots of each tomato line will be replicated twice at each site. The phenotypic data emphasizes fruit quality characteristics. Passport data is also collected for each site and year. These data will be used in combination with SNP genotyping to survey the distribution of phenotypic and genotypic variation among varieties and breeding lines.
Tomato	
Education/Outreach	
Tools	
Publications	
Impacts and Outcomes	
Links	
Downloads	
Contacts	

The SolCAP project is supported by the Agriculture and Food Research Initiative Applied Plant Genomics CAP Program of USDA's National Institute of Food and Agriculture.



United States Department of Agriculture
National Institute of Food and Agriculture

Cooperator Guides:

- [2009 California Fresh Market Material](#)
- [2009 California Processing Material](#)
- [2009 North Carolina Fresh Market Material](#)
- [2009 North Carolina Vintage Material](#)
- [2009 Ohio Processing Material](#)
- [2009 Ohio Vintage Material](#)
- [2010 California Fresh Market Material](#)
- [2010 California Processing Material](#)
- [2010 Florida Fresh Market Material](#)
- [2010 North Carolina Fresh Market Material](#)
- [2010 North Carolina Vintage Material](#)
- [2010 Ohio Processing Material](#)
- [2010 Ohio Vintage Material](#)

Transverse and Cross Section Fruit Images:

- [2009 California Fresh Market Material](#)
- [2009 California Processing Material](#)
- [2009 North Carolina Vintage Material](#)
- [2009 Ohio Processing Material](#)
- [2009 Ohio Vintage Material](#)

- Data available in coop guides (.xls format)
- Images available in zip folders – **WARNING – these are large!**

Locate Data – SGN Website

www.solgenomics.net

- Data will be available in a searchable database

The screenshot displays the Sol Genomics Network (SGN) website interface. At the top left is the SGN logo, a stylized sunburst. The main header reads "sol genomics network" and includes navigation links: "home | forum | contact | help | faq". Below the header is a search bar with the text "search" and "maps genomes tools" on either side, and a "sol search" button. To the right of the search bar are links for "log in" and "new user".

The main content area features several interactive tiles:

- Maps & Markers:** A tile showing a chromosome map with markers labeled "CT233", "CD15", and "C2_At4g15790".
- Genes:** A tile featuring a 3D DNA double helix icon.
- Phenotypes:** A tile showing a photograph of various colored peppers (red, green, yellow).
- Breeders Toolbox:** A tile with an icon of a red toolbox and several vegetables (tomatoes, eggplant, peppers).
- Genomes & Sequences:** A tile showing a colorful representation of a genome map with various colored bands.
- Pathways:** A tile displaying a chemical structure diagram of a sugar molecule, likely sucrose, with various functional groups labeled.

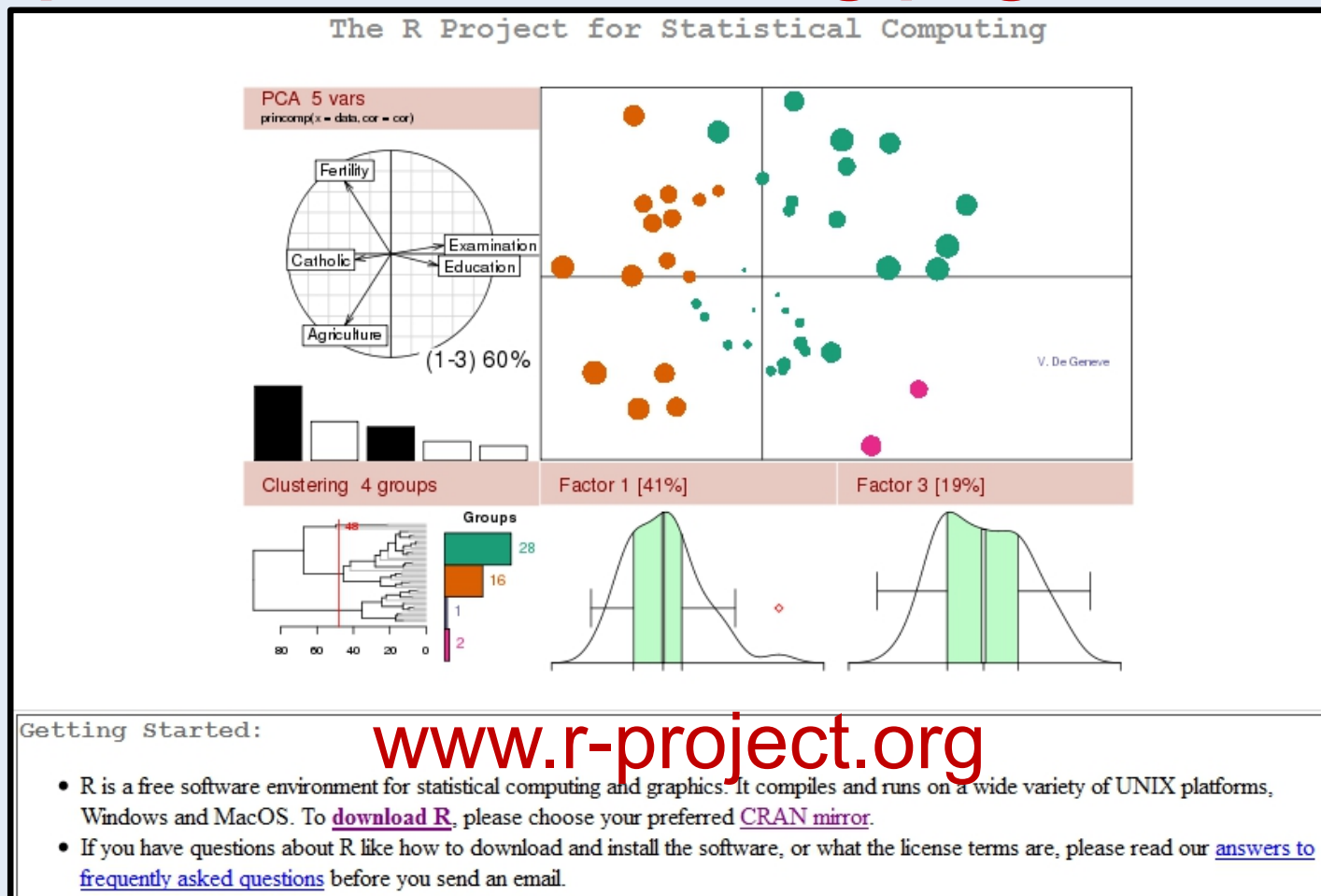
Sample Analysis Using Brix Data

- Data file ready for analysis: TBRTQuality.csv
- R Script used for analysis: TBRT2011.txt
- Available at: www.eXtension.org/pages/61006

Working with R

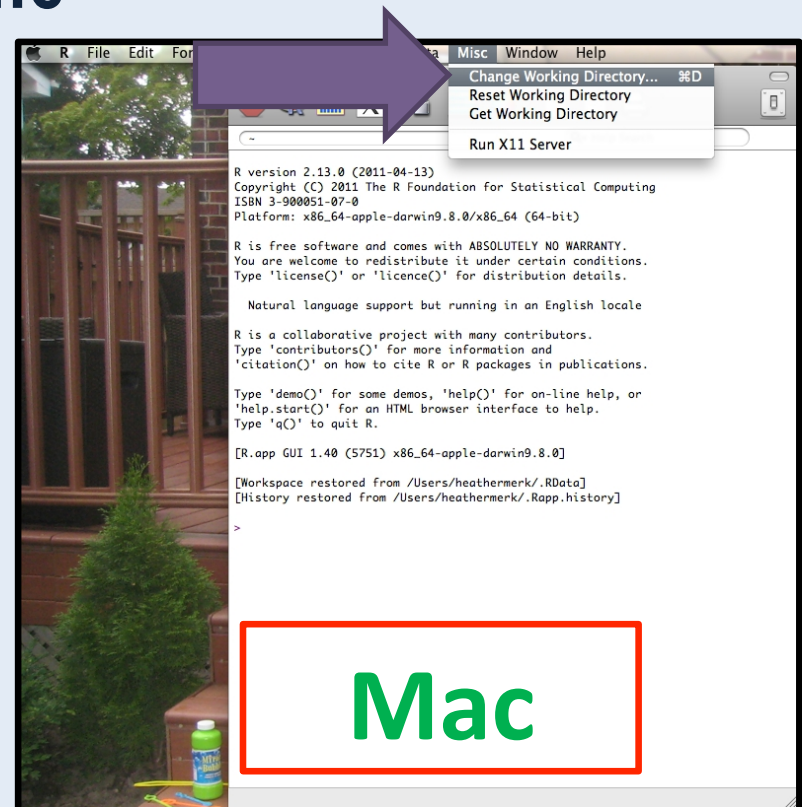
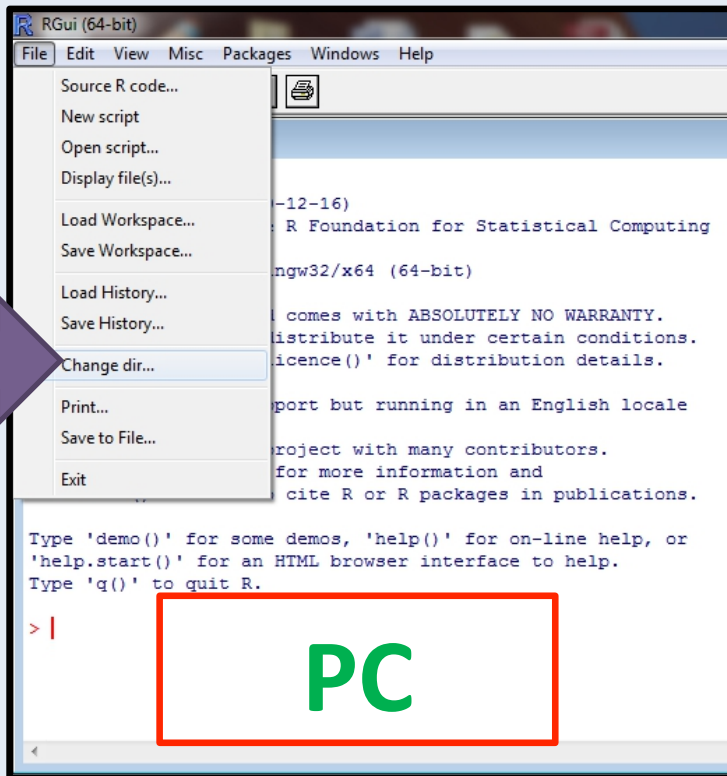
Introduction to R Webinar Recording

<http://www.extension.org/pages/60427>

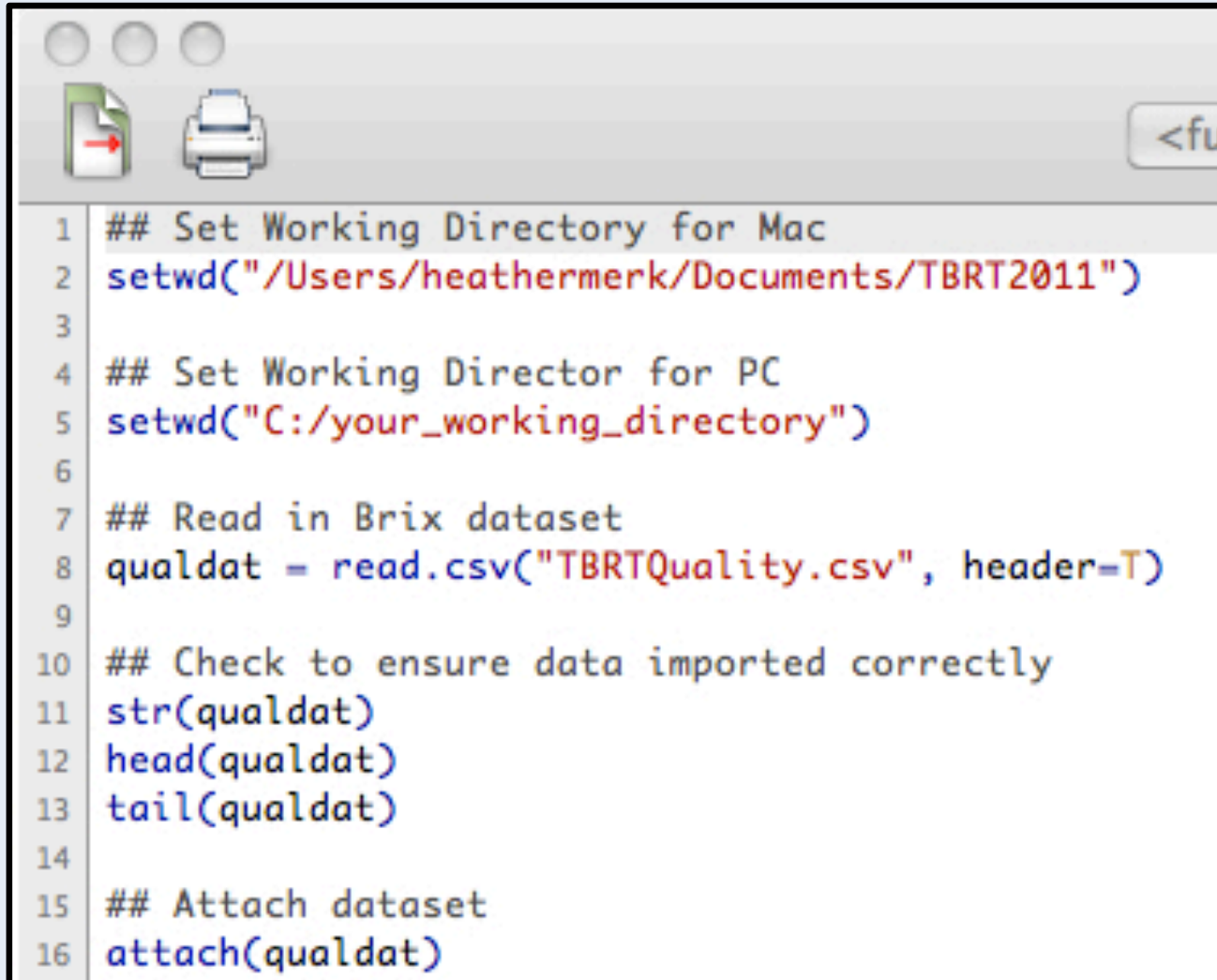


Setting Working Directory in R

- Simplifies commands for importing and exporting data by eliminating the need to write full path statements



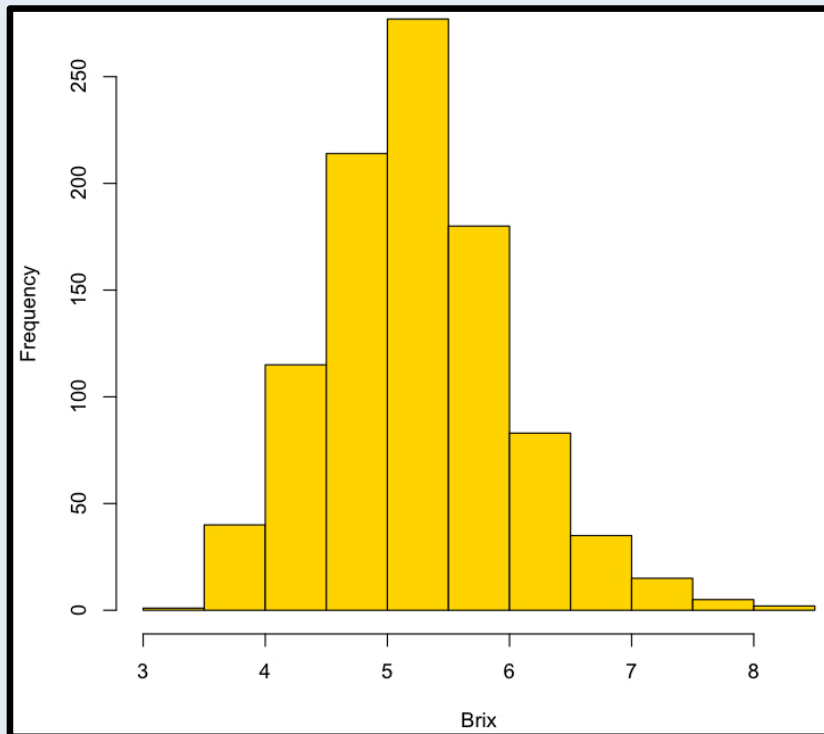
Importing & Checking Data



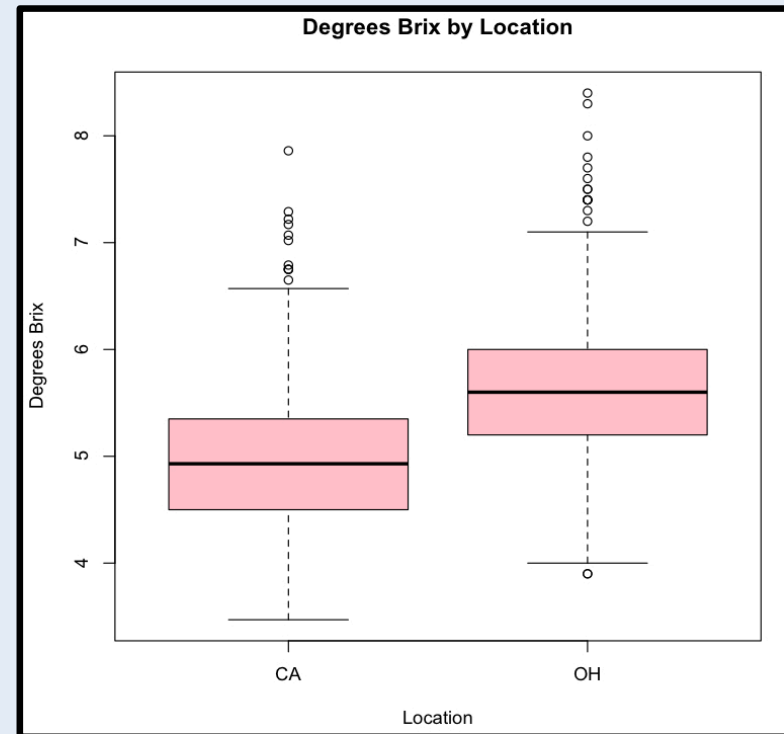
```
1 ## Set Working Directory for Mac
2 setwd("/Users/heathermerk/Documents/TBRT2011")
3
4 ## Set Working Director for PC
5 setwd("C:/your_working_directory")
6
7 ## Read in Brix dataset
8 qualdat = read.csv("TBRTQuality.csv", header=T)
9
10 ## Check to ensure data imported correctly
11 str(qualdat)
12 head(qualdat)
13 tail(qualdat)
14
15 ## Attach dataset
16 attach(qualdat)
```

Visualizing Data

```
hist(Brix, col="gold")
```



```
boxplot(Brix ~ Loc,  
        xlab="Location",  
        ylab="Degrees Brix", main  
        = "Degrees Brix by  
        Location", col="pink")
```



Multi-year, Multi-location Data

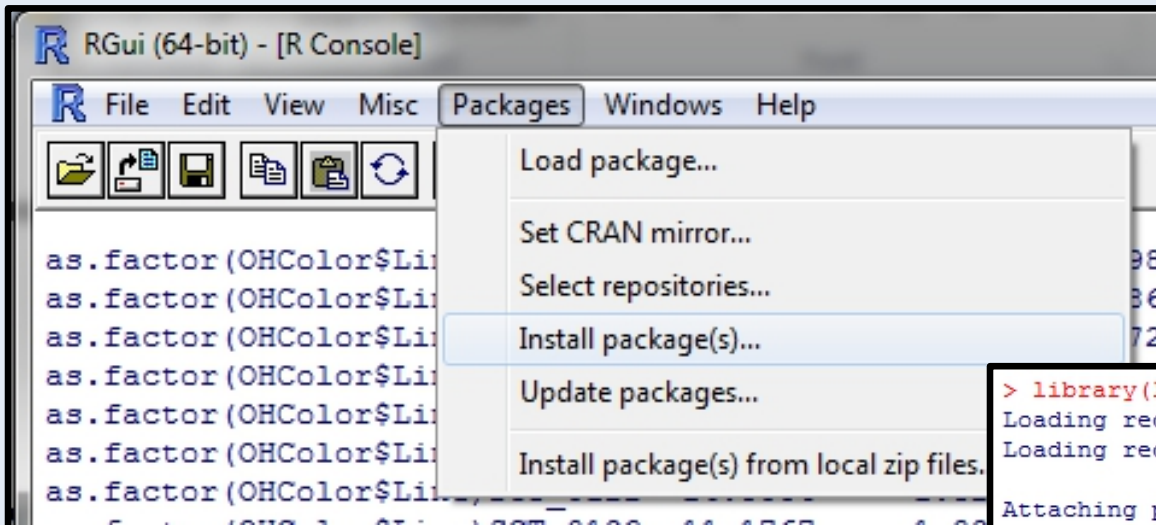
- Brix data from 143 lines collected in 2 years, 2 locations, 2 reps per location
- Model - (BRIX ~ (LINE) + (LOC) + (YEAR) + (REP%in%LOC:YEAR) + (LINE:LOC) + (LINE:YEAR))
 - Denote nesting using %in%
 - Denote interactions using a colon or asterix between terms

Treating Main Effects

- Need to make decisions of whether to treat main effects as fixed or random
 - This has implications for analysis
- Random effects – used when we believe our data are drawn from a subset representing all possible genotypes (locations, years, etc...)
- For many plant breeding applications we would consider main effects to be random, and would want to estimate the proportion of variance due to effects in our experimental design (e.g. estimates of heritability)
- Use lme4 package in R

Installing & loading lme4 Package

- Install - first time



- Load - every time
- `library(lme4)`

```
> library(lme4)
Loading required package: Matrix
Loading required package: lattice
Attaching package: 'Matrix'

The following object(s) are masked from 'package:base':

  det

Attaching package: 'lme4'

The following object(s) are masked from 'package:stats':

  AIC

Warning messages:
1: package 'lme4' was built under R version 2.12.2
2: package 'Matrix' was built under R version 2.12.2
> |
```


Calculating Variance Components

- `brixvarcomp = lmer(BRIX ~ (1 | LINE) + (1 | LOC) + (1 | YEAR) + (1 | REP%in%LOC:YEAR) + (1 | LINE:LOC) + (1 | LINE:YEAR))`
- `summary(brixvarcomp)`

```
Linear mixed model fit by REML
Formula: BRIX ~ (1 | LINE) + (1 | LOC) + (1 | YEAR) + (1 | REP %in% LOC:YEAR) + (1 |
LINE:LOC) + (1 | LINE:YEAR)
   AIC   BIC logLik deviance REMLdev
1758 1797   -871    1742    1742
Random effects:
Groups      Name      Variance   Std.Dev.
LINE:YEAR   (Intercept) 0.01021213 101055
LINE:LOC    (Intercept) 0.00948028  97367
LINE        (Intercept) 0.19193287 438101
YEAR        (Intercept) 0.03164132 177880
LOC          (Intercept) 0.18845950 434119
REP %in% LOC:YEAR (Intercept) 0.00056113  23688
Residual    (Intercept) 0.25781992 507760
Number of obs: 967, groups: LINE:LOC, 281; LINE, 143; YEAR, 2; LOC, 2; REP %in%
LOC:YEAR, 1

Fixed effects:
              Estimate Std. Error t value
(Intercept)    5.2817    0.3351   15.76
```

Estimating Heritability on a Line Mean Basis

- Heritability = $\text{var}(\text{LINE}) / [\text{var}(\text{LINE}) + \text{var}(\text{LINE:LOC})/2 + \text{var}(\text{LINE:YEAR})/2 + \text{var}(\text{RESIDUAL})/4]$
 $= 0.1919 / [0.1919 + (0.0095/2) + (0.0102/2) + (0.2578/4)]$
 $= 0.72$

```
Linear mixed model fit by REML
Formula: BRIX ~ (1 | LINE) + (1 | LOC) + (1 | YEAR) + (1 | REP %in% LOC:YEAR) + (1 | LINE:LOC) + (1 | LINE:YEAR)
AIC   BIC  logLik deviance REMLdev
1758 1797   -871    1742    1742
Random effects:
Groups      Name          Variance  Std.Dev.
LINE:YEAR   (Intercept)  0.01021213  0.101055
LINE:LOC    (Intercept)  0.00948028  0.097367
LINE        (Intercept)  0.19193287  0.438101
YEAR        (Intercept)  0.03164132  0.177880
LOC         (Intercept)  0.18845950  0.434119
REP %in% LOC:YEAR (Intercept)  0.00056113  0.023688
Residual                    0.25781992  0.507760
Number of obs: 967, groups: LINE:YEAR, 284; LINE:LOC, 281; LINE, 143; YEAR, 2; LOC, 2; REP %in% LOC:YEAR, 1
Fixed effects:
              Estimate Std. Error t value
(Intercept)  5.2817     0.3351   15.76
```

Selection

- When estimating heritability, ignore main environmental effects
- Justification for doing so is in part because assume that will account for these effects prior to selection
- Make selections based on BLUPs

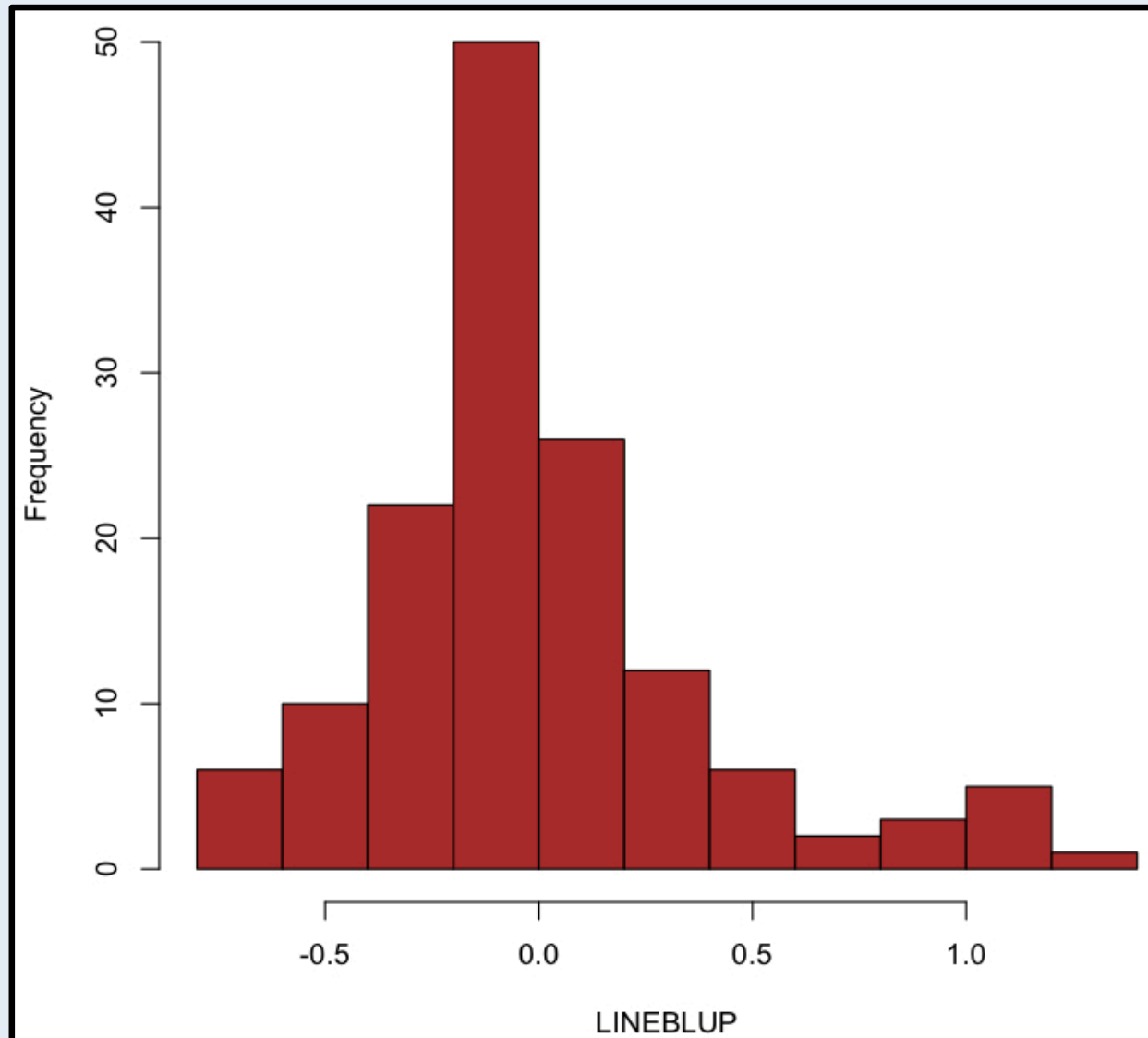
BLUPs

- Best Linear Unbiased Predictor
- Allows us to account for environmental factors in our model, missing data, etc..
- BLUPs shrink estimates toward the mean
- Calculate a BLUP for each line for each trait
- Make selections based on BLUPs

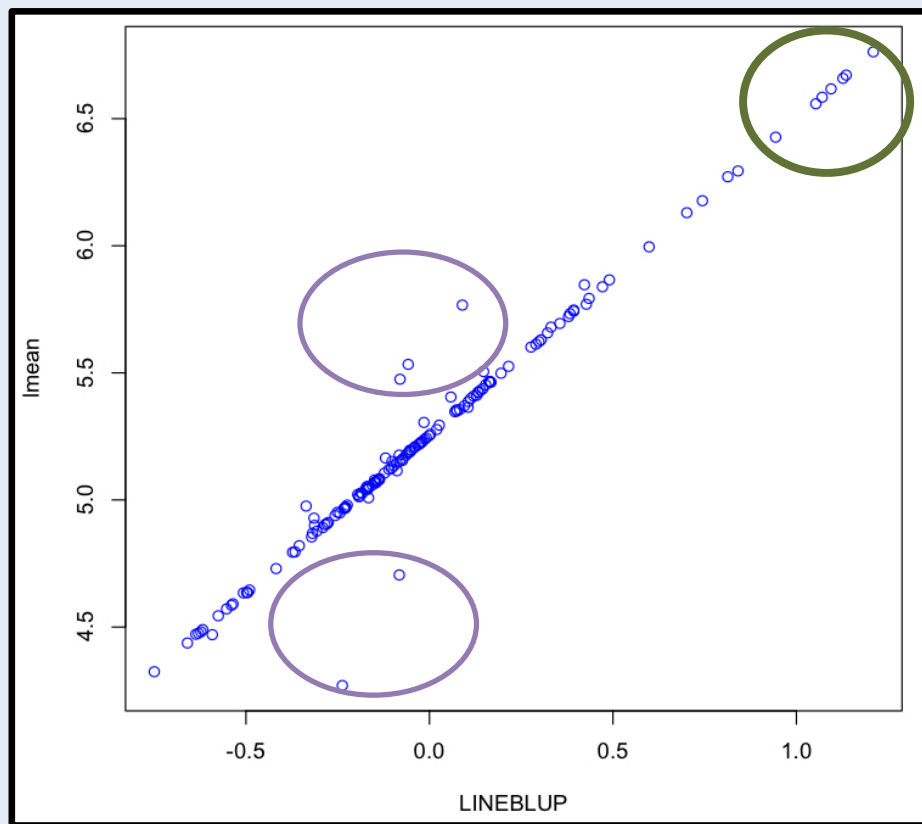
Estimating BLUP for Each line

```
> ## BLUPS
> # fit the model
> brixmodel = lmer(BRIX ~ (1|LINE) + (1|LOC) + (1|YEAR) + (1|REP%in%LOC:YEAR) + (1|LINE:LOC) +
(1|LINE:YEAR))
> # estimate BLUPS
> brixblup = ranef(brixmodel)
> # look at output structure
> str(brixblup)
List of 6
 $ LINE:YEAR      : 'data.frame':  284 obs. of  1 variable:
  ..$ (Intercept): num [1:284] -0.01287 0.00562 0.04129 -0.04924 0.00459 ...
 $ LINE:LOC       : 'data.frame':  281 obs. of  1 variable:
  ..$ (Intercept): num [1:281] -0.00806 0.00132 0.01163 -0.01901 0.0381 ...
 $ LINE          : 'data.frame':  143 obs. of  1 variable:
  ..$ (Intercept): num [1:143] -0.1363 -0.1494 -0.0535 -0.0388 1.1357 ...
 $ YEAR          : 'data.frame':   2 obs. of  1 variable:
  ..$ (Intercept): num [1:2] 0.125 -0.125
 $ LOC           : 'data.frame':   2 obs. of  1 variable:
  ..$ (Intercept): num [1:2] -0.306 0.306
 $ REP %in% LOC:YEAR: 'data.frame':  1 obs. of  1 variable:
  ..$ (Intercept): num -1.59e-13
- attr(*, "class")= chr "ranef.mer"
> # extract blup for line
> brixlineblup = brixblup$LINE
> # see the structure of the blup for each line
> str(brixlineblup)
'data.frame':  143 obs. of  1 variable:
 $ (Intercept): num  -0.1363 -0.1494 -0.0535 -0.0388 1.1357 ...
> # save the brixlineblup output to a separate .csv file
> write.csv(brixlineblup, file="BrixLineBLUPS.csv")
```

BLUP Distribution



BLUPs and line Means



- Highest values for lines developed for arid environment
- BLUP shrinks outliers toward the mean
- Brix and yield are significantly negatively correlated – this has implications for selection.....

Acknowledgements

Ohio State

- Dr. David Francis
- Dr. Sung-Chur Sim
- Debora Liabeuf
- Troy Aldrich
- Nancy Huarachi Morejon

UC Davis

- Dr. Allen Van Deynze
- Dr. Shawn Yarnes

Campbell's Soup

- Dr. Nankui Tong

University of Florida

- Dr. Jay Scott
- Dr. Sam Hutton

USDA/ARS Geneva

- Dr. Joanne Labate

Cornell – Long Island

- Dr. Meg McGrath

NC State

- Dr. Dilip Panthee

Michigan State

- Dr. C. Robin Buell
- Dr. Candy Hansey

SGN

- Dr. Lukas Mueller
- Dr. Naama Menda



United States
Department of
Agriculture

National Institute
of Food
and Agriculture