HW2

Santos Barrera

4/1/2020

## Normalize the path:

#1. Version control

[**https://github.com/Sybalemus1/Agro932Hw1.git**](https://github.com/Sybalemus1/Agro932Hw1.git)

#2. Identify three traits of interest showing different levels of heterosis

Traits selected according Flint-Garcia et al., 2009 data PltYield: Plant yield (gr/plant) LFLEN: Leaf Length (cm) TSLLEN: Tassel length (cm)

#3.Following the basic steps to estimate genetic variances

First, the linear model is written down

pijr = μ + fi + Ej + eijr

pijr:phenotype value μ: averall mean fi: effect of the ith father Ej: is the effect of the jth environment eijr: is the residual error ~𝑁𝑜𝑟𝑚𝑎𝑙(0,𝜎\_𝑔^2)

Second, phenotypic variance is written in terms of components. 𝜎2P=𝜎f+𝜎2E+𝜎^2e 𝜎2p=Phenotypic Variance 𝜎^2f=Genotpic Variance 𝜎^2E=Entironmental Variance 𝜎^2e=error Variance

Third, interpret the ANOVA table and variance components in terms of the covariances between relatives I have a question: For half-siblings (hybrids in Flint-Garcia et al., 2009 data) we have to consider the inbreeding coefficient (F), which is 1/2 for half-siblings, for the calculation of VA (additive variance). However, in your lecture and Rex Bernardo Book, we calculate it from the mean sq (MS) of the interaction gxe. What is the procedure for finding mean sq (MS) of the interaction gxe in R?. In this case do we treat the environment as random effect in the model? Because I couln’t calculate this value I can’t reporte the additive variance and therefore I can’t report narrow sense heredability h^2

VA=4/1+FVprogeny

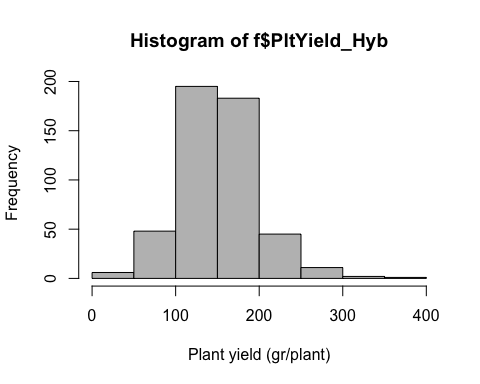
# Hybrids

Plant yield (gr/plant)

# library(lme4) when we account for randome effects   
library(Matrix)  
f <- read.delim("data/journal.pone.0007433.s001.txt", header=TRUE)  
# Convert missing data to NA  
f[f=="."] <- NA  
  
# four environments  
table(f$Env)

##   
## 02F 03M 03N 03W   
## 1 295 267 267 267

f$INBRED <- as.factor(f$INBRED)  
f$Env <- as.factor(f$Env)  
# tricky part, be careful:  
f$PltYield\_Hyb <- as.numeric(as.character((f$PltYield\_Hyb)))  
hist(f$PltYield\_Hyb, col="grey", xlab="Plant yield (gr/plant)")



fit\_PYH <- lm(PltYield\_Hyb ~ INBRED + Env, data=f)  
#plot (fit\_PYH)  
anova(fit\_PYH)

## Analysis of Variance Table  
##   
## Response: PltYield\_Hyb  
## Df Sum Sq Mean Sq F value Pr(>F)   
## INBRED 261 608627 2332 1.9277 2.441e-07 \*\*\*  
## Env 1 161885 161885 133.8275 < 2.2e-16 \*\*\*  
## Residuals 228 275801 1210   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

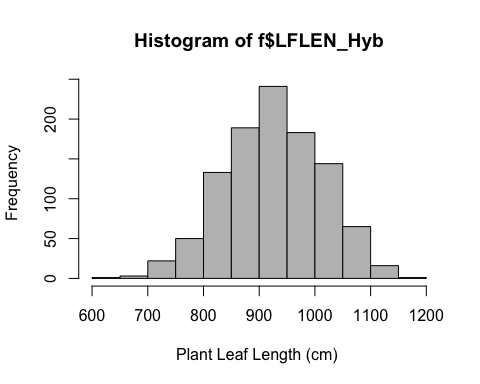
Vg= (2332 - 1210)/2 #only 2 locations for PltYield\_Hyb (North Carolina and Missouri)  
Ve = 1210  
  
H2 = Vg/(Vg + Ve)  
H2

## [1] 0.3167702

# Hybrids

Leaf Length (cm)

f$INBRED <- as.factor(f$INBRED)  
f$Env <- as.factor(f$Env)  
# tricky part, be careful:  
f$LFLEN\_Hyb <- as.numeric(as.character((f$LFLEN\_Hyb)))  
hist(f$LFLEN\_Hyb, col="grey", xlab="Plant Leaf Length (cm) ")



fit\_LFH <- lm(LFLEN\_Hyb~ INBRED + Env, data=f)  
#plot (fit\_LFH)  
anova(fit\_LFH)

## Analysis of Variance Table  
##   
## Response: LFLEN\_Hyb  
## Df Sum Sq Mean Sq F value Pr(>F)   
## INBRED 309 5057357 16367 11.449 < 2.2e-16 \*\*\*  
## Env 3 1797385 599128 419.107 < 2.2e-16 \*\*\*  
## Residuals 735 1050708 1430   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

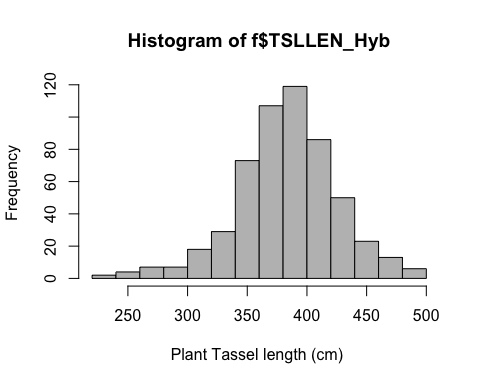
Vg= (16367 - 1430)/4  
Ve = 1430   
  
H2 = Vg/(Vg + Ve)  
H2

## [1] 0.7230963

# Hybrids

Tassel length (cm)

f$INBRED <- as.factor(f$INBRED)  
f$Env <- as.factor(f$Env)  
# tricky part, be careful:  
f$TSLLEN\_Hyb <- as.numeric(as.character((f$TSLLEN\_Hyb)))  
hist(f$TSLLEN\_Hyb, col="grey", xlab="Plant Tassel length (cm) ")



fit\_TSLH <- lm(TSLLEN\_Hyb~ INBRED + Env, data=f)  
#plot (fit\_TSLH)  
anova(fit\_TSLH)

## Analysis of Variance Table  
##   
## Response: TSLLEN\_Hyb  
## Df Sum Sq Mean Sq F value Pr(>F)   
## INBRED 309 675238 2185 5.3935 < 2.2e-16 \*\*\*  
## Env 1 204924 204924 505.7852 < 2.2e-16 \*\*\*  
## Residuals 233 94402 405   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Vg= (2185 - 405)/2 #only 2 locations for Tassel length (cm) (Florida and North Carolina)  
Ve = 405   
  
H2 = Vg/(Vg + Ve)  
H2

## [1] 0.6872587

### Flint-Garcia et al., 2009 data

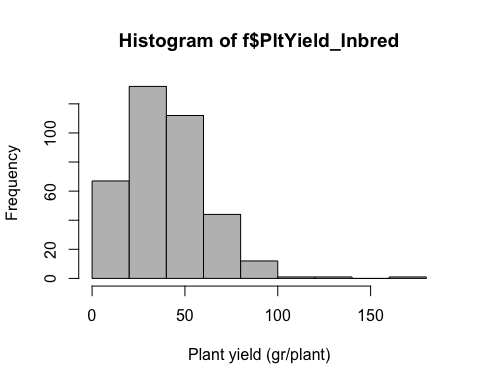
# Inbreed lines

# Plant yield (gr/plant)

f <- read.delim("data/journal.pone.0007433.s001.txt", header=TRUE)  
# Convert missing data to NA  
f[f=="."] <- NA  
  
# four environments  
table(f$Env)

##   
## 02F 03M 03N 03W   
## 1 295 267 267 267

f$INBRED <- as.factor(f$INBRED)  
f$Env <- as.factor(f$Env)  
# tricky part, be careful:  
f$PltYield\_Inbred <- as.numeric(as.character((f$PltYield\_Inbred)))  
hist(f$PltYield\_Inbred, col="grey", xlab="Plant yield (gr/plant)")



fit\_PYI <- lm(PltYield\_Inbred ~ INBRED + Env, data=f)  
#plot (fit\_PYI)  
anova(fit\_PYI)

## Analysis of Variance Table  
##   
## Response: PltYield\_Inbred  
## Df Sum Sq Mean Sq F value Pr(>F)   
## INBRED 227 124640 549.08 1.5295 0.003149 \*\*  
## Env 1 2203 2202.82 6.1362 0.014426 \*   
## Residuals 141 50617 358.99   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

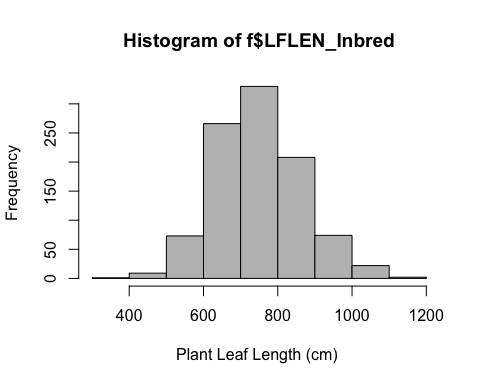
Vg= (549.08 - 358.99)/2 #only 2 locations for PltYield\_Hyb (North Carolina and Missouri)  
Ve = 358.99  
  
H2 = Vg/(Vg + Ve)  
H2

## [1] 0.2093341

# Inbreed lines

# Leaf Length (cm)

f$INBRED <- as.factor(f$INBRED)  
f$Env <- as.factor(f$Env)  
# tricky part, be careful:  
f$LFLEN\_Inbred <- as.numeric(as.character((f$LFLEN\_Inbred)))  
hist(f$LFLEN\_Inbred, col="grey", xlab="Plant Leaf Length (cm) ")



fit\_LFI <- lm(LFLEN\_Inbred~ INBRED + Env, data=f)  
#plot (fit\_LFI)  
anova(fit\_LFI)

## Analysis of Variance Table  
##   
## Response: LFLEN\_Inbred  
## Df Sum Sq Mean Sq F value Pr(>F)   
## INBRED 305 9546977 31302 13.859 < 2.2e-16 \*\*\*  
## Env 3 2106305 702102 310.853 < 2.2e-16 \*\*\*  
## Residuals 676 1526836 2259   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

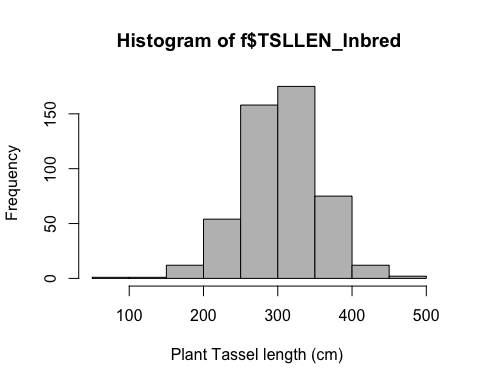
Vg= (31302 - 2259)/4  
Ve = 2259   
  
H2 = Vg/(Vg + Ve)  
H2

## [1] 0.7627039

# Inbreed lines

# Tassel length (cm)

f$INBRED <- as.factor(f$INBRED)  
f$Env <- as.factor(f$Env)  
# tricky part, be careful:  
f$TSLLEN\_Inbred <- as.numeric(as.character((f$TSLLEN\_Inbred)))  
hist(f$TSLLEN\_Inbred, col="grey", xlab="Plant Tassel length (cm) ")



fit\_TSLI <- lm(TSLLEN\_Inbred~ INBRED + Env, data=f)  
#plot (fit\_TSLI)  
anova(fit\_TSLI)

## Analysis of Variance Table  
##   
## Response: TSLLEN\_Inbred  
## Df Sum Sq Mean Sq F value Pr(>F)   
## INBRED 297 1115899 3757 5.8293 < 2.2e-16 \*\*\*  
## Env 1 99864 99864 154.9387 < 2.2e-16 \*\*\*  
## Residuals 191 123107 645   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Vg= (3757 - 645)/2 #only 2 locations for PltYield\_Hyb (Florida and North Carolina)  
Ve = 645   
  
H2 = Vg/(Vg + Ve)  
H2

## [1] 0.7069514

#4. Visualization

library(ggplot2)  
H2\_hyb <- c(0.32, 0.72, 0.68)  
H2\_Inb <- c(0.20, 0.76, 0.70)  
Trait <- c("Plant yield (gr/plant)", "Leaf Length (cm)", "Tassel length (cm) ")  
dat <- data.frame(Trait = Trait ,H2\_hyb = H2\_hyb, H2\_Inb = H2\_Inb)  
dat

## Trait H2\_hyb H2\_Inb  
## 1 Plant yield (gr/plant) 0.32 0.20  
## 2 Leaf Length (cm) 0.72 0.76  
## 3 Tassel length (cm) 0.68 0.70

ggplot(dat, aes(x= Trait, y=H2\_hyb, fill=Trait)) +  
 geom\_bar(stat="summary", fun.y="heritability")

## Warning: Ignoring unknown parameters: fun.y

## No summary function supplied, defaulting to `mean\_se()`



ggplot(dat, aes(x= Trait, y=H2\_Inb, fill=Trait)) +  
 geom\_bar(stat="summary", fun.y="heritability")

## Warning: Ignoring unknown parameters: fun.y

## No summary function supplied, defaulting to `mean\_se()`

