Compute genetic variances

Jinliang Yang

03-31-2022

Path Normalization

How to estimate genetic variances

Basic steps

1. Relative developed by some sort of mating design.

2. The progeny are evaluated in a set of environments.

3. Variance components are estimated from the mean squares in the analysis of variance.

4. The variance components are interpreted in terms of the covariances between relatives.

A commerical soybean breeding program

Season	Activity
Winter1a	(1) Grow 200 F2 populations (S0 generation)
	(2) Advance the S0 to the S1 generation by using single-seed descent method
Winter1b	(1) For each population, plant the S1 seeds
	(2) Save selfed seeds from 200-500 plants in each population
Summer1	(1) Evaluate 70,000 S2 families in unreplicated trials at 1-2 locations
	(2) Select the best 5,000 S2 families on the basis of yield trial data
	(3) Save selfed (i.e., S3) seeds of the best S2 families
Summer2	testing ...
Summer3	testing ...
Summer4	testing ...
Summer5	Yield trials of advanced lines at 20-50 locations
Fall	Release 0-5 lines as new cultivars

Bernardo, Table 1.1

A commerical soybean breeding program

- Suppose that 100 S2 families in **Summer1** are allowed to open pollinate.
- The open-pollinated seeds within each S2 familiy are bulked to form a half-sib family.
- An inbreeding coefficient of $F = 1/2$ among the parents of the half-sib families.

Next summer: - The 100 half-sib families (i.e., $n = 100$) are evaluated for their grain yield in a **randomized complete block design** with two replications (i.e., $r = 2$). - The experiment is grown in three environments (i.e., $e = 3$)

Plot the phentypic data

```
d <- read.csv("data/soybean_half-sib_yield.csv")
#install.packages("tidyr")
library(tidyr)
df <- gather(d, key="Env", value="yield", 2:4)
```

_

Using ggplot2

```
library(ggplot2)
ggplot(df, aes(x=yield, color=rep, fill=rep)) +
    geom_histogram(aes(y=..density..), position="identity", alpha=0.5)+
    geom_density(alpha=0.6)+
    facet_wrap(~ Env)+
    #scale_color_manual(values=c("#56B4E9", "#fe6f5e"))+
    labs(title="", y="Yield", x = "Density")+
    theme_classic() +
    guides(color=FALSE) +
    theme(plot.title = element_text(size=20, face = "bold"),
          axis.text=element_text(size=16, face="bold"),
          strip.text.y = element_text(size = 16, face = "bold"),
          axis.title=element_text(size=18, face="bold"),
          )
```

How to estimate genetic variances

Basic steps

1. Relative developed by some sort of mating design.

- Half-sib design with the S2 families

2. The progeny are evaluated in a set of environments.

- RCB design with 2 reps in 3 environments

3. Variance components are estimated from the mean squares in the analysis of variance.

4. The variance components are interpreted in terms of the covariances between relatives.

Half-sib design

$$p_{ijr} = \mu + f_i + l_j + f_i \times l_j + b_r + e_{ijr}$$

- where p_{ijr} is the phenotype value of the jth offspring of the ith father evaluated in the rth replication,
- f_i is the effect of the ith father,
- l_j is the effect of the jth environment (or location),
- $f_i \times l_j$ is the interaction effect of the ith father with the jth environment (or location),
- b_r is the effect of the rth replication,
- and e_{ijr} is the residual error. The e_{ijr} have expectation equal to zero.

Fit a linear model

```
fit <- lm(yield ~ fam + Env + fam:Env + rep, data=df)

summary(aov(fit))
```

The general framework

ANOVA table for one type of progeny (one-factor design)

Source	df	Observed MS	E(MS)
Environment	$e-1$		
Rep	$r-1$		
Progeny	$n-1$	$MS_{progeny}$	$V_e + rV_{PE} + reV_{progeny}$
Progeny x E	$(n-1)(e-1)$	MS_{PE}	$V_e + rV_{PE}$
pooled error	$(n-1)(r-1)e$	MS_{error}	V_e

$$V_{progeny} = \frac{MS_{progeny} - MS_{PE}}{re}$$

- MS_{error}: the mean squares for the pooled error
- MS_{PE}: mean squares for progeny × environment interaction
- $MS_{progeny}$: mean squares for progeny

The Soybean half-sib example

3. Variance components are estimated from the mean squares in the analysis of variance. ANOVA table for half-sib families.

Source	df	Observed MS	E(MS)
Environment	$e-1=2$		
Rep/E	$r-1=1$		
HS Families	$n-1=95$	$MS_{progeny}=255.84$	$V_e + rV_{PE} + reV_{progeny}$
HS F x E	$(n-1)(e-1)=190$	$MS_{PE}=7.77$	$V_e + rV_{PE}$
pooled error	$(n-1)(r-1)e=287$	$MS_{error}=6.67$	V_e

–

$$V_{progeny} = \frac{MS_{progeny} - MS_{PE}}{re}$$

```
out <- summary(aov(fit))[[1]]
vprogeny <- (out[1,3] - out[4,3])/(2*3)
```

The Soybean half-sib example

4. The variance components are interpreted in terms of the covariances between relatives.

- Half-sibs: $V_{progeny} = \frac{1+F}{4} V_A$
  - $V_A = \frac{4}{1+F} V_{progeny}$

–

Therefore,

$$V_A = \frac{4}{1+F} * 41.3 = 110$$

```
Va = 4/(1+1/2)*vprogeny
```

The variance of a progeny mean $V_{\bar{Y}}$

$V_{\bar{Y}}$ measures the sampling variation in the mean of a single group of individuals (e.g., half-sibs in this example).

- In this example, soybean yield is not measured for each plant in a plot.
- Rather, each plot is harvested by machine and the total yield from all the plants in each plot is recorded.
- The variance of the measurement for an individual plot is $V_e + rV_{PE}$, which is estimated by MS_{PE}.

Therefore,

$$V_{\bar{Y}} = \frac{MS_{PE}}{re} = \frac{V_e}{re} + \frac{V_{PE}}{e}$$

The variance of a progeny mean is equal to the variance of an individual plot divided by the number of observations for each group of progeny.

The Soybean half-sib example

ANOVA table for half-sib families.

Source	df	Observed MS	E(MS)
Environment	$e - 1 = 2$		
Rep/E	$r - 1 = 1$		
HS Families	$n - 1 = 95$	$MS_{progeny} = 255.84$	$V_e + rV_{PE} + reV_{progeny}$
HS F x E	$(n-1)(e-1) = 190$	$MS_{PE} = 7.77$	$V_e + rV_{PE}$
pooled error	$(n-1)(r-1)e = 287$	$MS_{error} = 6.67$	V_e

```
Vp = 7.77/(2*3)
```

Narrow-sense heritablity h^2

In this case, since individual-plant measurements are unavailable, the exact narrow-sense heritability (h^2) cannot be estimated.

But the h^2 on a progeny-mean basis can be estimated as

$$h^2_{HS} = \frac{V_{progeny}}{V_{progeny} + V_{\bar{Y}}}$$

```
h2 = vprogeny/(vprogeny + Vp)
```

Another example for inbred lines

Rodene, et. al., The Plant Phenome Journal, 2022.

- The inbred lines ($n = 230$) are assumed to be a random sample of genotypes from the population.
- Two environments (with or without N treatment, $e = 2$)
- Each with two replications ($r = 2$)

Another example for inbred lines

$$p_{ijk} = \mu + g_i + t_j + g_i \times t_j + r_k + e_{ijk}$$

- where p_{ijk} is the phenotype value of the ith genotype evaluated in the jth treatment with the kth rep,
- g_i is the effect of the ith genotype,
- t_j is the effect of the jth treatment (or environment),
- $g_i \times t_j$ is the interaction effect of the ith genotype with the jth treatment,
- r_k is the effect of the kth rep,
- and e_{ijk} is the residual error. The e_{ijk} have expectation equal to zero.

```
cc <- read.csv("data/ppj220030-sup-0002-tables1.csv")
table(cc$date)

### add replication information
cc$Rep <- "Rep2"
cc[cc$Row< 3000,] $Rep <- "Rep1"


j6 <- subset(cc, date %in% "July6")

fit <- lm(Canopy_Coverage ~ Genotype + Treatment + Genotype:Treatment + Rep,
          data=j6)
summary(aov(fit))
```

h^2 for Canopy Coverage

Source	df	Observed MS	E(MS)
Environment	$e - 1 = 1$		
Replications/E	$r - 1 = 1$		
Inbred lines	$n - 1 = 232$	$MS_{progeny} = 275$	$V_e + rV_{G\times E} + reV_{progeny}$
Inbreds x E	$(n-1)(e-1) = 224$	$MS_{PE} = 31$	$V_e + rV_{G\times E}$
pooled error	$(n-1)(r-1)e = 419$	$MS_{error} = 32$	V_e

- Inbred lines: $V_{progeny} = V_A$
  - $V_{progeny} = V_A = \frac{MS_{progeny} - MS_{PE}}{re} = \frac{275-31}{2\times 2} = 61$

The h^2 on a plot-mean basis can be estimated as

$$
\begin{aligned}
h^2 &= \frac{V_A}{V_A + V_{\bar{Y}}} \\
&= \frac{V_A}{V_A + V_{G\times E}/e + V_e/(re)} \\
&= \frac{61}{61 + 31/4} = 0.89
\end{aligned}
$$