Variance Partitioning, Estimating Genetic Variance, and Heritability

When beginning a breeding program, the breeder must first define the problems of importance. Next, he/she must determine whether there is genetic variation for the traits of highest priority within the available germplasm. Heritability is the proportion of phenotypic variation due to genetics. Measuring heritability therefore involves partitioning variance into genetic affects and environmental affects and is a key step in determining the best strategy for breeding.

The purpose of the next two class periods will be to review heritability (first introduced in Lynch and Walsh, Chapter 3) and introduce students to techniques for measuring heritability in the context of a breeding program. Data sets for estimation of heritability are provided.

Definitions (from various sources)

Heritability is a measure of the relative importance of heredity to the expression of a character.

Heritability in the broad sense is the ratio of genotypic variance to phenotypic variance.

Heritability in the narrow sense is the ratio of additive genetic variance to phenotypic variance.

Heritability is the regression of breeding value on phenotypic value.

Definitions consider different aspects of the heredity of a trait. Traits can be “(1) determined by genotype, or (2) transmitted from parent to offspring” (Nyquist, 1991).

Determined by Genotype

One definition of Heritability, Heritability in the broad sense (H), refers to heredity determined by genotype. Falconer’s term for broad sense heritability is “degree of genetic determination”.

Phenotype = Genotype + Environment. (equation from Lect1.)

The equation from Lect. I can also be written as:

\[ \text{VarPhenotype} = \text{VarGenotype} + \text{VarEnvironment} \]

\[ \text{VPhenotype} = \text{VGenotype} + \text{VEnvironment} \]

\[ s^2 P = s^2 G + s^2 E \]

Heritability in the broad sense (H) is:

\[ H = \frac{s^2 G}{s^2 P} \]

If we can estimate \( s^2 G \) and \( s^2 P \), we can estimate H. This can be done by partitioning the observed Phenotypic variation into Environment and Genotype components (we’ll come back to this).

Transmitted from parent to offspring:

Another definition of heritability stems from the meaning of heredity in the sense that a trait is passed from parent to offspring. This definition is more useful in answering the question “how well can we predict progeny performance based on the performance of the parent?”. This type of heritability is often called “narrow-sense heritability” (\( h^2 \)). Estimating \( h^2 \) requires that we further partition \( s^2 G \) into components (additive, dominance, epistatic).

\[ h^2 = \frac{s^2 G_{\text{Additive}}}{s^2 P} \]
Consider the following:

Aa → AA:Aa:aa (1:2:1)

AA and Aa have a trait value of 4 and aa has trait value of 0, the inheritance is dominant. With dominance a heterozygous individual that is selfed (or crossed to another Aa individual) will give rise to a population with a mean trait value of 3. In contrast, if each allele of A is worth 2 and inheritance is additive (AA has a value of 4, Aa of 2, and aa of zero), selfing an individual Aa with a trait value of 2 will give rise to a population with a mean value of 2. In both cases, the phenotype of the inbred line AA will be 4, but with additive inheritance the parents have better predictive value for the subsequent generation.

The symbol \( h^2 \) for heritability does not imply a square, but is used for historical reasons. Wright used ‘h’ as a regression coefficient between \( s^2 \) \( G_{\text{Additive}} \) and \( s^2 P \). He used ‘h’ in correlated response formulas, and as a path variable in tracing pedigrees. It is also a factor used to express genetic gain as a percent of a population mean. The symbol \( h^2 \) is therefore retained for narrow-sense heritability.

Appropriate questions which can be answered by estimation of heritability

1. Is there enough genetic variation within the germplasm to allow improvement in the traits of importance?
2. How extensively must the material be tested in order to identify superior parents?
3. Which populations are most promising as a source of improved breeding material?
4. What is the best breeding procedure to most efficiently lead to improved germplasm?
5. What type of variety is the most appropriate (hybrid, inbred, …)?
6. Will the same breeding procedure be effective for improving all traits?

Assumptions

1. Genetic components of variance estimated from mating designs can be equated to covariances among relatives if
   - Parents are a random selection of the genetic population
   - Experimental errors are independent
2. Translating covariances between relatives into additive, dominance, and epistasis requires a rigorous definition of the genetic population
   - Normal diploid inheritance
   - No environmental correlations among progenies (randomization)
   - Progenies are considered random
   - Non-inbred population
   - Linkage equilibrium

Limitations

1. BSH provides a relative magnitude of genetic and environmental variation in the germplasm pool. It is not an indication of progress that might be made within the population.
2. Heritability estimates apply only to the population and environments sampled.
3. Choosing populations depends on mean performance of the population & genetic variation within a population.
4. Estimation of additive and non-additive genetic variances requires the use of appropriate mating and environment (replication).
5. The simplest design which will provide information is preferred.
Figure 1. Schematic of a breeding program. The diagram, below, illustrates the structure of pedigree breeding program that progresses from F2 populations to inbred lines. Within the context of such a program there are several opportunities to estimate the genetic component of important traits. For example, the F2 populations (A) can be used to estimate genetic variance using Wright’s methods. These populations can also be used for genetic mapping studies. The regression of parent phenotype on offspring phenotype can provide an estimate of additive genetic variance (in the absence of selection) or realized genetic gain in the presence of selection (B and C. “Inferences concerning the genetic basis of quantitative traits can be extracted from phenotypic measures of the resemblance between relatives” [Lynch and Walsh, p. 50]). Finally, partitioning variance from multi-location and multi-year trials of varieties and lines can provide an estimate of the relative importance of genotype in the expression of a trait (D). Such information can be used to improve trait measurement and selection.

Experimental approaches to measuring broad sense heritability

Wright’s methodology estimates of Ve from pure lines (P1, P2, F1) and line crosses. Estimates of Vp = (Ve + Vg) from segregating populations. The foundation of this approach is based on a simple, single-locus model with two alleles (x1 and x2). The midparent is defined as m = (x1x1 + x2x2)/2 (i.e. the average of two parental classes for the inbred parents of an F2). If the heterozygote does not equal the midparent, some dominance exists.

-\[\begin{array}{c}
\text{x1x1} & \text{m} & \text{x1x2} & \text{x2x2} \\
\text{-a} & \text{+d} & \text{+a}
\end{array}\]

Segregation in the F2 is expected to be \(\frac{1}{4}\) x1x1 + \(\frac{1}{2}\) x1x2 + \(\frac{1}{4}\) x2x2

Substituting the values from the figure:
\( \frac{1}{4} (-a) + \frac{1}{2} d + \frac{1}{4} a = 1/2d \) (the average phenotypic value of the F2)

The contribution that each individual makes to the variance is the square of the deviation from the mean (Note: variance is the square of the standard deviation (e.g. \( \sigma^2 \)). So, for the whole population we multiply by the expected frequency to obtain:

\[
\frac{1}{4} (-a-\frac{1}{2}d)^2 + \frac{1}{2} (d-\frac{1}{2}d)^2 + \frac{1}{4} (a-\frac{1}{2}d)^2 = \frac{1}{2}a^2 + \frac{1}{4}d^2
\]

Note: the assumption of Hardy-Weinberg equilibrium.

English quantitative geneticists define \( a^2 \) as A and \( d^2 \) as D. Environmental variance is defined as E. So the total phenotypic variance is \( VF2 = \frac{1}{2} A + \frac{1}{4} D + E \).

Note that additive genetic variance \( V_A = \frac{1}{2} A \) and dominant genetic variance \( V_d = \frac{1}{4} D \) (the \( Va \) or \( \sigma^2a \) notation is used more commonly by American quantitative geneticists)

We have now seen how the total genetic variance can be estimated in two different ways. Remember as noted earlier that measuring BSH provides a relative magnitude of genetic and environmental variation in the germplasm pool. It is not an indication of progress that might be made within the population. The Narrow sense heritability, or additive genetic variance is a much better predictor of breeding value. (Question for class: in which cases is broad sense heritability an accurate predictor of performance?)

**Measures of Broad Sense Heritability from multi-location and multi-year trials.**

Broad sense heritability can be estimated through replication of a population in time and space

Experiments designed to estimate variance components to be used in broad sense heritability estimates must be grown in an adequate sample of environments

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Expected MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypes</td>
<td>n-1</td>
<td>( \sigma^2 + \rho \sigma^2(GYL) + \rho \sigma^2(GL) + \rho \sigma^2(GY) + \rho \sigma^2(G) )</td>
</tr>
<tr>
<td>Gen X year</td>
<td>(n-1)(y-1)</td>
<td>( \sigma^2 + \rho \sigma^2(GYL) + \rho \sigma^2(GY) )</td>
</tr>
<tr>
<td>Gen X Loc</td>
<td>(n-1)(p-1)</td>
<td>( \sigma^2 + \rho \sigma^2(GYL) + \rho \sigma^2(GL) )</td>
</tr>
<tr>
<td>GenX LocX year</td>
<td>(n-1)(p-1)(y-1)</td>
<td>( \sigma^2 + \rho \sigma^2(GL) )</td>
</tr>
<tr>
<td>Error</td>
<td>(r-1)(n-1)p</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

Main effects of Location, year, and rep are excluded because it assumed that selection will occur on means across location and years. Cotterill, P. P. (1987). On estimating heritability according to practical applications. Silvae Genetica 36:46-48.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Expected MS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotypes</td>
<td>n-1</td>
<td>( \sigma^2 + \rho \sigma^2(GYL) + \rho \sigma^2(GL) + \rho \sigma^2(GY) + \sigma^2(G) )</td>
</tr>
<tr>
<td>Error</td>
<td>(r-1)(n-1)</td>
<td>( \sigma^2 )</td>
</tr>
</tbody>
</table>

Appropriate estimate of heritability for two replicates:

\[
\frac{\sigma^2(G)}{\sigma^2(G) + \sigma^2(2GY) + \sigma^2(GL) + \sigma^2(GY)}
\]

Estimate which can be obtained from one year:

\[
\frac{\sigma^2(G) + \sigma^2(GYL) + \sigma^2(GL) + \sigma^2(GY)}{\sigma^2(G) + \sigma^2(2GY) + \sigma^2(GL) + \sigma^2(GY)}
\]

So, even though selection may be based on one year at one location it is necessary to measure heritability in >1 year and >1 location if the G x Y, G x L, and G x Y x L interactions are of importance.
*SAS code (2yearlyc.txt) */
data cdata;
    infile 'a:2yrlyc.csv' delimiter = ',' firstobs = 2;
    input gen rep year hplc uv;
proc glm;
class year rep gen;
model hplc uv = year rep(year) gen year*gen/ ss3;
means year rep(year) gen / lsd lines;
title 'ANOVA for lycopene data.';
proc varcomp;
class year rep gen;
model hplc uv = year rep(year) gen year*gen;
title 'Variance Components for lycopene data,';
run;

Notes on the SAS code

The program above uses a new procedure (to us), Proc varcomp. Proc varcomp is one of two methods we can use to estimate variance components (the other, Proc Mixed will be introduced later). You can specify effects as fixed by putting them first in the MODEL statement and indicating the number of fixed effects with the FIXED= option. Variance components are estimated for RANDOM effects. There are four methods of estimation that can be specified in the PROC VARCOMP statement by using the METHOD = option.

- **TYPE1** based on computation of the type 1 sums of squares
- **MIVQUE0** The default. Similar to type 1, but computationally faster.
- **ML** Maximum likelihood
- **REML** Restricted maximum likelihood (favored in breeding work)

Output = 2yearlyc.doc
Dependent Variable: UV

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>15</td>
<td>98.72185357</td>
<td>6.58145690</td>
<td>2.04</td>
<td>0.1094</td>
</tr>
<tr>
<td>Error</td>
<td>12</td>
<td>38.68361429</td>
<td>3.22363452</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>27</td>
<td>137.40546786</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square          | C.V. | Root MSE   | UV Mean  |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.718471</td>
<td>15.96816</td>
<td>1.7954482</td>
<td>11.24392857</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Type III SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td>1</td>
<td>1.95043214</td>
<td>1.95043214</td>
<td>0.61</td>
<td>0.4517</td>
</tr>
<tr>
<td>REP(YEAR)</td>
<td>2</td>
<td>11.54743571</td>
<td>5.77371786</td>
<td>1.79</td>
<td>0.2086</td>
</tr>
<tr>
<td>GEN</td>
<td>6</td>
<td>62.16554286</td>
<td>10.36092381</td>
<td>3.21</td>
<td>0.0404</td>
</tr>
<tr>
<td>YEAR*GEN</td>
<td>6</td>
<td>23.05844286</td>
<td>3.84307381</td>
<td>1.19</td>
<td>0.3732</td>
</tr>
</tbody>
</table>

General Linear Models Procedure
T tests (LSD) for variable: HPLC
NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.
Alpha= 0.05  df= 12  MSE= 3.120588
Critical Value of T= 2.18
Least Significant Difference= 2.7216
Means with the same letter are not significantly different.
T Grouping | Mean | N  | YEAR |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>9.7707</td>
<td>14</td>
<td>98</td>
</tr>
<tr>
<td>A</td>
<td>9.7236</td>
<td>14</td>
<td>97</td>
</tr>
</tbody>
</table>

T tests (LSD) for variable: UV
NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.
Alpha= 0.05  df= 12  MSE= 3.223635
Critical Value of T= 2.18
Least Significant Difference= 1.4786
Means with the same letter are not significantly different.
T Grouping | Mean | N  | YEAR |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>11.5079</td>
<td>14</td>
<td>97</td>
</tr>
<tr>
<td>A</td>
<td>10.9800</td>
<td>14</td>
<td>98</td>
</tr>
</tbody>
</table>

Level of Level of  -------------HPLC-------------  -------------UV-------------
REP   YEAR  N  Mean   SD    Mean   SD
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>97</td>
<td>7</td>
<td>8.9971429</td>
<td>2.32359430</td>
<td>11.0285714</td>
<td>2.04596700</td>
</tr>
<tr>
<td>2</td>
<td>97</td>
<td>7</td>
<td>10.4500000</td>
<td>2.74847837</td>
<td>11.9871429</td>
<td>1.72652555</td>
</tr>
<tr>
<td>1</td>
<td>98</td>
<td>7</td>
<td>9.1985714</td>
<td>2.85019548</td>
<td>10.2085714</td>
<td>3.12000229</td>
</tr>
<tr>
<td>2</td>
<td>98</td>
<td>7</td>
<td>10.3428571</td>
<td>1.20769359</td>
<td>11.7514286</td>
<td>1.93648676</td>
</tr>
</tbody>
</table>

General Linear Models Procedure
T tests (LSD) for variable: HPLC
NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.
Alpha= 0.05  df= 12  MSE= 3.120588
Critical Value of T= 2.18
Least Significant Difference= 2.7216
Means with the same letter are not significantly different.
T Grouping | Mean | N  | GEN |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>12.393</td>
<td>4</td>
<td>9242</td>
</tr>
<tr>
<td>A</td>
<td>11.898</td>
<td>4</td>
<td>9241</td>
</tr>
<tr>
<td>B</td>
<td>10.460</td>
<td>4</td>
<td>7983</td>
</tr>
<tr>
<td>B</td>
<td>9.030</td>
<td>4</td>
<td>9423</td>
</tr>
<tr>
<td>B</td>
<td>8.360</td>
<td>4</td>
<td>696</td>
</tr>
<tr>
<td>B</td>
<td>8.185</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>B</td>
<td>7.905</td>
<td>4</td>
<td>8245</td>
</tr>
</tbody>
</table>
T tests (LSD) for variable: UV  
NOTE: This test controls the type I comparisonwise error rate not the experimentwise error rate.  
Alpha= 0.05  df= 12  MSE= 3.223635  
Critical Value of T= 2.18  
Least Significant Difference= 2.7662  
Means with the same letter are not significantly different.  

<table>
<thead>
<tr>
<th>T Grouping</th>
<th>Mean</th>
<th>N</th>
<th>GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>13.378</td>
<td>4</td>
<td>9241</td>
</tr>
<tr>
<td>A</td>
<td>13.033</td>
<td>4</td>
<td>9242</td>
</tr>
<tr>
<td>B A</td>
<td>11.665</td>
<td>4</td>
<td>9423</td>
</tr>
<tr>
<td>B A</td>
<td>11.595</td>
<td>4</td>
<td>7983</td>
</tr>
<tr>
<td>B</td>
<td>9.925</td>
<td>4</td>
<td>696</td>
</tr>
<tr>
<td>B</td>
<td>9.720</td>
<td>4</td>
<td>38</td>
</tr>
<tr>
<td>B</td>
<td>9.393</td>
<td>4</td>
<td>8245</td>
</tr>
</tbody>
</table>

Variance Components for lycopene data  

Variance Components Estimation Procedure  
Class Level Information  

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td>2</td>
<td>97 98</td>
</tr>
<tr>
<td>REP</td>
<td>2</td>
<td>1 2</td>
</tr>
<tr>
<td>GEN</td>
<td>7</td>
<td>38 696 7983 8245 9241 9242 9423</td>
</tr>
</tbody>
</table>

Number of observations in data set = 28  

MIVQUE(0) Variance Component Estimation Procedure  

SSQ Matrix  

<table>
<thead>
<tr>
<th>Source</th>
<th>YEAR</th>
<th>REP(YEAR)</th>
<th>GEN</th>
<th>YEAR*GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td>196.000000000</td>
<td>98.000000000</td>
<td>0.000000000</td>
<td>28.000000000</td>
</tr>
<tr>
<td>REP(YEAR)</td>
<td>98.000000000</td>
<td>147.000000000</td>
<td>0.000000000</td>
<td>14.000000000</td>
</tr>
<tr>
<td>GEN</td>
<td>0.000000000</td>
<td>0.000000000</td>
<td>96.000000000</td>
<td>48.000000000</td>
</tr>
<tr>
<td>YEAR*GEN</td>
<td>28.000000000</td>
<td>14.000000000</td>
<td>48.000000000</td>
<td>52.000000000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>YEAR</th>
<th>REP(YEAR)</th>
<th>GEN</th>
<th>YEAR*GEN</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td>14.000000000</td>
<td>0.217800000</td>
<td>27.306050000</td>
<td></td>
</tr>
<tr>
<td>REP(YEAR)</td>
<td>21.000000000</td>
<td>83.903400000</td>
<td>94.480750000</td>
<td></td>
</tr>
<tr>
<td>GEN</td>
<td>24.000000000</td>
<td>326.43768571</td>
<td>248.66217143</td>
<td></td>
</tr>
<tr>
<td>YEAR*GEN</td>
<td>26.000000000</td>
<td>195.56134286</td>
<td>174.34883571</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>27.000000000</td>
<td>147.19837143</td>
<td>137.40546786</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Year</th>
<th>HPLC</th>
<th>UV</th>
</tr>
</thead>
<tbody>
<tr>
<td>YEAR</td>
<td>0.217800000</td>
<td>27.306050000</td>
<td></td>
</tr>
<tr>
<td>REP(YEAR)</td>
<td>83.903400000</td>
<td>94.480750000</td>
<td></td>
</tr>
<tr>
<td>GEN</td>
<td>326.43768571</td>
<td>248.66217143</td>
<td></td>
</tr>
<tr>
<td>YEAR*GEN</td>
<td>195.56134286</td>
<td>174.34883571</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>147.19837143</td>
<td>137.40546786</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variance Component</th>
<th>HPLC Estimate</th>
<th>UV Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Var(YEAR)</td>
<td>-0.39584226</td>
<td>-0.31733750</td>
</tr>
<tr>
<td>Var(REP(YEAR))</td>
<td>0.40924762</td>
<td>0.36429762</td>
</tr>
<tr>
<td>Var(GEN)</td>
<td>2.72723869</td>
<td>1.62946250</td>
</tr>
<tr>
<td>Var(YEAR*GEN)</td>
<td>-0.21398631</td>
<td>0.30971964</td>
</tr>
<tr>
<td>Var(Error)</td>
<td>3.12058810</td>
<td>3.22363452</td>
</tr>
</tbody>
</table>

How would we use these to estimate heritability?
Estimates of heritability from Parent-offspring regression.

Parent-offspring regression:

Another way is to estimate genetic components of variance from mating designs by using covariances among relatives.

A brief review of regression analysis (for a straight line) follows:

\[ Y' = a + bX \]

where \( Y \) and \( X \) are variables and \( Y' \) is an estimate of \( Y \) based on a value of \( X \).

The regression coefficient \( b \) is an expression of how much (on average) \( Y \) may be expected to change per unit change of \( X \).

\[ b = \frac{\text{Sum}(X_i - X_{ave})(Y_i - Y_{ave})}{\text{Sum}(X_i - X_{ave})^2} = \frac{\text{change in } X}{\text{change in } Y} \]

\( a \) = the \( Y \) intercept

The statistical correlation coefficient “\( r \)” measures how closely two sets of data are associated. It is without units and as the limits of –1.0 to +1.0. the regression coefficient, \( b \), and the correlation coefficient, \( r \), always have the same sign. The correlation coefficient of \( Y \) on \( X \) is defined as the linear change of \( Y \) in standard deviations, for each increase of one standard deviation in \( X \).

\[ \text{Cov}(X,Y) = \frac{\text{Sum}(X_i - X_{ave})(Y_i - Y_{ave})}{n - 1} \]

\[ r = \frac{\text{Cov}(X,Y)}{\text{Sdev}_x \times \text{Sdev}_y} \]

For PO regression

\[ h^2 = r = \frac{\text{Cov}(P,O)}{\left( \text{Var } P \times \text{Var } O \right)^{1/2}} \]

Note that \( r \) and \( b \) are related by:

\[ b = r(\text{Sdev}_x / \text{Sdev}_y) \]

When the variances of \( X \) and \( Y \) are equal, \( b = r \) and either can be used to estimate heritability. If the variances are unequal, standardized variables (sample has a mean of zero and a standard deviation of 1) can be used to insure that \( b = r \).

Corrections for inbreeding:

Parent-offspring regression is used to estimate narrow sense heritability, but as we see above the numerator often contains some contribution from variation due to dominance. In practice, regression coefficients are corrected based on the relationship between relatives used for the analysis. We will discuss this further on Monday.
SAS code for Parent offspring regression

```sas
data plants;
  infile 'a:f2f3lyc.csv' delimiter = ',' firstobs = 2;
  input plot ped gen lyc;
  proc sort data=plants;
    by ped;
  proc means noprint data=plants;
    where gen=2;
    by ped;
    var lyc;
    output out=f2means mean = m2lyc;
  proc means noprint data=plants;
    where gen=3;
    by ped;
    var lyc;
    output out=f3means mean = m3lyc;
  data both;
    merge f2means f3means;
    by ped;
  proc reg;
    model m2lyc = m3lyc;
    title 'parent offspring regression for lycopene content (proc reg)';
  proc plot;
    plot m2lyc*m3lyc = 'X' / overlay vpos=40 hpos=60;
  proc corr outp=b(type=cov);
    var m2lyc m3lyc;
    title 'parent offspring regression for lycopene content (proc corr)';
  run;

Note: the data merge statement creates a variable named "_type_" which will lead to a warning statement because proc reg has an option to use datasets of TYPE=CORR, TYPE=COV, and TYPE=SSCP as inputs. The warning is to let you know that there could be confusion if you were using these type statements. The warning does on affect the output. The error message can be avoided by adding ‘drop “_type_”’ at the end of the “data both” statement.
```
Sources of lecture:


Figure 8-1 from R.W. Allard 1960. Principles of Plant Breeding John Wiley and Sons, Inc.

Assignments

Papers for next class as an example of parent-offspring regression.


To focus on in the papers:

1) variance components for selfing generations derived from two inbred parents
2) correction for selfing species (there is a difference of opinion in the literature on this subject).
3) The interpretation of the parent-offspring regression
4) Differences between heritability in the broad sense and narrow sense

Analytical exercise

Exercises. Use the data sets provided to estimate heritability of lycopene content. Be prepared to discuss the relative advantages and disadvantages of different approaches and each data set.

Three data sets are provided. These files can be opened in EXCEL for viewing. Descriptions follow:

“2yearlyc.xls” contains replicated data for lycopene content from 7 genotypes. Each genotype is replicated twice in each of two years.

“f2f3lyc.xls” contains F2 data (8823 plant numbers) and F3 data (1300 plot numbers). Progenies are derived from the cross between 3259 and 8245. The F1 hybrid of these two inbred lines is 120. Data for the parents and their hybrid may be found in the “1yearlyc.xls” data set.

Describe three different methods for estimating the heritability of lycopene content based on the data sets. Describe the type of heritability you are estimating with each. Use any or all of these data sets to estimate heritability for lycopene content using one method.