Analysis of Diallel Mating Designs

Fikret Isik
North Carolina State University, Raleigh, USA

6.1 Introduction

6.1.1 Diallel mating designs

When the same parents are used as females and males in breeding, the mating design is called diallel. Here are some commonly used diallel mating designs in forestry:

Half diallel - Each parent is mated with every other parent, excluding selfs and reciprocals

F/M	1	2	3	4	5	6
1	.	x	x	x	x	x
2		.	x	x	x	x
3			.	x	x	x
4				.	x	x
5					.	x
6						.

Smart diallel - Parents are sorted for their breeding values from the best to the poorest and most crosses are made among the best.

F/M	1	2	3	4	5	6
1	.	x	.	x	.	x
2		.	x	.	x	.
3			.	x	.	.
4				.	.	.
5					.	.
6						.

Disconnected half diallel - The half-diallel mating is repeated for the second diallel group. Sometimes crosses are made between parents from two diallels to have connection between two groups.

		Diallel 1										
F/M	**1**	**2**	**3**	**4**	**5**	**|**	**6**	**7**	**8**	**9**	**10**	
1	**.**	**x**	**x**	**x**	**x**	**|**	**.**	**.**	**x**	**.**	**.**	
2		**.**	**x**	**x**	**x**	**|**	**.**	**.**	**.**	**x**	**.**	
3			**.**	**x**	**x**	**|**	**.**	**x**	**.**	**.**	**.**	
4				**.**	**x**	**|**	**.**	**.**	**.**	**.**	**.**	
5					**.**	**|**						
6						**|**	**.**	**x**	**x**	**x**	**x**	
7						**|**		**.**	**x**	**x**	**x**	
8						**|**			**.**	**x**	**x**	**Diallel 2**
9						**|**				**.**	**x**	
10						**|**					**.**	

There are many other combinations of diallel mating designs. See White et al. (2007??) to see details.

- **Advantages and drawbacks of diallel mating designs**
  - Diallel designs provide good evaluation of parents and full-sib families,
  - Provide estimates of both additive and dominance genetic effects,
  - Provide estimates of genetic gains from both additive and non-additive genetic variance,
  - When the number of parents mated increases, the number of crosses increases by 2N, where N is the number of parents and the design can be costly
  - Using the same parents as males and females make the mating design a little bit complicated to analyze

6.2 Example for Randomized Complete Blocks with Single-Tree Plots

- **Genetic materials**: 18 loblolly pine trees were mated to produce 40 full-sib families (crosses) for progeny testing.
- **Field design**: A randomized complete blocks design was used with single tree plot. One progeny of each cross was randomly assigned in a block. There were 25 blocks in one site. Thus, each cross had 25 progeny at one site. The experiment was replicated at six sites but for simplicity, we will be initially giving an example for one site.
- **The statistical Model**: The following linear mixed model was fitted to data to estimate variance components.

[1] $$Y_{ijkl} = \mu + B_i + G_j + G_k + S_{jk} + E_{ijkl}$$

where

Y_{ijkl} is the l-th observation of the i-th block for the *jk*-th cross;

μ is the overall mean;

B_i is the fixed effect of the i-th block, i=1 to b;

G_j or G_k is the random general combining ability (GCA) effect of the j-th female or the k-th male ~ Normally and Independently Distributed (NID) (0, σ^2_G), j, k=1 to p and $j<k$;

S_{jk} is the random specific combining ability (SCA) effect of the j-th and the k-th parents (j≠k) ~NID (0, σ^2_S);

E_{ijkl} is the random within plot error term ~NID (0, σ^2_E)

General combining (parents) effects, specific combining ability (crosses) effects, and the error term are considered random. The random effects are associated with zero mean and variance. The block effect is considered fixed. See Chapter 4 and 12 for discussions of random and fixed effects. We can write above linear model in a matrix form, which is shorter.

[2] $\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}$

where,

y is the vector of individual observations,

β is the vector of fixed-effects parameters (overall mean, and blocks),

γ is the vector of random-effects parameters including general combining ability (GCA) for female and male, and specific combining ability (SCA),

ε is an unknown random error vector

X is the known design matrix for the fixed effects

Z is the known design matrix for random effects

- The major assumption of the linear mixed model is that the random effects **γ** and error term **ε** are assumed to have normal distributions with 0 mean and variances.

$$E\begin{pmatrix}\gamma\\ \varepsilon\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix} \quad \text{Var}\begin{pmatrix}\gamma\\ \varepsilon\end{pmatrix}=\begin{pmatrix}\mathbf{G} & 0\\ 0 & \mathbf{R}\end{pmatrix}$$

- The second major assumption is that residuals have a normal distribution and they are independent of each other.

6.3 Implementation with SAS Mixed Procedure

ESO data set has 7 variables or columns (block, female, male, cross, tree, height) and 757 observations (rows). The first 10 observations of data are given below. Each tree has a unique number. Height of trees was measured at age six in meters.

block	female	male	cross	Tree	Height
1	2	1	2x1	2502	9.0

block	female	male	cross	Tree	Height
3	2	1	2x1	2554	11.0
4	2	1	2x1	2582	9.0
5	2	1	2x1	2612	8.7
6	2	1	2x1	2639	10.0
7	2	1	2x1	2670	9.6
8	2	1	2x1	2699	10.0
9	2	1	2x1	2729	10.2
10	2	1	2x1	2763	10.4
..	..	..	..	..	..

In diallel mating designs, the same parents are used as females and males in producing crosses (families). There are no specific SAS procedures or options to overlay design matrices of parents so we can obtain one GCA variance. Most of the SAS programming presented here is about creating the **Z** design matrix, which connects observations of individual trees to parents.

Code 1a: Creating Z design matrix for random effects (parents)

```
%LET dlset=hbook.eso ;


/* sort data and create a list of parents: PLIST */
PROC SORT DATA=&dlset;
  BY female male;    *****very important!!!;

*Create a list of female and male;
PROC SUMMARY DATA=&dlset NOPRINT;
  CLASS female male ;
  OUTPUT OUT=plist(where=(_type_=3));

TITLE 'List of females, males and number of trees per cross' ;
PROC PRINT DATA=plist noobs;
var female male _FREQ_;
RUN;


DATA parent;
  SET plist(rename=(female=parent))
      plist(rename=(male=parent));

PROC SUMMARY DATA=parent(keep=parent);
```

```
  CLASS parent;
  OUTPUT OUT=parent(where=(_type_=1));

DATA parent(drop=_type_ _freq_ pn);
  SET parent;
  pn+1;
  CALL SYMPUT('pn',compress(pn)); *get total number of
parents;

TITLE 'List of parents' ;
PROC PRINT DATA=parent;
RUN;
```

The above code is to create a list of parents. Explanations of some of the code are given below.

1. %LET: Is a macro variable. It helps to reduce typing longer names. For example, instead of typing the full name of the data set (hbook.diall), we can simple define the name as %LET **dset**=**hbook.eso** and use **&dlset** for the rest of the code.
2. CALL **SYMPUT**: creates a macro variable called **&np** (number of parents). We need this number (18) later to create the design matrix **Z** of random effects (i.e., parents)
3. Which part of the code to change? All you need to do is change the name of the data file yellow highlighted above. If you have different names for female and male in the data set, you should also change these names in above code. Otherwise the code gives error message.

Output 1a: The list of females, males, number of trees per cross (female x male) and the list of parents

List of females, males and number of trees per cross

female	male	_FREQ_
2	1	21
3	5	18
3	6	14
3	7	16
3	10	17
4	7	20
4	16	20
5	4	19
5	7	17

5	16	17
6	4	17
..	..	..
18	2	24

A partial printout of the data file ‘plist’ is given. The females and their crosses with males is given in the first two columns. The last column _FREQ_ is the number of trees or observations for each cross in the data. If there is one observation for a cross, it is likely that female or male id is a typo error.

List of parents

Parent
1
2
3
4
5
..
18

The females and males used in the diallel mating design were combined in one list, called ‘parents’.

Code 1b: Creating Z design matrix for random effects (continued…)

```
/* construct dummy variables p1-p19d */
PROC IML;
  USE &dlset;
  READ ALL VAR {female male} INTO d;
  CLOSE &dlset;
  n=NROW(d);

  *create a matrix (pn x 4) with parent, parent code (1-pn);
  USE parent;
  READ all var {parent} into p;
  CLOSE parent;
  pcode=CHAR(1:NROW(p),5,0)`; * 5 is the length;
  *** create pcode corresponding parent coding in dummy;
  p=p||pcode;
  PRINT n p ; *<--Check # observations and # of parents(pn);

  CREATE pcode FROM pcode [COLNAME={'p'}];
```

```
  APPEND FROM pcode ;
  CLOSE pcode;

  *create dummy variables;
  a=SHAPE(0,n,NROW(p));
  DO i=1 to n;
    DO k=1 to nrow(p);
      IF d[i,1]=p[k,1] | d[i,2]=p[k,1] then a[i,k]=1;
    END;
  END;
CREATE dummy from a;
APPEND FROM a;
CLOSE dummy;
QUIT;
```

Explanation of the code

1. You DO NOT need to change above code. The above code is to create a design matrix (Z matrix) for parents. The dimension of the Z matrix is **757 rows x 18 columns**. Number of columns in the matrix is 18; one column for each parent. Number of rows in the matrix would be the total number of observations (r = 757). The elements of the Z design matrix are either 1 or 0.

```
* Merge dummy variables with original data;
DATA &dlset;
 MERGE &dlset dummy;
PROC SORT DATA=&dlset; BY block cross;
RUN;

TITLE 'Data with dummy variables' ;
PROC PRINT DATA=&dlset (OBS=10) NOOBS;
VAR female male HEIGHT col1-col5;
RUN;
```

2. The above code is to merge dummy variables (the $\mathbf{Z}_{757 \times 18}$) with the original data set (**&dlset**). The female, male, Height, and the first 5 dummy columns are printed below.

Output 1: Z matrix overlays observations from the same parent (continued…)

Data with dummy variables

female	male	HEIGHT	COL1	COL2	COL3	COL4	COL5
P02	P01	29.5	1	1	0	0	0
P04	P07	26.5	0	0	1	0	0
P04	P08	32.5	0	0	1	0	0
P04	P11	33.5	0	0	1	0	0
P05	P08	30.5	0	0	0	1	0
P05	P17	34.0	0	0	0	1	0
P06	P17	27.5	0	0	0	0	1
P07	P05	28.5	0	0	0	1	0
P07	P06	32.0	0	0	0	0	1
P07	P11	29.5	0	0	0	0	0

What we have done so far is to create the **Z** design matrix for random effects GCA in the model and added this matrix to the original data set &dlset. Now we are ready to run the mixed model.

Code 1c: Running the linear mixed model

```
/* Run Proc Mixed on variable Height */

PROC MIXED DATA=&dlset COVTEST ASYCOV UPDATE;
CLASS block cross ;
  MODEL Height = block;
  RANDOM col1-col&pn/TYPE=TOEP(1);  * GCA effects ;
  RANDOM cross;                     * SCA effects ;
ODS OUTPUT COVPARMS=_varcomp ASYCOV=_cov;
RUN;
```

1. ASYCOV: The option produces the variances of variance components (diagonal elements) and the covariances (off diagonal elements) between them.
2. COVTEST: produces asymptotic standard errors and Wald Z-tests for the covariance parameter estimates (variance components).
3. CLASS statement: We list the factors (independent variables) after the CLASS statement. **block**, and **cross** are class (CLASS) variables in the model.
4. MODEL statement: The response variable Height is given after the MODEL statement. The **block** is a fixed effect and listed after the model statement. There is no need to list the intercept. The intercept (μ) is included in the model by default.
5. RANDOM **col1-col&pn** statement is the GCA effect. **col1-col&pn** is the design matrix or dummy variables we created in previous code (see IML code). This matrix of 0 and 1

relates individual Height values to the parents. Here, we are constructing our own columns of Z with continuous variables.

a. We are using the TYPE=TOEP(1) covariance structure to group parents together to have a common variance component. In another words, the option TYPE=TOEP(1) estimates a single variance component across all levels of parents. See SAS Mixed procedure manual for details.

6. RANDOM **cross** is the SCA effect. We also desire to have different covariance structures in different parts of **G**, thus we must use multiple RANDOM statements with different TYPE= options. For example, for the RANDOM **cross,** the default covariance structure (TYPE=VC) is preferred. The TYPE=VC (variance components) option models a different variance component for each random effect.
7. ODS OUTPUT: This is to create output tables (SAS data sets). Here we are creating the variance components (COVPARMS) and the covariances of variance components (ASYCOV). The name of the new tables would be **_covparms** and **_cov**.

Output 1: Mixed procedure output

Most of the output from the MIXED procedure is similar to the output in Chapter 4 (Code 1). Here, only a few important tables are interpreted.

```
                Model Information

Data Set                      HBOOK.ESO
Dependent Variable            HEIGHT
Covariance Structures         Banded Toeplitz, Variance Components
Estimation Method             REML
Residual Variance Method      Profile
Fixed Effects SE Method       Model-Based
Degrees of Freedom Method     Containment
```

1. Te **Model Information** is about statistical methods used to analyze data. The name of the data set and the dependent variable (Height) are listed. The method used for calculation of variance component is *Banded Toeplitz* because we used the TYPE=TOPE(1) method after the RANDOM statement to obtain one variance component for females and males.

```
            Class Level Information

Class     Levels    Values
```

block	25	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25
cross	40	P02xP01 P04xP06 P04xP07 P04xP08 P04xP11 P05xP08 P05xP17 P06xP05 P06xP08 P06xP17 P07xP05 P07xP06 P07xP11 P07xP17 P08xP17 P09xP02 P10xP02 P11xP05 P11xP06 P11xP08 P11xP17 P12xP02 P13xP02 P14xP02 P15xP04 P15xP05 P15xP06 P15xP07 P15xP08 P15xP11 P16xP02 P18xP04 P18xP05 P18xP06 P18xP07 P18xP08 P18xP11 P18xP15 P18xP17 P19xP02

2. **Class Level Information** table lists the independent variables and their levels. The **block** and **cross** effects are listed here but not the GCA effects. We understand from the table that there are 25 blocks and 40 crosses.

Dimensions

Covariance Parameters	3
Columns in X	26
Columns in Z	58
Subjects	1
Max Obs Per Subject	757

Number of Observations

Number of Observations Read	757
Number of Observations Used	757
Number of Observations Not Used	0

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	3628.50837773	

1	3	3555.16229153	0.00003964
2	1	3555.11481671	0.00000080
3	1	3555.11392458	0.00000000

Convergence criteria met.

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
Variance	0.4675	0.2159	2.17	0.0152
cross	0.1054	0.1269	0.83	0.2032
Residual	6.3994	0.3435	18.63	<.0001

3. **Covariance Parameter Estimates:** The 'Estimate' column is the variance components.
   a. The **Variance** is the GCA variance ($\sigma^2_G = 0.4675$)
   b. The **cross** is the SCA variance ($\sigma^2_S = 0.1054$)
   c. The **Residual** is the error variance ($\sigma^2_E = 6.3994$)

Asymptotic Covariance Matrix of Estimates

Row	Cov Parm	CovP1	CovP2	CovP3
1	Variance	0.04662	-0.00381	0.000259
2	cross	-0.00381	0.01610	-0.00609
3	Residual	0.000259	-0.00609	0.1180

4. **Asymptotic Covariance Matrix of Estimates:** The table is the variances of the variance components (diagonal values) and the covariances between variance components (off diagonal elements). For example, the variance of GCA variance is 0.04662, the covariance between GCA and SCA is -0.00381. These variances and covariances of variance components are needed to calculate standard error of heritability or any other function of variance components.

Fit Statistics	
-2 Res Log Likelihood	3555.1
AIC (smaller is better)	3561.1
AICC (smaller is better)	3561.1
BIC (smaller is better)	3555.1

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
block	24	693	4.53	<.0001

5. **Type 3 Tests of Fixed Effects**: Analysis of variance for fixed effects is given. Blocks are different at Pr<0.0001 level.

BOX 1: Causal variance components and heritabilities from a diallel mating design

Using controlled crosses such as diallels, we can obtain additive and dominance genetic variances from analysis of variance.

The variance explained by the general combining ability effects of parents (half-sibs) is a quarter of additive genetic variance.

Additive genetic variance: $\boldsymbol{\sigma^2_A} = 4^* \sigma^2_G$

$= 4^*0.4675 = 1.87$

The variance explained by the female and male interactions (specific combining ability) is one quarter of the dominance genetic variance.

Dominance genetic variance: $\boldsymbol{\sigma^2_D} = 4^* \sigma^2_S$

$= 4^*0.1054 = 0.42$

Phenotypic variance is the sum of the observational components of variance. Notice that

the variance of general combining ability (σ^2_G) is multiplied by 2 because Females and Males contribute ¼ of additive genetic variance to the total variance.

Phenotypic variance: $\boldsymbol{\sigma^2_P} = 2\sigma^2_G + \sigma^2_S + \sigma^2_E$

$= 2*0.4675 + 0.1054 + 6.3994 = 7.44$

Individual-tree narrow-sense heritability: $\mathbf{h^2_i} = \sigma^2_A / \sigma^2_P$

$= 1.87 \ / \ 7.44 \ = \mathbf{0.25}$

Individual-tree broad-sense heritability:

$\mathbf{H^2_i} = 4*(\sigma^2_G + \sigma^2_S) \ / \sigma^2_P$

$= 4*(0.4675 + 0.1054) \ / \ 7.44 = \mathbf{0.31}$

6.4 Using SAS/IML to Estimate Functions of Variance Components

For most of functions of variance components, such as narrow-sense heritability, you may use a spread sheet to do the calculations. However, for more complex or repeated calculations of the same functions, you may consider using software, such as SAS/IML. IML is part of SAS developed to do matrix calculations. For a step-by-step introduction of IML and simple examples about how to use it see Chapter 4.

Remember, we created a matrix of variance components and named it as *_varcomp* and a matrix of covariances of variance components and named it *_cov* in the MIXED procedure in Code 1c. These tables are saved in the WORK library of SAS. We need these tables to calculate heritability and standard error of heritability. Since the *_covparm* and *_cov* are not large, we can simply type the matrices in IML to do calculations.

Code 2a: Calculation of functions of variance components by typing the variance-covariance matrices

We would like to calculate additive genetic variance, phenotypic variance and heritability.

```
/* Heritability estimate  - 1 */
/* Start IML */

   PROC IML;
   _varcomp = {0.4675, 0.1054, 6.3994} ;

   Additive={4 0 0}*_varcomp ;


   /* Phenotypic variance */
   AV={2,1,1};
   Phenotypic=AV`*_varcomp ;

   /* Narrow-sense heritability */
   h2_ns=Additive/Phenotypic;

   /* Broad-sense heritability */
   Genetic ={4 4 0}*_varcomp ;
   h2_bs = Genetic /Phenotypic;


   PRINT
    _varcomp
     Additive [format=6.2]
     Phenotypic [format=6.2]
     h2_ns [format=6.2]
     h2_bs [format=6.2];

RUN;

QUIT;
```

Explanation of the code:

1. **_varcomp={0.4675, 0.1054, 6.3994}**: This is a row vector of variance components. We obtained variance components from the MIXED procedure and created a column vector with 3 rows.

2. **Additive={4 0 0}*_varcomp**: We would like to calculate additive genetic variance, which is four times of the GCA variance (4*0.4675). In order to multiply GCA variance with 4, we need to create a Row vector of coefficients {4 0 0}. The product of the row

vector of coefficients {4 0 0} and the vector of variance components {_varcomp} will give the additive genetic variance.

1. **AV={2,1,1}; Phenotypic=AV`*_varcomp ;** Remember, phenotypic variance is the sum of all variance components that contribute to the Expected Mean Square for the family effect. Notice that the transpose of AV vector is used in multiplication. Multiplying the *_varcomp* vector by the vector of coefficients {2, 1, 1} will give us the phenotypic variance.
2. PRINT: In order to see results, we use the PRINT option. Notice that there is no semicolon ';' after the PRINT option.
3. [format= **6.2**]: This is to set the column length to 6 and the number of decimals to 2 for the output.

Output 2a:

1

_VARCOMP	ADDITIVE	PHENOTYPIC	H2_NS	H2_BS
0.4675	1.87	7.44	0.25	0.31
0.1054				
6.3994				

Code 2b: Calculation of functions of variance components by using the saved output of Mixed procedure of SAS

```
proc iml ;

/*Create column vector of variance components */
USE _varcomp;
READ all var {Estimate} into VC;
CLOSE _varcomp;

/* Create matrix of covariances of variance components */
USE _cov;
```

```
READ all var {CovP1 CovP2 CovP3} into COV;
CLOSE _cov;

/* vector of coefficients for the numerator of heritability */
AU=SHAPE(0,nrow(VC),1);
AU[1,1]=1*4;

/* vector of coefficients for the denominator of heritability */
AV=SHAPE(1,nrow(VC),1);
AV[1,1]=2;

Total=VC[+,1]; *<-- Take the SUM of VC column vector to obtain
total observed variance;
phen=AV`*VC ; *<-- Phenotypic variance;

h2_i=AU`*VC/Phen ;   *<-- Heritability =Additive/Phenotypic;
VC_pct=VC/Total*100; *<-- Percentage of variances by each term;
Var_VC=VECDIAG(Cov); *<-- Variance of variances ;
SE_VC=sqrt(Var_VC);  *<-- Standard Errors of variances ;

* Delta method to estimate standard error of heritability;
var_U =AU`*Cov*AU ; *<---variance of numerator ;
var_V =AV`*Cov*AV ; *<---variance of denominator ;
cov_UV=AU`*Cov*AV ; *<--covariances between variances;


seh2_i=sqrt( (h2_i*h2_i) *
((var_U/(AU`*VC)**2)+(var_V/(AV`*VC)**2)
-(2*cov_UV/(AU`*VC)/(AV`*VC))));

PRINT
VC [format=6.3]
SE_VC [format=6.4]
VC_pct [format=6.1]
phen [format=6.3]
h2_i [format=6.3]
seh2_i [format=6.3] ;

RUN; QUIT;
```

Explanation of the code:

```
/*Create column vector of variance components */
USE _cov;
READ all var {CovP1 CovP2 CovP3} into COV;
CLOSE _cov;
```

Use the variance components created by the Mixed code and create a row vector.

```
/* vector of coefficients for the numerator of heritability */
AU=SHAPE(0,nrow(VC),1);
AU[1,1]=1*4;
```

This is to create a vector of coefficients for the numerator of heritability. The **SHAPE** function creates a matrix named AU, reads the number of rows in the VC matrix and assigns 1 to each element. The AU[1,1]=1*4 multiplies the second element of the row matrix by 4 to obtain additive genetic variance.

```
/* vector of coefficients for the denominator of heritability */
AV=SHAPE(1,nrow(VC),1);
AV[1,1]=2;
```

This is to create a vector of coefficients for the denominator of heritability. The **SHAPE** function creates a matrix named AV, reads the number of rows in the VC matrix and assigns 1 to each element. The AV[1,1]=1*2 multiplies the second element of the vector by 4 to obtain additive genetic variance.

The **VECDIAG** function takes the Diagonal of the COV matrix

Output 2b:

VC	SE_VC	VC_PCT	PHEN	H2_I	SEH2_I
0.468	0.2159	6.7	7.440	0.251	0.103
0.105	0.1269	1.5			
6.399	0.3435	91.8			

6.5 Example for Multiple Environments

- **Genetic materials**: 18 loblolly pine trees were mated to produce 40 full-sib families (crosses) for progeny testing.
- **Field design**: A randomized complete blocks design was used with single tree plots. One progeny of each cross was randomly assigned in a block. There were 25 blocks in one site (environments). Thus, each cross had 25 progeny at one site. The experiment was replicated at six sites.

6.5.1 The statistical model

The following linear mixed model was fitted to data to estimate variance components for multi environment diallel tests.

[1] $$Y_{ijklm} = \mu + S_i + B_{j(i)} + GCA_k + GCA_l + SCA_{kl} + S*GCA_{ik} + S*GCA_{il} + S*SCA_{ikl} + E_{ijklm}$$

where

Y_{ijklm} is the mth observation of the jth block for the klth cross in the ith Site;

μ is the overall mean;

S_i is the ith fixed Site (environment) effect, i=1 to t;

$B_{j(i)}$ is the fixed effect of the jth block within the ith Site, j=1 to b;

GCA_k ,GCA_l is the random **general combining ability (GCA) effect** of the kth female or the lth male ~Normally, Independently Distributed (NID) (0, σ^2_G), k, l=1 to p and $k<l$;

SCA_{kl} is the random **specific combining ability (SCA) effect** of the kth and the lth parents ($k \neq l$) ~NID (0, σ^2_S);

$S*GCA_{ik}$,$S*GCA_{il}$ is the random **GCA by Site Interaction** ~NID (0, σ^2_{TG});

$S*SCA_{ikl}$ is the random **SCA by Site interaction** effect ~NID (0, σ^2_{TS});

E_{ijklm} is the random error term ~NID (0, σ^2_E)

We can write above linear model in a matrix form, which is shorter.

[2] $\mathbf{y} = \mathbf{X\beta} + \mathbf{Z\gamma} + \boldsymbol{\varepsilon}$

where,

- $\mathbf{y}$ is the vector of individual observations,
- $\boldsymbol{\beta}$ is the vector of fixed-effects parameters (overall mean, site and blocks within site),
- $\boldsymbol{\gamma}$ is the vector of random-effects parameters including general combining ability (GCA) for female and male, specific combining ability (SCA), GCA x Site interaction, SCA x Site interaction.
- $\boldsymbol{\varepsilon}$ is an unknown random error vector
- $\mathbf{X}$ is the known design matrix for the fixed effects
- $\mathbf{Z}$ is the known design matrix for random effects

- The major assumption is that the random effects $\boldsymbol{\gamma}$ and error term $\boldsymbol{\varepsilon}$ are assumed to have normal distributions with 0 mean and variances.

$$E\begin{pmatrix}\gamma\\ \varepsilon\end{pmatrix}=\begin{pmatrix}0\\ 0\end{pmatrix} \quad \text{Var}\begin{pmatrix}\gamma\\ \varepsilon\end{pmatrix}=\begin{pmatrix}G & 0\\ 0 & R\end{pmatrix}$$

- The second major assumption is that residuals have a normal distribution and they are independent of each other. See Chapter 2 for full description of assumptions of a linear mixed model.

6.6 Implementation with SAS Mixed Procedure

In the diallels, the same parents are used as females and males. Thus, each parent contributes to the general combining ability variance (GCA) as females and males. SAS does not have a simple procedure to take into account the 'double' effects of parents. Instead, various SAS codes were developed to estimate a single GCA variance by aggregating the effects of parents (Xiang and Li 2001, Johnson and King 1998, Wu and Matheson 2000, 2001, Zang and Kang 1997).

The first 10 observations of the data are given below.

Diallel data from multi environments

Obs	site	block	female	male	treeID	HEIGHT	VOLUME
1	4	1	P02	P01	2502	30	1.37
2	4	3	P02	P01	2554	36	3.20
3	4	4	P02	P01	2582	30	1.42
4	4	5	P02	P01	2612	29	0.93
5	4	6	P02	P01	2639	33	2.36
6	4	7	P02	P01	2670	32	2.04
7	4	8	P02	P01	2699	33	2.85
8	4	9	P02	P01	2729	34	2.02
9	4	10	P02	P01	2763	34	1.91
10	4	11	P02	P01	2798	34	1.68

The following SAS code was modified from Xiang and Li (CJFR 2003) and Gary Hodge for a multi environment diallel data.

```
/* ANALYSIS OF DIALLEL DATA */

%let var1=height; * Set Var1 = variable for Analysis ;
%let ds=A; * Set data for Analysis ;

/* Generate a list of the Female parents */
PROC SORT data=&ds; by female;
data females;
set &ds;
  by female;
  if first.female;
  parent=female;
keep parent;

/* Generate a list of the Male parents */
PROC SORT data=&ds; by male;
data males;
  set &ds;
  by male;
  if first.male;
  parent=male;
keep parent;
```

```
/* Combine the two lists into one */
data parents;
 set females males;

/* Remove duplicate parent IDs from the list*/
PROC SORT; by parent;

data parents;
set parents;
by parent;
if first.parent;


/* Create total number of parents */
proc freq data=parents noprint;
tables parent / all ;
output out=numpar n;

data numpar;
  set numpar;
   call symput('numpar',n);
/* symput creates a macro variable called &numpar
   with value n=numpar, number of parents */

proc print; title 'numpar'; run;
```

numpar

Obs	N
1	19

The output shows that we have 10 parents in the data. They were used as female and male in the mating design.

```
/* Create variables for use in Proc Mixed to designate
   P1 P2 ... PN  for parents
   site*P1 site*P2 ... site*PN for Parent x Site Interactions
   in the Random statements */
data listpar;
length mixpar $400;  * set the length of column for P1, P2,.. ;
length mixpxt $1100; * set the length of column for site*P1,
site*P2,.. ;
mixpar='P1X';
mixpxt='site*P1X';

data listpar;
```

```
set listpar;
do i=2 to &numpar;
 mixpar = compress(mixpar||'P'||i)||'X';
 mixpxt = compress(mixpxt||'site*P'||i)||'X';
end;
output;
```

```
data listpar;
set listpar;
mixpar=translate(mixpar,' P','XP');
mixpxt=translate(mixpxt,' t','Xt');

proc print; title 'listpar'; run;
```

Obs	mixpar	mixpxt	i
1	P1 P2 P3 P4 P5 P6 P7 P8 P9 P10 P11 P12 P13 P14 P15 P16 P17 P18 P19	site*P1 site*P2 site*P3 site*P4 site*P5 site*P6 site*P7 site*P8 site*P9 site*P10 site*P11 site*P12 site*P13 site*P14 site*P15 site*P16 site*P17 site*P18 site*P19	20

The output shows the lists of the parent IDs' (dummy variable) and parent by site ID (dummy variable)

```
data mixlist; set listpar;
call symput('mixpar',mixpar);
call symput('mixpxt',mixpxt);
/* symput creates a macro variables called &mixpar which lists
the parents and a macro variable called &mixpxt which lists the
effects for site*parents. These lists will be used in Proc Mixed
to create the Z matrix for GCA and GCA x site */
run;
```

```
/* Create dummy variables for each parent
 to generate a Design Matrix for parents*/
PROC IML;
use parents;
read all var {parent} into P;
nparents=nrow(P);
close parents;

codes99='P1':'P99';
codes=codes99[1:nparents];
```

```
print codes P;

use &ds;
read all var {female male} into FM;
/* FM is the list of female and male parents for all
observations */
n=nrow(FM);
/* n = number of observations */

/* Create parent design matrix D (n rows x nparents columns)
   There is a 1 in the two colums corresponding to the male and
female parents */
  D=shape(0,N,nparents);
  do I=1 to N;
    do J=1 to nparents;
      if FM[I,1]=P[J,1] | FM[I,2]=P[J,1] then D[I,J]=1;
    end;
  end;

/* Create a SAS data set DUMMY from the design matrix D */
create DUMMY from D [colname=codes];
append from D;
quit;

/* Merge the dummy variables onto the diallel data set */
data &ds;
merge &ds DUMMY;
run;


proc print data=&ds (obs=5);
title 'Original data set and dummy variables';
run;
```

Original data set and dummy variables

Obs	Set	Block	Female	Male	Hybrid	Yield	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19
1	4	1	P02	P01	2502	30	1.37	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	4	3	P02	P01	2554	36	3.20	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	4	4	P02	P01	2582	30	1.42	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	4	5	P02	P01	2612	29	0.93	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	4	6	P02	P01	2639	33	2.36	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

The output shows the combined original data dummy variables.

```
/* Run Proc Mixed on variable var1 */

PROC MIXED data=&ds covtest noitprint ;
class site block female male ;
  model &var1=site block(site) / solution outpm=pm&var1;
     /* site and block are fixed effects. Blocks are nested */
  random &mixpar/type=toep(1) solution  ; /* GCA effects */
  random female*male;                     /* SCA effects */
  random &mixpxt/type=toep(1);     /* GCA x Site effects */
  random female*male*site;         /* SCA x Site effects */

ODS output covparms=_varcomp&var1 asycov=_cov ;
  /* Write the parameter estimates into a SAS data set */
ODS output solutionR=BLUP&var1;
  /* Write the BLUP solutions into a SAS data set */
ODS output solutionF=BLUE&var1;
  /* Write the BLUE solutions of fixed effects into a data*/

ods listing exclude solutionf solutionr;
ods html exclude solutionf solutionr;

run;
```

1. ODS LISTING EXCLUDE: This is to stop printing large data (predicted values fo fixed and randoem effects.
2. Here we are creating the variance components (COVPARMS) and the covariances of variance components (ASYCOV). The name of the new tables would be **_covparms** and **_cov**.

OUTPUT:

The Mixed Procedure

Model Information

Data Set	WORK.A
Dependent Variable	HEIGHT
Covariance Structures	Banded Toeplitz, Variance Components

Estimation Method	REML
Residual Variance Method	Profile
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Name of the data set, the response variable (HEIGHT) and error covariance structure (banded toeplitz) is summarized. Restricted maximum likelihood method is used to estimate variances.

Class Level Information

Class	Levels	Values
site	6	1 2 3 4 5 6
block	25	1 10 11 12 13 14 15 16 17 18 19 2 20 21 22 23 24 25 3 4 5 6 7 8 9
female	17	P02 P03 P04 P05 P06 P07 P08 P09 P10 P11 P12 P13 P14 P15 P16 P18 P19
male	10	P01 P02 P04 P05 P06 P07 P08 P11 P15 P17

Model information is summarized. There are 6 sites and 25 blocks in each site. 17 parents were used females and 10 were used as males.

Dimensions

Covariance Parameters	5
Columns in X	157
Columns in Z	413
Subjects	1
Max Obs Per Subject	4913

Covariance parameters are the random effects (GCA, SCA etc.). Dimension of incidence matrices ($X_{157x4913}$) and $Z_{413x4913}$) are given. Large Z matrix may substantially increase computation time.

Number of Observations

Number of Observations Read	4913
Number of Observations Used	4913
Number of Observations Not Used	0

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
Variance	0.3938	0.1720	2.29	0.0110
female*male	0.09642	0.04660	2.07	0.0193
Variance	0.1900	0.05036	3.77	<.0001
site*female*male	0.01476	0.04691	0.31	0.3765
Residual	7.1799	0.1508	47.62	<.0001

Covariance Parameters Estimates are variance components.
The first Variance is GCA, female*male is SCA, the second Variance is GCA*Site, site*female*male is SCA*Site.
Approximate standard errors of estimates, Z test with probability of Z were given.

Fit Statistics

-2 Res Log Likelihood	23591.6
AIC (smaller is better)	23601.6
AICC (smaller is better)	23601.6
BIC (smaller is better)	23591.6

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
site	5	114	361.68	<.0001
block(site)	144	4530	5.45	<.0001

Analysis of variance for fixed effects is given. The sites are significantly different. Similarly, blocks within are site different.

The Asymptotic Covariance Parameters table is the variances of the ESTIMATES (diagonal values) and the covariances between ESTIMATES (off diagonal elements). For example, the variance of GCA estimate is 3778.32, the covariance between GCA and SCA is -1.1663.

The program creates several data sets: Here is a list taken from the LOG window of SAS:

```
NOTE: The data set WORK.BLUEHEIGHT has 157 observations and 8 variables.
NOTE: The data set WORK.BLUPHEIGHT has 413 observations and 9 variables.
NOTE: The data set WORK._COV has 5 observations and 7 variables.
NOTE: The data set WORK._VARCOMPHEIGHT has 5 observations and 5 variables.
NOTE: The data set WORK.PMHEIGHT has 4913 observations and 33 variables.
```

WORK.BLUEHEIGHT is the BLUE of fixed effects
WORK.BLUPHEIGHT is the GCA values of parents and SCA values of crosses
WORK._COV is the table of variance and covariances of variance components
WORK._VARCOMPHEIGHT is the variance components
WORK.PMHEIGHT is the residual value associated with each individual tree

6.7 Genetic model, functions of variance components

The following functions of variance components (e.g., heritability) can be obtained from two tables WORK._VARCOMPHEIGHT and WORK._COV.

[3] Genetic variances and standard errors

Covariance among half-sibs is $\sigma_{HS} = \frac{1}{4}\sigma_A^2$

Covariance among full-sibs is $\sigma_{FS} = \frac{1}{4}\sigma_D^2$

$\sigma_A^2 = 4\sigma_G^2$ Additive genetic variance is 4 times of GCA variance

$Var(\sigma_A^2) = Var(4\sigma_G^2) = 16Var(\sigma_G^2)$ Variance of additive genetic variance

$SE(\sigma_A^2) = \sqrt{16Var(\sigma_G^2)}$ Standard error of additive genetic variance

$\sigma_D^2 = 4\sigma_S^2$ Non-additive genetic variance

$Var(\sigma_D^2) = Var[4(\sigma_S^2)]$ =16 [Var(σ_S^2)] Variance of non-additive genetic variance

$SE(\sigma_D^2) = \sqrt{16Var(\sigma_S^2)}$ Standard error of non-additive genetic variance

The variance of GCA variance [Var(σ^2_G)] comes from the output of SAS MIXED procedure. The table is called Asymptotic Covariance Matrix of Estimates. See an example in Code 2. We need the variances and covariances of variance components to calculate standard error of additive genetic variance or standard error of heritability.

[4] Total phenotypic variance and heritability

Total phenotypic variance

$\boldsymbol{\sigma^2_P} = 2\sigma^2_G + \sigma^2_S + 2\sigma^2_{GT} + \sigma^2_{ST} + \sigma^2_E$

Individual-tree narrow-sense heritability (for mass selection)

$$h_i^2 = \frac{4\sigma^2_G}{2\sigma^2_G + \sigma^2_S + 2\sigma^2_{GT} + \sigma^2_{ST} + \sigma^2_E}$$

Variance of heritability Var(h_i^2):

a) Dickerson approximation (Assuming σ^2_P is a constant):

$$\mathbf{Var(}\, h_i^2\mathbf{)} = \frac{16\mathrm{Var}(\sigma^2_G)}{(\sigma^2_P)^2}$$

b) Delta method (Assuming σ^2_P is random):

$$\mathrm{Var}(h_i^2) = \left(\frac{4\sigma^2_G}{\sigma^2_P}\right)\left[\frac{Var(4\sigma^2_G)}{(4\sigma^2_G)^2} + \frac{Var(\sigma^2_P)}{(\sigma^2_P)^2} - \frac{2Cov(4\sigma^2_G, \sigma^2_P)}{(4\sigma^2_G\sigma^2_P)}\right]$$

Broad-sense heritability

$$\mathrm{H}_\mathrm{i}^2 = \frac{4(\sigma^2_G + \sigma^2_S)}{2\sigma^2_G + \sigma^2_S + 2\sigma^2_{GT} + \sigma^2_{ST} + \sigma^2_E}$$

[5] Phenotypic variance and heritability of <u>half-sib family mean</u>

Phenotypic variance of half-sib family mean is

$$\mathrm{Var}(\overline{Y}..k.) = \sigma^2_{P\_HS} = \left(\frac{1}{p-1}\right)\left(p\sigma^2_G + \sigma^2_S + \frac{p\sigma^2_{GT}}{t} + \frac{\sigma^2_{ST}}{t} + \frac{\sigma^2_E}{tbn}\right)$$

where p is the number of parents used in the analysis. If p is large (p>20), then it can be ignored.

Half-sib family mean heritability

$$h^2_{HS} = \frac{\sigma^2_G}{\left(\frac{1}{p-1}\right)\left(p\sigma^2_G + \sigma^2_S + \frac{p\sigma^2_{GT}}{t} + \frac{\sigma^2_{ST}}{t} + \frac{\sigma^2_E}{tbn}\right)}$$

The variance of half-sib family mean heritability Var(h^2_{HS}), can be obtained applying general formula of Dickerson approximation or the Delta Method given above.

a) Assuming $\sigma^2_{P\_HS}$ is a constant (Dickerson approximation): Var(σ^2_G) / ($\sigma^2_{P\_HS}$)^2

b) Assuming $\sigma^2_{P\_HS}$ is random (Delta method):

$$\left(\frac{\sigma^2_G}{\sigma^2_{P\_HS}}\right)\left[\frac{Var(\sigma^2_G)}{(\sigma^2_G)^2}+\frac{Var(\sigma^2_{P\_HS})}{(\sigma^2_{P\_HS})^2}-\frac{2Cov(\sigma^2_G,\sigma^2_{P\_HS})}{(\sigma^2_G\sigma^2_{P\_HS})}\right]$$

[6] Phenotypic variance and heritability of <u>full-sib family mean</u>

Phenotypic variance of full-sib family mean is

$$\sigma^2_{P\_FS}=2\sigma^2_G+\sigma^2_S+\frac{2\sigma^2_{GT}}{t}+\frac{\sigma^2_{ST}}{t}+\frac{\sigma^2_{PLOT}}{tb}+\frac{\sigma^2_E}{tbn}$$

Heritability of full-sib family mean (narrow sense)

$$h^2_{FS}=\frac{2\sigma^2_G}{\sigma^2_{P\_FS}}$$

Heritability of full-sib family mean (broad sense)

$$h^2_{FS}=\frac{2\sigma^2_G+\sigma^2_S}{\sigma^2_{P\_FS}}$$

The variance of full-sib family mean heritability can be obtained applying general formula of Dickerson approximation or the Delta Method given above.

[7] Phenotypic variance and heritability of <u>within full-sib family</u>

Phenotypic variance of within full-sib family

$$\sigma^2_{P\_FSw}=\frac{(t-1)}{t}\left(2\sigma^2_{GT}+\sigma^2_{ST}+\frac{(b-1)\sigma^2_{PLOT}}{b}+\frac{(bn-1)\sigma^2_E}{bn}\right)$$

Heritability of within full-sib family (narrow sense)

$$h^2_{FSW}=\frac{2\sigma^2_G}{\sigma^2_{P\_FSw}}$$

Heritability of within full-sib family (broad sense)

$$H^2_{FSW}=\frac{2\sigma^2_G+3\sigma^2_S}{\sigma^2_{P\_FSw}}$$

The variance of within full-sib family heritabilities can be obtained applying general formula of Dickerson approximation or the Delta Method given above.

YOU MAY MODIFY THE IML CODE GIVEN IN 6.4 TO CALCULATE ABOVE PHENOTYPIC VARIANCES AND HERITABILITIES.

6.8 Breeding Values

Breeding value of a parent or half-sib family is 2 times of its general combining ability.

$BV = 2GCA$

Any cross between two parents (let's say F and M) has an expected breeding value, which is the sum of the GCA of F and M.

$$BV_{FM} = GCA_F + GCA_M$$

The expected full-sib family (cross) mean may deviate from above sum. This deviation is called **specific combining ability** (SCA) of two parents. Sometimes, the sum of three components is called **genetic value** of the cross:

$$GV_{FM} = GCA_F + GCA_M + SCA_{FM}$$

Where, GCAf, GCAm, and SCAfm are general combining ability of female, male and the specific combining ability of the cross between two.

BLUP individual-tree breeding value (IBV) is obtained by adding parental GCA estimates to the estimated within-family value (*A*w).

$IBV = GCAf + GCAm + Aw$

$Aw = 2\sigma_G^2 / \sigma_E^2 \, (\mathbf{y} - \mathbf{X}\hat{\mathbf{B}} - \mathbf{Z}\hat{\boldsymbol{\gamma}})$

The deviation ($\mathbf{R} = \mathbf{y} - \mathbf{X}\hat{\mathbf{B}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}$) is the residual, which is the difference between observed values ($\mathbf{y}$) and the Best Linear Unbiased Predicted values of fixed ($\mathbf{X}\hat{\mathbf{B}}$) and random ($\mathbf{Z}\hat{\boldsymbol{\gamma}}$) effects. The measured trait of a tree is adjusted for fixed and random effects in the model and

then multiplied by approximate within-family heritability ($2\sigma_G^2/\sigma_E^2$) to obtain within family deviation Aw (Xiang and Li 2001).

Now let's look at the BLUP values:

```
proc print data=BLUP&var1 (obs=25);
title 'GCA values';
run;
```

GCA values

Obs	Effect	site	female	male	Estimate	StdErr Pred	DF	tValue	Probt
1	P1				0.6556	0.4335	4530	1.51	0.1305
2	P2				0.4744	0.4097	4530	1.16	0.2469
3	P3				-0.3003	0.4301	4530	-0.70	0.4851
4	P4				0.4512	0.2836	4530	1.59	0.1117
5	P5				0.5621	0.2779	4530	2.02	0.0432
6	P6				0.01165	0.2749	4530	0.04	0.9662
7	P7				-1.2821	0.2796	4530	-4.59	<.0001
8	P8				-0.5041	0.2798	4530	-1.80	0.0716
9	P9				-0.00705	0.3974	4530	-0.02	0.9859
10	P10				-0.1480	0.3920	4530	-0.38	0.7057
11	P11				-0.1121	0.2749	4530	-0.41	0.6833
12	P12				0.003334	0.3943	4530	0.01	0.9933
13	P13				-0.1956	0.3927	4530	-0.50	0.6184
14	P14				-0.5703	0.3961	4530	-1.44	0.1501
15	P15				0.5034	0.2781	4530	1.81	0.0704
16	P16				0.7933	0.3904	4530	2.03	0.0422
17	P17				0.2051	0.2843	4530	0.72	0.4708
18	P18				-0.7839	0.2742	4530	-2.86	0.0043
19	P19				0.2433	0.3906	4530	0.62	0.5333
20	female*male		P02	P01	0.1605	0.2900	4530	0.55	0.5798
21	female*male		P03	P02	-0.07353	0.2897	4530	-0.25	0.7996
22	female*male		P04	P06	0.06612	0.2322	4530	0.28	0.7759
23	female*male		P04	P07	-0.00294	0.2361	4530	-0.01	0.9901
24	female*male		P04	P08	-0.02507	0.2384	4530	-0.11	0.9163
25	female*male		P04	P11	0.1122	0.2326	4530	0.48	0.6295

Observations from 1 to 19 are GCA values of parents (estimate), their standard errors (StdErrPred) were produced. Breeding value of a parent is 2 x GCA since a parent can contribute only 50% of its progeny genetics.

Observations starts from 20 are the SCA values of crosses. SCA of the cross can be added to the parental GCA values to calculate genetic value (GV) of a cross. For example, genetic value of cross P1 x P2;

GV = gcaf + gcam + sca = 0.6556 + 0.4744 + 0.1605

6.9 Literature

Johnson, G.R. and King, J.N. 1998. Analysis of half diallel mating designs I - a practical analysis procedure for ANOVA approximation. Silvae Genetica. 47(2-3): 74-79.

Manjit S. (ed). 2003. Handbook of formulas and software for plant geneticists and tree breeders. Food Products Press, New York. 347 p.

Wu, H.X. and Matheson, A.C. 2000. Analysis of half-diallel mating design with missing crosses: theory and SAS program for testing and estimating GCA and SCA fixed effects. Silvae Genetica 49:130-137.

Wu, H.X. and Matheson, A.C. 2001. Analysis of half-diallel mating design with missing crosses: theory and SAS program for testing and estimating GCA and SCA variance components. Silvae Genetica 50:265-271.

Xiang, Bin and Li, Bailian. Best linear unbiased prediction of clonal breeding values and genetic values from full-sib mating designs. Canadian Journal of Forest Research 33:2036–2043

Zhang, Y., and Kang, M.S. 1997. DIALLEL-SAS: A SAS program for Griffing's diallel analyses Agronomy Journal 89:176-182.

Weblinks for ASReml:
Supplier: http://www.vsn-intl.com/ASReml/index.htm
Forestry Examples: http://uncronopio.org/luis/asreml_cookbook.

Appendix:

Derivation of variance of phenotypic variance

(Assuming RCB design with row plots)

Variance of phenotypic variance

$\mathrm{Var}(\sigma^2_P)$: $\mathbf{Var(\sigma^2_P)} = \mathrm{Var}(2\sigma^2_G + \sigma^2_S + 2\sigma^2_{GT} + \sigma^2_{ST} + \sigma^2_{PLOT} + \sigma^2_E)$

$= \mathrm{Var}(2\sigma^2_G) + \mathrm{Var}(\sigma^2_S) + \mathrm{Var}(2\sigma^2_{GT}) + \mathrm{Var}(\sigma^2_{ST}) + \mathrm{Var}(\sigma^2_{PLOT}) + \mathrm{Var}(\sigma^2_E) +$

$2\ [\ \mathrm{Cov}(2\sigma^2_G,\sigma^2_S) + \mathrm{Cov}(2\sigma^2_G,2\sigma^2_{GT}) + \mathrm{Cov}(2\sigma^2_G,\sigma^2_{ST}) + \mathrm{Cov}(2\sigma^2_G,\sigma^2_{PLOT}) + \mathrm{Cov}(2\sigma^2_G,\sigma^2_E) +$

$\mathrm{Cov}(\sigma^2_S,2\sigma^2_{GT}) + \mathrm{Cov}(\sigma^2_S,\sigma^2_{ST}) + \mathrm{Cov}(\sigma^2_S,\sigma^2_{PLOT}) + \mathrm{Cov}(\sigma^2_S,\sigma^2_E) +$

$\mathrm{Cov}(2\sigma^2_{GT},\sigma^2_{ST}) + \mathrm{Cov}(2\sigma^2_{GT},\sigma^2_{PLOT}) + \mathrm{Cov}(2\sigma^2_{GT},\sigma^2_E) +$

$\mathrm{Cov}(\sigma^2_{ST},\sigma^2_{PLOT}) + \mathrm{Cov}(\sigma^2_{ST},\sigma^2_E) + \mathrm{Cov}(\sigma^2_{PLOT},\sigma^2_E)\ ]$

We assume variance components are not independent. That's why they have covariances. Again, the variances and the covariances of variance components are produced by SAS MIXED procedure. The name of the table output is 'Asymptotic Covariance Matrix of Estimates'.

Standard error of phenotypic variance SE(σ^2_P) is simply the square root of the variance.

Derivation of within family phenotypic variance

(Assuming RCB design with row plots)

We can easily derive within full-sib family phenotypic variance using the total phenotypic variance and variance of full-sib family means.

$$\sigma^2_{P\_i} - \sigma^2_{P\_FS} = 2\sigma^2_G + \sigma^2_S + 2\sigma^2_{GT} + \sigma^2_{ST} + \sigma^2_{PLOT} + \sigma^2_E - 2\sigma^2_G + \sigma^2_S + \frac{2\sigma^2_{GT}}{t} + \frac{\sigma^2_{ST}}{t} + \frac{\sigma^2_{PLOT}}{tb} + \frac{\sigma^2_E}{tbn}$$

$$(2\sigma^2_{GT} + \sigma^2_{ST} + \sigma^2_{PLOT} + \sigma^2_E) - (\frac{2\sigma^2_{GT}}{t} + \frac{\sigma^2_{ST}}{t} + \frac{\sigma^2_{PLOT}}{tb} + \frac{\sigma^2_E}{tbn})$$

After some algebra within full-sib family phenotypic variance would be

$$\sigma^2_{P\_w} = \frac{(t-1)}{t}\left(2\sigma^2_{GT} + \sigma^2_{ST} + \frac{(b-1)\sigma^2_{PLOT}}{b} + \frac{(bn-1)\sigma^2_E}{bn}\right)$$