FOR 728: Quantitative Forest Genetics Methods

(Fall 2009)

Fikret Isik,
North Carolina State University,
Department of Forestry and Environmental Resources

Lecture 9: Genetic Correlations and Correlated Response

Background

Phenotypic values of different traits in the same trees are often correlated, such as height and diameter. Environmental factors and genetic effects are two reasons for correlations. Similar to partitioning phenotypic variance, we can decompose phenotypic correlation (r_P) into its genetic (r_A) and environmental (r_E) components.

$$r_P = r_A + r_E$$
$$r_P = cov_{Pxy} / [\sigma^2_{Px}\, \sigma^2_{Py}]^{1/2} = cov_{Axy} / [\sigma^2_{Ax}\, \sigma^2_{Ay}]^{1/2} + cov_{Exy} / [\sigma^2_{Ex}\, \sigma^2_{Ey}]^{1/2}$$

where cov_{Axy} is genetic covariance, cov_{Exy} is environmental covariance, σ^2 _subscripts are additive genetic or environmental variances for traits X and Y. Genetic correlations can arise from two causes; (i) Pleiotropy: One gene or several genes may influence multiple traits, (ii) Gametic phase disequilibrium between genes affecting different traits. A set of closely linked genes present on one chromosome tend to be inherited together (not easily separable by recombination), i.e. gene A is affecting Trait X, gene B affecting trait Y only. If two genes are in linkage disequilibrium, a genetic covariance may arise between traits X and Y. This could be temporary when there is random mating.

We are already familiar with the decomposition of genetic variances for traits X and Y. Recall that genetic variance for traits X and Y can be portioned as follows:

$$\sigma^2_{Gx} = \sigma^2_{Ax} + \sigma^2_{Dx} + \sigma^2_{AA\,x} + \sigma^2_{AD\,x} + \sigma^2_{DD\,x} + \ldots$$
$$\sigma^2_{Gy} = \sigma^2_{Ay} + \sigma^2_{Dy} + \sigma^2_{AAy} + \sigma^2_{ADy} + \sigma^2_{DDy} + \ldots$$

Ignoring linkage disequilibrium, we can partition genetic covariances into components the same way as we did for genetic variances:

$$cov_{Gxy} = cov_A(1,2) + cov_D(1,2) + cov_{AA}(1,2) + cov_{AD}(1,2) + cov_{DD}(1,2) + \ldots$$

Where, 1 and 2 are two loci, cov_A is additive, cov_D is dominant genetic covariance etc. If additive and dominance genetic covariances can be estimated, then r_A, r_D and r_E can also be estimated.

$$r_P = cov_P / [\sigma^2_{Px}\, \sigma^2_{Py}]^{1/2}$$
$$cov_P = r_P\, \sigma_{Px}\, \sigma_{Py}$$

$$r_A = cov_A / [\sigma^2_{Ax}\, \sigma^2_{Ay}]^{1/2}$$
$$cov_A = r_A\, \sigma_{Ax}\, \sigma_{Ay}$$

$$r_E = cov_E / [\sigma^2_{Ex}\, \sigma^2_{EPy}]^{1/2}$$
$$cov_E = r_E\, \sigma_{Ex}\, \sigma_{Ey}$$

$$cov_P = cov_A + cov_E$$

$$r_P\, \sigma_{Px}\, \sigma_{Py} = r_A\, \sigma_{Ax}\, \sigma_{Ay} + r_E\, \sigma_{Ex}\, \sigma_{Ey}$$

Phenotypic correlation is function of genetic and environmental correlation. The expression can be simplified by substituting the square root of variances as suggested by Walsh.

$$h^2 = \sigma^2_A / \sigma^2_P$$
$$\sigma^2_A = h^2\, \sigma^2_P$$
$$\sigma_A = h\, \sigma_P$$

The square root of additive genetic variance (σ_A) is the product of square root of heritability (h) and the square root of phenotypic variance (σ_P).

$$\sigma^2_e = 1 - h^2$$
$$\sigma e = (1 - h^2)^{1/2}$$
$$\sigma^2_e = \sigma^2_E / \sigma^2_P$$
$$\sigma^2_E = \sigma^2_e\, \sigma^2_P$$
$$\sigma_E = \sigma_e\, \sigma_P$$

The σ^2_e is the remaining ratio of the total phenotypic variance after subtracting heritability. In another word, σ^2_e is the ratio of environmental variance and phenotypic variance. Phenotypic variance can be formulated as the product of σ_e and σ_P.

Phenotypic correlations can be written as the function of heritabilities, genetic and environmental correlations.

$$r_P = r_A\ h_x \sigma_{Px}\ h_y \sigma_{Py} + r_E\, [\sigma^2_{Ex} / \sigma^2_{Px})\, ([\sigma^2_{Ey} / \sigma^2_{Py})]^{1/2}$$
$$r_P = r_A\, h_x\, h_y + r_E\, [(1 - h^2_x)\, (1 - h^2_y)]^{1/2}$$

How phenotypic correlations change as the heritabilities increase or decrease?

Why r_A is important in quantitative genetics and in breeding?

- Use for indirect selection and predict correlated response (genetic gain). In some cases it could be expensive to measure a trait directly. If Y is an easily observed trait that is highly correlated with X, then we can improve Y instead of X, and hope to make positive change in X in the population.
- Develop selection indices to select for multiple traits simultaneously
- Determine extend of genotype-environment interaction to develop breeding strategies.
- Understand evolutionary process of traits

Regression of family means

If it is difficult to estimate r_A and test its significance, you can replace r_A with a correlation of family means. The rationale behind this is, if we keep increasing the number of measured individuals for a family, then the sampling error of the mean becomes so small that the phenotypic mean would be approximately equal to genetic mean. This approach could be biased if the heritabilities of two traits are low. See Lynch and Walsh 1998, p 636 for more details.

$$r_P = \frac{\mathrm{cov}(\bar{z}_x, \bar{z}_y)}{\sqrt{\sigma^2_{\bar{z}x}\sigma^2_{\bar{z}y}}} = r_A\left[\frac{\phi h_x h_y (r_z / r_A)}{\sqrt{(\phi h_x^2 + 1)(\phi h_y^2 + 1)}}\right]$$

Where,

$\mathrm{cov}(\bar{z}_x, \bar{z}_y)$ = covariance between family means for traits X and Y,

σ^2_{zx} and σ^2_{zy} = family variances for traits X and Y,

φ (phi) $=2\Theta(n-1)$ and 2Θ = coefficient of relationship (i.e, 0.25 for half-sibs and 0.5 for full-sibs), n is the number of progeny,

r_P is phenotypic correlation, h_x and h_y are square-root of heritabilities of traits X and Y.

- If heritability of X and Y are 1, then family-mean correlation is an unbiased estimate of genetic correlation (that is unlikely)
- If heritabilities are moderate i.e., 0.5, then family size should be large,
- The advantage of family-mean correlation is that it is a true product-moment correlation. It is always estimated between theoretical limits (-1, 1). More importantly, it is significance can be easily tested using standard tables of critical values.

The precision of genetic correlation (r_A)

Precise estimation of r_A requires many families each with a large number of progenies. The following formulas are approximate to estimate variation around r_A. When the population size small then the standard errors of r_A should be used with cautious because we do not know its distribution. With the increased power of computing, resampling methods have become available to better estimate approximate distributions of correlations.

1-An approximate standard error of genetic correlation (Falconer 1996):

$$SE(r_A) = \frac{1 - r_A^2}{\sqrt{2}}\sqrt{\frac{SE(h_x^2)SE(h_y^2)}{h_x^2 h_y^2}}$$

2-Delta method is considered a better way to estimate variance of ratios for unknown distributions (Lynch and Walsh 1998). See Appendix 1 in Lynch and Walsh for more details about the Delta method.

$$\mathrm{var}(r) = (r^2)\left[\frac{\mathrm{var}(\sigma_x^2)}{4\sigma_x^{2^2}} + \frac{\mathrm{var}(\sigma_y^2)}{4\sigma_y^{2^2}} + \frac{\mathrm{var}(\sigma_{xy})}{\sigma_{xy}^2} + \frac{2\,\mathrm{cov}(\sigma_x^2\sigma_y^2)}{4\sigma_x^2\sigma_y^2} - \frac{2\,\mathrm{cov}(\sigma_x^2\sigma_{xy})}{2\sigma_x^2\sigma_{xy}} - \frac{2\,\mathrm{cov}(\sigma_{xy}\sigma_y^2)}{2\sigma_{xy}\sigma_y^2}\right]$$

$\sigma^2_{\text{-}}$ = Variance,
$\mathrm{var}(\sigma^2_{\text{-}})$ = Variance of variance,
σ_{xy} = Covariance
σ^2_{xy} = Squared covariance,
$\mathrm{var}(\sigma_{xy})$ = Variance of the covariance
$\mathrm{cov}(\sigma^2_x\sigma^2_y)$ = Covariance between the variances of X and Y
$\mathrm{cov}(\sigma^2_x\sigma_{xy})$ = Covariance between the variance of X and the covariance,
$\mathrm{cov}(\sigma_{xy}\sigma^2_y)$ = Covariance between the covariance and the variance of Y

Correlated response and indirect selection

The mean of trait X in a breeding population can change in two ways:
1) *as a direct response to selection on X (R_x).* The response (R_x) in directly selected character X is;

$$R_x = i_x\, h_x\, \sigma_{Ax}$$
$$= i_x\, (\sigma_{Ax} / \sigma_{Px})\, \sigma_{Ax}$$
$$= i_x\, \sigma^2_{Ax} / \sigma_{Px}$$

2) *as a correlated, indirect, response to direct selection on Y (CR_x).* When X and Y are genetically correlated, selection on X will result in change in Y too. Such a change in the unselected trait (Y) is correlated response. The response to selection of trait X is (by definitation) the mean breeding value of selected individuals. Thus, the change in trait Y in response to selection on X is the regression of breeding value of Y on the breeding value of X (Walsh lecture notes, page 13, Falconer and MacKay page 317). The slope (b) of the regression is;

$$b_{Ax|Ay} = \mathrm{cov}_A / \sigma^2_{Ax}$$
$$= [r_A\, \sigma_{Ax}\, \sigma_{Ay}] / \sigma^2_{Ax} \qquad \text{(remember } \mathrm{cov}_A = r_A\, \sigma_{Ax}\, \sigma_{Ay}\text{)}$$
$$= r_A\ \sigma_{Ay} / \sigma_{Ax}$$

The regression of breeding values of Y on breeding values of X;

$$Y = b_{Ax|Ay}\, X$$
$$Y = r_A\ \sigma_{Ax}/\sigma_{Ay}\, X$$

$$CR_y = b\, R_x$$
$$= b\, [i_x\, h_x\, \sigma_{Ax}]$$
$$= (r_A\ \sigma_{Ay} / \sigma_{Ax})\, [i_x\, h_x\, \sigma_{Ax}]$$
$$= \mathbf{i_x\, h_x\ r_A\, \sigma_{Ay}} \qquad \text{(substituting } \sigma_{Ay} \text{ with } h_y\, \sigma_{Py} \text{ gives)}$$
$$= i_x\, h_y\, h_x\, r_A \sigma_{Py}$$

The $h_x h_y r_A$ is the co-heritability of trait X and Y. Where, i_x is the selection intensity.

Selection efficiency

If traits X and Y are genetically correlated, and if trait X is difficult, expensive and time consuming to measure, then we may make selection on Y to improve the mean response in X. Sometimes indirect selection of X could be more efficiently than direct selection.

$$CR_x = i_y h_y r_A \sigma_{Ax}, \quad R_x = i_x h_x \sigma_{Ax}$$

$$\begin{aligned} E &= CR_x / R_x \\ &= i_y h_y r_A \sigma_{Ax} \ / \ i_x h_x \sigma_{Ax} \quad \text{(remember that } h_x = \sigma_{Ax} / \sigma_{Px}) \\ &= i_y h_y r_A \ / i_x h_x \quad \text{(assuming } i_y = i_x) \\ &= r_A h_y / h_x \end{aligned}$$

Selection efficiency E can be greater than 1 if $h_y > h_x$ and if r_A is high.

Example: (Isik and Li 2003. Canadian J of Forest Research 33:2426-2435).Wood density of trees was measured indirectly using a drilling tool called the Resistograph. The actual of wood density was also measured. The objective of the research was to develop indirect efficient wood density assessment of trees in tree improvement programs. Researchers estimated the following genetic parameters: $h^2_{i\,(density)} = 0.61$, $h^2_{i\,(resi)} = 0.29$, $r_A = 0.95$, $h^2_{f\,(density)} = 0.81$, $h^2_{f\,(resi)} = 0.79$

Assuming the same selection intensity for two methods, calculate efficiency of indirect selection both for individual tree and for family means.
Solution:

E_{tree} = `0.95*sqrt(0.29) / sqrt(0.61) = 0.655`
E_{family} = `0.95*sqrt(0.79) / sqrt(0.81) = 0.94`

Multivariate trait selection response

Selection response for one trait is

$$\begin{aligned} R &= i h^2 \sigma_P = h^2 S \\ &= (\sigma^2_A / \sigma^2_P) S \\ &= \sigma^2_A \sigma^{2\,-1}_P S \end{aligned}$$

Where, S is the selection differential. In multi trait selection there are multi genetic and phenotypic variances. Suppose there are n traits, their selection differentials S would be a vector. For two characters;

$$\mathbf{S} = \begin{bmatrix} S_1 \\ S_2 \end{bmatrix},$$

$$\mathbf{G} = \begin{bmatrix} \sigma^2_{Ax} & cov_{Axy} \\ cov_{Axy} & \sigma^2_{Ay} \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \sigma^2_{Px} & cov_{Pxy} \\ cov_{Pxy} & \sigma^2_{Py} \end{bmatrix}$$

Where **G** is the additive genetic variance-covariance matrix of two traits, **P** is the phenotypic variance and covariance matrix. The response of selection for multi traits becomes;

$$\mathbf{R} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S}$$

Decomposition of mean cross-products using ANOVA approach

Just like decomposition of observed phenotypic variance into causal components we can do the same for phenotypic covariance. This time we use cross-products [$CP = (x_i - \bar{x})(y_i - \bar{y})$] between two traits. An example from half-sib families, single site, single-tree plots.

Source	df	MCP	Exp. Mean Cross-Products
Rep	b-1	MCP_b	-
Fam	f-1	MCP_f	$\sigma_e + bn\,\sigma_{fxy}$
Error	fb(n-1)	MCP_e	σ_e

Genetic model:

Component	σ_A	σ_D	σ_{AA}	σ_{AD}	σ_{DD}	σ_E
σ_{fxy}	1/4	0	1/16	0	0	0
σ_{wxy}	3/4	1	15/16	1	1	1

Covariance explained by family effect is 0.25 of the additive genetic covariance.

$\sigma_{fxy} = (MCP_f - MCP_e) / nb$

$\sigma_{fxy} = 1/4\sigma_A + 1/16\sigma_A ...$

Genetic correlation:

$r_A = \sigma_{fxy} / [\sigma^2_{fx}\sigma^2_{fy}]^{1/2}$

1- The Variance of MS:

$Var(MS_f) = 2MS^2_f / (df_f+2)$

2- The Variance of MCP:

$Var(MCP_f) = [(MS_{fx},MS_{fy}) + MCP^2_{fxy}] / (df_f+2)$

3- The Covariance of MS_x and MS_y:

$Cov(MS_{fx},MS_{fy}) = (2MCP^2_{gxy}) / (df_f+2)$

4- The Covariances of MS and MCP:

$Cov(MS_{fx},MCP_{fxy}) = (2MS_{fx}*MCP_{fxy}) / (df_f+2)$

$Cov(MS_{fy},MCP_{fxy}) = (2MS_{fy}*MCP_{fxy}) / (df_f+2)$

The general formula for the variance of r_A

$$Var(r_A) = (r_A)^2 [var(\sigma^2_x) / 4(\sigma^2_x)^2 + var(\sigma_{xy})/(\sigma_{xy})^2 + var(\sigma^2_y)/4(\sigma^2_y)^2 + 2cov(\sigma^2_x\sigma^2_y)/4\sigma^2_x\sigma^2_y - 2cov(\sigma^2_x\sigma_{xy})/2\sigma^2_x\sigma_{xy} - 2cov(\sigma^2_y\sigma_{xy})/2\sigma^2_y\sigma_{xy}]$$

$SE(r_A)= [var(r_A)]^{1/2}$

An example from factorial mating design:

Linear Model: $Y_{ijlk} = mu + R_l + Fi + M_j + FM_{ij} + e_{ijlk}$

Source	df	MCP	Expected MCP	Estimate
Rep	r-1	MCP_r	-	-
F	f-1	MCP_f	$\sigma_e + nb\ \sigma_{fm} + bnm\ \sigma_f$	$\sigma_f= (MCP_f - MCP_{fm}) / bnm$
M	m-1	MCP_m	$\sigma_e+ nb\ \sigma_{fm} + bnf\ \sigma_m$	$\sigma_m= (MCP_m- MCP_{fm}) / bnf$
FM	(f-1)(m-1)	MCP_{fm}	$\sigma_e + nb\ \sigma_{fm}$	$\sigma_{fm}= (MCP_{fm}- MCP_e) / bn$
Err	subt.	MCP_e	σ_e	
TOT	rfm(n-1)			

Genetic model:

Components	σ_A	σ_D	σ_{AA}	σ_{AD}	σ_{DD}	σ_E
σ_f	1/4	0	1/16	0	0	0
σ_m	1/4	0	1/16	0	0	0
σ_{fm}	0	1/4	1/8	1/8	1/16	0
σ_w	1/2	3/4	3/4	7/8	15/16	1

Genetic correlations:

Female $r_A = \sigma_{fxy} / [\sigma^2_{fx}\sigma^2_{fy}]^{1/2}$

Male $r_A = \sigma_{mxy} / [\sigma^2_{mx}\sigma^2_{my}]^{1/2}$

Combined Female and Male:

$r_A = (\sigma_{fxy}+\sigma_{mxy}) / [(\sigma^2_{fx}+\sigma^2_{my})(\sigma^2_{fx}+\sigma^2_{my})]^{1/2}$

Dominance genetic correlation:

$r_D = \sigma_{fmxy} / [\sigma^2_{fmx}\sigma^2_{fmy}]^{1/2}$

Calculation of Genetic Correlations Using SAS Procedures

(1) Creating a dummy variable and running univariate models

Estimation of genetic variances for traits X and Y is straightforward using univariate models. To obtain genetic covariance between two traits we can create a dummy variable. Remember that the variance of a summation of two traits is equal to their variances, plus 2 times of covariance:

$$var(x+y) = var_x + var_y + 2cov_{xy}$$

To obtain 2 cov_{xy}, we need to create a dummy variables as Z = X+Y . Run a univariate model for dummy Z and obtain family variance. Using genetic variances from univariate analysis of X, Y and Z, we can easily derive cov_{xy}.

$\sigma_{xy} = [\sigma^2_{fz} - (\sigma^2_{fx} + \sigma^2_{fy})] / 2$

```
/* Creating a dummy variable */

data C;
  set VP (where=(test=9801));
```

```
   HTVAL=Height4+VALUE ; * This is our dummy variable ;
run;


/* Using VARCOMP procedure to obtain family variances for x, y and dummy */

proc varcomp data=c method=reml ;
      class rep family;
      model ht value htval= rep family/fixed=1;
run;
```

Partial OUTPUT
Trait: Height4

Variance Component	Estimate
Var(rep)	263.024
Var(family)	**136.7**
Var(Error)	1182.0

Asymptotic Covariance Matrix of Estimates

	Var(rep)	Var(family)	Var(Error)
Var(rep)	4517.3	6.69499	-13.22934
Var(family)	6.69499	566.46176	-36.61546
Var(Error)	-13.22934	-36.61546	932.59003

Trait: VALUE

Variance Component	Estimate
Var(rep)	6.23151
Var(family)	**1.697**
Var(Error)	22.04402

Asymptotic Covariance Matrix of Estimates

	Var(rep)	Var(family)	Var(Error)
Var(rep)	2.47752	0.0032394	-0.0049609
Var(family)	0.0032394	0.11056	-0.01262
Var(Error)	-0.0049609	-0.01262	0.32819

Trait: HTVAL

Variance Component	Estimate

Var(rep)	331.25648
Var(family)	153.1
Var(Error)	1421.7

Asymptotic Covariance Matrix of Estimates

	Var(rep)	Var(family)	Var(Error)
Var(rep)	7143.3	14.16394	-20.99875
Var(family)	14.16394	746.33177	-55.60797
Var(Error)	-20.99875	-55.60797	1371.4

Genetic covariance between HT and VALUE:

cov_f =[153.1- (1.697+136.7)] / 2 = 7.35

Genetic correlation between HT and VALUE:

r_A= 7.35 / $(1.697*136.7)^{1/2}$ =0.48

This method may give r_A outside of theoretical values, because it involves genetic variances that are estimates and are associated with standard errors.

(2) Estimation of r_A with MANOVA option in GLM

```
proc glm data=A ;
      class rep family;
      model height4 value = rep family;
       random family ;
      manova h=family / printh printe;
run;
```

The program by default produces univariate analysis of variance for each trait.

The GLM Procedure

Dependent Variable: HT

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	153	1258886.873	8228.019	7.06	<.0001

Error	2948	3434099.999	1164.891

R-Square	Coeff Var	Root MSE	ht Mean
0.268249	18.57571	34.13051	183.7373

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rep	34	755153.4671	22210.3961	19.07	<.0001
family	119	523218.7528	4396.7962	3.77	<.0001

Dependent Variable: **value**

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	153	26720.40131	174.64315	8.17	<.0001
Error	2948	62997.98228	21.36974		

R-Square	Coeff Var	Root MSE	value Mean
0.297825	27.77316	4.622741	16.64464

Source	DF	Type III SS	Mean Square	F Value	Pr > F
rep	34	19445.94572	571.93958	26.76	<.0001
family	119	7548.17003	63.43000	2.97	<.0001

Source	Type III Expected Mean Square
rep	Var(Error) + Q(rep)
family	Var(Error) + 25.761 Var(family)

The PRINTE option in the MANOVA statement displays the elements of the error matrix, also called the Error Sums of Squares and Cross-products matrix. The diagonal elements of this matrix are the error SS from the corresponding univariate analyses.

E = Error SSCP Matrix

	height4	value
height4	3434099.9988	346578.40021
value	346578.40021	62997.982279

Partial Correlation Coefficients from the Error SSCP Matrix / Prob > |r|

DF = 2948	height4	value
height4	1.000000	0.745129
		<.0001
value	0.745129	1.000000
	<.0001	

The PRINTE option also displays the partial correlation matrix associated with the E matrix. In this example, HT and the VALUE are highly correlated.

The PRINTH option produces the SSCP matrix for the hypotheses being tested (family). The diagonal elements of this matrix are the model sums of squares from the corresponding univariate analyses.

H = Type III SSCP Matrix for family

	ht	value
ht	523218.75	52648.56
value	52648.5	7548.17

Using the above outputs we can easily derive genetic covariance between two traits and their genetic variances. Family=120, Rep=35.

Source	df	SSCP	MCP	Expected MCP	Estimate
Rep	r-1	----	---		----
Family	119	52648	442	$\sigma_e + bn\ \sigma_f$	12.5
Error	2948	346578	118	σ_e	

The MCP of Family:

MCP_f =SSCP / df =52648/119 = 442

The variance of family MCP:

$Var(MCP_f) = [(MS_{fx}*MS_{fy}) + MCP_{fxy}* MCP_{fxy}] / (DFf+2)$

$Var(MSP_f) = [(4397*63) + (442*442)] / (119+2) = 278625$

SE of Family $MCP_f = (278625)^{1/2} = 527$

Genetic covariance between HT and VALUE: $cov_f = (442-118) / 26 = 12.5$

Genetic variance for HT: $\sigma^2_{fHT} = (4397-1165) / 26 = 124$

Genetic variance for VALUE: $\sigma^2_{fVALUE} = (63-21) / 26 = 1.615$

Genetic Correlation Between HT and VALUE: $rA = 12.5/(124*1.615)^{1/2} = 0.88$

(3) Multivariate model with SAS Proc Mixed

```
/* Multivariate Model for Single site */
ods listing exclude solutionf solutionr ;
ods html exclude solutionf solutionr ;

 proc mixed data=A covtest asycov scoring=1;
   Class trait rep family tree;
   model y =trait rep ;
     random trait /type=toep(1) sub=family g gcorr;
     repeated /type=toep(1) sub=family*tree;
 ods output covparms=_varcomp asycov=_cov ;
run;
```

A partial output from analysis:

Estimated R Matrix for family 1-1518

Row	Col1	Col2
1	0.7243	0.5410
2	0.5410	0.7449

Estimated R Correlation Matrix for family 1-1518

Row	Col1	Col2
1	1.0000	0.7365

2	0.7365	1.0000

Estimated G Matrix

Row	Effect	family	trait	Col1	Col2
1	trait	1-1518	1	0.07720	0.05904
2	trait	1-1518	2	0.05904	0.05781

Estimated G Correlation Matrix

Row	Effect	family	trait	Col1	Col2
1	trait	1-1518	1	1.0000	0.8838
2	trait	1-1518	2	0.8838	1.0000

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	family	0.07720	0.01375	5.62	<.0001
UN(2,1)	family	0.05904	0.01144	5.16	<.0001
UN(2,2)	family	0.05781	0.01119	5.17	<.0001

Example for half-sib families:

SAS Proc Mixed can be used to estimate genetic correlations among traits and their standard errors. Let's say we have 120 half-sib families of loblolly pine and they were tested at two locations using RCB design single-tree plots. Two traits were measured e.g., var1 var2. We want to estimate genetic correlation between var1 and var2. The original data are as follows;

```
data B;
input location $ rep family $ tree var1 var2 ;
datalines ;
AL 1 F1 1 128 214
AL 1 F1 2 107 165
AL 1 F1 3 187 162
. . . . .
;
```

For bivariate/multivariate analysis, a special data structure is needed. First, variables for each individual tree should be put into one column and must be named as Y . Second, a dummy variable named TRAIT needs to be created.

```
data A;
```

```
set B(rename=(var1=y))
b(rename=(var2=y));
if var1=. then trait=1 ;
if var2=. then trait=2 ;
drop var1 var2;
run;
```

When a portion of the data is printed using the Proc PRINT of SAS, the output will be as follows:

```
title 'Arranged data';
proc print data=A (obs=5);
run;
```

Partial arranged data output

Obs	site	block	family	tree	height	fork	color	qual	fork4	y	trait
1	9801	1	1	34	202	0	2	3	0	202.00	1
2	9801	10	1	70	140	0	2	2	0	140.00	1
3	9801	11	1	27	164	1	2	3	1	164.00	1
4	9801	12	1	65	166	0	2	2	0	166.00	1
5	9801	13	1	2	198	1	2	2	1	198.00	1
6	9801	14	1	20	152	1	2	2	1	152.00	1

Each tree now has two observations rather than one, e.g. tree 34 of family 1-1518 has 4.91 and 3.06. For simplicity, I included 2 traits; however, you may include 3 or more traits.

The following code fits a mixed model to data.

```
proc mixed data=A covtest asycov scoring=1;
   Class trait site block family tree;
   model y = trait site block(site) ;
     random trait /type=un subject=family g gcorr;
     repeated /type=un  subject=family*site*tree r rcorr;
ods output covparms=_varcomp asycov=_cov ;
run;
```

1. The fixed effect class variables [TRAIT, SITE, BLOCK(SITE)] are listed after the dependent variable Y. We want to model different means for the multivariate observations, hence the inclusion of TRAIT in the MODEL statement.
2. The variables TRAIT (which is a dummy), FAMILY and FAMILY*SITE*TREE are random.

3. The G option requests genetic covariance-variance matrix and the GCORR requests genetic correlation matrix between two traits.
4. The REPEATED statement here is used to define error (**R**) variance-covariance structure (Var(**y**)=**ZGZ**^{T}+**R**). TYPE=UN specifies that the error variance and covariance matrix is unstructured. We use the SUBJECT= option to instruct the procedure that which sets of observations from this subject are correlated. All observations having the same level of the GROUP effect have the same covariance parameters. The TYPE= options defines the relationships between these selected set of observations. In other application, such heterogeneous models, the REPEATED statement allows fitting heterogeneous variance models using the GROUP option. For example, if there is a large difference between two sites, then, instead of using transformations, we can fit a heterogeneous variance model to the data as REPEATED / GROUP=site. Here the repeated statement replaces the RANDOM statement.
5. The R matrix requests the first block diagonal matrix and RCORR requests within family correlation matrix. The **r** requests that blocks of the estimated R matrix be displayed. The first block determined by the SUBJECT= effect is the default displayed block. RCORR=value-list produces the correlation matrix corresponding to blocks of the estimated R matrix. Complete independence is assumed across subjects; therefore, the SUBJECT= option produces a block-diagonal structure in R with identical blocks. The TYPE=UN (unstructured) option is useful for correlated random coefficient models.
6. Finally, two output files (variance components, covariance matrix) were requested using the ODS statements of SAS. Here, COVPARMS is the SAS table name of the variance components; _VARCOMP is a given name (you may give a different name). ASYCOV is the variance-covariance matrix of variance components.
7. The TYPE=UN option requests an unstructured covariance matrix for each SUBJECT= FAMILY. This structure does not make any assumptions of covariances. The covariance matrix for three traits would be as follows:

```
Cov = { 11   12   13
        12   22   23
        13   23   33};
```

Depending on the data, you may fit different covariance structures, such as compound symmetry (CS) or autoregressive (AR) etc.

A partial output of the Mixed Procedure is given below:

The Mixed Procedure

Model Information

Data Set	WORK.A
Dependent Variable	y
Covariance Structure	Unstructured
Subject Effects	family, site*family*tree
Estimation Method	REML

Residual Variance Method	None
Fixed Effects SE Method	Model-Based
Degrees of Freedom Method	Containment

Class Level Information

Class	Levels	Values
trait	2	1 2
site	4	9801 9802 9803 9804
block	35	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 26 27 28 29 30 31 32 33 34 35
family	120	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 26 27 28 29 30 31 32 33...

Dimensions

Covariance Parameters	6
Columns in X	143
Columns in Z Per Subject	2
Subjects	120
Max Obs Per Subject	218

Number of Observations

Number of Observations Read	11787
Number of Observations Used	11778
Number of Observations Not Used	9

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	113331.69046094	
1	2	112661.35777224	0.00000033
2	1	112661.34411503	0.00000001

3	1	112661.34347914	0.00000000

Convergence criteria met.

Estimated G Matrix

Row	Effect	trait	family	Col1	Col2
1	trait	1	1	71.7416	97.0170
2	trait	2	1	97.0170	657.33

Estimated G Correlation Matrix

Row	Effect	trait	family	Col1	Col2
1	trait	1	1	1.0000	0.4468
2	trait	2	1	0.4468	1.0000

The elements of the G MATRIX are genetic (FAMILY) variances and covariance of two variables. By requesting GCORR option after the RANDOM statement, the code produces genetic correlation between two traits (0.4468). We can calculate genetic correlation using the G MATRIX components as r_A = 97.01 / sqrt(71.74*657.33).

Covariance Parameter Estimates

Cov Parm	Subject	Estimate	Standard Error	Z Value	Pr Z
UN(1,1)	family	71.7416	10.5217	6.82	<.0001
UN(2,1)	family	97.0170	39.8345	2.44	0.0149
UN(2,2)	family	657.33	232.09	2.83	0.0023
UN(1,1)	site*family*tree	864.72	11.4421	75.57	<.0001
UN(2,1)	site*family*tree	-138.18	144.20	-0.96	0.3379
UN(2,2)	site*family*tree	714.03	210.06	3.40	0.0003

The above table is covariance parameters.
71.7 =1/4 additive genetic variance for var1,
97.0 =1/4 additive genetic covariance between var1 and var2,
657.3 =1/4 additive genetic variance for var2
The approximate standard errors of genetic variances and the covariance were given after the estimates, and the Z values and Probability test statistics.

Asymptotic Covariance Matrix of Estimates

Row	Cov Parm	CovP1	CovP2	CovP3	CovP4	CovP5	CovP6
1	UN(1,1)	110.71	132.47	158.97	-1.3724	-13.3739	-17.3885
2	UN(2,1)	132.47	1586.79	3219.00	1.7286	-20.0467	-37.3361
3	UN(2,2)	158.97	3219.00	53867	-117.18	35.7764	30.7322
4	UN(1,1)	-1.3724	1.7286	-117.18	130.92	-19.6660	1.6981
5	UN(2,1)	-13.3739	-20.0467	35.7764	-19.6660	20793	-6095.12
6	UN(2,2)	-17.3885	-37.3361	30.7322	1.6981	-6095.12	44124

Variances of variance components and variance of genetic covariance is given in the table of Asymptotic Covariance Matrix above. The diagonal elements are the variances of variance for trait1, variance of the genetic covariance, and variance of variance for trait 2. The off diagonal elements are covariances between variances and the covariance. For example `132.47` is the covariance between the genetic variance for trait 1 and the genetic covariance.

Using the Covariance Parameters and Asymptotic Covariance Matrix Tables we can easily calculate standard error of genetic correlation. The method used here is the DELTA, which is simply taking the variances of functions (see Lynch and Walsh 1998 for details).

```
/* You do NOT need to change the following lines, except, If the name of
the genotype in your data is not 'family' then change it accordingly */

/* Start IML */

    proc iml ;

/* Go to output file '_varcomp', and create a 3-row vector for 'family'
group only. The 1st estimate (row) is genetic variance of Trait1, the 2nd
is the genetic Covariance, and the 3rd is genetic variance of Trait2 */

    use _varcomp;
    read all var {Estimate} where(Subject="family") into _varcomp;
    close _varcomp;

/* Create Asymptotic Covariance Matrix for 'family' group, select the first
3x3 block matrix of variances and covariances between estimates */

     use _cov;
     read all var {CovP1 CovP2 CovP3} into _cov;
```

```
     close _cov;

/* Genetic correlation */

     r = _varcomp[2,1]/sqrt(_varcomp[1,1]*_varcomp[3,1]);

/* Standard error of genetic correlation */

     a =_cov[1,1]/(4*(_varcomp[1,1])**2) ;
     ab=_cov[2,2]/((_varcomp[2,1])**2) ;
     b =_cov[3,3]/(4*(_varcomp[3,1])**2) ;

     c1=(2*_cov[1,3])/ (4*(_varcomp[1,1]*_varcomp[3,1]));
     c2=(2*_cov[1,2])/ (2*(_varcomp[1,1]*_varcomp[2,1]));
     c3=(2*_cov[2,3])/ (2*(_varcomp[2,1]*_varcomp[3,1]));

** Variance of genetic correlation;
     var_r=(r*r)*(a+b+ab+c1-c2-c3);

** Standard error of genetic correlation
     SE_r =sqrt(var_r) ;;

     print r var_r SE_r ;

     run; quit;
```

OUTPUT of the IML code

R	VAR_R	SE_R
0.446755	0.0274051	0.1655449

Calculation of Genetic Correlations Using ASReml

The syntax for specifying a multivariate linear model in ASReml is

Y-variates ~ fixed [!r random] [!f sparse fixed]

Y-variates is a list of traits,

Fixed, random and sparse fixed are as in the univariate case (see Chapter 6) but involve the special term Trait and interactions with Trait.
The design matrix for Trait has a level (column) for each trait.

- Trait by itself fits the mean for each variate,
- In an interaction Trait.Fac fits the factor Fac for each variate and Trait.Cov fits the covariate Cov for each variate.
-

EXAMPLE: Half-sib family, 4 sites, single-tree plots
(C:\RESEARCH\Handbook\Chapter 4 - Half-sibs)

Command file:

```
------------------------------------------------------------------------
Half-sib family, RCBD with single-tree plots, multi locations
 site        4 !I
 block      35
 family    120
 tree
 height
 fork
 color
 qual
 value
 fork4
 height4
C:\RESEARCH\Handbook\Data\HS.csv  !skip 1
# Multivariate analysis to estimate genetic correlations
value height4 ~ Trait Trait.site Trait.site.block !r Trait.family
Trait.site.family

1 2 2
11787 0 ID
Trait 0 US
3*0

Trait.family 2
Trait 0 CORR
0.708
0.400 70.5 !GP
family 0 ID

Trait.site.family 2
Trait 0 US
3*0
site.family 0 ID
------------------------------------------------------------------------
```

Specifying multivariate variance structures in ASReml

A more sophisticated error structure is required for multivariate analysis.
In case of bivariate model, Trait is used instead of mu. **Trait** is the multivariate version of mu. It creates a vector holding the overall means for each trait included in the analysis.

In order to estimate covariances in the bivariate model, you need to supply starting variances and covariances.
ERROR header line (1 2 2)
1 = there is ONE independent error structure (# of environments)
2 = the error structure is the product of TWO matrices (it is generally 2), the error structure for the residual must be specified as two-dimensional with independent records and an unstructured variance matrix across traits
2 = is number of G header line (number of random effects). In our case, there are TWO covariance structures to define, FAMILY and SITE.FAMILY
# ERROR structure line

- the R structure definition line for units, that is, 11787 0 ID, could be replaced by 0 or 0 0 ID; this tells ASReml to fill in the number of units and is a useful option when the exact number of units in the data is not known to the user. It tells ASReml to count # of observations to design ID matrix
- The error variance matrix is specified by the model Trait 0 US. The 0 between Trait and US is a place holder that we will use when dealing with spatial analysis. US is unstructured variance
- 3*0 to create initial values. The initial values are for the lower triangle of the (symmetric) matrix specified row-wise. Finding reasonable initial values can be a problem. If initial values are written on the next line in the form q * 0 where q is $t(t+1)=2$ and t is the number of traits, ASReml will take half of the phenotypic variance matrix of the data as an initial value

Family (G Matrix structure line)

- The second structure is FAMILY 2. The first matrix has a 2 x 2 dimension as determined by Trait. The matrix is a correlation matrix (CORR). You can use a US matrix as well, but correlation matrices are easier to run.
- 0.708 0.400 70.5 are starting variances and the CORRELATION for trait1 and trait2. The correlation (0.400) is just a guess
- !GP attempts to keep the parameter in theoretical limit

Partial output

```
Source                  Model  terms     Gamma     Component    Comp/SE   % C
 Residual         UnStructured  1  1   16.0845       16.0845      74.42   0 U
 Residual         UnStructured  2  1   68.6547       68.6547      53.02   0 U
 Residual         UnStructured  2  2   865.299       865.299      74.80   0 U
 Trait.family       UnStructured 1 1  0.716301      0.716301       5.69   0 U
 Trait.family       UnStructured 2 1   4.26258       4.26258       4.45   0 U
 Trait.family       UnStructured 2 2   70.4063       70.4063       6.55   0 U
 Trait.site.family UnStructured 1 1  0.287621      0.287621       4.01   0 U
 Trait.site.family UnStructured 2 1   1.31084       1.31084       3.13   0 U
 Trait.site.family UnStructured 2 2   11.6331       11.6331       3.29   0 U
 Covariance/Variance/Correlation Matrix UnStructured Residual
   16.08      0.5819
   68.65       865.3
 Covariance/Variance/Correlation Matrix UnStructured Trait.family
  0.7163      0.6002
   4.263       70.41
 Covariance/Variance/Correlation Matrix UnStructured Trait.site.family
  0.2876      0.7166
   1.311       11.63
```

More info:
ASReml Cookbook. Luis Apiolaza, http://uncronopio.org/ASReml/HomePage
ASReml Manual, Chapter 8, Multivariate analysis