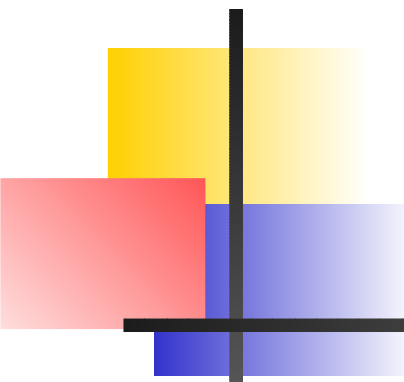
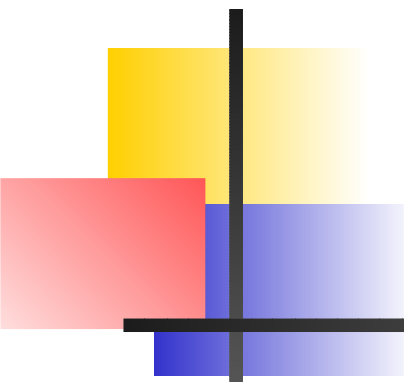




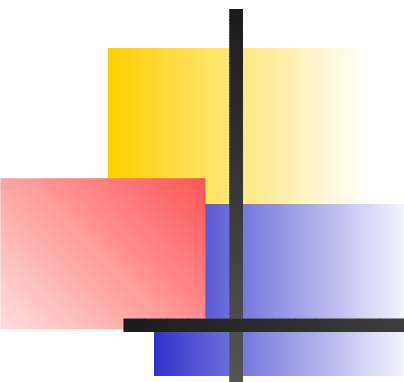
Exercises

- Coopworth data set - see Reference manual
- Five traits with varying amounts of data.
- No depth of pedigree (dams not linked to sires)
- Do univariate analyses
- Do bivariate analyses.
- Use COOP data set and attempt multivariate models.



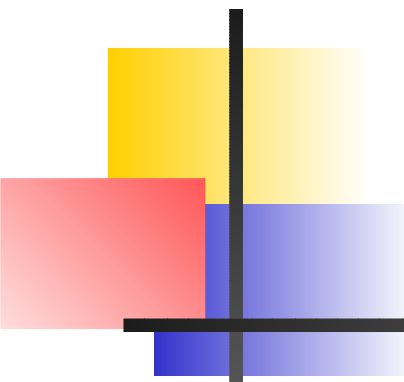
Multivariate analysis

- Modelling becomes more difficult as number of traits increases - there may be variance but the matrix may not be positive definite (covariances too big).
 - Maybe do bivariate pairs first
 - ! GP constrain positive definite - will use EM updates if AI updates generate NPD matrix
 - try FA or CHOL reduced parameterization
 - try Singular XFA or CHOL parameterization



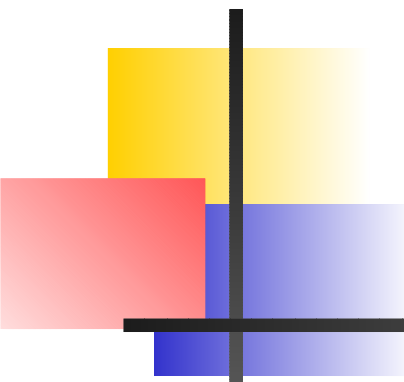
Usual process

- Univariate analysis of each trait.
 - Identify an appropriate fixed model
 - Check for outliers and problems with data structure
 - Sort out fixed model appropriate for each trait
 - Ensure there is variance at each level: if there is no sire variance in a univariate model, ASReml will not be able to estimate it in a multivariate model



Output: ANOVA

■	ANOVA	NumDF	DenDF	F-incr	Pro
9	Trait	2	4.9	5019.79	<.00
10	Tr.line	4	5.3	2.65	0.15
12	at(Tr,2).Age	1	58.5	6.33	0.01



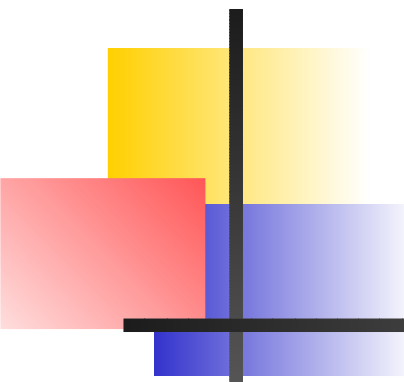
Output:structures

- Covar/Var/Corr UnStructured

132.4	-0.3319
-98.03	659.0

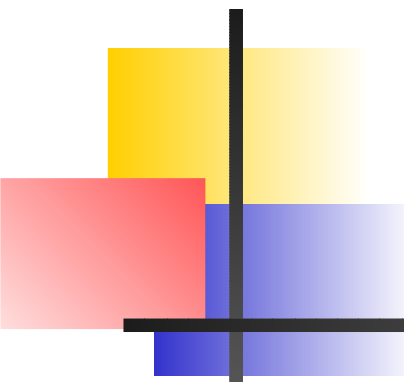
Covar/Var/Corr UnStructured

27.20	-0.2354
-12.20	98.72



Output:components

■ Source	Model	terms	Gamma	Component	C/SE	%	C
Residual	UnStru	1 1	132.370	132.370	5.29	0	U
Residual	UnStru	2 1	-98.0288	-98.0288	-2.27	0	U
Residual	UnStru	2 2	659.005	659.005	5.19	0	U
Tr.sire	UnStru	1 1	27.2002	27.2002	1.02	0	U
Tr.sire	UnStru	2 1	-12.1986	-12.1986	-0.30	0	U
Tr.sire	UnStru	2 2	98.7165	98.7165	0.88	0	U

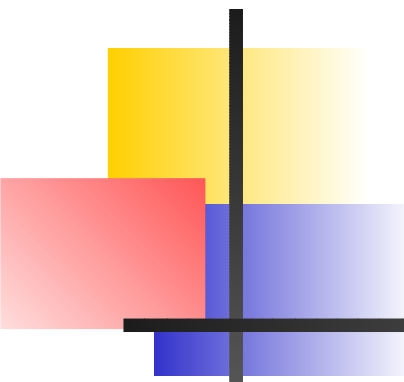


Output:convergence

- 1 LogL=-466.835 S2=1.0000 123 df
2 LogL=-451.914 S2=1.0000 123 df
3 LogL=-437.150 S2=1.0000 123 df
4 LogL=-428.395 S2=1.0000 123 df
5 LogL=-427.247 S2=1.0000 123 df
6 LogL=-427.201 S2=1.0000 123 df
7 LogL=-427.201 S2=1.0000 123 df

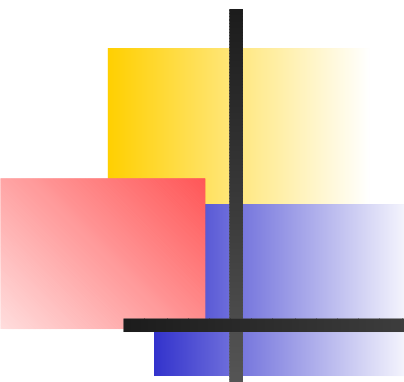
- Analysis on Variance scale

$$V = \sigma^2(\mathbf{R} + \mathbf{ZGZ}') \text{ with } \sigma^2 = 1.$$



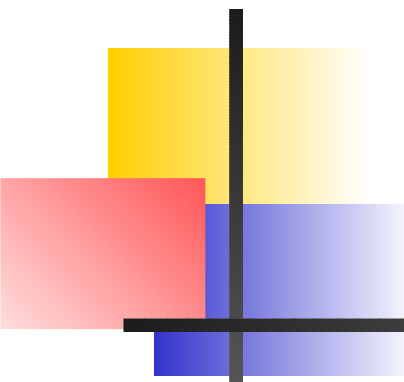
Missing data

- Multivariate analysis, with US error variance matrix, ASReml will automatically handle missing values in the traits.



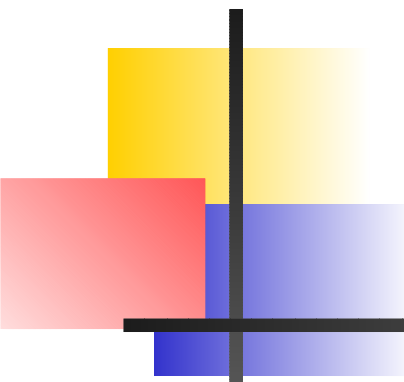
Notes

- R structure is 65×2 because data is ordered traits within records
- G structure is 2×9 because Tr . sire means the effects are ordered sires within traits
- Initial values must be supplied but are often difficult to guess. Specifying them as zeros ($3 * 0$) tells ASReml to work out some values from the data.



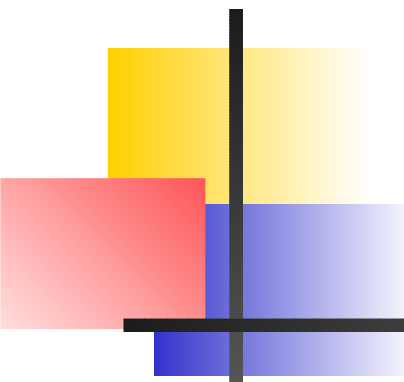
1 G Structure

```
■ ADG WT ~ Trait Tr.line,  
                                at(Tr,2).Age !r Tr.sire  
  
1 2 1  
0      # 65 records  
Trait 0 US  
3*0  
Tr.sire 2          # <======  
Tr 0 US           # <======  
3*0              # <======  
sire 0 ID         # <=====
```



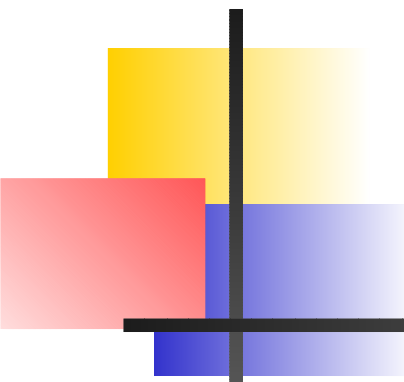
1 R structure

```
■ ADG WT ~ Trait Tr.line,  
                                at(Tr,2).Age !r Tr.sire  
  
1 2 1  
0      # 65 records # <=====  
Trait 0 US          # <=====  
3*0                # <=====  
Tr.sire 2  
Tr 0 US  
3*0  
sire 0 ID
```



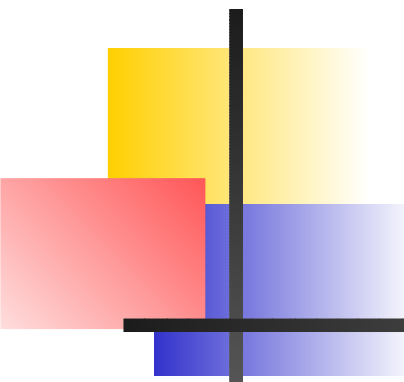
Variance header line

- ```
ADG WT ~ Trait Tr.line,
 at(Tr,2).Age !r Tr.sire
1 2 1 # <=====
0 # 65 records
Trait 0 US
3*0
Tr.sire 2
Tr 0 US
3*0
sire 0 ID
```



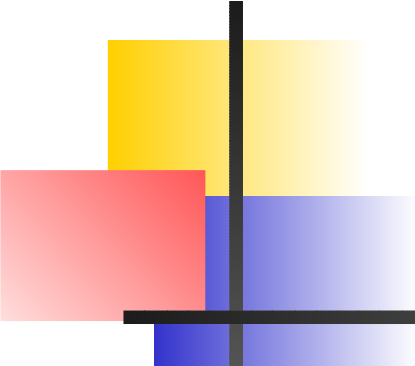
# Bivariate continued

- Now stack WT effects below ADG effects:  
 $u = \text{vec}(U)$   
 $\text{var}(u) = \Sigma_C \otimes \Sigma_R$
- Essential to get order correct: second term is nested within first.
- Adding part 3 to the Harvey job: the model  
!PART 3  
ADG WT ~ Trait Tr.line,  
          at(Tr,2).Age !r Tr.sire



# Bivariate Analysis

- With Harvey data, could analyse ADG and WT - to get two sets of sire effects and two sets of residuals.
- Consider sire effects ( $U$ ) in a  $9 \times 2$  table.  
Rows are independent:  $\Sigma_R = \mathbf{I}_9$   
Columns are correlated but have different variances:  $\Sigma_C = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}$



# ASReml tutorial

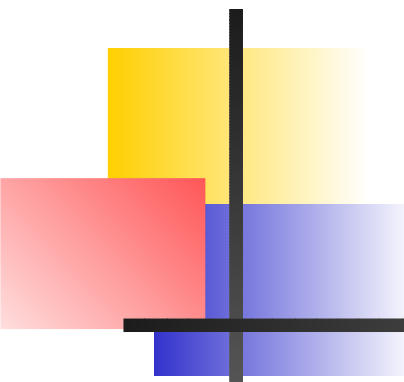
---

## B4 Bivariate Analysis

Arthur Gilmour

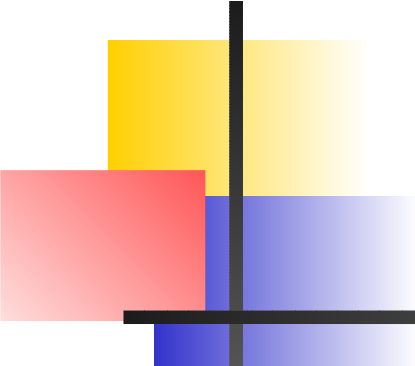


NSW DEPARTMENT OF  
**PRIMARY INDUSTRIES**



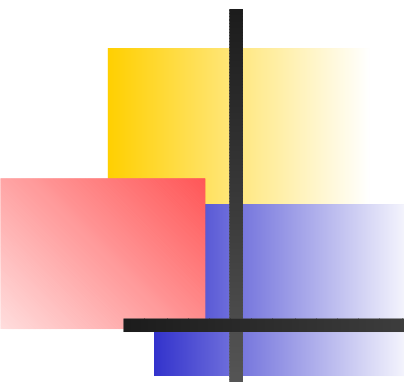
# Sire model → Animal model

- Solving the model in terms of genetic components can lead to further problems.  $\sigma_E^2 = \sigma_e^2 - 3\sigma_s^2$  may be negative and not estimable (ASReml requires a positive residual variance)
- $h^2 = 4\sigma_s^2 / (\sigma_e^2 + \sigma_s^2)$  is  $\leq 4$  but if the genetic model is correct, should be less than 1.
- I.e. Our model may not adequately represent the variation in the data leading to unacceptable genetic parameter estimates.



# Estimability of Components

- Henderson's method III equated the Sire Mean Square (SMS) to its expectation  $\sigma_e^2 + k\sigma_s^2$
- While  $\text{SMS} \geq 0$ ,  
 $\sigma_s^2 = (\text{SMS} - \text{EMS})/k$  is  $\geq -\text{EMS}/k$
- In REML, often constrain components to be positive (equivalent to estimating  $\sigma_e^2$  after dropping sire from the model).



# Other genetic components

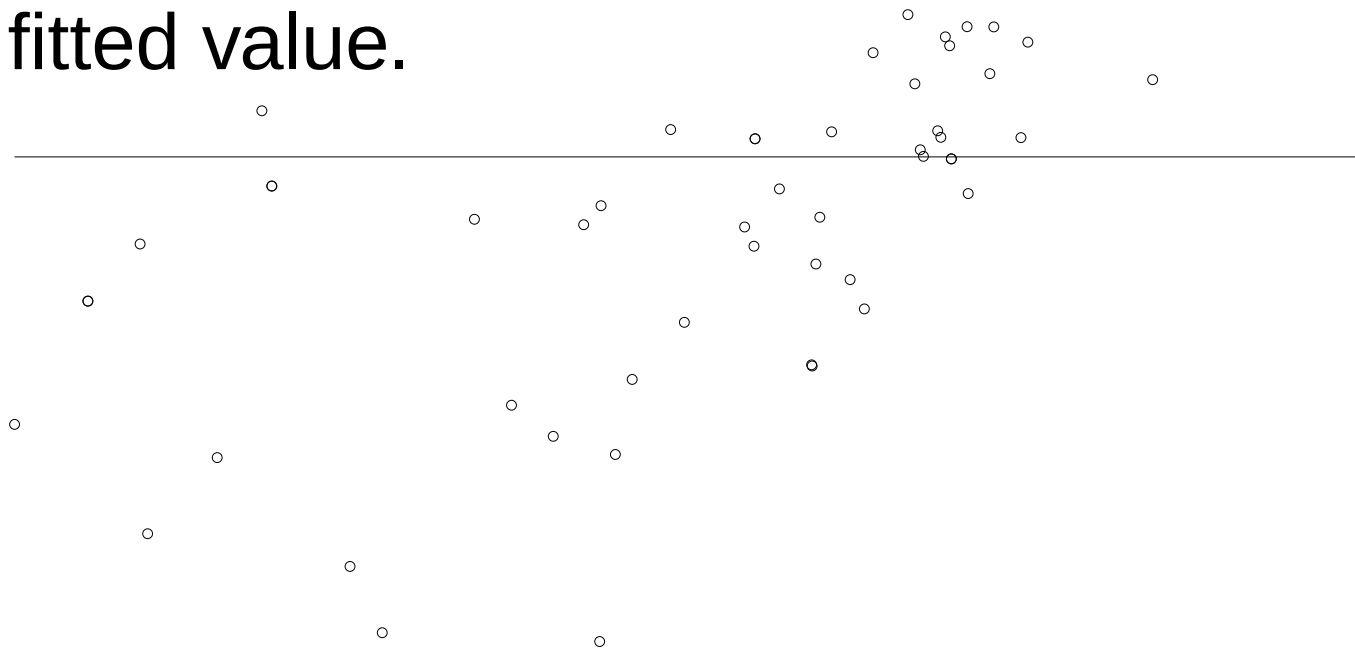
---

- Genetic maternal definition `DAM !P` and include DAM in the model
- Permanent environment effect  
Use `ide(animal)`
- Maternal environment effect  
Use `ide(DAM)`

# Plot of Residuals

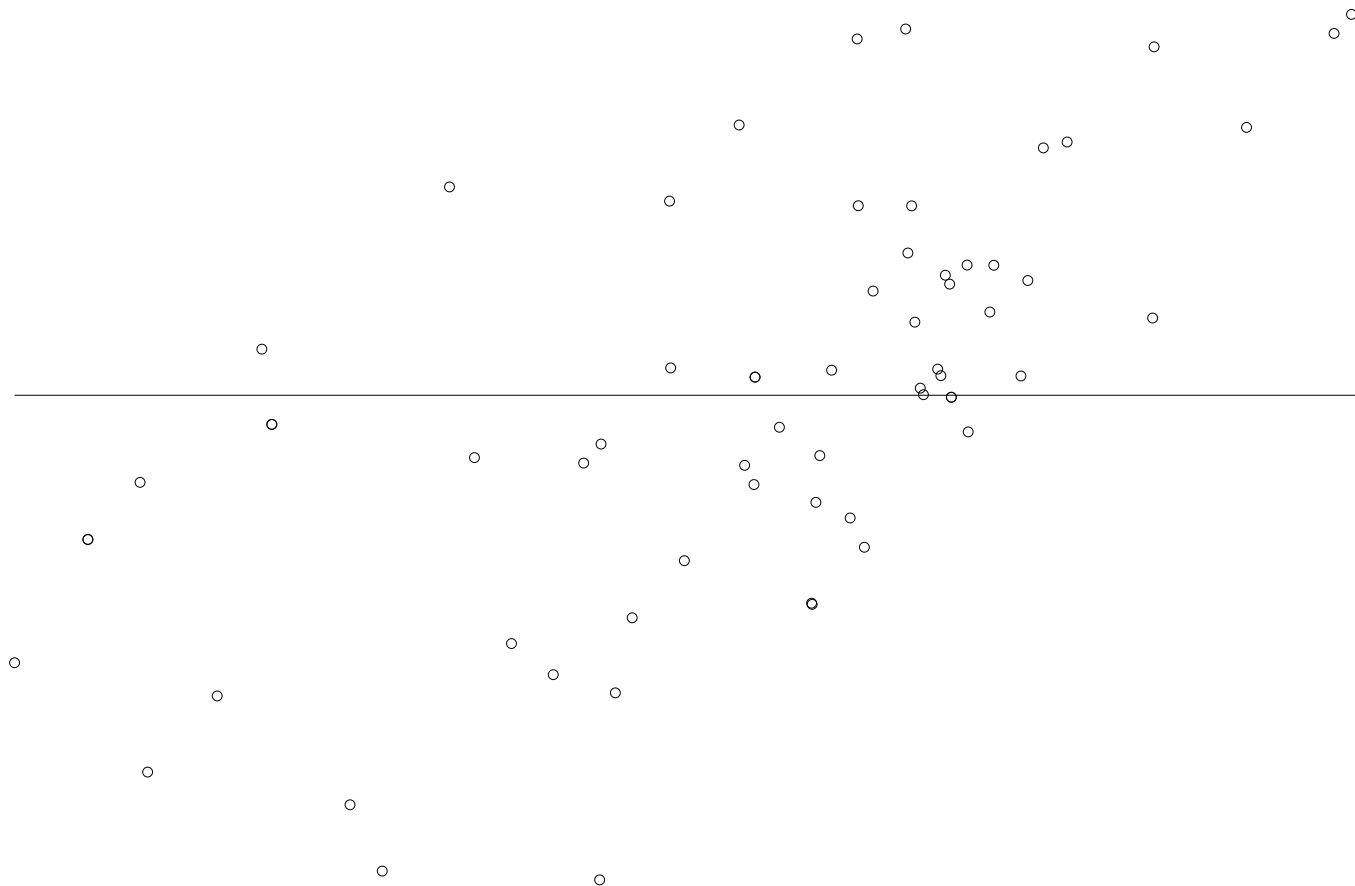
Harvey Test data - Animal Model    Residuals vs Fitted values  
Residuals (Y) -8.84: 6.95    Fitted values (X) 148.88: 199.43

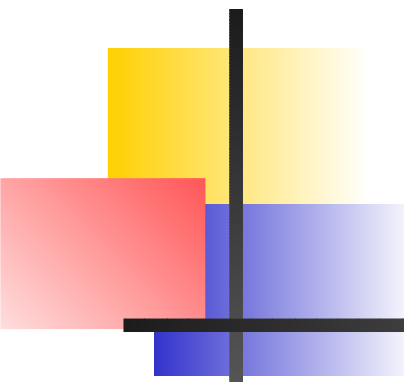
- Apparent mean variance relationship arises because the animal model adds a proportion of the residual (from the sire model) into the fitted value.



# Plot of Residuals

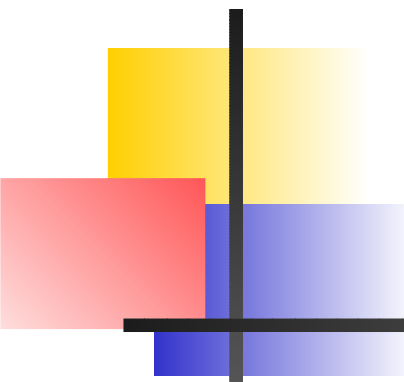
Harvey Test data - Animal Model    Residuals vs Fitted values  
Residuals (Y) -8.84: 6.95    Fitted values (X)    148.88: 199.43





# Results - Effects

|   |   |          | Estimate          | StandErr | T-val |
|---|---|----------|-------------------|----------|-------|
| ■ | 5 | DamAge 1 | -1.47752          | 1.88080  | -0.79 |
|   | 4 | line 2   | -14.4126          | 6.28603  | -2.29 |
|   |   | 3        | 6.46567           | 5.29341  | 1.22  |
|   | 9 | mu 1     | 183.489           | 9.31871  | 19.69 |
|   | 1 | animal   | 74 effects fitted |          |       |

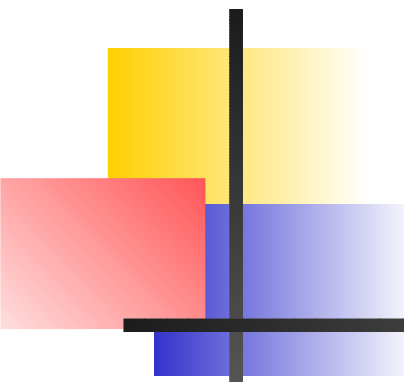


# Results - ANOVA

---

| ■ ANOVA  | NumDF | DenDF | F-incr  | Prob  |
|----------|-------|-------|---------|-------|
| 9 mu     | 1     | 5.9   | 5906.95 | <.001 |
| 4 line   | 2     | 5.9   | 6.19    | 0.035 |
| 5 DamAge | 1     | 57.8  | 0.62    | 0.435 |

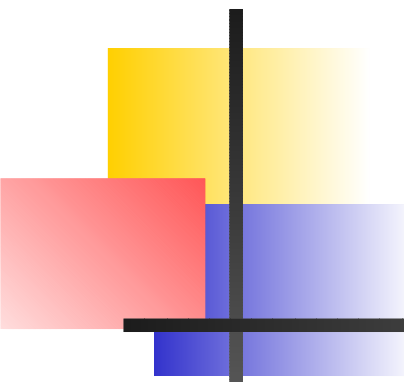
is the same as the Sire model



# Results - components

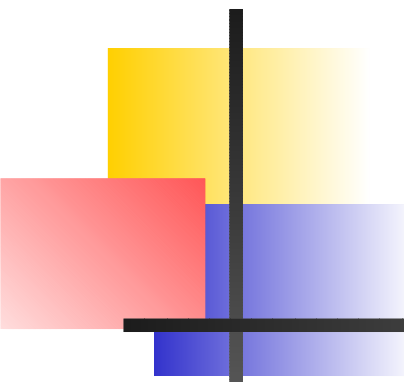
■ 5 LogL=-188.781 S2=54.092 61df 1.957  
6 LogL=-188.777 S2=47.314 61df 2.411  
7 LogL=-188.777 S2=46.100 61df 2.506  
Final parameter values 2.508

| Source   | terms | Gamma  | Component | C/SE |
|----------|-------|--------|-----------|------|
| animal   | 74 74 | 2.5086 | 115.578   | 1.04 |
| Variance | 65 61 | 1.0000 | 46.0723   | 0.51 |



# Running the model in ASReml

- !REDO !ARG !? 1=Sire 2=Animal  
Harvey Test data - Sire and Animal Model  
animal !P sire 9 dam line 3 # Added  
DamAge ADG Age WT  
harvey.dat # Pedigree file line a  
harvey.dat !DOPART \$1 # Data  
!PART 1  
ADG ~ mu line DamAge !r sire  
!PART 2  
ADG ~ mu line DamAge !r animal



# Variance model

- Fitting an animal model for the harvey data

$$\mathbf{V} = \sigma_E^2 \mathbf{I}_{65} + \sigma_A^2 \mathbf{Z} \mathbf{A} \mathbf{Z}'$$

$$\text{where } \mathbf{Z} = (\mathbf{0}_{65 \times 9} \mathbf{I}_{65})$$

Thus  $\mathbf{V}$  simplifies to

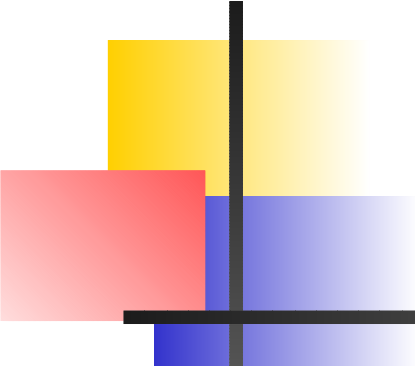
$$\sigma_E^2 \mathbf{I}_{65} + \sigma_A^2 (0.25 \mathbf{Z}_s \mathbf{Z}_s' + 0.75 \mathbf{I}_{65})$$

$= (\sigma_E^2 + 0.75 \sigma_A^2) \mathbf{I}_{65} + 0.25 \sigma_A^2 \mathbf{Z}_s \mathbf{Z}_s'$  which is exactly the same as under the sire variance model.

# reducing to

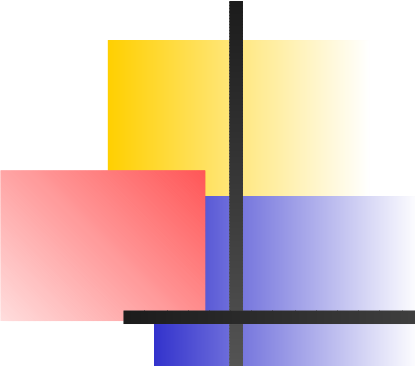
- $A_i^{-1} = \begin{pmatrix} A^{-1} + pqp' & pq \\ qp' & q \end{pmatrix}$  where  $p$  is zero except for two  $\frac{1}{2}$ 's in parental rows,  
 $q = (1 + f_i - p'Ap)^{-1} = (1 - (a_{s,s} + a_{d,d})/4)^{-1}$   
so that it just requires keeping  $\text{diag}(A)$ .
- For the case of a sire model (dams unknown, sires not inbred),

$$A = \begin{pmatrix} I_9 & 0.5Z'_s \\ 0.5Z_s & 0.25Z_sZ'_s + 0.75I_{65} \end{pmatrix}$$



# Relationship Matrix

- The diagonal of the relationship matrix ( $a_{i,i}$ ) is  $1 + f$  where  $f = a_{s,d}/2$  is the inbreeding coefficient.
- The relationship of an animal with other animals is the average of its parental values.
- When parents are listed before progeny, this gives a straight forward way to calculate relationships.
- However, the inverse relationship matrix is in fact easier to calculate and more sparse.



# ASReml tutorial

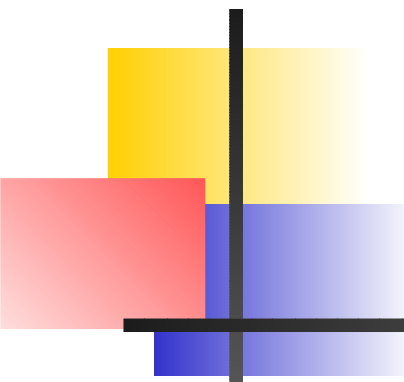
---

## B3 Animal model

Arthur Gilmour



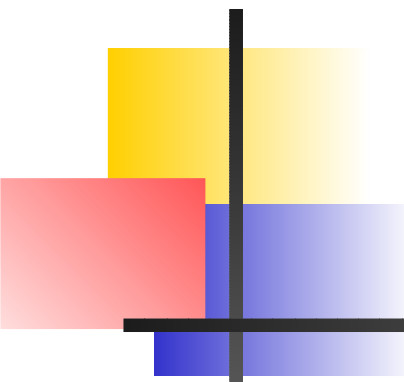
NSW DEPARTMENT OF  
**PRIMARY INDUSTRIES**



# Line+Sire

■ predict Sire !present Sire Line

|                                  |          |          |
|----------------------------------|----------|----------|
| 3.0000                           | 173.4734 | 4.1885 E |
| 4.0000                           | 157.4867 | 3.6989 E |
| 5.0000                           | 167.7099 | 3.8875 E |
| 6.0000                           | 183.5371 | 3.8512 E |
| 7.0000                           | 186.8966 | 3.4811 E |
| 8.0000                           | 183.8807 | 3.6774 E |
| 9.0000                           | 179.5918 | 3.4811 E |
| Overall Stnd Error of Diff 5.146 |          |          |

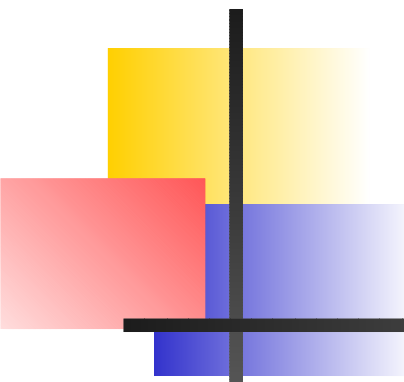


# Line+Sire

- predict Sire !present Sire Line

Line is average of combinations present  
DamAge is evaluated at 4.3846

| Sire   | Pred_Value | Stand_Error | Ecode |
|--------|------------|-------------|-------|
| 1.0000 | 176.9570   | 3.5634      | E     |
| 2.0000 | 180.6023   | 3.5962      | E     |
| 3.0000 | 173.4734   | 4.1885      | E     |
| 4.0000 | 157.4867   | 3.6989      | E     |
| 5.0000 | 167.7099   | 3.8875      | E     |
| 6.0000 | 183.5371   | 3.8512      | E     |



# Predict

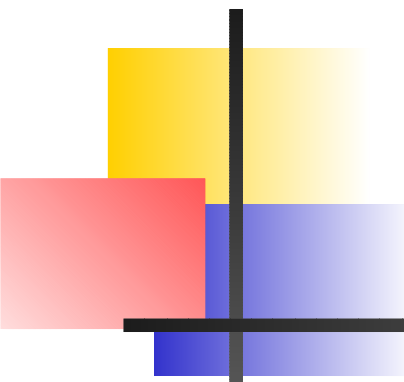
- predict Line

Predicted values of ADG

DamAge evaluated at average value 4.

Sire is ignored in the prediction

| Line    | Predicted_Value    | Stand_Error | Ec |
|---------|--------------------|-------------|----|
| 1.0000  | 177.0109           | 4.0193      | E  |
| 2.0000  | 162.5983           | 4.8301      | E  |
| 3.0000  | 183.4766           | 3.4415      | E  |
| Overall | Stnd Error of Diff | 5.851       |    |



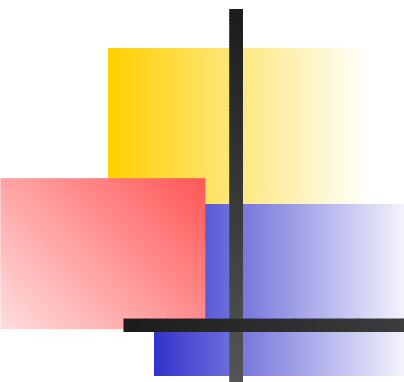
# Fixed effects

| ■ | Term   | Level | Effect | SE    |
|---|--------|-------|--------|-------|
|   | DamAge | 1     | -1.478 | 1.881 |
|   | Line   | 1     | 0.000  | 0.000 |
|   | Line   | 2     | -14.41 | 6.286 |
|   | Line   | 3     | 6.466  | 5.293 |
|   | mu     | 1     | 183.5  | 9.319 |

■ Notes:

Terms are in reverse order (ASReml solves from bottom)

Line-1 effect is singular.

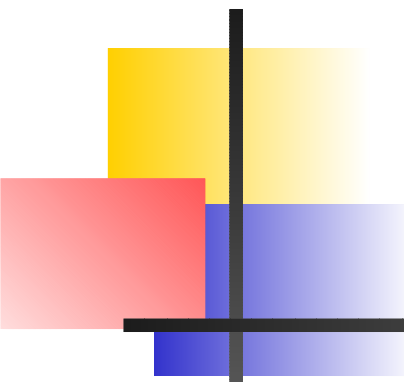


# ANOVA

- Degrees of Freedom and Stratum Variances

|          |       |         |         |       |
|----------|-------|---------|---------|-------|
|          | 5.93  | 341.034 | 7.2     | 1.0   |
|          | 55.07 | 132.756 | 0.0     | 1.0   |
| ANOVA    | NumDF | DenDF   | F-incr  | Prob  |
| 9 mu     | 1     | 5.9     | 5906.95 | <.001 |
| 4 Line   | 2     | 5.9     | 6.19    | 0.035 |
| 5 DamAge | 1     | 57.8    | 0.62    | 0.435 |

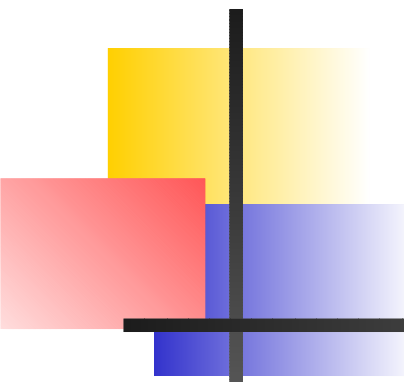
- Damage is NS; Line is significant tested against Sire variance.



# Sire BLUPS

- from the .sln file – sum to zero within lines

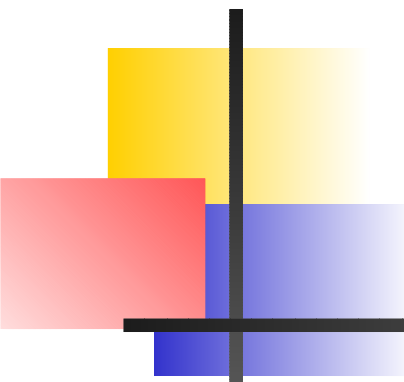
|      |   |             |       |
|------|---|-------------|-------|
| Sire | 1 | -0.5386E-01 | 4.137 |
| Sire | 2 | 3.591       | 4.142 |
| Sire | 3 | -3.538      | 4.269 |
| Sire | 4 | -5.112      | 4.468 |
| Sire | 5 | 5.112       | 4.468 |
| Sire | 6 | 0.6051E-01  | 4.055 |
| Sire | 7 | 3.420       | 3.914 |
| Sire | 8 | 0.4042      | 3.990 |
| Sire | 9 | -3.885      | 3.914 |



# Genetic parameters

■      3 PhenVar    1            161.7            35.07  
         4 GenVar   1            115.6            110.8  
Heritability = GenVar 4/PhenVar 3  
                 =            0.7150    0.5877

Notice: The parameter estimates are  
         followed by their  
         approximate standard errors



# Genetic components

- Create a PIN file (harvey.pin)

```
1 is Sire component
```

```
2 is Residual
```

```
F PhenVar 1 2 #3 is Sire + Residual
```

```
F GenVar 1*4. #4 is Sire x 4.0
```

```
H Herit 4 3 #Heritability is GenVar/Phe
```

- Run using ASReml -p harvey

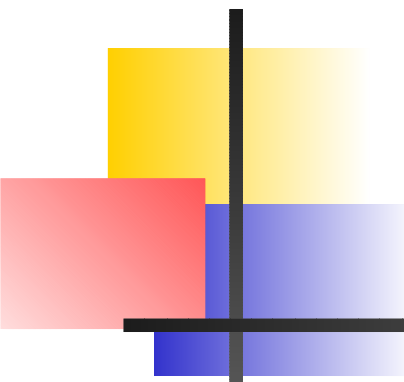
This extracts the variance components from the .asr file and their variances from the .vvp file and computes the requested quantities

# Components

■ 5 LogL=-188.777 S2= 132.76 61 df 0.2176 1.000

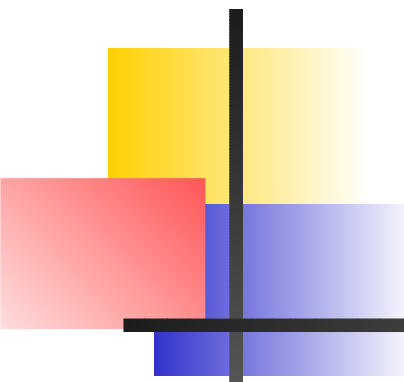
| Source   | terms | Gamma | Component | Comp/SE | % C      |
|----------|-------|-------|-----------|---------|----------|
| Sire     | 9     | 9     | 0.217651  | 28.8946 | 1.04 0 P |
| Variance | 65    | 61    | 1.00000   | 132.756 | 5.25 0 P |

■ Notice  $\gamma = 0.21765$ ,  $\sigma_e^2 = 132.756$ ,  
 $\sigma_s^2 = \gamma\sigma_e^2 = 28.8946$ .



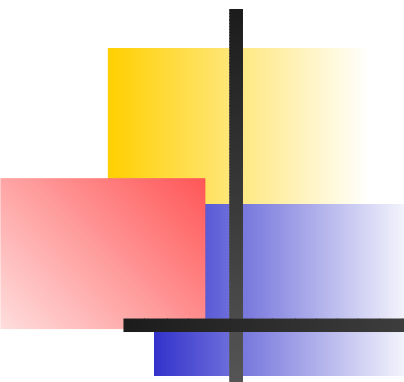
# Random design

- $Z$  has 9 columns being zeros except for 8 8 5 8 7 6 8 7 8 1's respectively indicating which sire
- The variance model is  $\sigma_e^2(I + \gamma Z Z')$  where  $\sigma_s^2 = \gamma \sigma_e^2$
- The genetic model
$$\sigma_A^2 = 4\sigma_s^2; \sigma_s^2 = 0.25\sigma_A^2$$
$$\sigma_E^2 = \sigma_e^2 - 3\sigma_s^2; \sigma_e^2 = \sigma_E^2 + 0.75\sigma_A^2$$



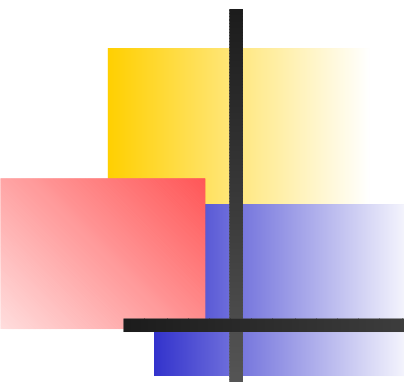
# Fixed Design

- $X$  has 5 columns  
mu is a column of ones  
line-1 has 21 1's, 15 0's, 29 0's  
line-2 has 21 0's, 15 1's, 29 0's  
line-3 has 21 0's, 15 0's, 29 1's  
DamAge (covariate) has vector of dam ages
- This design has 1 singularity because the three line columns sum to give the mu column.
- ASReml will set  $\tau_2 = 0$



# Summary

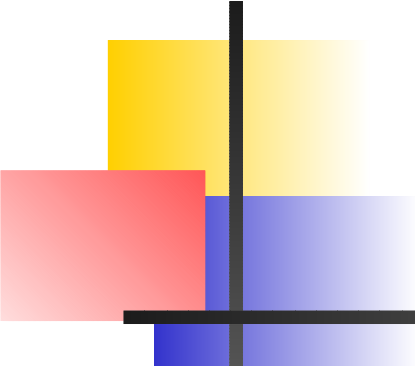
| ■ | Model term | Size    | #mv | #00 | MinNon0 | Mean   | MaxNon0 |
|---|------------|---------|-----|-----|---------|--------|---------|
| 1 | ID         |         | 0   | 0   | 101.0   | 133.0  | 165.0   |
| 2 | Sire       | 9       | 0   | 0   | 1       | 5.0154 | 9       |
| 3 | Dam        |         | 0   | 65  | 0.000   | 0.000  | 0.000   |
| 4 | Line       | 3       | 0   | 0   | 1       | 2.1231 | 3       |
| 5 | DamAge     |         | 0   | 0   | 3.000   | 4.385  | 5.000   |
| 6 | ADG        | Variate | 0   | 0   | 144.0   | 176.6  | 206.0   |
| 7 | Age        |         | 0   | 0   | 337.0   | 416.8  | 498.0   |
| 8 | InitialWT  |         | 0   | 0   | 144.0   | 241.1  | 300.0   |



## 2.2 Sire model

---

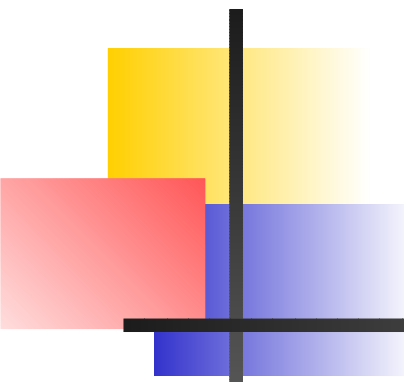
- Harvey Test data - Sire Model  
animal sire 9 dam line 3  
DamAge ADG Age WT  
harvey.dat  
ADG ~ mu line DamAge !r sire



# Harvey.dat summary

---

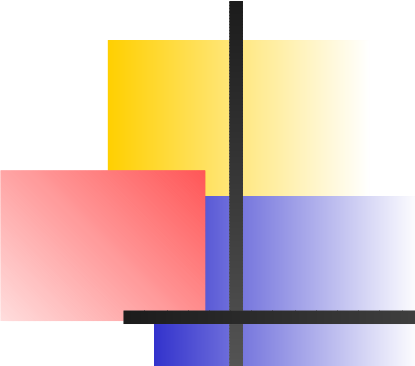
- 9 sires representing 3 sire lines  
(1 2 3) (4 5) (6 7 8 9)  
8 8 5   8 7   6 8 7 8 records (65)
- Data set originally distributed with Harvey's program



# Harvey.dat

■ animal sire dam

|     | line | Dam | Age | ADG | Age | WT  |     |   |
|-----|------|-----|-----|-----|-----|-----|-----|---|
| 101 | 1    | 0   | 1   | 3   | 192 | 390 | 224 | 1 |
| 102 | 1    | 0   | 1   | 3   | 154 | 403 | 265 | 1 |
| 103 | 1    | 0   | 1   | 4   | 185 | 432 | 241 | 1 |
| 104 | 1    | 0   | 1   | 4   | 183 | 457 | 225 | 1 |
| 105 | 1    | 0   | 1   | 5   | 186 | 483 | 258 | 1 |
| 106 | 1    | 0   | 1   | 5   | 177 | 469 | 267 | 1 |



# ASReml tutorial

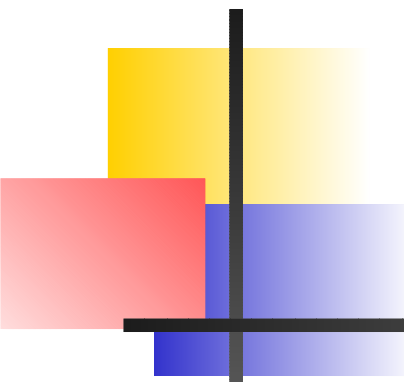
---

## B2 Sire Model

Arthur Gilmour



NSW DEPARTMENT OF  
**PRIMARY INDUSTRIES**



# The challenge

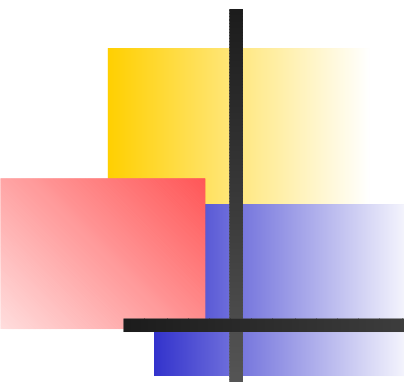
---

- to define  $X$   $Z$   $R$  and  $G$  to obtain the desired analysis.

$$R = \sigma^2 I$$

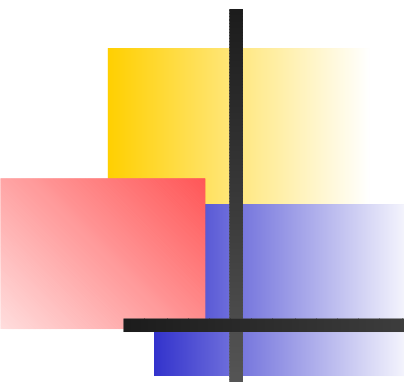
$$G_i = \sigma^2 \gamma_i I_i$$

- Often several ways of writing equivalent models.



# Differentiation

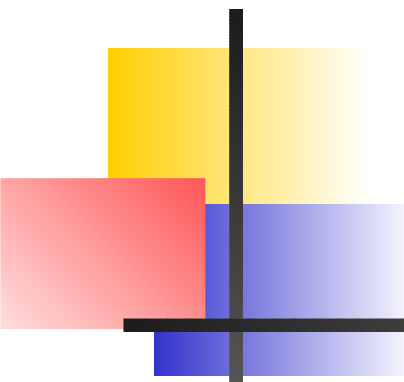
- $l_R = -(\log|\mathbf{C}| + \log|\mathbf{R}| + \log|\mathbf{G}| + \nu \log \sigma^2 + \mathbf{y}'\mathbf{P}\mathbf{y}/\sigma^2)/2$
- $\partial l_R / \partial \phi_i =$   
 $-( -\mathbf{C}_i \mathbf{C}^{Z'Z} + 0 + \text{tr}(\mathbf{G}_i \mathbf{G}^{-1}) + 0 - \mathbf{y}'\mathbf{P}\mathbf{y}_i / \sigma^2 ) / 2$   
 $\mathbf{y}_i = \mathbf{Z}\mathbf{G}_i \mathbf{G}^{-1} \tilde{\mathbf{u}}$
- $\partial l_R / \partial \kappa_j =$   
 $-( -\mathbf{C}_j \mathbf{C}^{-1} + \text{tr}(\mathbf{R}_j \mathbf{R}^{-1}) + 0 + 0 - \mathbf{y}'\mathbf{P}\mathbf{y}_j / \sigma^2 ) / 2$   
 $\mathbf{y}_j = \mathbf{R}_j \mathbf{R}^{-1} \tilde{\boldsymbol{\eta}}$



# Two forms

---

- $V = \sigma^2(R + ZGZ')$   
default for univariate single site analyse.  
 $R$  defined as a correlation matrix  
 $G$  defined as a variance ratio
- $V = R + ZGZ'$   
used for multivariate and multisite analyses.



# REML

■ Let  $C = \begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix}$

$$P = R^{-1} - R^{-1}WC^{-1}W'R^{-1}$$

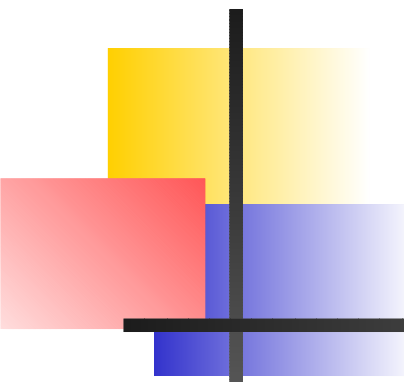
$$l_R = -(\log|C| + \log|R| + \log|G| + \nu \log \sigma^2 + \mathbf{y}'P\mathbf{y}/\sigma^2)/2$$

# Mixed model equations

- Mixed model

$$y = X\tau + Zu + \epsilon$$

$$\begin{pmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{pmatrix} \begin{pmatrix} \tau \\ u \end{pmatrix} = \begin{pmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{pmatrix}$$



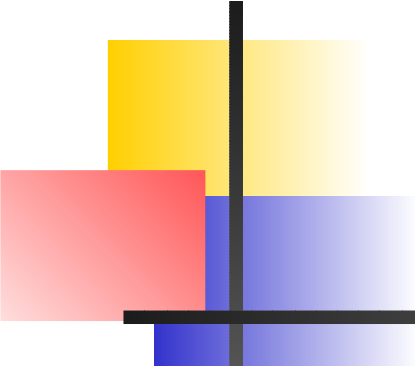
# Matrix operations

- Matrix multiplication – must be conformable

$$\begin{pmatrix} a & b & c \\ d & e & f \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = \begin{pmatrix} aA + bB + cC \\ dA + eB + fC \end{pmatrix}$$

- Direct product

$$\begin{pmatrix} a \\ b \end{pmatrix} \otimes [A \ B \ C] = \begin{pmatrix} aA & aB & aC \\ bA & bB & bC \end{pmatrix}$$



# Matrix operations

- Transpose

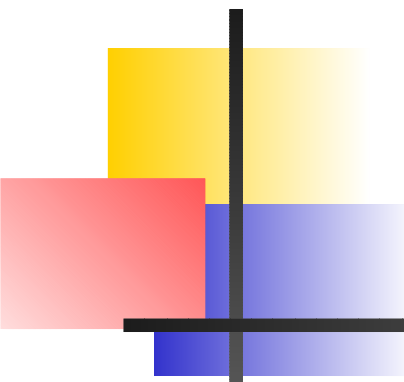
$$[a \ b \ c]' = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

- Addition - matrices of the same order are added element by element

$$\begin{bmatrix} 1 & 2 \end{bmatrix} + \begin{bmatrix} 3 & 4 \end{bmatrix} = \begin{bmatrix} 4 & 6 \end{bmatrix}$$

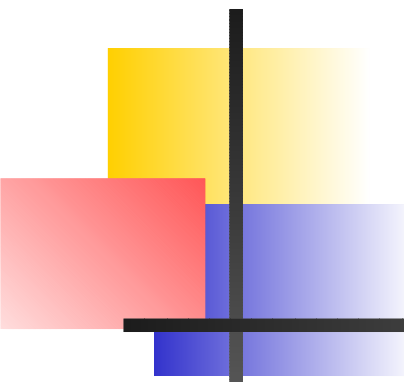
- Multiplication by scalar

$$3 \begin{bmatrix} 1 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 6 \end{bmatrix}$$



# Some matrices

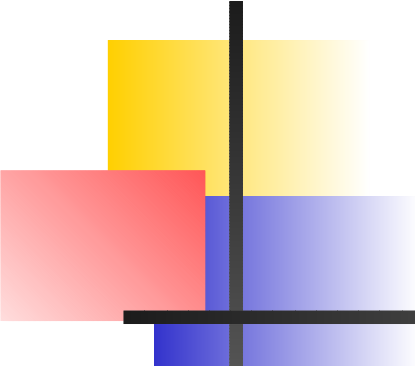
- A *matrix* is a rectangular array of numbers
  - $X$  is the design matrix for fixed effects
  - $Z$  is the design matrix for random effects
  - $W = [X \ Z]$  is the whole design matrix
  - $G = \text{var}(u)$
  - $R = \text{var}(\epsilon)$
  - $A$  is a relationship matrix



# Some vectors

---

- A *vector* is a column of numbers
  - $y$  is the response variable
  - $\hat{\tau}$  is the fixed effects
  - $\tilde{u}$  is the random effects
  - $\tilde{\epsilon}$  is the residuals



# ASReml tutorial

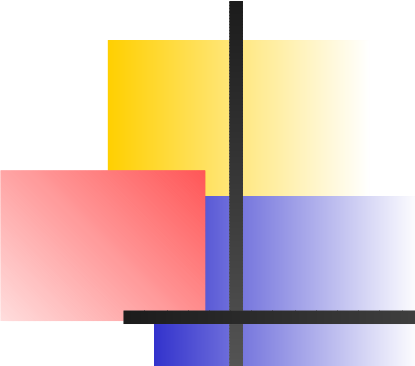
---

## B1 A matrix refresher

Arthur Gilmour



NSW DEPARTMENT OF  
**PRIMARY INDUSTRIES**



# ASReml tutorial

---

## B1 A matrix refresher

Arthur Gilmour



NSW DEPARTMENT OF  
**PRIMARY INDUSTRIES**