Linear Mixed Models: Review

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Zu} + \boldsymbol{e}$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \sum_{i=1}^{r} \boldsymbol{Z}_i \boldsymbol{u}_i + e$$

$$\mathrm{E}(\boldsymbol{y}) = \boldsymbol{X\beta}$$

$$\boldsymbol{V} = \mathrm{var}(\boldsymbol{y}) = \sum_{i=0}^{r} \boldsymbol{Z}_i \boldsymbol{D}_i \boldsymbol{Z}_i'$$

$$\boldsymbol{u} = (\boldsymbol{u}_1' \ \boldsymbol{u}_2' \ \ldots \ \boldsymbol{u}_r')'$$

$$\boldsymbol{D} = \mathrm{var}(\boldsymbol{u}) = \bigoplus_{i=1}^{r} \boldsymbol{D}_i$$

$$\boldsymbol{R} = \mathrm{var}(\boldsymbol{e}) = \boldsymbol{I}_N \otimes \boldsymbol{\Sigma}$$

Prediction

- Considerable attention is focused on the problem of how to **estimate** fixed values
  - $\boldsymbol{K}'\boldsymbol{\beta}$
  - σ_i^2
  - $\sigma_a^2/(\sigma_a^2 + \sigma_e^2)$

- What if we want to **predict** the value of a random variable

  -
    $$\boldsymbol{L}' \begin{pmatrix} \boldsymbol{\beta} \\ \boldsymbol{u} \end{pmatrix}$$
  - Interested in the difference between two sires
  - How much "better" is line 1 of variety A compared to line 2 of variety C?

- Minimize the difference between the realized value of our random variable and our prediction based on the data.

Best Linear Unbiased Predictor

$$\boldsymbol{w} = \boldsymbol{L}'_X\boldsymbol{\beta} + \boldsymbol{L}'_Z\boldsymbol{u}$$

1. Linear

$$\tilde{\boldsymbol{w}} = \boldsymbol{a} + \boldsymbol{B}\boldsymbol{y}$$

2. Unbiased

$$\mathrm{E}(\tilde{\boldsymbol{w}}) = \boldsymbol{L}'_X\boldsymbol{\beta}$$

3. Best "Minimum prediction error variance"

$$\mathrm{var}[\boldsymbol{a}'(\tilde{\boldsymbol{w}} - \boldsymbol{w}))]$$

or

$$\mathrm{E}\left[(\tilde{\boldsymbol{w}} - \boldsymbol{w})'\boldsymbol{A}(\tilde{\boldsymbol{w}} - \boldsymbol{w})\right]$$

with $\boldsymbol{A}$ positive definite

BLUP of w

$$\tilde{\boldsymbol{w}} = \boldsymbol{L}'_X \widehat{\boldsymbol{\beta}} + \boldsymbol{L}'_Z \widehat{\boldsymbol{u}}$$

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{y}$$

$$\widehat{\boldsymbol{u}} = \boldsymbol{C}\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}})$$

$$\boldsymbol{C} = \mathrm{cov}(\boldsymbol{u}, \boldsymbol{y}')$$

Mixed Model Equations

$$\begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{X} & \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} + \boldsymbol{D}^{-1} \end{pmatrix} \begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{y} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{y} \end{pmatrix}$$

Variance

- The variance of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{u}}$ can be obtained directly from the mixed model equations.

$$\begin{pmatrix} \boldsymbol{C}^{XX} & \boldsymbol{C}^{XZ} \\ \boldsymbol{C}^{ZX} & \boldsymbol{C}^{ZZ} \end{pmatrix} = \begin{pmatrix} \boldsymbol{X}'\boldsymbol{R}^{-1}\boldsymbol{Z} \\ \boldsymbol{Z}'\boldsymbol{R}^{-1}\boldsymbol{Z} \end{pmatrix}$$

$$\operatorname{var}\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} - \boldsymbol{u} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}^{XX} & \boldsymbol{C}^{XZ} \\ \boldsymbol{C}^{ZX} & \boldsymbol{C}^{ZZ} \end{pmatrix}$$

$$\operatorname{var}\begin{pmatrix} \widehat{\boldsymbol{\beta}} \\ \widehat{\boldsymbol{u}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{C}^{XX} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{D} - \boldsymbol{C}^{ZZ} \end{pmatrix}$$

Linear Mixed Models in ASREML

Command File (.as file)

- The command file has three main parts
  1. Data input and transformation
  2. Model specification
  3. Variance specification

- Return to the experiment comparing the oxygen consumption of four species of frog.(Zar (1999))[4]

[4] Original data was obtained from http://www.statsci.org/data/general/frogs.html

Frog example

- Dependent Variable
  - Oxygen Consumption
- Fixed Effects
  - Species (4 levels)
  - Temperature (Low and High)
  - Exercise (Rest and Exercise)
- Random effects
  - Frog (nested w/in Species and Temperate)
  - Error

Data

```
Subject Species Temperature     Rest    Exercise
1       1       Low     0.107   0.152
2       1       Low     0.114   0.163
3       1       High    0.133   0.194
...
```

- `Rest` Resting oxygen consumption
- `Exercise` Exercise oxygen consumption
- One record/subject
- Tab delimited

Data input

ASReml

```
Oxygen consumption of frogs
 Frog *
 Species * !I
 Temp * !A
 Rest # Resting oxygen consumption
 Exercise # Exercise oxygen consumption
frogs.txt !SKIP 1
```

Model Specification

Model Equation

```
Rest Exercise ~ mu Trait*Species*Temp !r Frog.Species.Temp
```

- `*` Expands into main affects and interactions.
- `Trait` Automatic variable

Variance structure

```
0 0 1
Frog.Species.Temp 1
0 0 I 1
```

R header Line

`0 0 1` or `s d g`

- `s` Number of heterogeneous $\boldsymbol{R}$ groups (Usually 1)

  Zero in our case says $\boldsymbol{e} \sim \mathrm{N}(\mathbf{0}, \boldsymbol{I}\sigma^2)$

- `d` Number terms in the direct product (Usually 1 or 2)

  Zero in our case since we are using the default covariance structure

- `g` Number of random effects

  We have one random effect

Random effects

- Header line `Effect parts`
  - `Frog.Species.Temp 1`
  - Effect `Frog.Species.Temp`
  - 1 one part

- Structure line `n l M <initial values>`
  - `n=0` number of levels (Zero tells ASReml to figure it out)
  - `l=0`
  - `M=I` Covariance structure (In this case an identity
  - 1 Initial estimate

ASReml (.asr file)

```
  1 LogL= 68.4455     S2= 0.20438E-04     16 df    1.000      1.000
  2 LogL= 68.6518     S2= 0.18062E-04     16 df    1.000      1.324
  3 LogL= 68.8008     S2= 0.15811E-04     16 df    1.000      1.793
  4 LogL= 68.8462     S2= 0.14259E-04     16 df    1.000      2.288
  5 LogL= 68.8465     S2= 0.14126E-04     16 df    1.000      2.340
 Final parameter values                        1.0000     2.3407

 Source                  Model  terms     Gamma     Component    Comp/SE   % C
 Variance                   32     16   1.00000      0.141262E-04   2.00   0 P
 Frog.Species.Temp   identity   128   2.34071      0.330653E-04   1.622402 U
Warning: LogL Converged; Parameters Not Converged
```

Single Trait Animal Model

- Weight data on 386 mice
- Five generations
- Birth date covariate
- Missing data

Data input

- Tab delimited file with a header line

```
Anim	Sire	Dam	Day	Gen	Gender	2Wk	3Wk
6515	5002	4636	31	0	F	104	150
6696	4531	5306	32	0	F	112	164
```

- Body Weights collected to the nearest .1 of the gram

```
Single Trait Analysis  Title
 Anim * !P
 Sire * !P
 Dam * !P
 Day  Covariate or Dependent Variable
 Generation * !I  Integer Factor
 Gender * !A  Alpha Factor
 BW2
```

- Title line
- Fields indented by a space
- * Number of levels determined by ASREML

```
 BW12
M42.dat  !SKIP 1 # Pedigree file
M42.dat !SKIP 1 !MAXIT 100 !EXTRA 5 #Data file
```

- Not indented
- File name
- `!Skip 1` skip header record
- `!MAXIT 100` Maximum number of iterations 100
- `!EXTRA 5` Do an extra 5 iterations after convergence

Pedigree

- Individual, Sire, Dam
- Individual, Sire, Maternal Grand Sire `!MGS`
- May be Alphanumeric `!ALPHA`
- Missing parents coded as zero
- Pedigree fields denoted by `!P`

```
Single Trait Analysis
 Anim * !P
 Sire * !P
 Dam * !P
 Day
 Generation * !A
 Gender * !I
 BW2
.
.
M42.dat  !SKIP 1 # Pedigree file
M42.dat !SKIP 1 !MAXIT 100 !EXTRA 5 #Data file
```

- Not indented

Transformations

```
 BW4_5
 BW6  !/10
 BW8  !/10
 BW12 !/10
 ADG !=V11 !-V10 !/14
 day !=V4 !/365
M42.dat  !SKIP 1 # Pedigree file
```

$$BW6/ = 10$$

$$ADG = (BW8 - BW12)/14$$

$$day = DAY/365$$

Model Specification

- Fixed
  - Generation, Gender, and Generation*Gender
    `A*B` $\equiv$ `A B A.B`
  - Day(Generation) covariate
    `Generation/Day`
- Random
  - Animal $\mathrm{N}(\mathbf{0}, \boldsymbol{A}\sigma_A^2)$

```
M42.dat !SKIP 1 !MAXIT 100 !EXTRA 5 #Data file
BW6 ~ mu Generation*Gender Generation.day  !r Anim
0 0 1        # Groups Direct_Prod Random_Effects
Anim 1       # Term Direct_Prod
0 0 AINV 1   # Size Sort Structure Initial_Val
```

- Model equation
- Residual and Number of Random effects
- Random effects

Results

```
Model term       Size Type    COL   Minimum     Mean       Maximum   #zero #miss
 1 Anim          419 Factor    1       51    235.2692        419         0     0
 2 Sire          309 Factor    2        9    147.2308        309         0     0
 3 Dam           324 Factor    3        3    153.6868        324         0     0
 4 Day                         4   149.0       313.7       479.0         0     0
 5 Generation      5 Factor    5        2      3.5220          5         0     0
 6 Gender          2 Factor    6        1      1.3736          2         0     0
 7 BW2                         7   63.00       100.8       124.0         0     0
 8 BW3                         8   83.00       152.8       201.0         0     0
 9 BW4_5                       9   216.0       274.4       352.0         0   270
10 BW6             1 Variate  10   14.60       29.97       41.50         0     0
11 BW8                        11   11.30       32.40       47.30         0     0
12 BW12                       12   24.30       35.34       50.40         0   284
13 ADG                        13-0.8857      0.1737      0.7429         1     0
14 day                        14 0.4082      0.8596       1.312         0     0
15 mu              1 Constant Term
16 Generation.Ge  10 Interaction  5 Generation:   5   6 Gender         :    2
17 Generation.da   5 Interaction  5 Generation:   5  14 day            :    1
  419  Ainverse       1.0000
```

```
NOTICE:    11 (more) singularities,
  1 LogL=-519.561     S2=  4.2063        352 df    1.000      1.000
  2 LogL=-519.429     S2=  4.1229        352 df    1.000      1.055
  3 LogL=-519.134     S2=  3.8923        352 df    1.000      1.222
  4 LogL=-518.913     S2=  3.6032        352 df    1.000      1.468
  5 LogL=-518.854     S2=  3.3588        352 df    1.000      1.713
  6 LogL=-518.854     S2=  3.3592        352 df    1.000      1.712
  7 LogL=-518.854     S2=  3.3591        352 df    1.000      1.712
  8 LogL=-518.854     S2=  3.3592        352 df    1.000      1.712
  9 LogL=-518.854     S2=  3.3592        352 df    1.000      1.712
 10 LogL=-518.854     S2=  3.3592        352 df    1.000      1.712
 11 LogL=-518.854     S2=  3.3592        352 df    1.000      1.712
Final parameter values                          1.0000     1.7122

Source                Model  terms     Gamma     Component    Comp/SE   % C
Variance                364    352   1.00000       3.35915       4.28   0 P
Anim                Ainverse   419   1.71225       5.75170       3.84   0 U
```

```
Analysis of Variance                      DF       F-incr
  5 Generation                             3         9.77
  6 Gender                                 1       335.48
 16 Generation.Gender                      3         2.20
 17 Generation.day                         4        12.26

                    Estimate       Standard Error     T-value      T-prev
 17 Generation.day
                       2   -12.1514         34.7302         -0.35
                       3   -64.1232         82.9213         -0.77      -0.58
                       4    285.773         41.3023          6.92       3.80
                       5   -14.1994         43.4579         -0.33      -5.09
 16 Generation.Gender
                      11   0.501526        0.810515          0.62
                      13  -0.698691        0.787780         -0.89
                      15   -1.57254        0.778862         -2.02
  6 Gender
M                           5.57881        0.562568          9.92
  5 Generation
                       2    38.6446         59.8426          0.65
                       3   -290.988         44.0447         -6.61      -4.64
                       4    11.4335         57.9476          0.20       4.41
 15 mu
                      23    34.5298         14.5634          2.37
  1 Anim                                 419 effects fitted
3  possible outliers: see .res file
```

```
SLOPES FOR LOG(ABS(RES)) on LOG(PV) for Section   1
 -0.15

                                    ** *
                                   *** **
                                   ******
                                   ******
                                   ******
                                 ********
                                 ******** * *
                                 ************
                               **************
                               **************
                              ***************
                             ****************
                        *    ******************
                        **** *******************
*          *    *  **  ***************************    *

Min Mean Max -8.0925     0.34521E-12  3.8418     omitting     0 zeros

STND RES  210 -4.75
STND RES  295 -6.71
STND RES  352 -3.76
```

209	29.789	-0.1889	1.699
210	25.038	-5.738	1.699
211	30.665	0.8347	1.699
...			
294	28.530	-0.4301	1.650
295	22.693	-8.093	1.665
296	28.926	-0.3260	1.665
...			
351	25.686	-0.3862	1.607
352	27.540	-4.540	1.682
353	30.953	-0.2526	1.682

Anim	Sire	Dam	Day	Gen	Gender	2Wk	3Wk	4.5Wk	6Wk	8Wk	12Wk
...											
6323	3962	3313	362	3	M	106	147	.	296	339	.
6324	3962	3313	362	3	M	107	147	.	193	237	.
6325	3962	3313	362	3	M	118	162	.	315	338	.
...											
8778	6971	6523	469	4	F	093	140	.	281	333	.
8981	7172	6207	469	4	M	080	120	.	146	113	.
8982	7172	6207	469	4	M	079	126	.	286	333	.
...											
9638	6202	6526	473	4	F	091	138	.	253	271	.
9641	6813	7188	473	4	M	091	119	.	230	300	.
9642	6813	7188	473	4	M	092	127	.	307	311	.

Heritability (.pin file)

Source	Model	terms	Gamma	Component	Comp/SE
Variance	364	352	1.00000	3.35915	4.28
Anim	Ainverse	419	1.71225	5.75170	3.84

$$\sigma_1^2 = \sigma_E^2 \quad \sigma_2^2 = \sigma_A^2$$

- Linear Functions

$$\sigma_{nvc+1}^2 = \sum_{i=1}^{nvc} \sigma_i^2 \lambda_i$$

- Ratios

$$\theta = \sigma_k^2 / \sigma_\ell^2$$

mouse_42_st.pin

`F Add. 2`	$\sigma_3^2 = \sigma_A^2 = \sigma_1^2$
`F Env. 1`	$\sigma_4^2 = \sigma_E^2 = \sigma_2^2$
`F Phen. 1 + 2 * 1`	$\sigma_5^2 = \sigma_P^2 = \sigma_1^2 + \sigma_2^2 \times 1$
`H h2 2 5`	$h^2 = \sigma_2^2 / \sigma_5^2$

- `F` linear functions of variance components $\sigma_P^2 = \sigma_A^2 + \sigma_E^2 * 1$
- `H` ratios of variance components $h^2 = \sigma_A^2 / \sigma_P^2$

Results

```
ASreml -P mouse_42_st
```

```
3 Add.  2         5.752       1.498
4 Env.  1         3.359      0.7848
5 Phen.  1        9.111      0.9741
  h2           = Anim        2/Phen.  1   5=         0.6313  0.1095
```

Effect	Variance	Standard Error
Add.	5.752	1.498
Env.	3.359	0.9741
Phen.	9.111	0.9741
h^2	0.6313	0.1095

Correlated Random Effects

- Direct and Maternal Genetic Effects
- Random Regression
- Model Specification `![ Animal Dam !]`
- Variance Specification

Direct and Maternal genetic effects

- Fixed
  - Generation, Gender, and Generation*Gender
  - Day(Generation) covariate
- Random
  - Animal Dam $\mathrm{N}(\mathbf{0}, \boldsymbol{G} \bigotimes \boldsymbol{A})$
  - ide(Dam) $\mathrm{N}(\mathbf{0}, \boldsymbol{I}\sigma_L^2)$

```
BW6 ~ mu Generation*Gender Generation.day,
!r ![ Anim Dam !] ide(Dam)
```

Variance Structures

```
0 0 2
Anim 2
2 0 US 1 .1 1 !GP
0 0 AINV
ide(Dam) 1
0 0 I 1
```

Comments: line 3: #σ_1^2 σ_{12} σ_2^2 *Lower Triangle by row*; line 4: #*No initial value*

ASReml mouse_42_dm

```
  1 LogL=-522.333     S2=  3.5466        352 df
  2 LogL=-519.787     S2=  3.4493        352 df    :   1 components constrained
Warning: EM updates for  1 positive definite US structure(s).
  3 LogL=-514.341     S2=  3.0163        352 df1
  4 LogL=-514.167     S2=  2.9896        352 df
  5 LogL=-514.164     S2=  2.9944        352 df
  6 LogL=-514.158     S2=  3.0088        352 df
  7 LogL=-514.151     S2=  3.0301        352 df
  8 LogL=-514.148     S2=  3.0561        352 df
  9 LogL=-514.148     S2=  3.0661        352 df
 10 LogL=-514.148     S2=  3.0694        352 df
 11 LogL=-514.148     S2=  3.0705        352 df
 12 LogL=-514.148     S2=  3.0709        352 df
 13 LogL=-514.148     S2=  3.0710        352 df
 14 LogL=-514.148     S2=  3.0710        352 df
Warning:   1 variance structures were modified to make them positive definite
          ASREML may have fixed the  structure [flagged by B]
          and may not have converged to  a maximum likelihood solution.
          Suggestion: rerun with  -C option
         or maybe try an ANTE or XFA structure
```

Source	Model	terms	Gamma	Component	Comp/SE	%	C
Variance	364	352	1.00000	3.07102	2.21	0	P
Anim	UnStruct	1	1.73603	5.33138	1.91	0	P
Anim	UnStruct	1	-0.482726	-1.48246	-0.83	0	P
Anim	UnStruct	2	0.202433	0.621676	0.32	0	P
ide(Dam)	identity	419	0.425535	1.30683	1.07	0	U

Covariance/Variance/Correlation Matrix UnStructured

1.736	-0.8143	σ_1^2	ρ_{12}
-0.4827	0.2024	σ_{12}	σ_2^2

Multiple Trait Models

- Multiple dependent variables
- `Trait` automatic variable
- $\boldsymbol{e} \sim \mathrm{N}(\mathbf{0}, \bigotimes \boldsymbol{R}_i)$
- Trait specific models (`at()`)
- Correlations (`R c12 v1 v2`)

Multiple Dependent Variables

- Dependent Variables
  - Body Weight at 6 weeks
  - Body Weight at 8 weeks

```
BW6 BW8 ~ Trait*Generation*Gender Trait.Generation.day  mv ,
!r ![ Trait.Anim Trait.Dam !]
1 2 1
0 0 I
2 0 US 3.885 3.271 6.963 !GP
Trait.Anim 2
4 0 US 3.724 4.356 6.372 .1 .1 1.927 .1 .1 1.414 1.329 !GP
0 0 AINV
```

Results

```
Source                Model  terms     Gamma     Component    Comp/SE   % C
Residual            UnStruct     1   2.50146       2.50146       0.00   0 P
Residual            UnStruct     1   2.12496       2.12496       4.42   0 P
Residual            UnStruct     2   5.39939       5.39939       8.51   0 P
Trait.Anim          UnStruct     1   6.61415       6.61415       0.00   0 P
Trait.Anim          UnStruct     1   6.78293       6.78293       7.67   0 P
Trait.Anim          UnStruct     2   9.55886       9.55886      13.94   0 P
Trait.Anim          UnStruct     1  -2.30962      -2.30962      -7.73   0 P
Trait.Anim          UnStruct     2  -4.00834      -4.00834      -2.06   0 P
Trait.Anim          UnStruct     3   2.66842       2.66842       1.66   0 P
Trait.Anim          UnStruct     1  -1.84080      -1.84080      -1.09   0 P
Trait.Anim          UnStruct     2  -4.29757      -4.29757      -1.90   0 P
Trait.Anim          UnStruct     3   2.66994       2.66994       1.88   0 P
Trait.Anim          UnStruct     4   3.45231       3.45231       1.85   0 P
Warning: Code B - fixed at a boundary (!GP)       F - fixed by user
              ? - liable to change from P to B    P - positive definite
              C - Constrained by user (!VCC)      U - unbounded
              S - Singular Information matrix
S means there is no information in the data for this parameter.
Very small components with Comp/SE ratios of zero sometimes indicate poor
          scaling.  Consider rescaling the design matrix in such cases.
```

```
Covariance/Variance/Correlation Matrix UnStructured
 2.501     0.5782
 2.125      5.399
Covariance/Variance/Correlation Matrix UnStructured
 6.614     0.8531    -0.5498    -0.3852
 6.783      9.559    -0.7937    -0.7481
-2.310     -4.008      2.668     0.8797
-1.841     -4.298      2.670      3.452
```

σ^2_{a6}

σ_{a86} σ^2_{a8}

σ_{m6a6} σ_{m6a8} σ^2_{m6}

σ_{m8a6} σ_{m8a8} σ_{m86} σ^2_{m8}

```
Analysis of Variance                   DF     F-incr
  5 Generation                          3      10.49
  6 Gender                              1     363.81
 16 Trait.Generation                    3       3.31
 17 Trait.Gender                        1      23.65
 18 Generation.Gender                   3       1.67
 19 Trait.Generation.Gender             3       1.71
 21 Trait.Generation.day                8       6.53
```

- As the models become more complicated the choice of starting values becomes more important

- Simple single trait models

- Complex single trait models

- Simple two trait models

- Complex two trait models

Functions of Variance Components

- Group terms together
  - Maternal $\sigma^2_{m6} = \sigma_{14}$ $\sigma_{m68} = \sigma_{15}$ $\sigma^2_{m8} = \sigma_{16}$

    ```
    F M1 9  #14:16
    F M12 12
    F M2 13
    ```

  - Direct*Maternal $\sigma_{m6a6} = \sigma_{17}$ $\frac{\sigma_{m6a8}+\sigma_{m8a6}}{2} = \sigma_{18}$ $\sigma_{m8a8} = \sigma_{19}$

    ```
    F AM1 7  17:19
    F AM12 8 *.5  10*.5
    F AM22 11
    ```

- Phenotypic Variances and Covariances $\sigma^2_{p6} = \sigma_{20}$ $\sigma_{p68} = \sigma_{21}$ $\sigma^2_{p8} = \sigma_{22}$

```
F Phen. 1:3 + 4:6 + 14:16 +17:19 #20:22
```

- Correlations

$$\frac{\sigma_{xy}}{\sqrt{\sigma^2_x \sigma^2_y}}$$

  – `R x xy y`

```
R Dr 4 5 6
R Mr 14 15 16
R Pr 20 21 22
```

```
14 M1  9          2.671       1.328
15 M12 12         2.669       1.423
16 M2 13          3.451       1.865
17 AM1  7        -2.310       1.591
18 AM12  8       -2.921       1.695
19 AM22 11       -4.296       2.259
20 Phen.  1       9.470       1.168
21 Phen.  2       8.652       1.133
22 Phen.  3       14.11       1.424
   h2D1         = Trait.An   4/Phen.  1  20=        0.6971  0.2229
   h2D2         = Trait.An   6/Phen.  3  22=        0.6772  0.1912
   Dr           = Trait.An/SQR[Trait.An*Trait.An]=  0.8533  0.0918
   Mr           = M12 12  /SQR[M1  9   *M2 13   ]=  0.8790  0.1284
   Pr           = Phen.  2/SQR[Phen.  1*Phen.  3]=  0.7484  0.0328
```

Trait Specific Models

- Sometimes will want to include an effect only for a subset of traits
  - `at(f,l)`
    * 1 when factor `f` is at level `l`
    * 0 otherwise

Trait	Trt	Trt	at(Trait,1).Trt
1	A	(1 0 0)	(1 0 0)
1	B	(0 1 0)	(0 1 0)
1	C	(0 0 1)	(0 0 1)
2	A	(1 0 0)	(0 0 0)
2	B	(0 1 0)	(0 0 0)

Model

- Traits
  - 6 week weight
  - ADG between 6 and 8 weeks

- Fixed effects
  - Generation*Gender
  - mv (missing values)
  - day(Generation) covariate

- Random effects
  - Direct and Maternal effects for 6 week weight
  - Direct effect for ADG

$$\begin{pmatrix} \sigma^2_{a6} & \sigma_{a68} & \sigma_{a6m6} \\ \sigma_{a86} & \sigma^2_{a8} & \sigma_{a8m6} \\ \sigma_{m6a6} & \sigma_{m6a8} & \sigma^2_{m6} \end{pmatrix} \otimes \boldsymbol{A}$$

```
BW6 ADG ~ Trait*Generation*Gender Trait.Generation.day   ,
!r    ![ Trait.Anim   at(Trait,1).Dam !] !f mv
1 2 1
0 0 I
2 0 US 3.885 3.271 6.963 !GP
Trait.Anim 2
3 0 US 1 .01 2 .1 .1 3 !GP
0 0 AINV
```

Solution file .sln

```
  Trait.Anim              1.004636          0.1936       2.440
  Trait.Anim              1.005002          0.1936       2.440
...
  Trait.Anim              1.009757         -1.659        1.795
  Trait.Anim              1.009758         -1.294        1.795
  Trait.Anim              2.004636         -0.4696E-01  0.1062
  Trait.Anim              2.005002         -0.4696E-01  0.1062
...
  Trait.Anim              2.009757          0.5524E-01  0.8915E-01
  Trait.Anim              2.009758          0.3535E-01  0.8915E-01
  at(Trait,1).Dam         1.004636          0.7488       1.596
  at(Trait,1).Dam         1.005002          0.7488       1.596
...
  at(Trait,1).Dam         1.009757         -0.1611       1.450
  at(Trait,1).Dam         1.009758         -0.1009       1.450
```

```
  1 LogL=-62.3268     S2=  1.0000        748 df
  2 LogL=-62.3268     S2=  1.0000        748 df
  3 LogL=-62.3268     S2=  1.0000        748 df

 Source               Model  terms     Gamma     Component    Comp/SE   % C
 Variance               772    748   1.00000       1.00000       1.44   0 C
 Residual           UnStruct     1   2.71788       2.71788       0.00   0 P
 Residual           UnStruct     1 -0.380548E-01 -0.380548E-01  -0.23   0 P
 Residual           UnStruct     2  0.197066E-01  0.197066E-01   0.01   0 P
 Trait.Anim         UnStruct     1   6.16375       6.16375       0.00   0 P
 Trait.Anim         UnStruct     1  0.360471E-01  0.360471E-01   0.06   0 P
 Trait.Anim         UnStruct     2  0.122103E-01  0.122103E-01   2.69   0 P
 Trait.Anim         UnStruct     1  -2.23893      -2.23893      -1.42   0 P
 Trait.Anim         UnStruct     2 -0.117956     -0.117956      -1.82   0 P
 Trait.Anim         UnStruct     3   2.73618       2.73618       2.05   0 P
 Warning: Code B - fixed at a boundary (!GP)       F - fixed by user
 ...
 Covariance/Variance/Correlation Matrix UnStructured
  2.718    -0.1644
-0.3805E-01 0.1971E-01
 Covariance/Variance/Correlation Matrix UnStructured
  6.164     0.1314    -0.5452
 0.3605E-01 0.1221E-01-0.6453
 -2.239    -0.1180      2.736
```

Random Regression

- General term used when the model includes random covariates
- Common examples
  - Polynomials
    Often use orthogonal polynomials for numerical reasons
  - Splines
    Piecewise polynomials
    Knots where the pieces join

Polynomials

- Arrange data with one record per Weight and create a “Time” covariate

Anim	Sire	Dam	Day	Gen	Gender	Week	Weight
6515	5002	4636	31	0	F	2	104
6515	5002	4636	31	0	F	3	150
6515	5002	4636	31	0	F	4.5	226
6515	5002	4636	31	0	F	6	.
6515	5002	4636	31	0	F	8	.
6515	5002	4636	31	0	F	12	.
6696	4531	5306	32	0	F	2	112
6696	4531	5306	32	0	F	3	164
6696	4531	5306	32	0	F	4.5	222
6696	4531	5306	32	0	F	6	.

Data Specification

```
Two Trait Analysis
 Anim 300 !P
 Sire 300 !P
 Dam 300 !P
 Day
 Generation * !I
 Gender * !A
 Week
 Weight  !/10
 day !=V4 !/365
M42.dat  !SKIP 1 # Pedigree file
M42_rr.dat !SKIP 1 !MAXIT 100 #Data file
```

Legendre Polynomials

- `leg(x,p)`: construct Legengre polynomials of order p
  - Linear transformation of

$$\begin{pmatrix}1 & x & x^2 & \cdots & x^p\end{pmatrix} \rightarrow$$

$$\begin{pmatrix}P_0(x) \, P_1(x) \, P_2(x) & \cdots & P_p(x)\end{pmatrix}$$

  - p< 0 drop the constant term
  - Mathematically equivalent to a standard polynomial
  - Advantages are numerical

Model

- Dependent Variable: Weight
- Fixed Effects:
  - Week*Generation*Gender contemporary groups
  - day(Generation*Week) covariates
- Random effects
  - Third order Legengre polynomials direct effects

- Residual
  - Second order Legendre polynomial
  - Heterogeneous Variance

Variance Structure

$$G = \begin{pmatrix} P_0(2) & P_1(2) & P_2(2) & P_3(2) \\ P_0(3) & P_1(3) & P_2(3) & P_3(3) \\ & & \vdots & \\ P_0(12) & P_1(12) & P_2(12) & P_3(12) \end{pmatrix} \begin{pmatrix} \sigma^2_{a0} & \sigma_{a01} & \sigma_{a02} & \sigma_{a03} \\ \sigma_{a10} & \sigma^2_{a1} & \sigma_{a12} & \sigma_{a13} \\ \sigma_{a20} & \sigma_{a21} & \sigma^2_{a2} & \sigma_{a23} \\ \sigma_{a30} & \sigma_{a31} & \sigma_{a32} & \sigma^2_{a3} \end{pmatrix} \begin{pmatrix} P_0(2) & P_0(3) & \cdots & P_0(12) \\ P_1(2) & P_1(3) & \cdots & P_1(12) \\ P_2(2) & P_2(3) & \cdots & P_2(12) \\ P_3(2) & P_3(3) & \cdots & P_3(12) \end{pmatrix}$$

$$= \sum_{i=0}^{3} \sum_{j=0}^{3} \left[\boldsymbol{P}_i \boldsymbol{P}'_j \right] \sigma_{aij}$$

$$\boldsymbol{P}'_j = \begin{pmatrix} P_j(2) & P_j(3) & \cdots & P_j(12) \end{pmatrix}$$

$$\boldsymbol{R} = \sum_{i=0}^{2} \sum_{j=0}^{2} \left[\boldsymbol{P}_i \boldsymbol{P}'_j \right] \sigma_{peij} + \mathrm{Diag}(\sigma^2_{ri})$$

```
Weight ~ fac(Week).Generation*Gender fac(Week).Generation.day   ,
  !r   leg(Week,3).Anim  ide(Anim).leg(Week,2) !f mv
1 2  2
386 0 I
6 0 DIAG 1 2 3 4 5 6 !S2==1 !GP
leg(Week,3).Anim 2
4 Week US !GP !+10
16
9 6
0 0 1
0 0 0 .7
419 0 AINV
ide(Anim).leg(Week,2) 2
419 0 IDEN
3 0 US !GP !+6
1
0 1
0 0 1
```

```
Source                  Model  terms     Gamma      Component    Comp/SE   % C
Variance                 2316   1624   1.00000        1.00000       1.19   0 F
Residual              DIAGonal     1  0.100000E-06  0.100000E-06    0.00   0 B
Residual              DIAGonal     2  0.568810      0.568810       0.57   0 P
Residual              DIAGonal     3   1.13715        1.13715       0.83   0 P
Residual              DIAGonal     4   2.35523        2.35523       0.00   0 P
Residual              DIAGonal     5   2.14743        2.14743       9.59   0 P
Residual              DIAGonal     6   1.14547        1.14547       1.32  -1 P
leg(Week,3).Anim      UnStruct     1   7.15231        7.15231       4.92   0 P
leg(Week,3).Anim      UnStruct     1   3.82721        3.82721       2.99   0 P
leg(Week,3).Anim      UnStruct     2   2.39690        2.39690       1.47   0 P
leg(Week,3).Anim      UnStruct     1 -0.467647      -0.467647      -1.09   0 P
leg(Week,3).Anim      UnStruct     2 -0.290318      -0.290318      -1.13   0 P
leg(Week,3).Anim      UnStruct     3  0.166023       0.166023       1.33   0 P
leg(Week,3).Anim      UnStruct     1  -1.10017       -1.10017      -2.80   0 P
leg(Week,3).Anim      UnStruct     2 -0.735352      -0.735352      -2.57   0 P
leg(Week,3).Anim      UnStruct     3 -0.599325E-01 -0.599325E-01  -0.41   0 P
leg(Week,3).Anim      UnStruct     4  0.957631       0.957631       4.27   0 P
ide(Anim).leg(Week,2UnStruct     1   4.46775        4.46775       3.98   0 P
ide(Anim).leg(Week,2UnStruct     1   2.69488        2.69488       3.18   0 P
ide(Anim).leg(Week,2UnStruct     2   2.17389        2.17389       3.01   0 P
ide(Anim).leg(Week,2UnStruct     1 -0.153679      -0.153679      -0.43   0 P
ide(Anim).leg(Week,2UnStruct     2  0.392403E-01  0.392403E-01   0.12   0 P
ide(Anim).leg(Week,2UnStruct     3  0.284981       0.284981       1.33   0 P
Warning: Code B - fixed at a boundary (!GP)       F - fixed by user
              ? - liable to change from P to B    P - positive definite
```

```
              C - Constrained by user (!VCC)      U - unbounded
              S - Singular Information matrix
 S means there is no information in the data for this parameter.
 Very small components with Comp/SE ratios of zero sometimes indicate poor
           scaling.  Consider rescaling the design matrix in such cases.
 Covariance/Variance/Correlation Matrix UnStructured
  7.152     0.9243    -0.4292    -0.4204
  3.827      2.397    -0.4602    -0.4854
-0.4676    -0.2903     0.1660    -0.1503
 -1.100    -0.7354    -0.5993E-01 0.9576
 Covariance/Variance/Correlation Matrix UnStructured
  4.468     0.8647    -0.1362
  2.695      2.174     0.4985E-01
-0.1537     0.3924E-01 0.2850
```

.res File

```
 leg(Week,3)                          has  4  levels
     0.500     0.707   -1.225    1.581   -1.871
     1.000     0.707   -1.118    1.187   -0.998
     1.500     0.707   -1.012    0.828   -0.318
     2.000     0.707   -0.905    0.505    0.186
     2.500     0.707   -0.799    0.218    0.533
     3.000     0.707   -0.692   -0.033    0.742
     3.500     0.707   -0.586   -0.248    0.830
     4.000     0.707   -0.479   -0.427    0.818
     4.500     0.707   -0.373   -0.571    0.722
     5.000     0.707   -0.266   -0.678    0.562
     5.500     0.707   -0.160   -0.750    0.356
     6.000     0.707   -0.053   -0.786    0.122
     6.500     0.707    0.053   -0.786   -0.122
     7.000     0.707    0.160   -0.750   -0.356
     7.500     0.707    0.266   -0.678   -0.562
     8.000     0.707    0.373   -0.571   -0.722
...
    11.500     0.707    1.118    1.187    0.998
    12.000     0.707    1.225    1.581    1.871
```

```
leg(Week,2)                       has  3  levels
    0.500    0.707   -1.225    1.581
    1.000    0.707   -1.118    1.187
    1.500    0.707   -1.012    0.828
    2.000    0.707   -0.905    0.505
    2.500    0.707   -0.799    0.218
    3.000    0.707   -0.692   -0.033
    3.500    0.707   -0.586   -0.248
    4.000    0.707   -0.479   -0.427
    4.500    0.707   -0.373   -0.571
    5.000    0.707   -0.266   -0.678
    5.500    0.707   -0.160   -0.750
    6.000    0.707   -0.053   -0.786
    6.500    0.707    0.053   -0.786
    7.000    0.707    0.160   -0.750
    7.500    0.707    0.266   -0.678
    8.000    0.707    0.373   -0.571
    8.500    0.707    0.479   -0.427
    9.000    0.707    0.586   -0.248
    9.500    0.707    0.692   -0.033
   10.000    0.707    0.799    0.218
   10.500    0.707    0.905    0.505
   11.000    0.707    1.012    0.828
   11.500    0.707    1.118    1.187
   12.000    0.707    1.225    1.581
```

Covariance Matrix

- `.rsv` The restart values file is convenient place to grab the variance components.

```
 4045 30 985748 4673
   0.000000       0.000000       0.000000       0.000000       0.000000     -0.1000000E-36 -0.1000000E-36   1.0
  0.1000000E-06  0.5688098       1.137154       2.355232       2.147434       1.145474       7.152305       3.8
   2.396900     -0.4676474     -0.2903185      0.1660231      -1.100170     -0.7353516     -0.5993249E-01  0.95
   4.467746       2.694879       2.173887     -0.1536786      0.3924026E-01  0.2849811
RSTRUCTURE                1   2
 VARIANCE                 1   1   0 1.00000
 STRUCTURE              386   0   0
 STRUCTURE                6   6   80.100000E-060.568810     1.13715     2.35523
 2.14743     1.14547
ide(Anim).leg(Week,2      2   0
 STRUCTURE              419   0   0
 STRUCTURE                3   6   9 4.46775     2.69488     2.17389    -.153679
0.392403E-010.284981
...
```

varfunc.sas

```
data leg3;
    input x p0 p1 p2 p3;
    cards;
     0.500    0.707   -1.225    1.581   -1.871
     1.000    0.707   -1.118    1.187   -0.998
     1.500    0.707   -1.012    0.828   -0.318
...
    12.000    0.707    1.225    1.581    1.871
data leg2;
    input x p0 p1 p2;
    cards;
     0.500    0.707   -1.225    1.581
     1.000    0.707   -1.118    1.187
...
    12.000    0.707    1.225    1.581
```

```
proc iml;

     start vec2var(vector);
       nd=ncol(vector);
       n=(-1+sqrt(1+8*nd))/2;
       matrix=j(n,n,0);
       idx=0;
       do i=1 to n;
           do j=1 to i;
               idx=idx+1;
               matrix[i,j]=vector[idx];
               matrix[j,i]=vector[idx];
               end;
           end;
           return(matrix);
     finish vec2var;
```

```
use leg3;
read all var {p0 p1 p2 p3} into P3;
read all var {x};
use leg2;
read all var {p0 p1 p2} into P2;
print x P3 P2;
xidx=2*{2 3 4.5 6 8 12};
xp=[xidx,];
print tp;
```

```
    RC={0.1000000E-06  0.5688098       1.137154
        2.355232       2.147434       1.145474 };
    GPC={7.152305       3.827207
        2.396900     -0.4676474     -0.2903185      0.1660231
        -1.100170     -0.7353516     -0.5993249E-01  0.9576310};
    RPC={4.467746       2.694879       2.173887     -0.1536786
        0.3924026E-01  0.2849811};
    RD=diag(RC);
    GP=vec2var(GPC);
    RP=vec2var(rpc);
    GCov=P3[xidx,]*GP*P3[xidx,]`;
    RCov=P2[xidx,]*Rp*P2[xidx,]`+Rd;
    print Gcov RCov;
    quit;
run;
```

Splines

- Smooth piecewise cubic polynomials which join at knots
- Knots at endpoints (3 knots at 2, 8, and 12 weeks)

```
!SPLINE spl(Week,3) 2 8 12
```

- `spl(x,k)`: construct spline with k knots
  - k-2 terms
    $$\boldsymbol{Z}_i = \begin{pmatrix} Z_{i3} & \cdots & Z_{ik} \end{pmatrix}$$
  - need to include the $a + b\,x$ linear component
    $$\begin{pmatrix} 1 & X_i \end{pmatrix}$$

```
M42_rr.dat !SKIP 1 !MAXIT 100 #Data file
!SPLINE spl(Week,3) 2 8 12
Weight ~ fac(Week).Generation*Gender fac(Week).Generation.day   ,
  !r   ![ Anim Week.Anim !] spl(Week,3).Anim  ![ ide(Anim) Week.ide(Anim) !],
   spl(Week,3).ide(Anim) !f mv
1 2 4
386 0 IDEN
6 0 DIAG 1 2 3 4 5 6 !S2==1 !GP
Anim 2
2 Week US !GP !+3
16
9 6
419 0 AINV
spl(Week,3).Anim 1
0 0 AINV 1 !GP
ide(Anim) 2
2 Week US !GP !+3
1
0 1
419 0 I
spl(Week,3).ide(Anim) 1
0 0 I 1
```

Variance Function

$$G = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a21} & \sigma_{a2}^2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix}' + Z \begin{pmatrix} \sigma_{as}^2 & & & \\ & \sigma_{as}^2 & & \\ & & \ddots & \\ & & & \sigma_{as}^2 \end{pmatrix} Z'$$

$$G = \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix} \begin{pmatrix} \sigma_{a1}^2 & \sigma_{a12} \\ \sigma_{a21} & \sigma_{a2}^2 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 1 & 3 \\ \vdots & \\ 1 & 12 \end{pmatrix}' + ZZ'\sigma_{as}^2$$

Genetic Merit

$$\boldsymbol{a}_i = \boldsymbol{1} a_{i1} + \boldsymbol{x} a_{i2} + \boldsymbol{Z} \begin{pmatrix} a_{i3} \\ \vdots \\ a_{ik} \end{pmatrix}$$

- `.res` contains the $\boldsymbol{Z}$ and $\boldsymbol{x}$ values
- `.sln` contains the solutions

predfun.sas

```
data spl3;
    int=1;
    input x sp1;
    cards;
0.500    2.547
1.000    2.026
1.500    1.506
....
11.500    0.831
12.000    1.477

data sln int slope spline;
length level $ 15;
length factor $20;
infile "h:/Mouse/mouse_42_spline.sln" expandtabs;
input factor $ level $ sol se;
output sln;
if factor="Anim" then output int;
if factor="Week.Anim" then output slope;
if factor="spl(Week,3).Anim" then output spline;
run;
```

```
proc iml;
    use  spl3;
    read all into Z;
    use sln;
    read all;
    use int;
    read all var sol into aint;
    read all var level into Anim;
    use slope;
    read all var sol into aslope;
    use spline;
    read all var sol into aspline;
    GM=aint||aslope||aspline;
    Zx=Z[2 3 4.5 6 8 12*2,];
    print Zx;
    GW=GM*Zx‘;
    week=W2 W3 W4_5 W6 W8 W12;
    print GW[rowname=Anim colname=week];
    create Gdata from GW[colname=week];
    append from GW;;
    create Anim var Anim;
    append;
    quit;
run;
```

```
data Gdata;
    merge Anim Gdata;
run;
```

```
                                    GW
                W2        W3      W4_5        W6        W8       W12

4636 0.1474189 0.0339654  -0.11938 -0.229476 -0.263939  0.077499
5002 0.1474189 0.0339654  -0.11938 -0.229476 -0.263939  0.077499
6515 0.2948575   0.06794 -0.238763  -0.45896 -0.527868 0.1551275
5306 -0.088076 -0.282496 -0.561779 -0.809342 -1.057036 -1.251446
4531 -0.088076 -0.282496 -0.561779 -0.809342 -1.057036 -1.251446
6696 -0.176172 -0.564971 -1.123482  -1.61857 -2.113953  -2.50293
4486 0.0914031 0.2157538 0.3876952 0.5221701 0.6041539 0.4126174
                   .
                   .
                   .
```