Introduction to Mixed Models

- Linear Models
- Fixed and Random Effects
- Fixed Models
  - Hypothesis testing
  - Estimation
- Mixed Models
  - Hypothesis testing
  - Prediction

Linear Models

- Needed to correct for unbalancedness in data
  - Different sires in different herds
  - Account for accuracy of information

Genetic level confounded with herds

Problem:

The contemporaries of some animals may have higher genetic mean than of others

Example

progeny means from 4 sires in 2 herds

sire	***1***	***2***	***3***	***4***	5: link sire
herd 1	**325**	**275**	**-**	**-**	*325*
herd2	**-**	**-**	**325**	**275**	*375*

Genetic level confounded with herds

Problem 3:

The contemporaries of some animals may have higher genetic mean than of others

Example

progeny means from 4 sires in 2 herds

sire	***1***	***2***	***3***	***4***	5: link sire
herd 1	**325**	**275**	**-**	**-**	*325*
herd2	**-**	**-**	**325**	**275**	*375*

Genetic level confounded with herds

Conclusion

Need links between herds (reference sires)

Need a simultaneous evaluation of all herd and sire effects

The argument for BLUP sire evaluation…..1970ties

Another example

Cow	Breed	Feeding regime	Weight (kg)
1	Angus	intensive	494
2	Angus	intensive	556
3	Angus	extensive	542
4	Hereford	extensive	473
5	Hereford	intensive	632
6	Hereford	extensive	544

Order data

	Intensive	Extensive	Mean
Angus	494 556	542	531
Hereford	632	473, 544	550
Mean	561	520	

Using a Linear Model

[X y] =

1	1	1	494
1	1	1	556
1	1	-1	542
1	-1	-1	473
1	-1	1	632
1	-1	-1	544

$$y = X b + e$$

$$= m + breed + feed$$

class variables

b = inv(X'*X)*X'*y =

540.1667	
-18.3750	Angus = - Hereford
26.6250	Intensive = - Extensive

Using a Linear Model

[X y] =

1	18	494
1	21	556
1	19	542
1	17	473
1	23	632
1	19	544

$$y = X\,b + e$$

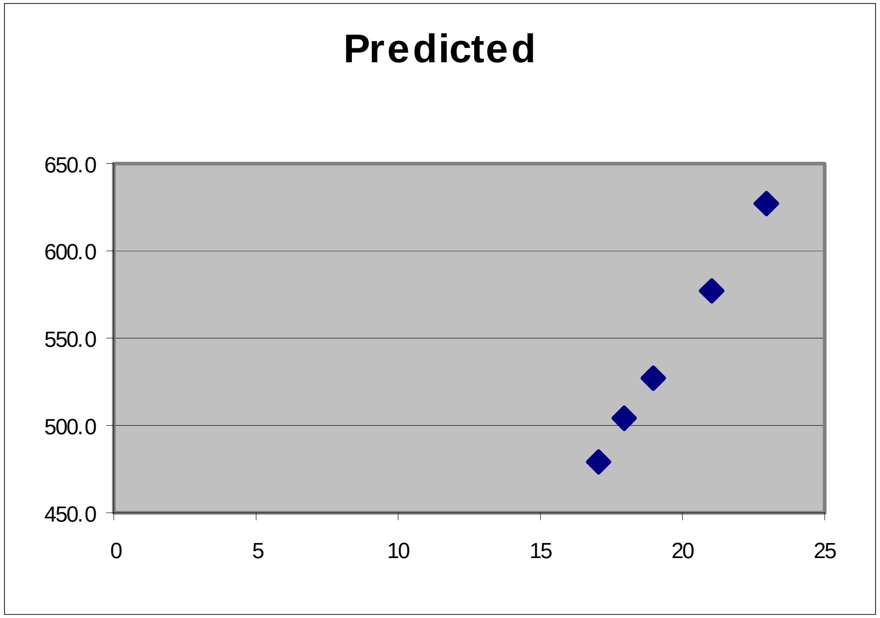
$$= m + age$$

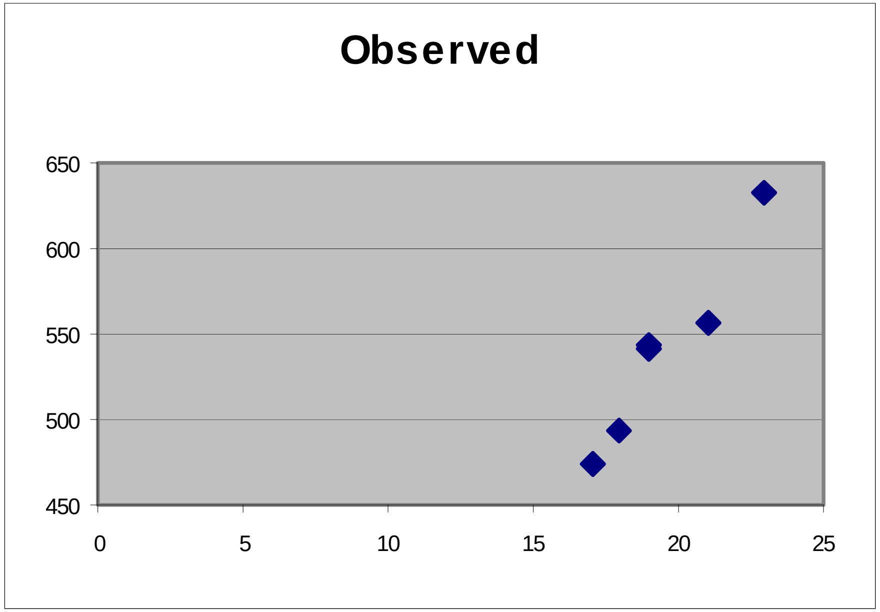
continuous variables

b = inv(X'*X)*X'*y =

59 intercept: weight at 0 months
24 slope: weight change per month

A linear regression model





Using a Linear Model

X =

1	1	1	18
1	1	1	21
1	1	-1	19
1	-1	-1	17
1	-1	1	23
1	-1	-1	19

b = inv(X'*X)*X'*y =

-11.3522	weight at age = 0
-0.6981	breed effect
-12.2642	feeding effect
28.2830	age effect

y = X b + e

= m + breed + feed + age

class variables

and

continuous variables

Fixed and Random Effects

Fixed Effects

- Defined classes, comprise all the possible levels of interest
- Number of levels relatively small and confined to this number after repeated sampling.
- E.g. sex, age, breed, contemporary group

Random Effects

- Levels that are considered to be drawn from an infinite large population of levels
- E.g. animal effects

The linear model

- Equation $y = Xb + Zu + e$
- Expectations and Variance Structures
- Assumptions and restrictions
  - Residuals IID?
  - Random sampling of u?

Expectations and Variance Structures

$$E\begin{pmatrix} y \\ u \\ e \end{pmatrix} = \begin{pmatrix} Xb \\ 0 \\ 0 \end{pmatrix}$$

$$\mathrm{Var}\begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix}$$

$$\mathrm{var}(y) = V = ZGZ' + R$$

Estimating Fixed effects

$b = (X'X)^{-1} X'y$ Ordinary Least Squares

$V = I \sigma_e^2$

$b = (X'V^{-1} X)^{-1} X'V^{-1} y$ Weighted Least Squares

$V = D \sigma_e^2$

$b = (X'V^{-1} X)^{-1} X'V^{-1} y$ Generalized Least Squares

$V = V$

A simple example of variance structure

animal	obs'n
1	9
1	11
2	10

$$Z = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

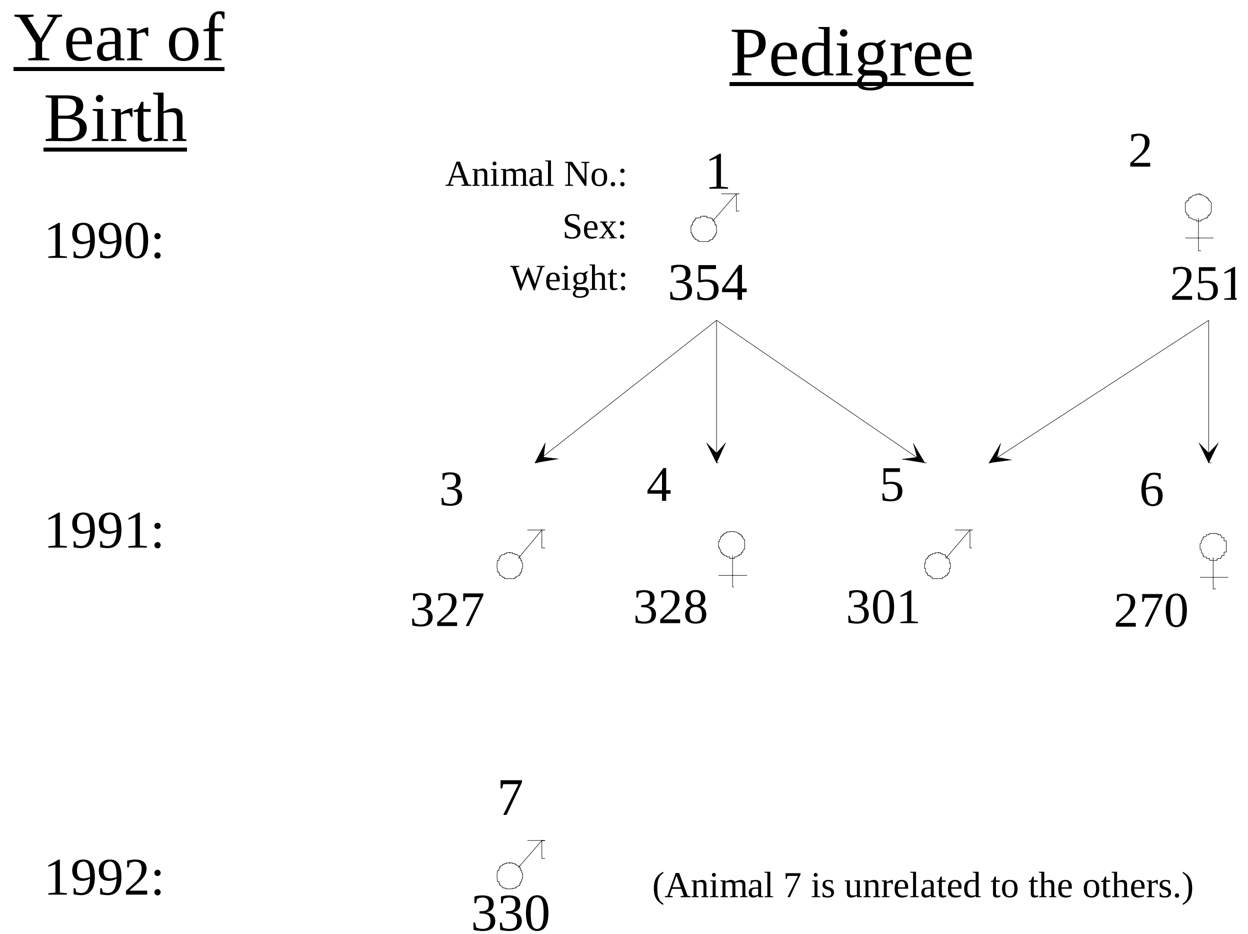
$$Z'Z = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \qquad ZZ' = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \sigma^2_{anm}$$

var(y) =ZGZ' + R $= \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma^2_{anm} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \sigma^2_e$

Estimability

example:



$$Y = X b + e$$

$$\begin{pmatrix} 354 \\ 251 \\ 327 \\ 328 \\ 301 \\ 270 \\ 330 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} b_{mean} \\ b_{1990} \\ b_{1991} \\ b_{1992} \end{pmatrix} + \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \end{pmatrix}$$

X is dependent

X'X can not be inverted

Can only estimate 3 parameters from 3 means

solutions:

$$\hat{b} = X'X^{-}X'Y$$

$$X'X = \begin{pmatrix} 7 & 2 & 4 & 1 \\ 2 & 2 & 0 & 0 \\ 4 & 0 & 4 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \text{ and } \quad X'y = \begin{pmatrix} 2161 \\ 605 \\ 1226 \\ 330 \end{pmatrix}$$

Need restriction to solution

$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix}$$

e.g. put effect of 1992 to zero

$$\hat{b} \quad = \quad (X'X)^{-1} \quad X'Y \quad = \quad \text{solutions}$$

$$\hat{b} = \begin{pmatrix} 7 & 2 & 4 \\ 2 & 2 & 0 \\ 4 & 0 & 4 \end{pmatrix}^{-1} \begin{pmatrix} 2161 \\ 605 \\ 1226 \end{pmatrix} = \begin{pmatrix} 330 \\ -27.5 \\ -23.5 \end{pmatrix}$$

There are more solutions possible

General mean zero	**First year zero**	**Last year zero**	**Sum of years to zero**
μ = 0	**μ = 302.5**	**μ = 330**	**μ = 313**
1990 = 320.5	**1990 = 0**	**1990 = -27.5**	**1990 = -10.5**
1991 = 306.5	**1991 = +4**	**1991 = -23.5**	**1991 = -6.5**
1992 = 330	**1992 = +27.5**	**1992 = 0**	**1992 = 17**

→ estimable functions are unchanged

- expected value of an observation
- difference between years

fitting mean and year effect

$$\hat{b} = \begin{pmatrix} 330 \\ -27.5 \\ -23.5 \end{pmatrix} \quad \boxed{-27.5, -23.5} \rightarrow 4.0$$

the mean of 1992
the effect of year 1990 (relative to 1992)
the effect of year 1991 (relative to 1992)

fitting mean, year effect and sex

X	$\hat{b}$	meaning
$\begin{pmatrix} 1 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 285.7 \\ -5.3 \\ -1.3 \\ 44.3 \end{pmatrix}$	the mean of females in 1992 the effect of year 1990 (relative to 1992) the effect of year 1991 (relative to 1992) the effect of males (relative to females)

$\boxed{-5.3, -1.3} \rightarrow 4.0$

Note: year 1992 appears not so good after all!

Conclusion

- Linear models are a powerful, and relatively simple way to correct for different fixed effects in unbalanced designs

Connectedness and Confounding

Year \ sex	Male	Steer	Female
1990	1	0	1
1991	2	0	2
1992	0	1	0

Connectedness and Confounding

1990	Male	316
1990	Female	314
1990	Male	312
1990	Male	324
1991	Female	311
1991	Male	312
1991	Female	293
1991	Female	304

model statement: weight ~ mu con(sex) con(year)

con(year)	1	2.06	2.06	[DF F_i F_a]
con(sex)	1	4.36	1.19	[DF F_i F_a]

model statement: weight ~ mu con(year) con(sex)

6 con(sex)	1	1.19	1.19	[DF F_i F_a]
5 con(year)	1	5.23	2.06	[DF F_i F_a]

Exmp4.dat

15	109	287
17	116	298
18	119	306
18	116	303
19	117	302
19	119	312
20	121	316
21	122	324

weight ~ mu height age

1 age	1	3.03	3.03	[DF F_inc F_all]
2 height	1	70.50	1.67	[DF F_inc F_all]

weight ~ mu age height

2 height	1	1.67	1.67	[DF F_inc F_all]
1 age	1	71.87	3.03	[DF F_inc F_all]

Hypothesis Testing

- H'b = c
- Or: H'b-c = 0

$$F = \frac{s / r(H')}{SSE / (N - r(X))}$$

where s = (H'b-c)'(H'CH)$^{-1}$(H'b-c)

and C = (X'V $^{-1}$ X)$^{-1}$

Accuracy of estimates

- $V(K'b) = K'(X'V^{-1}X)^{-} K \; se2$

Mixed model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

var(u) = G

var(e) = R

var(y) = ZGZ’ + R

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

simple version

$\text{var(u)} = A\,\sigma_a^2$

$\text{var(e)} = I\,\sigma_e^2$

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + \lambda A^{-1} \end{bmatrix} \begin{bmatrix} b \\ u \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix}$$

In the MMM we estimate

BLUE(b) = β

= (X' V-1X)-1 X' V-1y

is a GLS estimate

BLUP(u)= û

$=(Z' R^{-1} Z+G^{-1})^{-1} Z' R^{-1}(y-X\beta)$

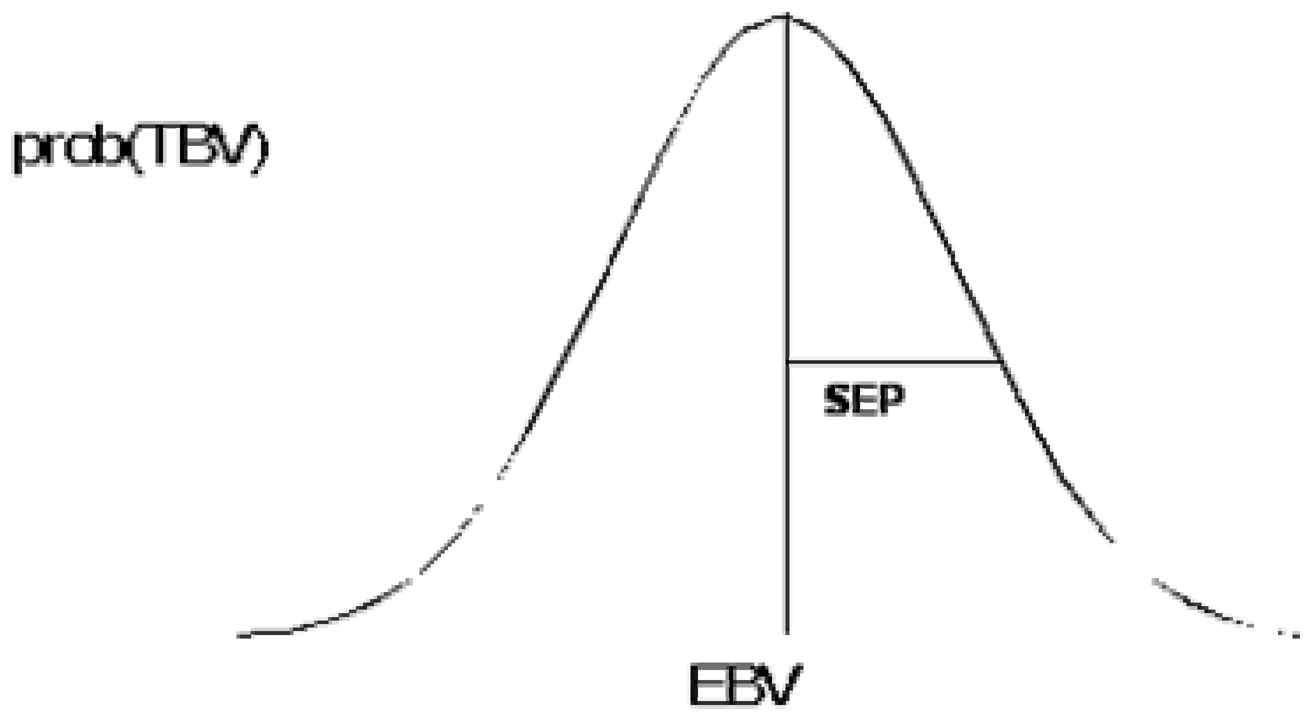
Accuracy of Estimates

$$Var(b) = C_{XX}$$

$$V(\hat{u}) \quad = G - C_{ZZ}$$

And $$V(u - \hat{u}) = C_{ZZ}$$

Prediction Error variance



Hypothesis Testing in Mixed Models

- Not well defined
- Can
  - Ignore random effects
  - Treat them as fixed
  - Estimate them