Models

LRS

CGIL

July-Aug 2012

A Complete Linear Model

1. The equation
2. Expectations and variances
3. Assumptions and limitations

All 3 parts must be present, to understand results from an analysis of the data.

A Traditional Model

Obs. $=$ Fixed Factors $+$ Random Factors $+$ Residual

Fixed Factors

- Age - finite possibilities
- Gender, Diet, Parity, Birth type
- Cage, Tank, Times milked, Breed
- Years

Random Factors

- Animals, Contemporary groups, Litters
- Dams, Herds, Flocks
- Anything that can be visualized to have been sampled randomly from a conceptually large population.

A Traditional Model

Fixed Factors

If you want to make inferences about these specific levels of a factor, such as males are always heavier than females, then that factor should be fixed.

Random Factors

If you want to extrapolate to all flocks of sheep, even those not included in your data, then flocks should be considered random.

A Traditional Model

- Factors should be randomly associated with other factors in the model. The concept of complete random sampling of observations should be true.
- If records of poorer animals are removed from the data then this could bias estimates of differences within factors.

Part 1 - Equation

An example equation is written like

$$y_{ijklm} = A_i + B_j + X_k + HYS_l + c_m + e_{ijklm},$$

where

- y_{ijklm} is a 200-d weaning weight on a calf,
- A_i is the age of the dam (in years), either 2, 3, 4, or 5 and greater,
- B_j is a breed of calf effect,
- X_k is a gender of calf effect (male or female),
- HYS_l is a herd-year-season of birth effect, with three seasons per year (i.e. Nov-Feb, Mar-Jun, and Jul-Oct),
- c_m is a calf additive genetic effect, and
- e_{ijklm} is a residual effect.

Part 2 - Expectations, Variances

$$
\begin{aligned}
E(y_{ijklm}) &= A_i + B_j + X_k \\
E(HYS_l) &= 0 \\
E(c_m) &= 0 \\
E(e_{ijklm}) &= 0 \\
Var(HYS_l) &= 0.15\ \sigma_y^2 \\
Var(c_m) &= 0.35\ \sigma_y^2 \\
Var(e_{ijklm}) &= 0.50\ \sigma_y^2
\end{aligned}
$$

Part 3 - Assumptions, Limitations

- There are no interactions between age of dam, breed of calf, or gender of calf.
- The weaning weights have been properly adjusted to 200-d of age.
- No maternal effects, no heterosis effects.
- Age of dam is known.
- All calves in the same herd-year-season were raised and managed in the same manner.
- All natural births, data covers same years in each herd, all calves weaned near 200-d.
- Pedigrees are known, so additive relationships included.

Model Building

- Read the literature, make a list of potential factors from each paper.
- Discuss factors with advisor, other colleagues.
- Are all factors recorded in your data?
- Try preliminary analyses to test factors, interactions in your data.
- Hypothesize a model, try it out.
- Try another model.

The process is iterative: propose, test, evaluate, AND propose, test, evaluate.

Part 1 - Equation

Temperament scores on dairy cows:

$$y_{ijklm} = YS_i + c_j + A_k + s_l + e_{ijklm},$$

where

- y_{ijklm} is a score from 1 (easy to handle) to 40(very difficult), for temperament at feeding,
- YS_i is a year-season of calving effect,
- c_{ij} is a contemporary group effect,
- A_k is an age group of cow effect,
- s_l is a sire of cow effect,
- e_{ijklm} is a residual effect.

Part 2 - Expectations, Variances

$$
\begin{aligned}
E(y_{ijklm}) &= YS_i + A_k \\
E(c_{ij}) &= 0 \\
E(s_l) &= 0 \\
E(e_{ijklm}) &= 0 \\
Var(c_{ij}) &= 0.15\ \sigma_y^2 \\
Var(s_m) &= 0.06\ \sigma_y^2 \\
Var(e_{ijklm}) &= 0.79\ \sigma_y^2
\end{aligned}
$$

Part 3 - Assumptions, Limitations

- Cows were at same stage of lactation.
- Scoring was done in same manner for all cows.
- Sires were unrelated to each other.
- Sires were mated randomly to dams.
- Only one offspring per dam.

Part 1 - Equation

Ultrasound loin depth of sheep at 100 days:

$$y_{ijklm} = YM_i + FYS_j + L_k + a_l + e_{ijklm},$$

where

y_{ijklm} is a loin depth in mm on a lamb at 100 d,

YM_i is a year-month of birth effect,

FYS_j is a flock within year-season effect,

L_k is a litter effect,

a_l is a lamb additive genetic effect,

e_{ijklm} is a residual effect.

Part 2 - Expectations, Variances

$$
\begin{aligned}
E(y_{ijklm}) &= YM_i \\
E(FYS_j) &= 0 \\
E(L_k) &= 0 \\
E(a_l) &= 0 \\
E(e_{ijklm}) &= 0 \\
Var(FYS_j) &= 0.15\ \sigma_y^2 \\
Var(L_k) &= 0.06\ \sigma_y^2 \\
Var(a_l) &= 0.30\ \sigma_y^2 \\
Var(e_{ijklm}) &= 0.49\ \sigma_y^2
\end{aligned}
$$

Part 3 - Assumptions, Limitations

- Lambs close to 100-d age, no age adjustments needed.
- No differences among makes of ultrasound machines.
- Lambs are related, **A** to be used.
- Flock divided into management groups, identified.
- Only one record per lamb.
- No breed differences.
- Age of dam and maternal effects not important.

Matrix Notation for Models

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

where

$\mathbf{y}$ is the observation vector, $N \times 1$

$\mathbf{b}$ is a vector of levels of fixed factors, $p \times 1$

$\mathbf{u}$ is a vector of levels of random factors, $q \times 1$

$\mathbf{e}$ is a vector of residual effects, $N \times 1$

$\mathbf{X}, \mathbf{Z}$ are design (incidence) matrices

$$\begin{aligned} E(\mathbf{y}) &= \mathbf{Xb} \\ E(\mathbf{u}) &= \mathbf{0}_{q\times 1} \\ E(\mathbf{e}) &= \mathbf{0}_{N\times 1} \end{aligned}$$

Matrix Notation

$$\begin{aligned} Var(\mathbf{u}) &= \mathbf{G} \\ Var(\mathbf{e}) &= \mathbf{R} \\ Var(\mathbf{y}) &= \mathbf{ZGZ}' + \mathbf{R} \end{aligned}$$

Example

Horse	Colour	Height
1	Chestnut	10.5
2	Black	11.0
3	Black	9.6
4	Chestnut	10.2
5	Grey	10.0

$$Height_{ij} = \mu + Colour_i + e_{ij}$$

$$E(y_{ij}) = \mu + C_i$$

$$\mathbf{b} = \begin{pmatrix} \mu \\ C_c \\ C_b \\ C_g \end{pmatrix}, \quad \mathbf{u} = na$$

Design Matrix

$$\mathbf{y} = \begin{pmatrix} 10.5 \\ 11.0 \\ 9.6 \\ 10.2 \\ 10.0 \end{pmatrix} = \mathbf{Xb} + \mathbf{e} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu \\ C_c \\ C_b \\ C_g \end{pmatrix} + \mathbf{e}$$

```
y = matrix(data=c(10.5,11.0,9.6,10.2,10.0),ncol=1)
X = matrix(data=c(1,1,0,0,
                  1,0,1,0, 1,0,1,0, 1,1,0,0,
         1,0,0,1),byrow=TRUE,ncol=4)
```

Design Function

```
desgn = function(v,nc){
  if(is.numeric(v)){
   va=v
   mrow = length(va)
   mcol = max(va)   }
 if(is.character(v){
   vf = factor(v)
   va = as.numeric(vf)
   mrow = length(v)
   mcol = length(levels(vf)) }
 if(nc > mcol)mcol = nc
 X = matrix(data=c(0),nrow=mrow,ncol=mcol)
 for( i in 1:mrow){
    ic=va[i]
    X[i,ic]=1 }
 return(X) }
```

Usage

```
 colour = c("C","B","B","C","G")
  Xc = desgn(colour,0)
  Xm = rep(1,5)
  X = cbind(Xm,Xc)
# OR
  X = matrix(data=c(1,1,0,0,
                   1,0,1,0, 1,0,1,0, 1,1,0,0,
          1,0,0,1),byrow=TRUE,ncol=4)
```