OHP Picture 1

①

Recap   2024-05-13

☐ Repeated Observations
=> New class of models: Linear Mixed Effects
                                     Model (LME)

Before : Fixed Linear Effects Models
    • Regression coefficients ⎫ fixed effects
    • Factor effects   ⎬ not associated
                                         with variation

    • Random error /residuals
        ↳ variation; can model by $var(e) \Rightarrow I \cdot \bar{\sigma}_e^2$

LME: Additional random effect in the
     model
     => Account for specific structure
         in variation pattern of data
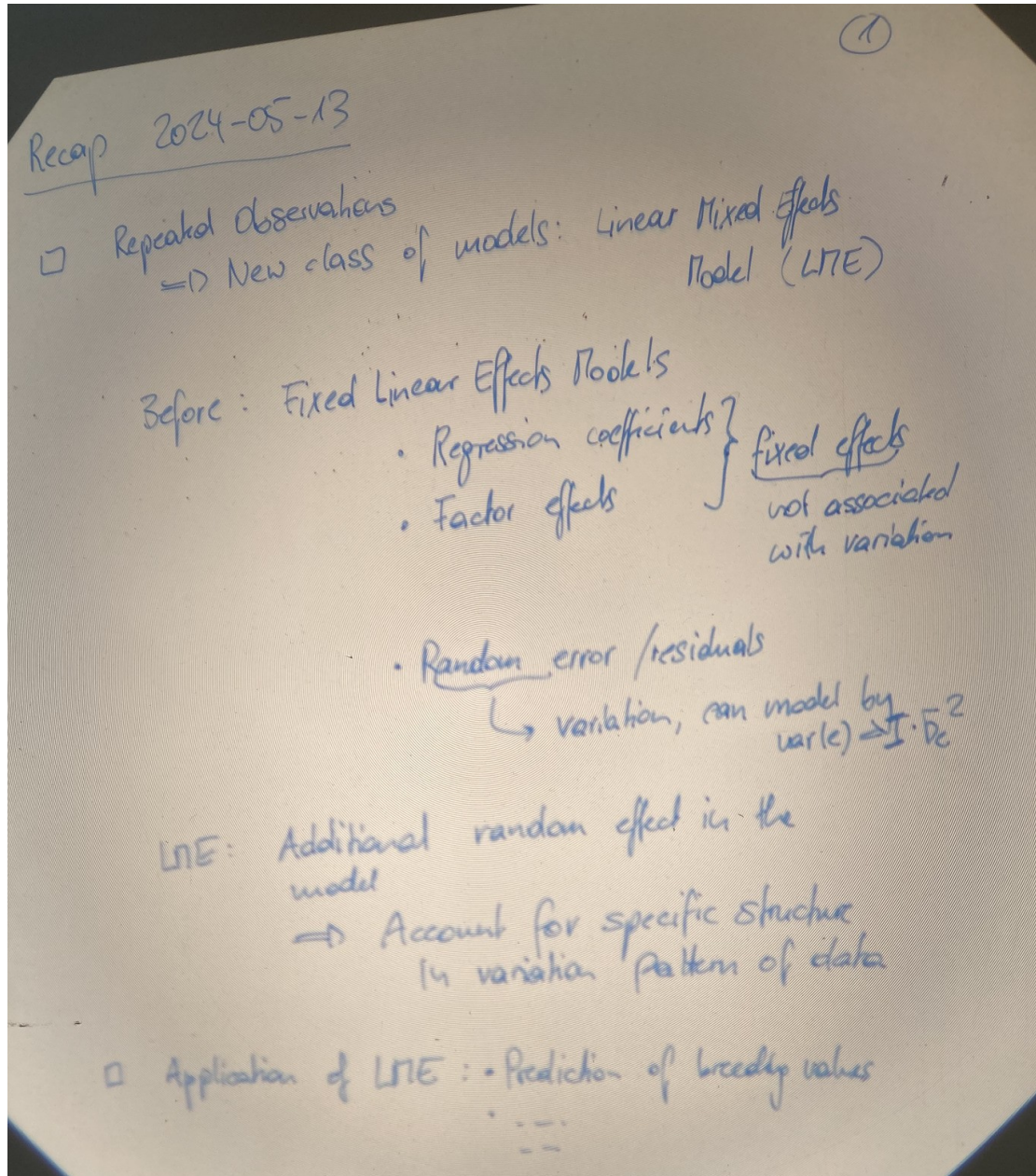
☐ Application of LME : • Prediction of breeding values

OHP Picture 2

②

□ Parents pass random sample of alleles to offspring. ⇒ Selection on phenotype y_i is inefficient

□ Selection criterion: "Value" of the alleles passed from parents to offspring ⇒ Breeding value

□ Breeding value (u) cannot be observed; what is commonly recorded are phenotypic measurements.

□ Connection between phenotypic records and unknown breeding values ?

□ Decompose phenotypes $y_i = g_i + e_i$ ⟶ Environment of animal i

Genotype of animal i

□ Differentiate effect of g_i on phenotype:

breeding value ← { • additive effect of a single allele on phenotype
   • dominance effect: single-locus genotype on phenotype
   • Interaction between different loci

3

OHP Picture 3

□ Decomposition of genotype:
$$g_i = BV_i + D_i + I_i$$

↓ Breeding value

□ Insert into model for y_i:

$$y_i = \mu + g_i + e_i$$
$$= \mu + BV_i + D_i + I_i + e_i$$

□ For selecting parents only BV_i is of interest

→ Regroup:
$$y_i = \mu + BV_i + \underline{D_i + I_i} + e_i$$
$$y_i = \mu + BV_i + e_i^*$$

□ Split environment into fixed known part and random unknown part

e.g. · herd
· season
· days in milk

$$y_i = \mu + BV_i + e_i^*$$
$$y_i = x_i^T b + u_i + e_i \to \text{Aggregation over all animals}$$

OHP Picture 4

④

□ $y_i = x_i^v b + u_i + e_i$; for animal

□ Aggregate over all animals in population and use matrix-vector notation

$$y = Xb + Zu + e$$

→ length q with random breeding values
→ length n with random residuals
→ length p with fixed effects
→ length n with observations

□ LME with random terms u and e as a consequence also y is random

□ In LME, expected values and variance-covariance matrices of all random terms must specified

▷ vector u of breeding values are defined as deviations ⟹ $E[u]^v = \begin{bmatrix} E(u_1) \\ E(u_2) \\ \vdots \\ E(u_q) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \underline{0}$

▷ residuals e : $E[e] = \underline{0}$

▷ $E[y] = E[Xb + Zu + e] = E[Xb] + E[Zu] + E[e]$
$= X E[b] + Z E[u] + E[e] = Xb$

5

⑤

□ Variance–Covariance Matrix

▸ Residuals: $\text{var}(e) = \begin{bmatrix} \text{var}(e_1) & \text{cor}(e_1,e_2) & \cdots \\ \text{cov}(e_2,e_1) & \text{var}(e_2) & \cdots \\ & & \ddots & \\ & & & \text{var}(e_n) \end{bmatrix}$

$\text{var}(e_1) = \text{var}(e_2) = \ldots = \text{var}(e_n) = \bar{\sigma}_e^2$ "Residual variance component"

$\text{cor}(e_1,e_2) = \ldots \quad \varnothing$

$\Rightarrow \text{var}(e) = I * \bar{\sigma}_e^2 = R$

↳ Identity matrix

estimated from data

extracted $\bar{\sigma}_e$ as standard error in output from $\text{Lm}()$

▷ Breeding values u:

$\text{var}(u) = \begin{bmatrix} \text{var}(u_1) & \text{cov}(u_1,u_2) & \cdots \\ \text{cov}(u_2,u_1) & \text{var}(u_2) & \\ & & \ddots \\ & & & \text{var}(u_q) \end{bmatrix}$

$\text{var}(u) = U = A * \bar{\sigma}_u^2$

↙   additive genetic variance

relationship matrix

▸ $\text{var}(y) = V = ZUZ^T + R$

OHP Picture 6

△ Covariance :
$$cov(u, y^T) = cov(u, [Xb + Zu + e]^T)$$
$$= cov(u, b^T X^T) + cov(u, u^T Z^T)$$
$$+ cov(u, e^T)$$

$$= 0 + cov(u, u^T) \cdot Z^T + cov(u, e^T)$$
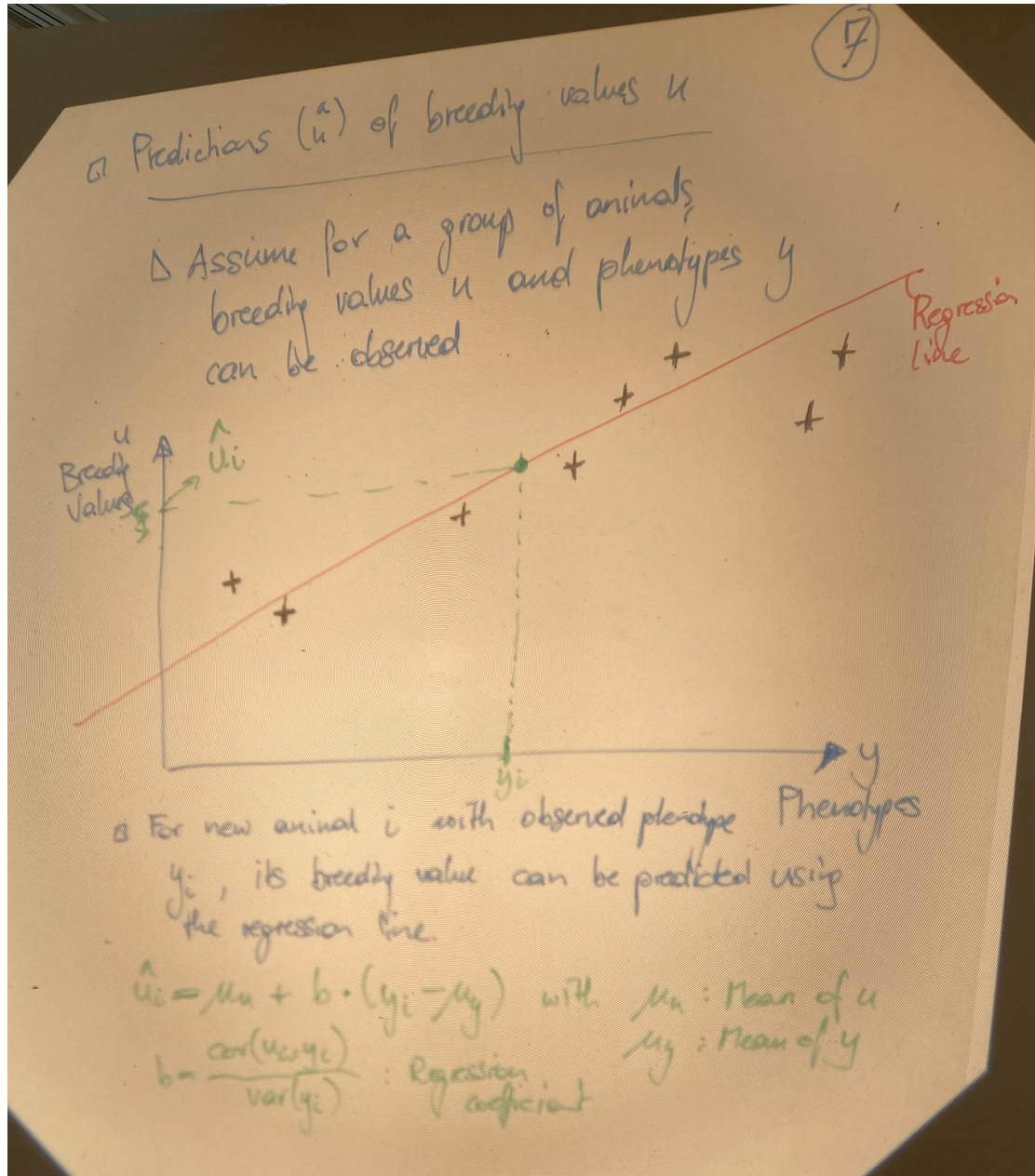
$$= 0 + var(u) \cdot Z^T + 0$$

$$= U \cdot Z^T$$

▢ Model $y = Xb + Zu + e$

known → ⟶ unknown

Goal : Use known components of a dataset
to get estimates ($\hat{b}$) of fixed effects b
and predicted breeding values ($\hat{u}$)

▢ Estimates $\hat{b} = (X^T V^{-1} X)^- X^T V^{-1} y$ ⌐see FLEM
↳ Least Squares

7

⑦

a) Predictions ($\hat{u}$) of breeding values u

△ Assume for a group of animals,
breeding values u and phenotypes y
can be observed

B For new animal i with observed phenotype
y_i, its breeding value can be predicted using
the regression line.

$$\hat{u}_i = \mu_u + b \cdot (y_i - \mu_y) \quad \text{with} \quad \mu_u : \text{Mean of } u$$
$$\mu_y : \text{Mean of } y$$

$$b = \frac{cov(u_i, y_i)}{var(y_i)} : \text{Regression coefficient}$$

OHP Picture 8

□ Generalisation

$$\hat{u}_i = \mu_u + b(y_i - \mu_y)$$

vectors $\vec{u}, e$ and y follow multivariate-normal distribution, then

$$\hat{u}_i = E(u_i | y_i) = E(u_i) + \frac{cov(u_i, y_i)}{var(y_i)} \{y_i - E(y_c)\}$$

□ Aggregation for all q animals: $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_i \\ \vdots \\ u_q \end{bmatrix}$

$$\hat{u} = E(u|y) = \underline{E(u)} + cov(u, y^T) \cdot var(y)^{-1} \cdot (y - E(y))$$

$$= \underline{0} + u \cdot Z^T \cdot V^{-1} \cdot (y - Xb)$$

$$= u \cdot Z^T \cdot V^{-1} (y - Xb)$$

unknown, replace it with $\hat{b}$

$$\hat{u} = u \cdot Z^T \cdot V^{-1} (y - X\hat{b})$$

□ $\hat{u}$ depends on V^{-1} which is difficult to compute $V = var(y)$ has dimensions $n \times n$. $\{n = 10^8$

OHP Picture 9

a Solutions via Mixed Model Equations (MME) ⑨

- Instead of solving $\hat{b} = (X^T V^{-1} X)^{-} X^T V^{-1} y$ and

$$\hat{u} = U Z^T V^{-1} (y - X\hat{b})$$

the following system of equations lead to comparable solutions

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + U^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

if $R = I \cdot \bar{\sigma}_e^2$, MME simplify to

$$R^{-1} = I \cdot \bar{\sigma}_e^{-2}$$

$$\begin{bmatrix} X^T I \bar{\sigma}_e^{-2} X & X^T I \bar{\sigma}_e^{-2} Z \\ Z^T I \bar{\sigma}_e^{-2} X & Z^T I \bar{\sigma}_e^{-2} Z + U^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T I \bar{\sigma}_e^{-2} y \\ Z^T I \bar{\sigma}_e^{-2} y \end{bmatrix} \Bigg| \cdot \bar{\sigma}_e^2$$

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda A^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix} \quad \{ \lambda = \bar{\sigma}_e^2 / \sigma_u^2 \}$$

10

OHP Picture 10

Examples

⊳ Sire model

LME where only sire get breeding values

$\left\{ \begin{array}{l} \text{Livestock Breeding} \\ \text{and Genomics} \checkmark \\ \quad \text{Fall 2024; Fri 9-12} \end{array} \right.$

⟹ Model :

$$y = Xb + Zs + e$$

Sire breeding values

$$E \begin{bmatrix} y \\ s \\ e \end{bmatrix} = \begin{bmatrix} Xb \\ \emptyset \\ \emptyset \end{bmatrix}$$

$$var \begin{bmatrix} y \\ s \\ e \end{bmatrix} = \begin{bmatrix} V & ZU & R \\ UZ^T & U & \emptyset \\ R & \emptyset & R \end{bmatrix}$$

$R = I \cdot \bar{\sigma}_e^2$

$U = A_s * \bar{\sigma}_s^2$; with A_s being the sire relationship matrix and $\bar{\sigma}_s^2$ is the sire variance component.

$V = ZUZ^T + R$

□ Special case: When sire are unrelated, then

$A_s = I$; identity matrix

□ Estimates $\hat{b}$ and predictions $\hat{s}$ are obtained via mixed model equations.

11

OHP Picture 11

Animals : LME, but u contains breeding values for all animals

△ Model : $y = Xb + Zu + e$, vector with g breeding values

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

$$var\begin{bmatrix} y \\ u \\ e \end{bmatrix} = \begin{bmatrix} V & ZU & R \\ UZ^T & U & 0 \\ R & 0 & R \end{bmatrix}$$

$R = I \cdot \sigma_e^2$

$U = A \cdot \sigma_u^2$ with A being the numerator relationship matrix and σ_u^2 the additive genetic variance

$V = ZUZ^T + R$

△ Predictions $\hat{u}$ and estimates $\hat{b}$ obtained from MME