Analysis of repeated records

Filippo Biscarini

(Biostatistician, bioinformatician and quantitative geneticist) **CNR**, Milan (Italy)

Repeated records?

- multiple milk production measurements for cows
- multiple blood pressure measurements of patients
- multiple Covid-10 test results for individuals
- repeated measurements of field coverage (crop growth)
- repeated measurements of root development (rhizotrons)
- etc.

Repeated records?

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

Repeated records - analysis

There are tons of ways to analyse repeated records

We'll see just one possible approach: the so called

- **repeatability model** (a.k.a. **repeated-records/measurements model**)

Originated - and still popular- in genetics, specifically plant and animal breeding

It can be applied to many other data problems with repeated measurements, though

Repeatability model

Borrowing from genetics:

$$\mathrm{P} = (\mathrm{G}) + \mathrm{PE} + \mathrm{E}$$

$$\sigma^2_P = (\sigma^2_G) + \sigma^2_{PE} + \sigma^2_E$$

$$\mathrm{P} = \mathrm{PE} + \mathrm{E}$$

$$\sigma^2_P = \sigma^2_{PE} + \sigma^2_E$$

Repeatability model

From the biological (physical) to the statistical model:

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

Linear mixed model:

- y, b, pe, e
- X, W
- Var(pe) = $\mathbf{I\sigma}_{pe}^2$
- Var(e) = $\mathbf{I\sigma}_{e}^2$

Repeatability model

MME: mixed model equations

y = Xb + Wpe + e

$$\begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{pe}^2} \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

$$\begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}'\mathbf{X} & \mathbf{X}'\mathbf{W} \\ \mathbf{W}'\mathbf{X} & \mathbf{W}'\mathbf{W} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{pe}^2} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{W}'\mathbf{y} \end{bmatrix}$$

Repeatability model: illustration

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

- same data as previous slide
- $y = Xb + Wpe + e$
- $y = ?$
- $X = ?$ [parity, HYS]
- $W = ?$

Repeatability model: illustration

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

$$y = Xb + Wpe + e$$

y = [you tell me!]

Repeatability model: illustration

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

$$y = Xb + Wpe + e$$

X = [you tell me!]

Repeatability model: illustration

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

$$y = Xb + Wpe + e$$

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Repeatability model: illustration - MME

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

y = Xb + Wpe + e

- **X'X**?

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Repeatability model: illustration - MME

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

y = Xb + Wpe + e

- **W'W**?

$$\mathbf{W} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Repeatability model: illustration - MME

Cow	Sire	Dam	Parity	HYS	Fat yield (kg)
4	1	2	1	1	201
4	1	2	2	3	280
5	3	2	1	1	150
5	3	2	2	4	200
6	1	5	1	2	160
6	1	5	2	3	190
7	3	4	1	1	180
7	3	4	2	3	250
8	1	7	1	2	285
8	1	7	2	4	300

HYS, herd–year–season. [From Mrode, 2013]

y = Xb + Wpe + e

- **W'X**?

$$\mathbf{X} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Coefficient of repeatability

$$\text{rep} = \frac{\sigma_{pe}^2}{\sigma_{pe}^2 + \sigma_e^2}$$

- **repeatability**: correlation between records of an individual
- permanent environmental effects

Repeatability model: extension to genetics

y = Xb + Zg + Wpe + e

$$\begin{bmatrix} \mathbf{X'X} & \mathbf{X'Z} & \mathbf{X'W} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{K}^{-1}\frac{\sigma_e^2}{\sigma_g^2} & \mathbf{Z'W} \\ \mathbf{W'X} & \mathbf{W'Z} & \mathbf{W'W} + \mathbf{I}\frac{\sigma_e^2}{\sigma_{pe}^2} \end{bmatrix} \cdot \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{g}} \\ \hat{\mathbf{pe}} \end{bmatrix} = \begin{bmatrix} \mathbf{X'y} \\ \mathbf{Z'y} \\ \mathbf{W'y} \end{bmatrix}$$

- covariance matrix **K**
- can be any covariance structure (e.g. spatial)