Applied Statistics with R: Mixed models

Cristina Sartori & Hugo Toledo-Alvarado

20-22 February 2018

cristina.sartori@unipd.it

https://github.com/Hugo-Toledo/Applied-Statistics-R-UNIPD

Università degli Studi di Padova

DAFNAE Department of Agronomy, Food, Natural resources, Animals and Environment

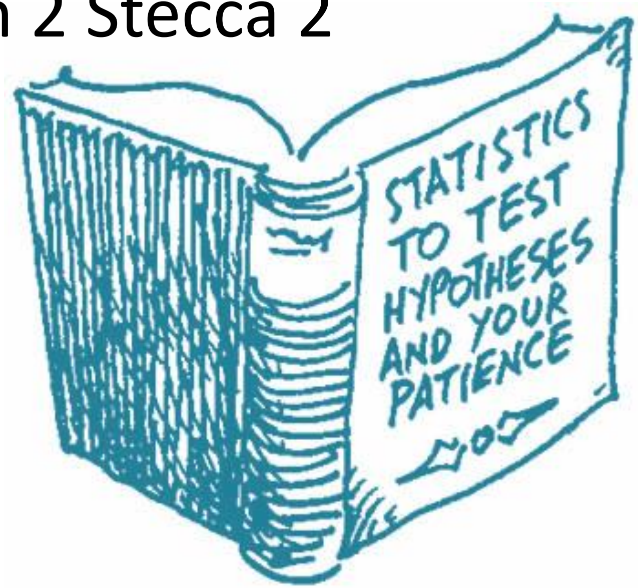
Ph.D. Course Animal & Food Science — PhD Course in Animal and Food Science

General information on the course

- Part III of the optional course «Applied Statistics with R: Mixed Models» for PhD students

- *When/Where:*
- February 22nd, 2-5 pm / Room 2 Stecca 1
- February 22nd, 11am-1.30 pm / Room 1 Stecca 2
- February 22nd, 2.30-6.30 pm / Room 2 Stecca 2
  - …sorry for the strange timetable and rooms,
  - we have to deal with the "Scegli il tuo domani" event…

- **9 hours**



General information about the course

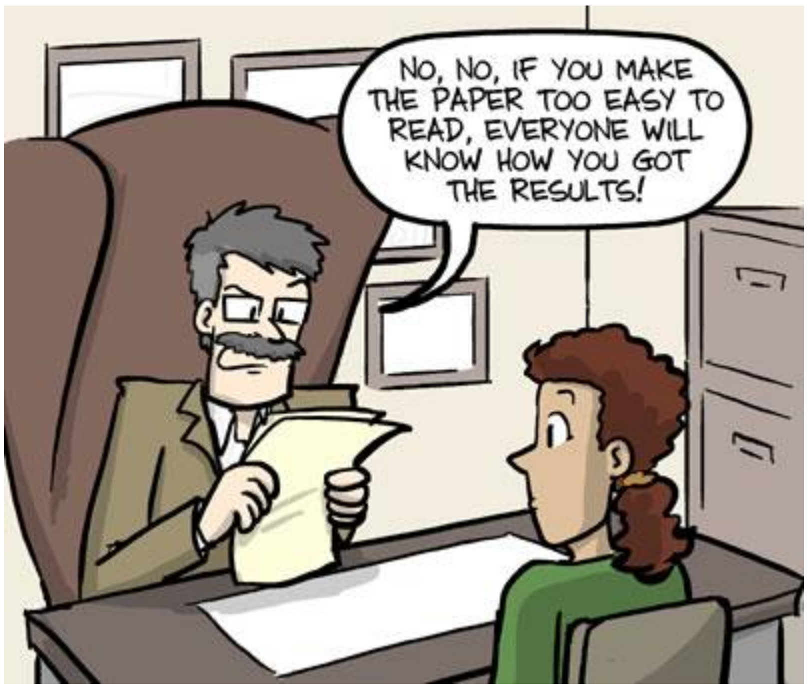
- *The program in brief*

1. Mixed models theory (with Matrix Notation)
2. Mixed models analysis with R
3. Mixed models for experimental design (in brief)
4. Mixed models and longitudinal data: designs with repeated measurements

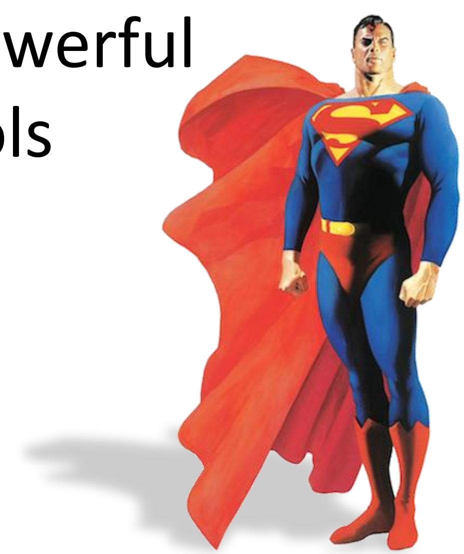
1. Mixed models theory

(with Matrix Notation)

Mixed models theory



- ✓ Mixed models: broadly used in a wide range of research fields
- ✓ Pretty complex but extremely powerful statistical tools



Introduction: fixed effects vs. random effects

Mixed models:

"Linear models including both fixed and random effects"

Fixed effects vs. random effects

Fixed effects vs. random effects

✓ Fixed effects:

- Levels of interest are selected by a non random-process and included in the study
- Inferences are to be made only to those levels included in the study (=as they were all the levels existing in a population)

E.g.: sex (2 levels: male, female)

E.g.: number of parity (3 levels: 1^{st}, 2^{nd} , 3^{rd})
(if you are interest to the first 3 parities
If a level is scarcely represented, then different levels may be grouped in 1:
E.g.: number of parity (3 levels: 1^{st}, 2^{nd} , 3^{rd})

1^{st} parity

2^{nd} parity

3^{rd} parity

Fixed effects vs. random effects

✓ Random effects:

- Levels consist of a random sample of levels from a population of possible levels
- Inferences are about the population of levels, not just the subset of levels included in the study

E.g.: herd (100 levels of 1000, for example)

E.g.: animal (1000 individuals of a population of 100.000)

Fixed effects vs. random effects

✓ Mixed models:

- Models in which some factors are fixed effects and other factors are random effects

Random effect:

E.g.: herd (100 levels of 1000, for example)

Fixed effect:

E.g.: number of parity (3 levels: 1^{st}, 2^{nd} , 3^{rd})














Fixed effects vs. random effects

✓ Mixed models:

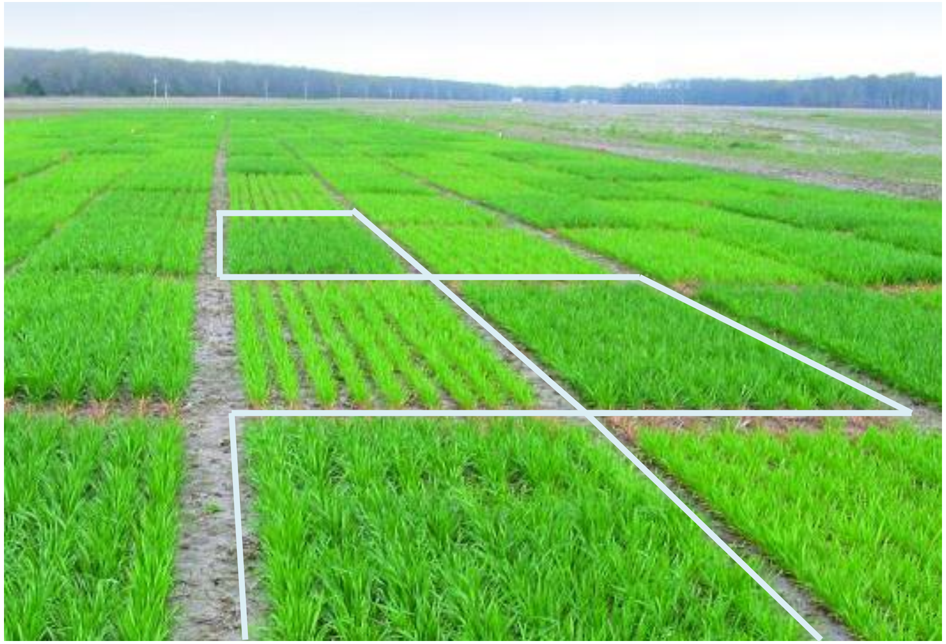
- E.g.: Crop experiments:

Random effect:

E.g.: blocks of similar plot

Fixed effect:

E.g.: fertilizer (3 levels)

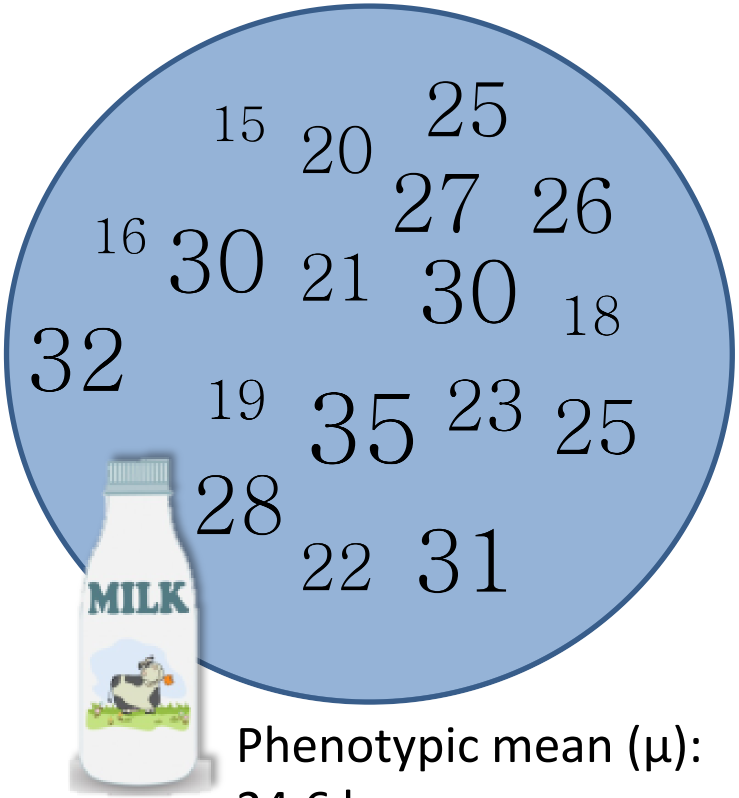


Fixed effects vs. random effects

✓ **Fixed effects:**	✓ **Random effects:**
• Interest "per se" (e.g. interest in estimating the differences among levels of the effects, or the value of the level) (Searle, Casella, & McCulloch, 1992)	• No interest in the single levels but in the variability of underlying population (e.g. variability of herd effect, among herds) (Searle, Casella, and McCulloch, 1992)
• "When a sample exhausts the population, the corresponding variable (= effect) is fixed" (Green & Tukey, 1960)	• "When the sample is a small (i.e., negligible) part of the population, the corresponding variable (=effect) is random" (Green &Tukey, 1960)
• In experimental designs, they represent the true effect of an intervention	• Goal is generally to estimate the variance among levels of the effect
• Their value is the same in every study (if experiment is repeated, levels are the same)	• Conclusions apply to the reference population

Fixed effects vs. random effects

- ✓ **Fixed effects:** interest in the estimation of the levels of the effect



Phenotypic mean (μ): 24.6 kg

- ✓ 1st parity: -5 kg
- ✓ Lsmean: 19 kg


- ✓ 2nd parity: +0.5 kg
Lsmean: 21 kg

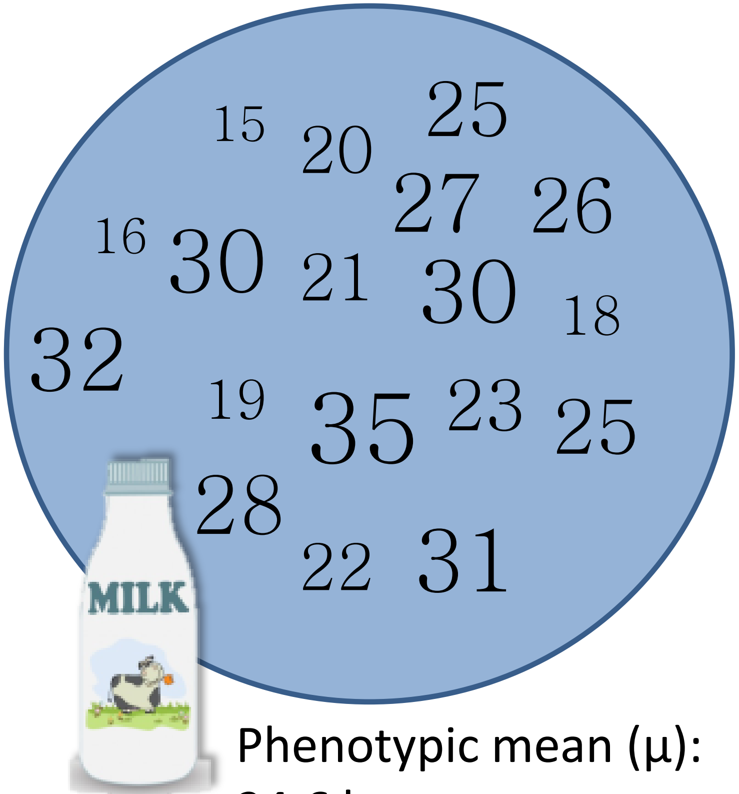


- ✓ 3rd parity: +3 kg
- ✓ Ls mean: 28 kg

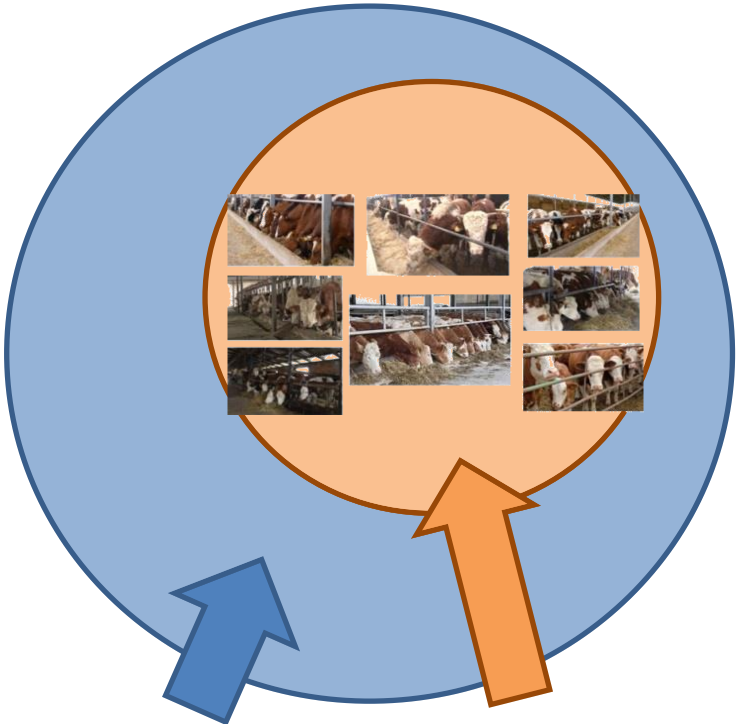
(Values by chance, not calculated)

Fixed effects vs. random effects

✓ **Random effects:** interest in the estimation of the variability of the effect



Phenotypic mean (μ): 24.6 kg



Phenotypic variability (σ^2_P): 33.31 kg^2

Herd variability (σ^2_P): 18.5 kg^2

Fixed effects vs. random effects

- If a factor is included as **fixed effect**, the model assumes that the **factor is totally explained by the levels included**, whereas if it is a **random effect**, the model takes into account that there **may be other levels explaining the factor** that are not included

- If the analysis aims to evaluate the **differences between the single levels of the effect**, then this effect should be included as fixed

- If **binary dependent variables** are used, fixed and random effects provide different results

- In case of **continuous variables**, results are often the same

(but… let's consider the next slide…)

Fixed effects vs. random effects

- In case of **continuous variables**, results are similar

10 cows

Var. Herd

10 cows

Var. Herd

10 cows

Var. Herd

Same n° records, same variance:
Fixed effect ~ Random effect

- If an effect is **random** and when **between-levels variance is large**, within levels variance become less important, and large and small levels tend to be weighted equally

2 cows

Var. Herd

10 cows

Var. Herd

100 cows

Var. Herd

Different n° records, (almost) same variance:
Better Random effect

- If there is **no heterogeneity**, fixed and random effects → similar results

10 cows

Var. Herd

10 cows

Var. Herd

10 cows

Var. Herd

Same n° records, different variance:
Better Random effect

Fixed effects vs. random effects

- In general, fixed and random effects may be distinguished as follows:

	Fixed	Random
Few levels of factor	X	
The number of levels is high ($\rightarrow\infty$)		X
Levels do not vary in different experiments	X	
Inference may be extended to uninvestigated levels		X
Levels of a factor are chosen by chance (or sampling, data editing)		X

Introduction: general linear models

General linear model (GLM)

- ✓ Model in which a dependent variable is though to be some linear combination of one or more variables that may be both **qualitative** (categorical effects or factors) or **quantitative** (continuous variables).

- ✓ It is a **generalization of multiple linear regression** model and is suited to implement any parametric statistical test with one dependent variable, **including ANOVA** and designs with a mixture of qualitative and quantitative variables.

- ✓ **Es: Meat production (Y) = sex (categorical, 2 levels) + weight (continuous)**

  ANOVA + LINEAR REGRESSION → GLM

- ✓ Explanatory variables are included as **FIXED EFFECTS**

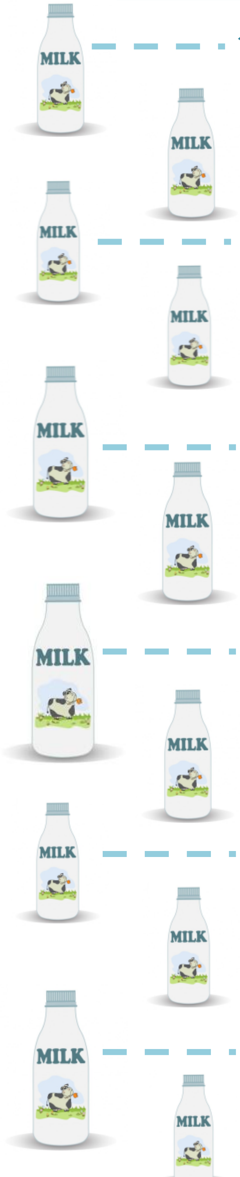
Mixed model:

- ✓ Linear model including both **FIXED** (as in ANOVA) and **RANDOM EFFECTS**

Before talking about Mixed Models & Random Effects… →
something about Fixed Effects & General linear model

Don't be frightened by the Matrix!!!!!!!!

Fixed effects: the general linear model



Milk yield of cow_1

Milk yield of cow_2

Milk yield of cow_3

Milk yield of cow_4

Milk yield of cow_5

Milk yield of cow_6

Milk yield of cow_7

Milk yield of cow_8

Milk yield of cow_9

Milk yield of cow_{10}

Milk yield of cow_{11}

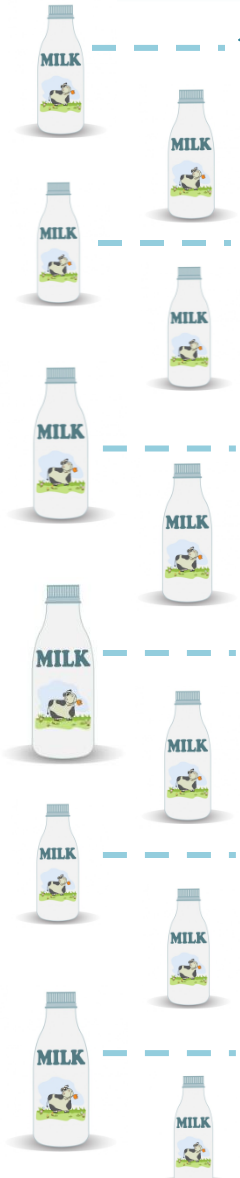
Milk yield of cow_{12}

✓ 1^{st} parity

✓ 2^{nd} parity

✓ 3^{rd} parity

Fixed effects: the general linear model

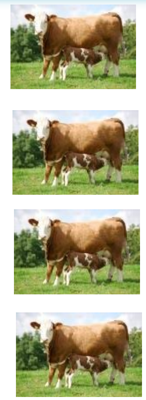


✓ Milk yield of cow_1 ✓ 1^{st} parity

✓ Milk yield of cow_2 ✓ 1^{st} parity

✓ Milk yield of cow_3 ✓ 1^{st} parity

✓ Milk yield of cow_4 ✓ 1^{st} parity



✓ Milk yield of cow_5 ✓ 2^{nd} parity

✓ Milk yield of cow_6 ✓ 2^{nd} parity

✓ Milk yield of cow_7 ✓ 2^{nd} parity

✓ Milk yield of cow_8 ✓ 2^{nd} parity

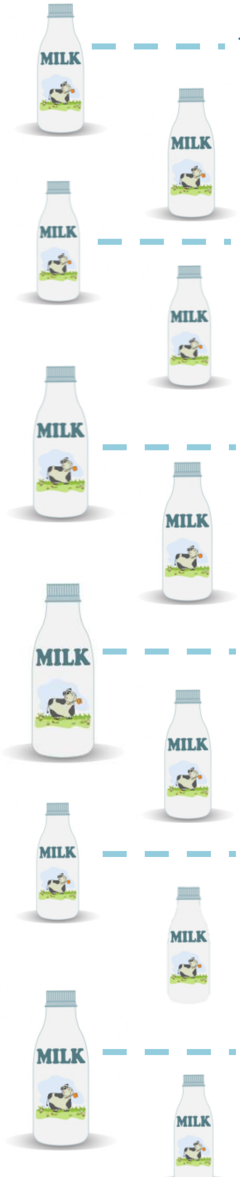
✓ Milk yield of cow_9 ✓ 3^{rd} parity

✓ Milk yield of cow_{10} ✓ 3^{rd} parity

✓ Milk yield of cow_{11} ✓ 3^{rd} parity

✓ Milk yield of cow_{12} ✓ 3^{rd} parity

Fixed effects: the general linear model



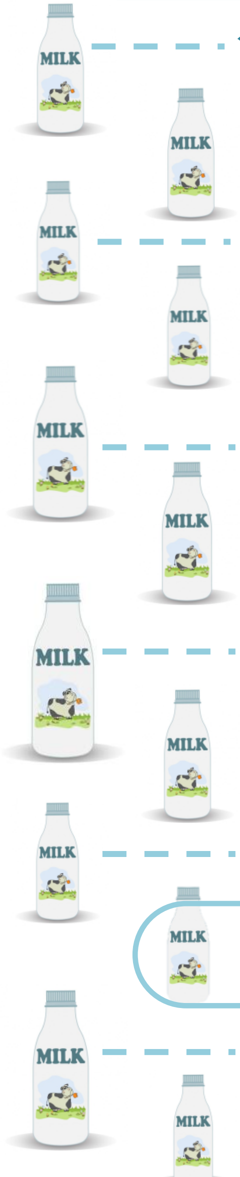
- $Y_1 = \mu + 1^{st}$ parity $+ e_1$
- $Y_2 = \mu + 1^{st}$ parity $+ e_2$
- $Y_3 = \mu + 1^{st}$ parity $+ e_3$
- $Y_4 = \mu + 1^{st}$ parity $+ e_4$
- $Y_5 = \mu + 2^{nd}$ parity $+ e_5$
- $Y_6 = \mu + 2^{nd}$ parity $+ e_6$
- $Y_7 = \mu + 2^{nd}$ parity $+ e_7$
- $Y_8 = \mu + 2^{nd}$ parity $+ e_8$
- $Y_9 = \mu + 3^{rd}$ parity $+ e_9$
- $Y_{10} = \mu + 3^{rd}$ parity $+ e_{10}$
- $Y_{11} = \mu + 3^{rd}$ parity $+ e_{11}$
- $Y_{12} = \mu + 3^{rd}$ parity $+ e_{12}$

- E.g. of data

Milk (Y)	µ	parity	e
22	24	-3	1
18	24	-3	-3
21	24	-3	0
23	24	-3	2
23	24	0.5	-1.5
26	24	0.5	1.5
22	24	0.5	-2.5
27	24	0.5	2.5
30	24	2.5	3.5
25	24	2.5	-1.5
23	24	2.5	-3.5
28	24	2.5	1.5

Effect of 1^{st} parity= -3
Effect of 2^{nd} parity= +0.5
Effect of 3^{rd} parity= +2.5

Fixed effects: the general linear model



$Y_1 = \mu + 1^{st}\text{ parity} + e_1$

$Y_2 = \mu + 1^{st}\text{ parity} + e_2$

$Y_3 = \mu + 1^{st}\text{ parity} + e_3$

$Y_4 = \mu + 1^{st}\text{ parity} + e_4$

$Y_5 = \mu + 2^{nd}\text{ parity} + e_5$

$Y_6 = \mu + 2^{nd}\text{ parity} + e_6$

$Y_7 = \mu + 2^{nd}\text{ parity} + e_7$

$Y_8 = \mu + 2^{nd}\text{ parity} + e_8$

$Y_9 = \mu + 3^{rd}\text{ parity} + e_9$

$Y_{10} = \mu + 3^{rd}\text{ parity} + e_{10}$

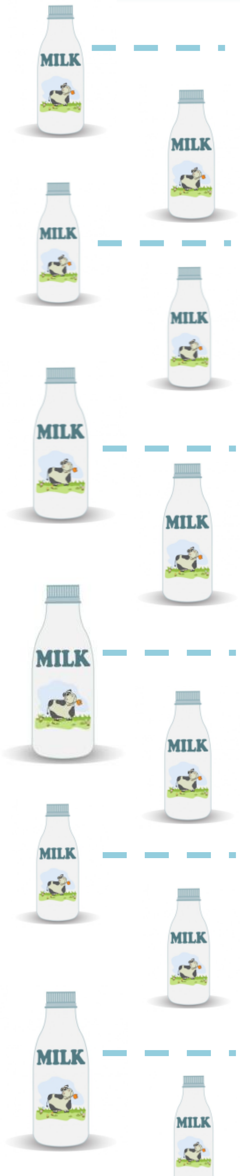
$Y_{11} = \mu + 3^{rd}\text{ parity} + e_{11}$

$Y_{12} = \mu + 3^{rd}\text{ parity} + e_{12}$

μ=overall mean
Let's call:
1^{st} parity=α_1
2^{nd} parity=α_2
3^{rd} parity=α_3
α_1, α_2 and α_3 are 3 levels of the «parity» effect (or factor)

This obs. includes the level α_3 of the effect «parity», thus here α_1 and $\alpha_2 = 0$

Fixed effects: the general linear model



✓ $Y_1 = 1 \cdot \mu + 1 \cdot \alpha_1 + 0 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_1$

✓ $Y_2 = 1 \cdot \mu + 1 \cdot \alpha_1 + 0 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_2$

✓ $Y_3 = 1 \cdot \mu + 1 \cdot \alpha_1 + 0 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_3$

✓ $Y_4 = 1 \cdot \mu + 1 \cdot \alpha_1 + 0 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_4$

✓ $Y_5 = 1 \cdot \mu + 0 \cdot \alpha_1 + 1 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_5$

✓ $Y_6 = 1 \cdot \mu + 0 \cdot \alpha_1 + 1 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_6$

✓ $Y_7 = 1 \cdot \mu + 0 \cdot \alpha_1 + 1 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_7$

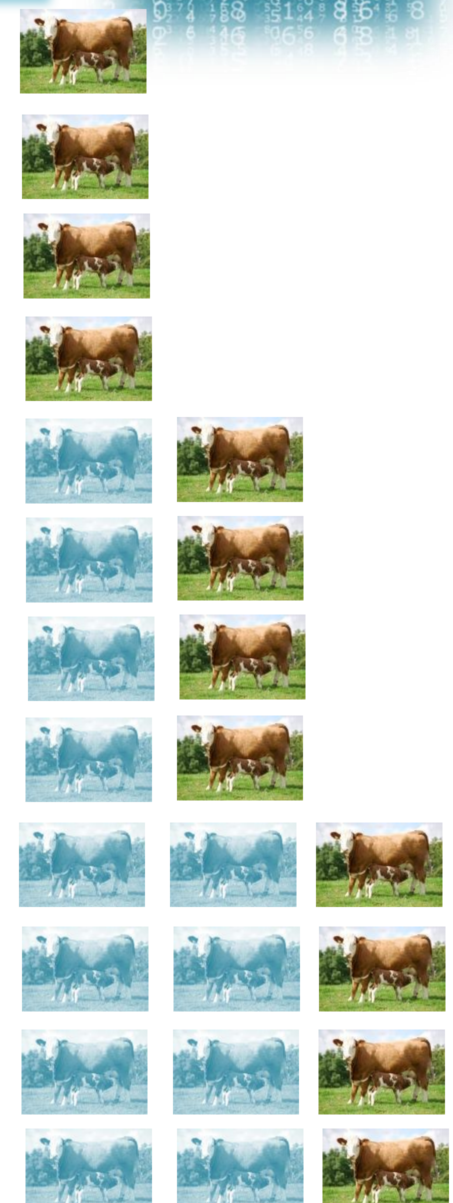
✓ $Y_8 = 1 \cdot \mu + 0 \cdot \alpha_1 + 1 \cdot \alpha_2 + 0 \cdot \alpha_3 + e_8$

✓ $Y_9 = 1 \cdot \mu + 0 \cdot \alpha_1 + 0 \cdot \alpha_2 + 1 \cdot \alpha_3 + e_9$

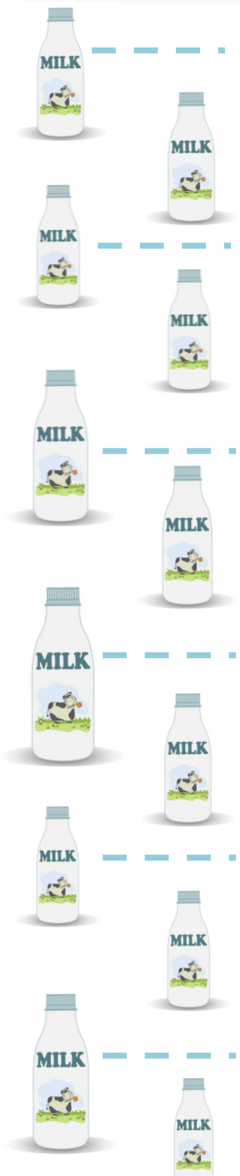
✓ $Y_{10} = 1 \cdot \mu + 0 \cdot \alpha_1 + 0 \cdot \alpha_2 + 1 \cdot \alpha_3 + e_{10}$

✓ $Y_{11} = 1 \cdot \mu + 0 \cdot \alpha_1 + 0 \cdot \alpha_2 + 1 \cdot \alpha_3 + e_{11}$

✓ $Y_{12} = 1 \cdot \mu + 0 \cdot \alpha_1 + 0 \cdot \alpha_2 + 1 \cdot \alpha_3 + e_{12}$



Fixed effects: the general linear model



$$Y_1 = 1\cdot\mu + 1\cdot\alpha_1 + 0\cdot\alpha_2 + 0\cdot\alpha_3 + e_1$$

$$Y_2 = 1\cdot\mu + 1\cdot\alpha_1 + 0\cdot\alpha_2 + 0\cdot\alpha_3 + e_2$$

$$Y_3 = 1\cdot\mu + 1\cdot\alpha_1 + 0\cdot\alpha_2 + 0\cdot\alpha_3 + e_3$$

$$Y_4 = 1\cdot\mu + 1\cdot\alpha_1 + 0\cdot\alpha_2 + 0\cdot\alpha_3 + e_4$$

$$Y_5 = 1\cdot\mu + 0\cdot\alpha_1 + 1\cdot\alpha_2 + 0\cdot\alpha_3 + e_5$$

$$Y_6 = 1\cdot\mu + 0\cdot\alpha_1 + 1\cdot\alpha_2 + 0\cdot\alpha_3 + e_6$$

$$Y_7 = 1\cdot\mu + 0\cdot\alpha_1 + 1\cdot\alpha_2 + 0\cdot\alpha_3 + e_7$$

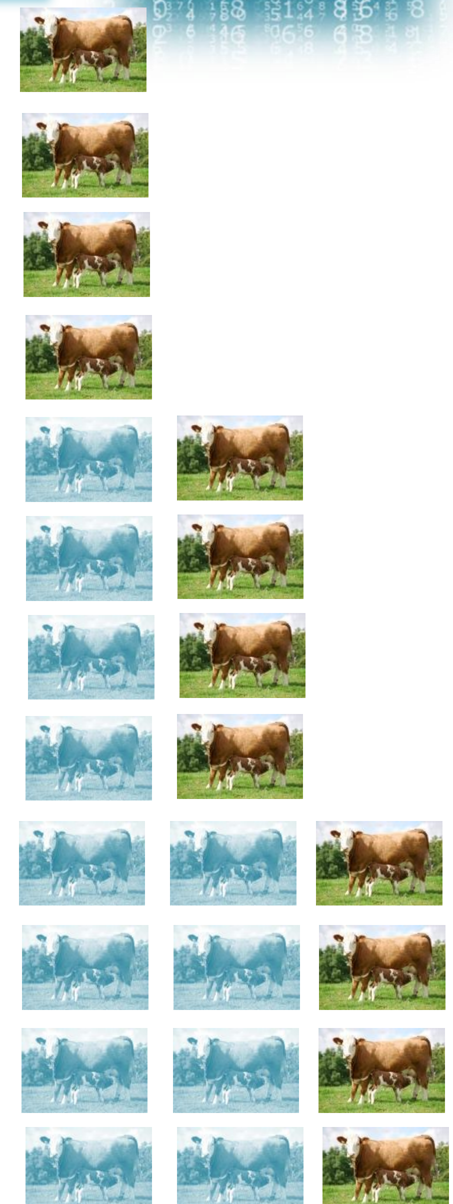
$$Y_8 = 1\cdot\mu + 0\cdot\alpha_1 + 1\cdot\alpha_2 + 0\cdot\alpha_3 + e_8$$

$$Y_9 = 1\cdot\mu + 0\cdot\alpha_1 + 0\cdot\alpha_2 + 1\cdot\alpha_3 + e_9$$

$$Y_{10} = 1\cdot\mu + 0\cdot\alpha_1 + 0\cdot\alpha_2 + 1\cdot\alpha_3 + e_{10}$$

$$Y_{11} = 1\cdot\mu + 0\cdot\alpha_1 + 0\cdot\alpha_2 + 1\cdot\alpha_3 + e_{11}$$

$$Y_{12} = 1\cdot\mu + 0\cdot\alpha_1 + 0\cdot\alpha_2 + 1\cdot\alpha_3 + e_{12}$$



Fixed effects: the general linear model

$$y_i = \mu + \alpha_i + e_i$$

Model in linear notation

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{bmatrix}$$

Model in matrix notation

y Vector of observations

X Incidence matrix for fixed effects

b

e Vector of residual errors

b Vector of parameters (μ, α_1, α_2, α_3)

Fixed effects: the general linear model

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

X Incidence matrix for fixed effects

- ✓ Matrix of values for the explanatory variables in the assumed linear model;
- ✓ The rows represent the observations and the columns the levels of the factors (e.g.: overall mean, and the 3 levels of parity)
- ✓ It is also called ”Design Matrix” because it traditionally describes the design of the experiment.

Focus on: Matrix Product

$$\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}$$

- ✓ How does the incidence matrix **X** relate to the vector of parameters β?

- ✓ Product of 2 matrices:

$$\begin{bmatrix} 3 & 1 & 10 \\ 2 & 2 & 8 \end{bmatrix} \cdot \begin{bmatrix} 50 & 60 \\ 100 & 80 \\ 30 & 40 \end{bmatrix} = \begin{bmatrix} (3\cdot 50 + 1\cdot 100 + 10\cdot 30) & (3\cdot 60 + 1\cdot 80 + 10\cdot 40) \\ (2\cdot 50 + 2\cdot 100 + 8\cdot 30) & (2\cdot 60 + 2\cdot 80 + 8\cdot 40) \end{bmatrix}$$

2x3 *(2 rows x 3 columns)*   3x2

$$= \begin{bmatrix} 550 & 660 \\ 540 & 600 \end{bmatrix}$$

2x2

The elements of the 1st row of the 1st matrix are multiplied by the elements of the 1st column of the 2nd matrix

Rows and columns of the product of 2 matrices are equal to the num rows of the 1st and num columns of the 2nd:

(es: $A_{2X3} \cdot B_{3X2} = AB_{2x2}$; $B_{2X3} \cdot A_{3X2} = BA_{3x3}$)

Focus on: Matrix Product

✓ How does the incidence matrix **X** relate to the vector of parameters **β**?

$$
\begin{bmatrix}
1 & 1 & 0 & 0\\
1 & 1 & 0 & 0\\
1 & 1 & 0 & 0\\
1 & 1 & 0 & 0\\
1 & 0 & 1 & 0\\
1 & 0 & 1 & 0\\
1 & 0 & 1 & 0\\
1 & 0 & 1 & 0\\
1 & 0 & 0 & 1\\
1 & 0 & 0 & 1\\
1 & 0 & 0 & 1\\
1 & 0 & 0 & 1
\end{bmatrix}
\cdot
\begin{bmatrix}
\mu\\
\alpha_1\\
\alpha_2\\
\alpha_3
\end{bmatrix}
$$

b: 4x1

X: 12x4

✓ Product of 2 matrices:

$$
\begin{bmatrix}
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3\\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3
\end{bmatrix}
$$

Xb: 12x1

Fixed effects: the general linear model

✓ $$y_i = \mu + \alpha_i + e_i$$

Model in linear notation

Model in matrix notation

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix}}_{\mathbf{y}} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}}_{\mathbf{X}} \cdot \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\mathbf{b}} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{bmatrix}}_{\mathbf{e}}$$

y Vector of observations

X Incidence matrix for fixed effects

e Vector of residual errors

b Vector of parameters (μ, α_1, α_2, α_3)

The general linear model

✓ $y_i = \mu + \alpha_i + e_i$ ⟵ Model in linear notation

✓ $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$ ⟵ Model in matrix notation

y Vector of observations

X Incidence matrix for fixed effects

b Vector of parameters (μ, α_1, α_2, α_3)

e Vector of residual errors

- ✓ General linear model including an intercept (μ), fixed effects (b_i), and a random residual term (variability of each observation)
- ✓ **Fixed effects (or factors)** may be **categories** (es.: num. parities), **continuous variables** (es.: the body weight of each cow evalued), may have **interactions** (es.: num. parities*body weight) etc…
- ✓ You should have learned a lot about fixed effects in previous parts of the course

The general linear model

- The general linear model may also include (or be just) one or more **continoius variables** (as in linear or in multiple linear regression models) as **fixed effects**:
  - $Y_1 = \mu + 1^{st}\text{ parity} + BW_1 + e_1$
  - $Y_2 = \mu + 1^{st}\text{ parity} + BW_2 + e_2$
  - $Y_3 = \mu + 1^{st}\text{ parity} + BW_3 + e_3$
  - $Y_4 = \mu + 1^{st}\text{ parity} + BW_4 + e_4$
  - $Y_5 = \mu + 2^{nd}\text{ parity} + BW_5 + e_5$
  - $Y_6 = \mu + 2^{nd}\text{ parity} + BW_6 + e_6$
  - $Y_7 = \mu + 2^{nd}\text{ parity} + BW_7 + e_7$
  - $Y_8 = \mu + 2^{nd}\text{ parity} + BW_8 + e_8$
  - $Y_9 = \mu + 3^{rd}\text{ parity} + BW_9 + e_9$
  - $Y_{10} = \mu + 3^{rd}\text{ parity} + BW_{10} + e_{10}$
  - $Y_{11} = \mu + 3^{rd}\text{ parity} + BW_{11} + e_{11}$
  - $Y_{12} = \mu + 3^{rd}\text{ parity} + BW_{12} + e_{12}$

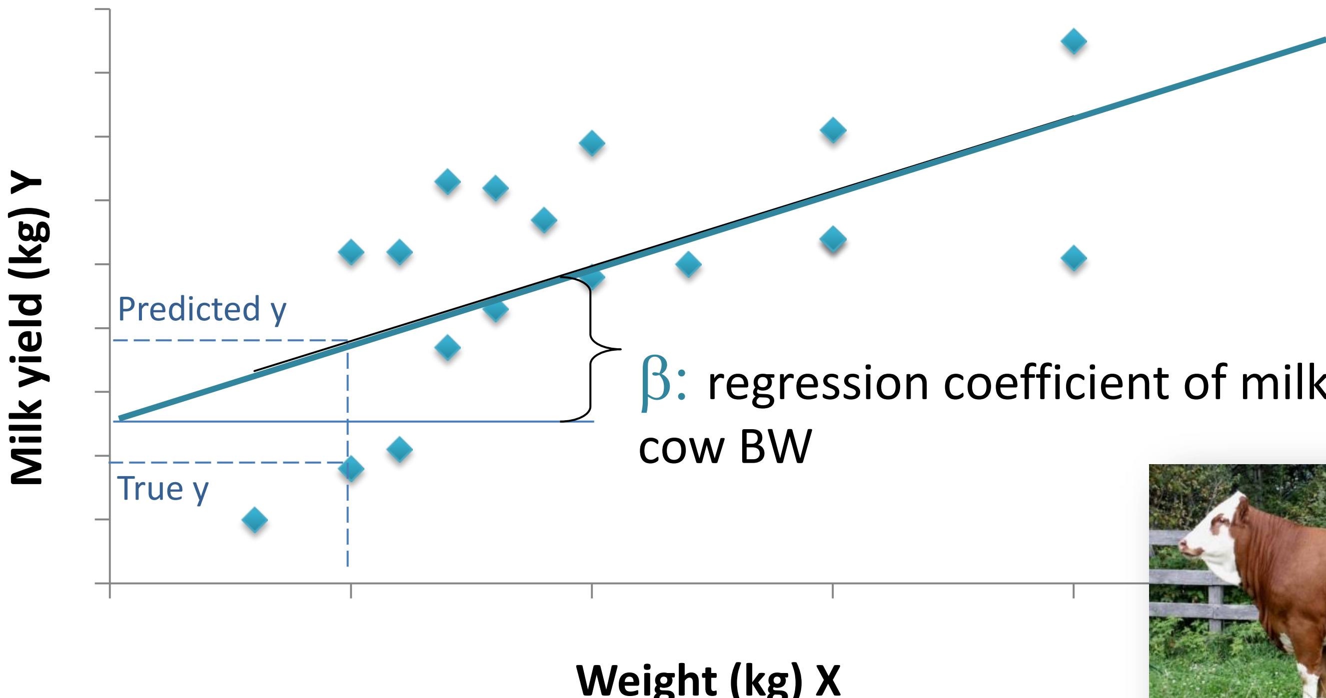
BW_i: individual body weight of cow i
BW assumes different values in each cow (x_1, x_2, x_3…)

The general linear model: linear regression

✓ $y_i = \mu + \alpha_i + \beta x_i + e_i$

Model in linear notation
(includes $\beta =$ regression slope and x_i = individual weight of cow i)

✓ **y_i can be predicted by βx_i**



The general linear model: linear regression

✓ $y_i = \mu + \alpha_i + \beta x_i + e_i$

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

Model in matrix notation

y

Vector of observations

e

Vector of residual errors

X

Incidence matrix
for fixed effects

b

Vector of parameters
(μ, α_1, α_2, α_3, β)

Now b includes also the
linear regression slope β

Fixed effects: the general linear model

✓ $y_i = \mu + \alpha_i + \beta x_i + e_i$

Model in linear notation

Model in matrix notation

$$\underset{\mathbf{y}}{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \\ \mathbf{y}_4 \\ \mathbf{y}_5 \\ \mathbf{y}_6 \\ \mathbf{y}_7 \\ \mathbf{y}_8 \\ \mathbf{y}_9 \\ \mathbf{y}_{10} \\ \mathbf{y}_{11} \\ \mathbf{y}_{12} \end{bmatrix}} = \underset{\mathbf{X}}{\begin{bmatrix} 1 & 1 & 0 & 0 & \mathbf{x}_1 \\ 1 & 1 & 0 & 0 & \mathbf{x}_2 \\ 1 & 1 & 0 & 0 & \mathbf{x}_3 \\ 1 & 1 & 0 & 0 & \mathbf{x}_4 \\ 1 & 0 & 1 & 0 & \mathbf{x}_5 \\ 1 & 0 & 1 & 0 & \mathbf{x}_6 \\ 1 & 0 & 1 & 0 & \mathbf{x}_7 \\ 1 & 0 & 1 & 0 & \mathbf{x}_8 \\ 1 & 0 & 0 & 1 & \mathbf{x}_9 \\ 1 & 0 & 0 & 1 & \mathbf{x}_{10} \\ 1 & 0 & 0 & 1 & \mathbf{x}_{11} \\ 1 & 0 & 0 & 1 & \mathbf{x}_{13} \end{bmatrix}} \cdot \underset{\mathbf{b}}{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{bmatrix}} + \underset{\mathbf{e}}{\begin{bmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \mathbf{e}_3 \\ \mathbf{e}_4 \\ \mathbf{e}_5 \\ \mathbf{e}_6 \\ \mathbf{e}_7 \\ \mathbf{e}_8 \\ \mathbf{e}_9 \\ \mathbf{e}_{10} \\ \mathbf{e}_{11} \\ \mathbf{e}_{12} \end{bmatrix}}$$

y Vector of observations

X Incidence matrix for fixed effects

e Vector of residual errors

b Vector of parameters (μ, α_1, α_2, α_3, β)

Solving the general linear model

✓ $\mathbf{y = Xb + e}$

y

Vector of observations

e

Vector of residual errors

X

Incidence matrix
for fixed effects

b

Vector of parameters (μ, α_1, α_2, α_3, β)

✓ Solving the model: estimating the **solutions** for the parameters α_1, α_2, α_3, β that are unknown

✓ First step: defining the **expected values** for the terms of the model

The general linear model: residual error

✓ $\mathbf{y = Xb + e}$

y

Vector of observations

X

Incidence matrix
for fixed effects

b

Vector of parameters (μ, α_1, α_2, α_3, β)

e

Vector of residual errors

$$\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{bmatrix}$$

Expected value (Mean) of **e** is zero → E(**e**)=0 → residuals are the deviations from the mean that are unique for each observation (=individual variability of data) and cannot be explained by factors (e.g. by parity num.). They sum at zero, have mean of zero and they have a own variability → Var(**e**)=σ_e^2

$$\mathbf{E(e)=0} \qquad \mathbf{Var(e)=R = \sigma_e^2 \cdot I}$$

(Homoscedasticy assumption: variance of **e** equal for all the observations; I is an identity matrix)

$$e_i \sim N(0, \sigma_e^2)$$

The residuals are normally distributed

Focus on: Identity Matrix

$E(e)=0$; $Var(e)=R = \sigma_e^2 \cdot I$

$$R=\sigma_e^2\cdot\begin{bmatrix}1&0&0&0&0&0&0&0&0&0&0&0\\0&1&0&0&0&0&0&0&0&0&0&0\\0&0&1&0&0&0&0&0&0&0&0&0\\0&0&0&1&0&0&0&0&0&0&0&0\\0&0&0&0&1&0&0&0&0&0&0&0\\0&0&0&0&0&1&0&0&0&0&0&0\\0&0&0&0&0&0&1&0&0&0&0&0\\0&0&0&0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&0&1&0&0&0\\0&0&0&0&0&0&0&0&0&1&0&0\\0&0&0&0&0&0&0&0&0&0&1&0\\0&0&0&0&0&0&0&0&0&0&0&1\end{bmatrix}$$

$$R=\begin{bmatrix}\sigma_e^2&0&0&0&0&0&0&0&0&0&0&0\\0&\sigma_e^2&0&0&0&0&0&0&0&0&0&0\\0&0&\sigma_e^2&0&0&0&0&0&0&0&0&0\\0&0&0&\sigma_e^2&0&0&0&0&0&0&0&0\\0&0&0&0&\sigma_e^2&0&0&0&0&0&0&0\\0&0&0&0&0&\sigma_e^2&0&0&0&0&0&0\\0&0&0&0&0&0&\sigma_e^2&0&0&0&0&0\\0&0&0&0&0&0&0&\sigma_e^2&0&0&0&0\\0&0&0&0&0&0&0&0&\sigma_e^2&0&0&0\\0&0&0&0&0&0&0&0&0&\sigma_e^2&0&0\\0&0&0&0&0&0&0&0&0&0&\sigma_e^2&0\\0&0&0&0&0&0&0&0&0&0&0&\sigma_e^2\end{bmatrix}$$

Identity matrix

Square matrix with ones on the main diagonal and zeros elsewhere

$$I_1=[1],\ I_2=\begin{bmatrix}1&0\\0&1\end{bmatrix},\ I_3=\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix},\ \cdots,\ I_n=\begin{bmatrix}1&0&0&\cdots&0\\0&1&0&\cdots&0\\0&0&1&\cdots&0\\\vdots&\vdots&\vdots&\ddots&\vdots\\0&0&0&\cdots&1\end{bmatrix}$$

The general linear model: predicted values

✓ $\mathbf{y = Xb + e}$

y

Vector of observations

e

Vector of residual errors

X

Incidence matrix
for fixed effects

b

Vector of parameters (μ, α_1, α_2, α_3, β)

Since E(**e**)=0 → Expected value (Mean) of **y** is **Xb**
→ E(**y**)=**Xb**

Each line is the **predicted value** for the corresponding observation y_i:
e.g.: I predict that the milk yield of cow 1 is
$1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_1$

$$\begin{bmatrix}
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_1 \\
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_2 \\
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_3 \\
1\cdot\mu+1\cdot\alpha_1+0\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_4 \\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_5 \\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_6 \\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_7 \\
1\cdot\mu+0\cdot\alpha_1+1\cdot\alpha_2+0\cdot\alpha_3+\beta\cdot x_8 \\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3+\beta\cdot x_9 \\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3+\beta\cdot x_{10} \\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3+\beta\cdot x_{11} \\
1\cdot\mu+0\cdot\alpha_1+0\cdot\alpha_2+1\cdot\alpha_3+\beta\cdot x_{12}
\end{bmatrix}$$

The general linear model: predicted values

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

y — Vector of observations

X — Incidence matrix for fixed effects

b — Vector of parameters (μ, α_1, α_2, α_3, β)

e — Vector of residual errors

Since $E(\mathbf{e})=0$ → Expected value (Mean) of **y** is **Xb**
→ $E(\mathbf{y})=\mathbf{Xb}$

Parameters of β (μ, α_1, α_2, α_3, β) may be estimated as **BLUE**

Best Linear Unbiased Estimates

Best Linear Unbiased Estimates

- Best → because they minimize error variance
- Linear → they are linear functions of the data
- Unbiased → their expected mean is equal to what they are estimating
- Estimates → they can be estimated (typically using ordinary least square method)

- **Estimation of b parameters:**

✓ $\mathbf{Y = Xb + e}$

$E(y)=\mathbf{Xb}$ ➡ ✓ $\mathbf{b = (X^TX)^{-1} \cdot (X^TY)}$

Matrix transpose:

If A=[2 3 5
4 6 8], A^T =[2 4
3 6
5 8]

Inverse of a matrix (if A is the matrix, the inverse corresponds to 1/A – but it is not a simple divisions of cells such as 1/cell value)

➡ ✓ $\mathbf{V_b = (X^TX)^{-1} \cdot \sigma_e^2}$

(Co)variance estimates of **b** parameters

(Homoscedasticy assumption = no variability among residuals)

Focus on: Matrix transposition

- Matrix transposition:

$$\begin{bmatrix} 9 & 0 & 1 \\ 0 & 11 & 1 \\ 1 & 1 & 4 \\ 1 & 0 & 1 \end{bmatrix} A \xrightarrow{\text{Transpose}} \begin{bmatrix} 9 & 0 & 1 & 1 \\ 0 & 11 & 1 & 0 \\ 1 & 1 & 4 & 1 \end{bmatrix} A^T$$

- The product of a matrix con its matrix transpose is a square symmetric matrix ($A \cdot A^T = AA^T$; $A^T \cdot A = A^TA$; $AA^T \neq A^TA$)

$$\begin{bmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 0 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} (1+4+1) & (3+0-1) \\ (3+0+1) & (9+0+1) \end{bmatrix} = \begin{bmatrix} 6 & 2 \\ 2 & 10 \end{bmatrix} \quad \mathbf{AA^T}$$

$$\begin{bmatrix} 1 & 3 \\ 2 & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1+9 & 2+0 & -1+3 \\ 2+0 & 4+0 & -2+0 \\ -1+3 & -2+0 & 1+1 \end{bmatrix} = \begin{bmatrix} 10 & 2 & 2 \\ 2 & 4 & -2 \\ 2 & -2 & 2 \end{bmatrix} \quad \mathbf{A^TA}$$

Focus on: Matrix inversion

- Matrix inversion:

$$\mathbf{B} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$ If $AD - BC \neq 0$, then B has an inverse, denoted B^{-1}

$$\mathbf{B}^{-1} = \frac{1}{AD-BC} \begin{bmatrix} D & -B \\ -C & A \end{bmatrix}$$

- (if **B** is the matrix, the inverse corresponds to $^1/_B$ → but it is not a simple division of cells such as 1/cell value)

$$\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix}^{-1} = \frac{1}{4x6-7x2} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \frac{1}{10} \begin{bmatrix} 6 & -7 \\ -2 & 4 \end{bmatrix} = \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix}$$

- The product of a matrix by its inverted matrix is an identity matrix ($\mathbf{B} \cdot \mathbf{B}^{-1} = \mathbf{I}$)

$$\begin{bmatrix} 4 & 7 \\ 2 & 6 \end{bmatrix} \begin{bmatrix} 0.6 & -0.7 \\ -0.2 & 0.4 \end{bmatrix} = \begin{bmatrix} 4 \times 0.6 + 7 \times -0.2 & 4 \times -0.7 + 7 \times 0.4 \\ 2 \times 0.6 + 6 \times -0.2 & 2 \times -0.7 + 6 \times 0.4 \end{bmatrix} = \begin{bmatrix} 2.4 - 1.4 & -2.8 + 2.8 \\ 1.2 - 1.2 & -1.4 + 2.4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Lsmeans & Test on Fixed effects

- In most of cases, the interest in GLM is not just to estimate the solutions of the fixed effects, but to compute the **Lsmeans** of the (categorical) fixed effects and to perform the **ANOVA test on fixed effects**

- **Lsmeans:** the means that each level of a target fixed effect could have if the other fixed effects of the model were set to zero (= if they were equal among individuals) → in Lsmeans, the values of the means of each level depend only on that effect that is considered

- **ANOVA test on fixed effects:** Analysis of variance to test if the variability of the effects significantly affect the phenotypic variance

- **N.B.:** generally, with a categorical effect, a level of the effect is set to zero to solve the equation and estimate the solutions for that effect. Different programs (e.g., R, SAS) set different effects to zero as a default, there is not a general rule

Lsmeans & Test on Fixed effects: example

✓ e.g. Dataset:

milk	parity	weight
22	1	600
18	1	570
21	1	580
23	1	602
23	2	650
26	2	600
22	2	630
27	2	580
30	3	690
25	3	610
23	3	590
28	3	640

✓ **Lsmeans:**

parity	Lsmean	SE
1	21.7404	1.4085
2	24.4016	1.2654
3	25.8580	1.3738

✓ **ANOVA on fixed effects:**

✓ (neither parity nor weight significantly affect the phenotype)

	F	Df	Pr > (F)
parity	1.91	2	0.2097
weight	1.41	1	0.2695

	Estimate	Std. Error	t value
(Intercept)	6.85123	16.6089	0.41
parity 1	-4.1176	2.13224	-1.93
parity 2	-1.4564	1.84359	-0.79
parity 3	0.0000	.	.
weight	0.0311	0.02618	1.19

✓ **Solutions for fixed effects**

✓ Parity 3 set to zero to solve the equations

✓ N.B. the intercept is the value of Y when all the X are 0; roughly its meaning corresponds to the phenotypic mean, but its value depends on the effects considered

Introduction: fixed effects vs. random effects

Mixed models:

“Linear models including both fixed and random effects”

Mixed models

- *Linear model including both FIXED (as in ANOVA) and RANDOM EFFECTS:*
- Linear regressions in which the **explanatory variables** affecting the dependent variable(s) are a **mixture of both 'fixed' and 'random' effects** (Pinheiro & Bates 2000; McCulloch & Searle 2001)
- **Fixed effects** of mixed models **affect the mean of a distribution (of phenotypes)**. Each factor level may be estimated
- **Random effects** are used to describe factors with multiple levels sampled from a population of possible values, for which **the analysis provides an estimate of the variance of the effects rather than a parameter for each factor level** (that is provided for fixed effects)**.** Random effects therefore **influence the variance of the trait**.

Mixed models

- ✓ **Random effects:** “**the analysis provides an estimate of the variance of the effects rather than a parameter for each factor level** (that is provided for fixed effects)”

Fixed effect:

Estimation of the levels (b):

$\alpha_1, \alpha_2, \alpha_3$

$\boldsymbol{\alpha_1 = -2}$

- ✓ 1st parity

$\boldsymbol{\alpha_2 = 0.5}$

- ✓ 2nd parity

$\boldsymbol{\alpha_3 = 1.5}$

- ✓ 3rd parity

Random effect:

Estimation of the Variance (G):

G

Mixed models equation

A mixed model is written as follows:

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

y
Vector of observations

X
Incidence matrix for fixed effects

b
Vector of parameters associated with the fixed factors (e.g., μ, α_1, α_2, α_3, β)

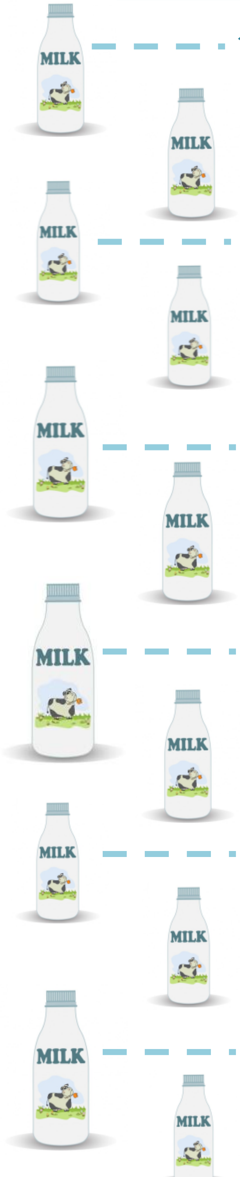
Z
Incidence matrix for random effects (+ random matrices if +1 random effects)

u
Vector of parameters associated with the random factors (e.g., u_1, u_2)

e
Vector of residual errors

- ✓ **Both u and e are vectors of RANDOM effects** (the residual error is a random effect)
- ✓ Parameters are typically estimated using the **restricted maximum likelihood estimator** (REML), but other estimators may be applied (e.g., ML)

Random effects



- Milk yield of cow_1
- Milk yield of cow_2
- Milk yield of cow_3
- Milk yield of cow_4
- Milk yield of cow_5
- Milk yield of cow_6
- Milk yield of cow_7
- Milk yield of cow_8
- Milk yield of cow_9
- Milk yield of cow_{10}
- Milk yield of cow_{11}
- Milk yield of cow_{12}

- Herd effect: cows belong to 2 herds sampled from a wider herd population of the breed

- Cows 1, 2, 5, 6, 9, 10 from herd 1
- Cows 3, 4, 7, 8, 11, 12 from herd 2

Mixed models

Model in matrix notation

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

y

Vector of observations

X

Incidence matrix for fixed effects

b

Vector of fixed effects (μ, α_1, α_2, α_3, β)

Z

Incidence matrix for random effects

u

Vector of random effects (u_1, u_2)

e

Vector of random residual errors

Mixed models

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

$$
\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix}}_{\mathbf{y}}
=
\underbrace{\begin{bmatrix}
1 & 1 & 0 & 0 & x_1 \\
1 & 1 & 0 & 0 & x_2 \\
1 & 1 & 0 & 0 & x_3 \\
1 & 1 & 0 & 0 & x_4 \\
1 & 0 & 1 & 0 & x_5 \\
1 & 0 & 1 & 0 & x_6 \\
1 & 0 & 1 & 0 & x_7 \\
1 & 0 & 1 & 0 & x_8 \\
1 & 0 & 0 & 1 & x_9 \\
1 & 0 & 0 & 1 & x_{10} \\
1 & 0 & 0 & 1 & x_{11} \\
1 & 0 & 0 & 1 & x_{13}
\end{bmatrix}}_{\mathbf{X}}
\cdot
\underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{bmatrix}}_{\mathbf{b}}
+
\underbrace{\begin{bmatrix}
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1 \\
1 & 0 \\
1 & 0 \\
0 & 1 \\
0 & 1
\end{bmatrix}}_{\mathbf{Z}}
\cdot
\underbrace{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{\mathbf{u}}
+
\underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{bmatrix}}_{\mathbf{e}}
$$

Model in matrix notation

y Vector of observations

X Incidence matrix for fixed effects

b Vector of fixed effects (μ, α_1, α_2, α_3, β)

Z Incidence matrix for random effects

u Vector of random effects (u_1, u_2)

e Vector of residual errors

Means and variances for y = Xb + Zu + e:

Means: $\mathbf{E(u)=0}$, $\mathbf{E(e)=0}$ — Means for random effects of **u** and **e** are zero

$\mathbf{E(y)=Xb}$ — The mean of observations **y** is due to fixed effects

Variances: $\mathbf{Var(e)= R = \sigma_e^2 \cdot I}$ (R is the covariance matrix for the residuals; under homoscedasticy the residual variance σ_e^2 multiplies an identiy matrix I)

$\mathbf{Var(u)=G}$ (e.g. $G=\sigma_u^2 \cdot I$) (Depending on the random effect considered, the covariance matrix may assume different forms and values)

$\mathbf{Var(y)=V=ZGZ^T+R}$ — The variance of **y** is due to random effects

Transpose of the incident matrix Z

Means and variances for y = Xb + Zu + e:

✓ Hence, phenotypes and random effects follow these distributions:

$$y_i \sim N(Xb, V)$$

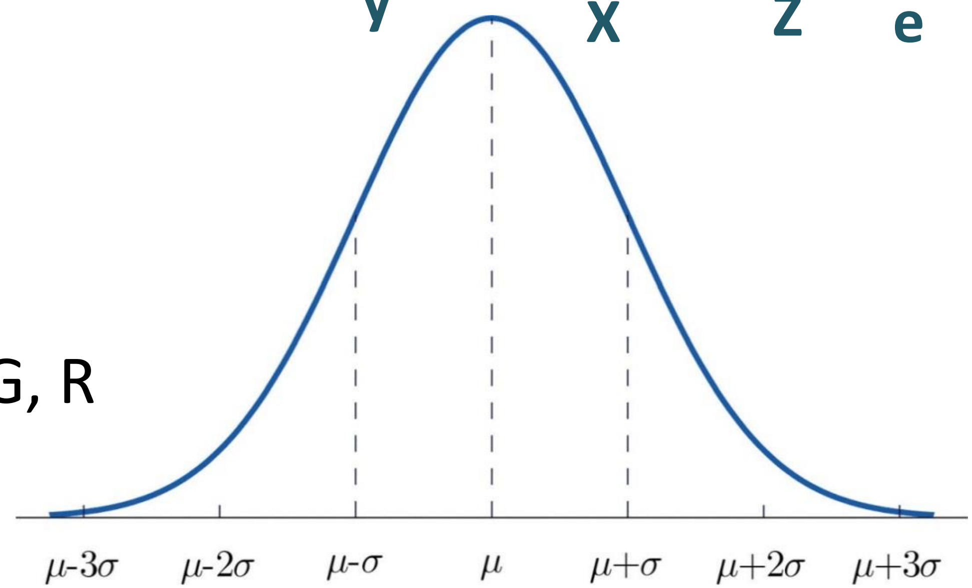
$$u_i \sim N(0, G)$$

$$e_i \sim N(0, R)$$

Variance of y:

$$V = ZGZ^T + R$$

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \end{bmatrix}}_{y} = \underbrace{\begin{bmatrix} 1 & 1 & 0 & 0 & x_1 \\ 1 & 1 & 0 & 0 & x_2 \\ 1 & 1 & 0 & 0 & x_3 \\ 1 & 1 & 0 & 0 & x_4 \\ 1 & 0 & 1 & 0 & x_5 \\ 1 & 0 & 1 & 0 & x_6 \\ 1 & 0 & 1 & 0 & x_7 \\ 1 & 0 & 1 & 0 & x_8 \\ 1 & 0 & 0 & 1 & x_9 \\ 1 & 0 & 0 & 1 & x_{10} \\ 1 & 0 & 0 & 1 & x_{11} \\ 1 & 0 & 0 & 1 & x_{13} \end{bmatrix}}_{X} \cdot \underbrace{\begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \\ \beta \end{bmatrix}}_{b} + \underbrace{\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix}}_{Z} \cdot \underbrace{\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}}_{u} + \underbrace{\begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \end{bmatrix}}_{e}$$



N = Normal distributions with mean Xb or 0 and variances V, G, R

Mixed models

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

y
Vector of observations

e
Vector of residual errors

X
Incidence matrix for fixed effects

b
Vector of parameters associated with the fixed factors (e.g., μ, α_1, α_2, α_3, β)

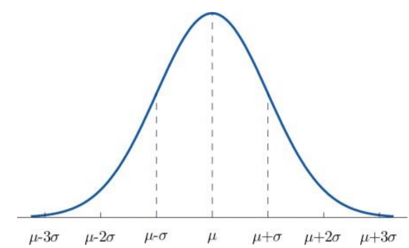
Z
Incidence matrix for random effects (+ random matrices if +1 random effects)

u
Vector of parameters associated with the random factors (e.g., u_1, u_2)

- ✓ Observe y, X, Z
- ✓ Estimate fixed effects (**b**)
- ✓ Estimate random effects (**u**)

Variances & estimates

- ✓ In the analysis of experiments, interest is in **estimation of treatment effects (fixed effects), b,** taking account of the variance matrix V (variance of random effects is considered because it influences the total phenotypic variance).



- ✓ In some applications (e.g. **quantitative genetics**), there is interest in the **estimation of the variances and covariances in G**, adjusting the data for the fixed effects **b**.

- ✓ In some cases, also there is an **interest in estimating the levels random effects** (e.g.: to predict the **genetic merit of individuals** in animal breeding → breeding values are the solutions of the additive genetic component random effect)

(R. Thompson 2008, Proc.R.Soc.B)

BLUE vs. BLUP!

✓ **Fixed effects** → Best linear unbiased estimates **(BLUEs)**

(e.g., μ, α_1, α_2, $\alpha_{3,}$ β)

✓ **Random effects** → Best linear unbiased predictions **(BLUPs)**

(e.g., u_1, u_2)

(Charles Roy Henderson, 1950)

- BLUPs → similar to BLUEs of fixed effects. The distinction arises because it is conventional to talk about estimating fixed effects but predicting random effects, but the two terms are otherwise equivalent.
- *"Predictions"* → in animal breeding random effects were usually genetic merits, which could be used to predict the quality of offspring (Robinson, 1991).
- Equations for "fixed" effects and for the random effects are though different.
- Typically, parameters associated with random effects are unknown (= variances of the random effects and residuals) →In mixed models analysis random effects parameters are estimated and plugged into the predictor (BLUP) → solutions for the random effects

BLUE & BLUP!

✓ In mixed models, **solutions** for fixed (b) and random (u) parameters are:

Best Linear Unbiased Estimator (BLUE) for Fixed Effects:

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

(e.g., μ, α_1, α_2, α_3, β)

Best Linear Unbiased Predictor (BLUP) for Random Effects:

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

(e.g., u_1, u_2)

After estimation of the fixed effects («adjustment» of data for the fixed effects), prediction of the random effects

✓ $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \mathbf{R}$

✓ Fixed effects & random effects are calculated by solving the matrices equations

Estimation of Variance Components

- **Before estimating fixed and random effects**, it is necessary to **estimate the Variance Components** of the model:

$$y_i \sim N(Xb, V) \qquad V = ZGZ^T + R \qquad u_i \sim N(0, G) \qquad e_i \sim N(0, R)$$

$$\underbrace{G = \sigma_u^2 \cdot I \quad G = \sigma_u^2 \cdot A \ldots}_{\text{Different structures, depending on the effect}} \qquad R = \sigma_e^2 \cdot I$$

- Most used methods of estimation:

REML (Restricted Maximum Likelihood)

(Patterson & Thompson, 1971; Harville, 1977)

Bayesian Inference (MCMC, Gibbs Sampling, Metropolis-Hasting)

(Zellner, 1971; Gianola & Fernando, 1986)

Restricted Maximum Likelihood (REML)

- ✓ **ML:** The likelihood function allows to find the parameter values (σ^2_u, σ^2_e, etc.) **maximizing the likelihood of obtaining the observations (y) given the parameters** (Maximum Likelihood, ML)

$$\boxed{L(\beta, V \mid x) = (2\pi)^{-n/2} |V|^{-1/2} \exp\left[-\frac{1}{2}(x-\beta)^T V^{-1} (x-\beta)\right]}$$

- ✓ **REML:** Method of choice for variance components; it **maximizes the portion of the likelihood function that does not depend on fixed effects** → transformation to remove fixed effects (**y**-**Xb**) and then perform ML

(Bruce Walsh Course, 2016)

Bayesian Inference

- ✓ **Bayesian Theorem:** description of the probability of an event (b_j), based on prior knowledge of conditions (A) that might be related to the event → the posterior distribution of parameters (b_j) due to the phenotypic data (A) is related to the prior knowledge of b_j and to the likelihood of obtaining A given b_j

$$\Pr(b_j \mid A) = \frac{\Pr(b_j)\Pr(A \mid b_j)}{\Pr(A)} = \frac{\Pr(b_j)\Pr(A \mid b_j)}{\sum_{i=1}^{n} \Pr(b_i)\,\Pr(A \mid b_i)}$$

This part is the likelihood function

A typical application in genetics is that A is some phenotype and b indexes some underlying (but unknown) genotype

- ✓ Marginal posterior distribution of parameters are thus obtained using the Gibbs sampler or equivalent algorithm (Geman & Geman, 1984; Gelfand & Smith, 1980)

(Bruce Walsh Course, 2016)

Model fitting

- ✓ **Model Selection:** A model is chosen among a number of candidate models by using some criteria → The goodness-of-fit is established using the likelihood (L) of the model and correcting for the number of estimated parameters (k):

- ✓ **-2Ln(L)** (criterion used but not penalizes for k)
- ✓ **AIC = Akaike's information criterion**

$$AIC = 2k - 2\,Ln(L)$$

- ✓ **BIC = Bayesian information criterion (Schwarz criterion)**

$$BIC = k*ln(n)/n - 2\,Ln(L)/n$$

- ✓ BIC penalizes free parameters more strongly than AIC (n is the number of observations)
- ✓ For all criteria, smaller value is better (see also further)

(Bruce Walsh Course, 2016)

Generalized linear models

- ✓ What described before assumes a **normal distribution** of parameters to be estimated (**y**, **b**, **u**) and **residuals**. But other models exist → non-linear functions, non-normality distribution of residuals

✓ GLM = general linear model

- Fixed effects

✓ Mixed models

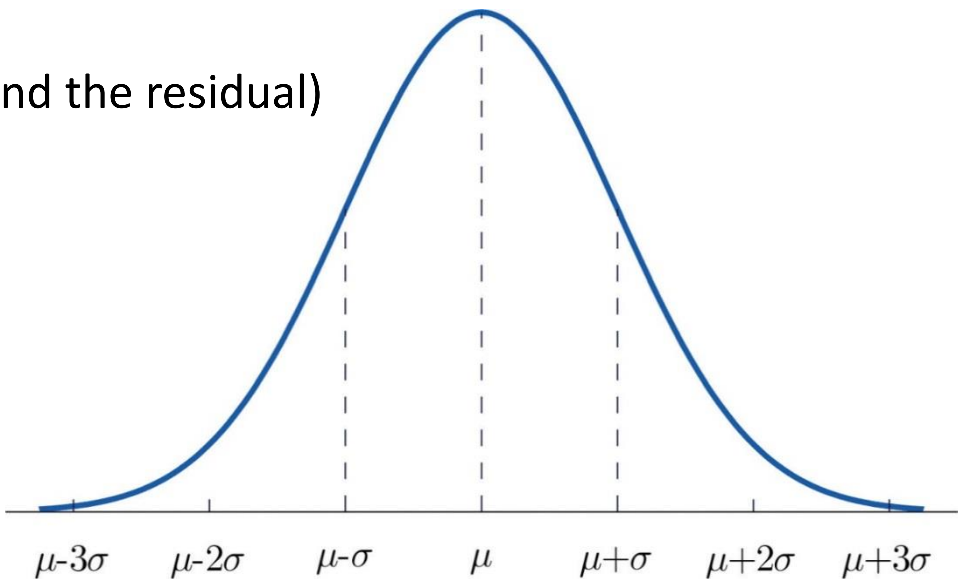
- Both fixed and random effects (beyond the residual)

✓ Mixture models

- A weighted mixture of distributions

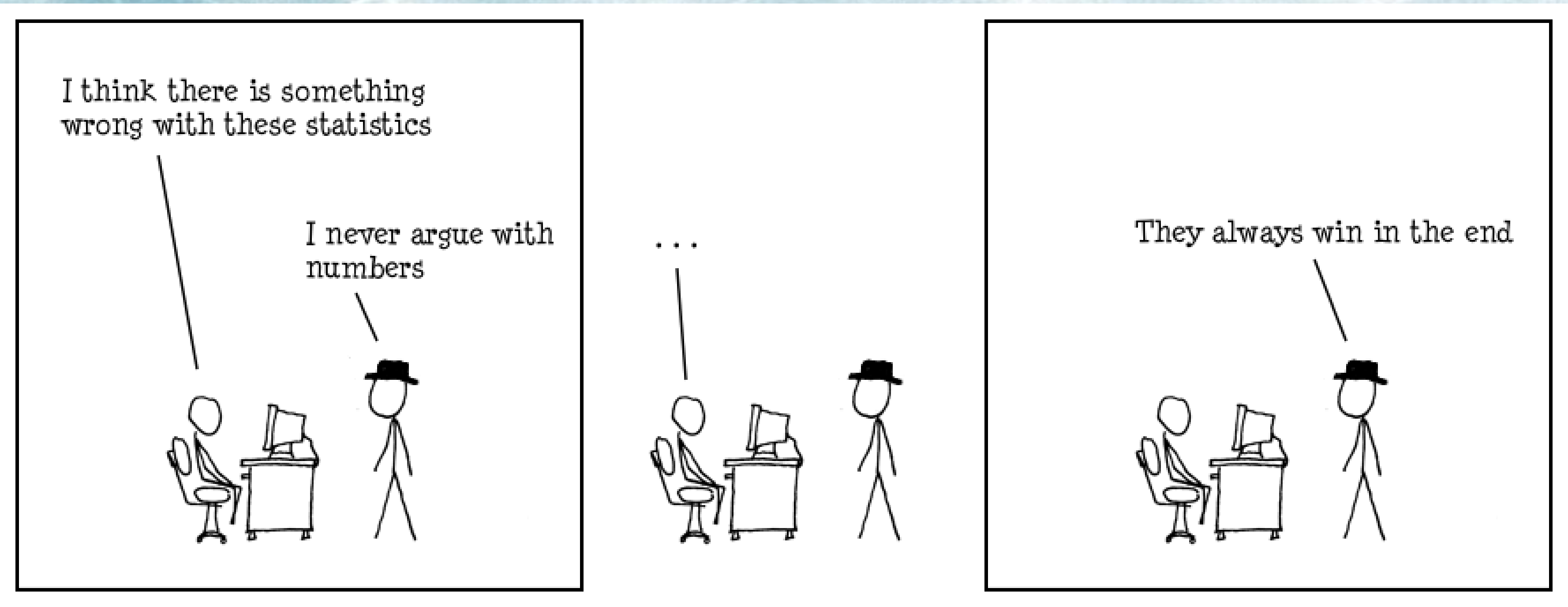
✓ Generalized linear models

- Nonlinear functions, non-normality

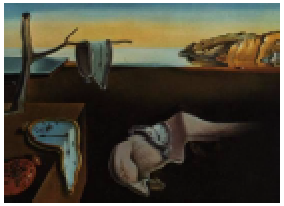


(Bruce Walsh Course, 2016)

2. Mixed Model Analysis with R



Mixed Model Analysis with R



Journal of Statistical Software

October 2015, Volume 67, Issue 1. doi:10.18637/jss.v067.i01

Fitting Linear Mixed-Effects Models Using lme4

Douglas Bates
University of Wisconsin-Madison

Martin Mächler
ETH Zurich

Benjamin M. Bolker
McMaster University

Steven C. Walker
McMaster University

Abstract

Maximum likelihood or restricted maximum likelihood (REML) estimates of the parameters in linear mixed-effects models can be determined using the `lmer` function in the lme4 package for R. As for most model-fitting functions in R, the model is described in an `lmer` call by a formula, in this case including both fixed- and random-effects terms.

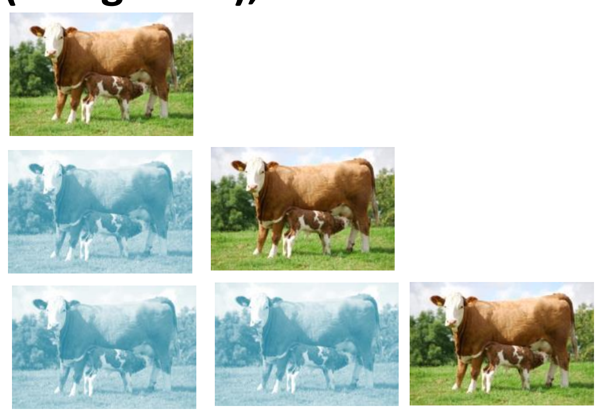
https://cran.r-project.org/web/packages/lme4/index.html

Example of mixed models with lme4

- ✓ The milk yield of 72 cows belonging to 3 parity orders and reared in 6 herds has been measured in the same test-day. Is the production between parity order significantly different, assuming herd as random effects and including also the individual BW as covariate within the mixed model?

- ✓ **Herd: random effect, 6 levels**

- ✓ **Parity: fixed effect (categorical), 3 levels**



- ✓ **Weight: fixed effect (covariate)**

Example of mixed models with lme4

✓ The milk yield of 72 cows belonging to 3 parity orders and reared in 6 herds has been measured in the same test-day. Is the production between parity order significantly different, assuming herd as random effects and including also the individual BW as covariate within the mixed model?

parity	herd 1		herd 2		herd 3		herd 4		herd 5		herd 6	
	weight	milk	weight	milk	weight	milk	weight	milk	weight	milk	weight	milk
1	600	22	580	21	590	21	610	18	620	23	620	16
1	570	18	602	23	610	22	570	19	600	24	600	17
1	590	20	590	19	580	22	600	19	590	23	580	16
1	610	21	590	20	610	21	620	17	600	24	590	17
2	650	23	630	22	650	31	630	27	660	23	650	26
2	600	26	580	27	620	22	620	28	620	23	630	17
2	660	31	620	20	630	21	630	29	650	33	640	27
2	640	31	630	30	660	21	620	29	630	24	650	16
3	690	30	590	23	670	33	680	27	660	34	680	26
3	610	25	640	28	650	32	650	28	640	34	680	26
3	660	21	660	19	660	31	650	27	660	34	650	27
3	670	31	660	30	650	32	660	27	640	33	630	26

Example of mixed models with lme4

- ✓ Mixed model:

Matrix notation

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

y is the vector of phenotypic observations of individual cow milk records, **b** is the vector of fixed effect of parity (3 levels), **u** is the vector of random effect herd (6 levels), **X** and **Z** are incidence matrices respectively referred to b and u, and e is the vector of random residuals. Distributions for random terms are **u**~N(0,**G**) and **e**~N(0,**R**), with $\mathbf{G}=\sigma_u^2 \cdot I$ and $\mathbf{R}=\sigma_e^2 \cdot I$

Linear notation

✓ $$y_{ijk} = \mu + \alpha_i + \beta x_j + u_k + e_{ijk}$$

y_{ijk} is the j-th individual milk record of cow j, μ is the overall mean, α_i is the i-th parity (3 levels), βx_j is the regression coefficient β of the weight x_j, u_k is the k-th herd and e_{ijl} is the residual term. Factors α_i and βx_j are treated as fixed effects, u_k as random term, as the residual.

Example of mixed models with lme4

✓ Exercise with a mixed model (file: cows.txt):

Excercise Cows

```
# rm(list = ls()) # Clean the workspace
# Read the data
cows<-read.table(file="C:/Users/toledo/Dropbox/UNIPD/Biostatistics Curse  R Spring 2018/curso STAT phD 2018 Mi
                 stringsAsFactors = TRUE,header = TRUE,sep = "\t")
cows$parity<-as.factor(cows$parity) # Set parity as factor
cows$herd<-as.factor(cows$herd)     # Set herd as factor
contrasts(cows$parity)<-contr.SAS   # Change the reference grid to SAS
contrasts(cows$herd)<-contr.SAS     # Change the reference grid to SAS

#install.packages("lme4")  # If necesary install the library
library(lme4)              # Call the library to the workspace
library(car)               # Call the library to the workspace
library(lsmeans)           # Call the library to the workspace
# Fit the model
lmm<-lmer(milk ~ parity + weight + (1 | herd) ,data = cows, REML = TRUE)
summary(lmm)            # Results of the mixed model
```

Parity & herd set as categorical effect (since they are numbers, otherwise they are read as a covariate)

Contrasts and solutions written as SAS (last level set to zero)

Example of mixed models with lme4

parity	weight	herd	milk
1	600	1	22
1	570	1	18
1	580	2	21
1	602	2	23
2	650	1	23
2	600	1	26
2	630	2	22
2	580	2	27
3	690	1	30
3	610	1	25
3	590	2	23
3	640	2	28
1	590	1	20
1	610	1	21
1	590	2	19
1	590	2	20
2	660	1	31
…			

Dataset to be imported (cows.txt)

(percorso.file: "../Dropbox/UNIPD/Biostatistics Curse R Spring 2018/curso STAT phD 2018 Mixed Models/data/cows.txt")

- ✓ **Parity: fixed effect (categorical), 3 levels**

- ✓ **Weight: fixed effect (covariate)**

- ✓ **Herd: random effect, 6 levels**

How to write the mixed model: Model 1

```
# Fit the model
lmm<-lmer(milk ~ parity + weight + (1 | herd) ,data = cows, REML = TRUE)
summary(lmm)                 # Results of the mixed model
```

`lmer`: function to fit **linear mixed model** (`glmer` fits generalized linear models and `nlmer` non linear models)

`parity, weight`: fixed effects

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

`(1 | herd)`: random effect (`1`: variance assumed as σ^2_{herd}***I**, with **I** = identity matrix)

`data`: dataset to be read for analysis

`REML`: Method used as a default to solve the mixed model (as opposed to the log-likelihood)

Default output of the mixed model

```
summary(lmm)              # Results of the mixed model
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: milk ~ parity + weight + (1 | herd)
##    Data: cows
##
## REML criterion at convergence: 388
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.5897 -0.5578  0.1287  0.7297  1.5893
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  herd     (Intercept)  3.407   1.846
##  Residual             11.831   3.440
## Number of obs: 72, groups:  herd, 6
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) 17.17861   14.98365   1.146
## parity1     -7.38790    1.63848  -4.509
## parity2     -2.85477    1.09722  -2.602
## weight       0.01732    0.02287   0.757
##
## Correlation of Fixed Effects:
##         (Intr) party1 party2
## parity1 -0.814
## parity2 -0.454  0.613
## weight  -0.998  0.795  0.425
```

- ✓ Num. of rounds performed by REML algorithm to reach convergence
- ✓ Residuals of the model (1Q: 1st quantile; 3Q; 3rd quantile)
- ✓ Variance estimates for random effects (herd and residual) via REML
- ✓ Solutions for fixed effects (**BLUEs**), including intercept. To solve the matrix, the level **`parity3`** was put as equal to zero
- ✓ Correlation of fixed effects, including intercept

Model fitting statistics

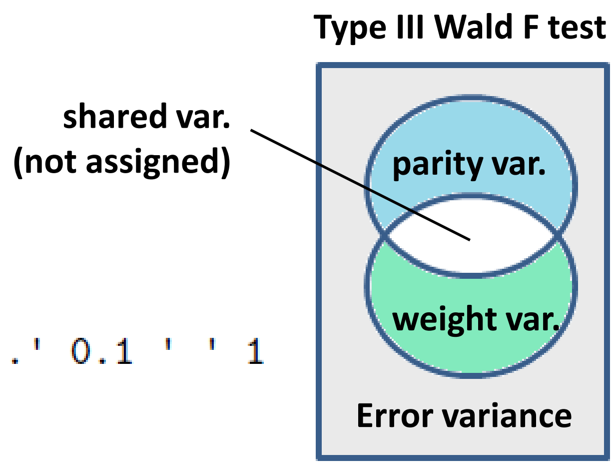
```
AIC(lmm)                    # Akaike's Information Criterion (small is better)

## [1] 400.0195

BIC(lmm)                    # ABayesian Information Criterion (small is better)

## [1] 413.6795
```

✓ **Goodness-of-fit statistics:** useful to compare a series of models differing for some effects included and therefore choose the best fitted one (let's see an example considering also the further model…)

ANOVA type III

```
Anova(lmm, type=3,test.statistic = "F") # ANOVA table SS type III
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: milk
##                    F Df Df.res    Pr(>F)
## (Intercept)   1.2837  1 66.293 0.2612908
## parity       10.0323  2 64.057 0.0001621 ***
## weight        0.5601  1 66.035 0.4568590
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



- ✓ Analysis of variance on fixed effects
- ✓ **`Type III Wald F test:`** each effect receives only the quote of its variance that is not shared with other effects
- ✓ **`Kenward-Roger df:`** method to compute the degrees of freedom
- ✓ **`F:`** value of F statistics (Effect variance/Error variance)
- ✓ **`Df:`** degrees of freedom (1 for intercept and covariates as weight; n-1 for effects)
- ✓ **`Df.res:`** degrees of freedom of error variance
- ✓ **`Pr(>F):`** Probability associated to the F statistics (significance threshold: P<0.05)

Least square means

```
lsmeans(lmm,"parity")       # LSM
```

```
##  parity   lsmean       SE    df lower.CL upper.CL
##  1      20.66521 1.252830 20.13 18.05291 23.27751
##  2      25.19834 1.037309 10.59 23.03543 27.36126
##  3      28.05311 1.187009 16.92 25.57806 30.52816
##
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
```

- ✓ Least Squares Means (`lsmeans`) for the levels of the target fixed effect(s) (here `parity`) as in GLM
  - they estimate the marginal means over a balanced population (**the mean of one effect as it were the only fixed effect in the model**)
  - roughly: target level to estimate is set at 1, the other levels of the effects at 0, and the n levels of the other effects at 1/n. By default, all covariate effects are set equal to their mean values
- ✓ **`lower.CL—upper.CL`:** range for the 95% of the target lsmean distributions

Incidence matrix X

```r
getME(lmm, name=c("X"))     #  Extract the matrix X
```

```
##      (Intercept) parity1 parity2 weight
## 1              1       1       0    600
## 2              1       1       0    570
## 3              1       1       0    580
## 4              1       1       0    602
## 5              1       0       1    650
## 6              1       0       1    600
## 7              1       0       1    630
## 8              1       0       1    580
## 9              1       0       0    690
## 10             1       0       0    610
## 11             1       0       0    590
## 12             1       0       0    640
## 13             1       1       0    590
## 14             1       1       0    610
## 15             1       1       0    590
## 16             1       1       0    590
## 17             1       0       1    660
## 18             1       0       1    640
...
## 71             1       0       0    650
## 72             1       0       0    630
```

✓ **getME:** Extract or Get Generalized Components from a Fitted Mixed Effects Model → this string provides the structure of the incidence matrix for fixed effects X

Incidence matrix X (72 lines, as the obs.)

Vector of fixed effects b

$$\begin{bmatrix} 1 & 1 & 0 & 0 & \mathbf{x}_1 \\ 1 & 1 & 0 & 0 & \mathbf{x}_2 \\ 1 & 1 & 0 & 0 & \mathbf{x}_3 \\ 1 & 1 & 0 & 0 & \mathbf{x}_4 \\ 1 & 0 & 1 & 0 & \mathbf{x}_5 \\ 1 & 0 & 1 & 0 & \mathbf{x}_6 \\ 1 & 0 & 1 & 0 & \mathbf{x}_7 \\ 1 & 0 & 1 & 0 & \mathbf{x}_8 \\ 1 & 0 & 0 & 1 & \mathbf{x}_9 \\ 1 & 0 & 0 & 1 & \mathbf{x}_{10} \\ 1 & 0 & 0 & 1 & \mathbf{x}_{11} \\ 1 & 0 & 0 & 1 & \mathbf{x}_{13} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\mu} \\ \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \boldsymbol{\alpha}_3 \\ \boldsymbol{\beta} \end{bmatrix}$$

Incidence matrix X

```
getME(lmm, name=c("X"))     # Extract the matrix X

##     (Intercept) parity1 parity2 weight
## 1             1       1       0    600
## 2             1       1       0    570
## 3             1       1       0    580
## 4             1       1       0    602
## 5             1       0       1    650
...

## 71            1       0       0    650
## 72            1       0       0    630
## attr(,"assign")
## [1] 0 1 1 2
## attr(,"contrasts")
## attr(,"contrasts")$parity
##   1 2
## 1 1 0
## 2 0 1
## 3 0 0
##
## attr(,"msgScaleX")
## character(0)
```

Reference grid for solutions and contrasts of the model
Solutions of the model obtained by setting at zero the last level of the fixed effect parity (parity3)
See the further command to extract the solutions of the fixed effects (BLUE)

Incidence matrix Z

```r
getME(lmm, name=c("Z"))    #  Extract the matrix Z
```

```
## 72 x 6 sparse Matrix of class "dgCMatrix"
##      1 2 3 4 5 6
## 1    1 . . . . .
## 2    1 . . . . .
## 3    . 1 . . . .
## 4    . 1 . . . .
## 5    1 . . . . .
## 6    1 . . . . .
## 7    . 1 . . . .
## 8    . 1 . . . .
## 9    1 . . . . .
## 10   1 . . . . .
## 11   . 1 . . . .
## 12   . 1 . . . .
## 13   1 . . . . .
## 14   1 . . . . .
. . . . . . . . .
## 70   . . . . 1 .
## 71   . . . . . 1
## 72   . . . . . 1
```

Incidence matrix Z
(72 lines, as the obs.)

$$\begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \end{bmatrix}$$

Vector of random effects u (6 levels, as the number of herds

Solutions for fixed & random effects

```r
getME(lmm, name=c("fixef"))#  Extract the BLUE

## (Intercept)     parity1     parity2      weight
## 17.17860547 -7.38789562 -2.85476524  0.01731762
ranef(lmm)                      # Extract the BLUP

## $herd
##    (Intercept)
## 1   0.19901915
## 2  -0.70047241
## 3   0.81174602
## 4  -0.04831009
## 5   2.30943820
## 6  -2.57142087
```

BLUE & BLUP: different words, same meaning

$$\begin{bmatrix} y_1\\ y_2\\ y_3\\ y_4\\ y_5\\ y_6\\ y_7\\ y_8\\ y_9\\ y_{10}\\ y_{11}\\ y_{12} \end{bmatrix} = \begin{bmatrix} 1&1&0&0&x_1\\ 1&1&0&0&x_2\\ 1&1&0&0&x_3\\ 1&1&0&0&x_4\\ 1&0&1&0&x_5\\ 1&0&1&0&x_6\\ 1&0&1&0&x_7\\ 1&0&1&0&x_8\\ 1&0&0&1&x_9\\ 1&0&0&1&x_{10}\\ 1&0&0&1&x_{11}\\ 1&0&0&1&x_{13} \end{bmatrix} \cdot \begin{bmatrix} \mu\\ \alpha_1\\ \alpha_2\\ \alpha_3\\ \beta \end{bmatrix} + \begin{bmatrix} 1&0\\ 1&0\\ 0&1\\ 0&1\\ 1&0\\ 1&0\\ 0&1\\ 0&1\\ 1&0\\ 1&0\\ 0&1\\ 0&1 \end{bmatrix} \cdot \begin{bmatrix} u_1\\ u_2 \end{bmatrix} + \begin{bmatrix} e_1\\ e_2\\ e_3\\ e_4\\ e_5\\ e_6\\ e_7\\ e_8\\ e_9\\ e_{10}\\ e_{11}\\ e_{12} \end{bmatrix}$$

Fixed effects: solutions as BLUE

Random effects: solutions as BLUP

- ✓ **`getME(lmm, name=c("fixef"))`**: extract the solutions for the fixed effects (***BLUE***) → an intercept is included as default in the model (and its solution is displayed). Then, solutions are expressed as deviations from the intercept
- ✓ **`ranef(lmm)`**: extract the solutions for the random effects (***BLUP***) → in a linear mixed model (lmm) they are the conditional means for random effects

Turn a covariate into a categorical effect

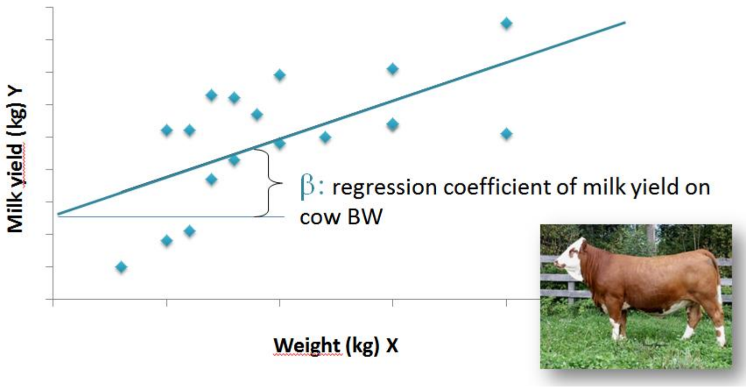
```
# Reshape Weight creating levels for weight
cows$weight_c<-NA
cows$weight_c[cows$weight <= 600]<-1
cows$weight_c[cows$weight > 600 & cows$weight <= 630]<-2
cows$weight_c[cows$weight > 630 & cows$weight <= 660]<-3
cows$weight_c[cows$weight > 660]<-4
table(cows$weight_c)
```

```
##  1  2  3  4
## 19 22 25  6
```

Weight converted into weight classes (**weight_c**)

Num. individuals by weight class

- ✓ In some models it is preferred to convert a covariate into a categorical effect, e.g., when the covariate distribution is not normal, or when the covariate does not exert a linear effect on the phenotype
- ✓ e.g.: let's convert the weight (BW) covariate into weight classes:



Class	BW interval
1	BW ≤ 600 kg
2	600 kg < BW ≤ 630 kg
3	600 kg < BW ≤ 630 kg
4	BW > 660 kg

Model 2: mixed model including the weight_c

```r
cows$weight_c<-as.factor(cows$weight_c) # Set Variable as Factor
contrasts(cows$weight_c)<-contr.SAS       # Set Contrasts as SAS
# Fit the model and test contrasts
lmm<-lmer(milk ~ parity + weight_c + (1 | herd) ,data = cows, REML = TRUE)
summary(lmm)              # Results of the mixed model
```

Weight_c set as factor (since they are numbers, otherwise they are read as a covariate)

Contrasts written as SAS (last level set to zero)

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: milk ~ parity + weight_c + (1 | herd)
##    Data: cows
##
## REML criterion at convergence: 371.5
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -2.6320 -0.4947  0.1119  0.7345  1.8331
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  herd     (Intercept)  3.546   1.883
##  Residual             11.548   3.398
## Number of obs: 72, groups:  herd, 6

## Fixed effects:
##             Estimate Std. Error t value
## (Intercept)   29.584      1.614  18.326
## parity1       -6.365      1.583  -4.021
## parity2       -1.918      1.185  -1.618
## weight_c1     -2.968      2.196  -1.351
## weight_c2     -3.347      1.978  -1.692
## weight_c3     -1.091      1.659  -0.658
```

- ✓ Num. of rounds performed by REML algorithm to reach convergence
- ✓ Residuals of the model (1Q: 1st quantile; 3Q; 3rd quantile)
- ✓ Variance estimates for random effects (herd and residual) via REML
- ✓ Solutions for fixed effects (**BLUEs**), including intercept. To solve the matrix, the level **`parity3`** was put as equal to zero

Model fitting statistics: Model 1 vs. Model 2

MODEL 2

```
AIC(lmm)                    # Akaike's Information Criterion (small is better)
```

```
## [1] 387.5244
```

```
BIC(lmm)                    # Bayesian Information Criterion (small is better)
```

```
## [1] 405.7377
```

- ✓ Goodness-of-fit statistics: now it is possible to compare the AIC & BIC statistics with the same values obtained by running the previous model:

MODEL 1

```
AIC(lmm)                    # Akaike's Information Criterion (small is better)
```

```
## [1] 400.0195
```

```
BIC(lmm)                    # ABayesian Information Criterion (small is better)
```

```
## [1] 413.6795
```

- ✓ Lower values for the new model (Model 2): the model including weight as weight_c fits better the data

Model fitting statistics: Model 1 vs. Model 2

```r
ranef(lmm)                # Extract the BLUP
```

```
## $herd
##    (Intercept)
## 1   0.003092077
## 2  -0.549440320
## 3   0.779687456
## 4   0.157743112
## 5   2.333933433
## 6  -2.725015758
```

Solutions for the random effects (***BLUP***)

```r
Anova(lmm, type=3,test.statistic = "F") # ANOVA table SS type III
```

```
## Analysis of Deviance Table (Type III Wald F tests with Kenward-Roger df)
##
## Response: milk
##                    F Df Df.res    Pr(>F)
## (Intercept) 332.9284  1 38.208 < 2.2e-16 ***
## parity        8.5457  2 62.023 0.0005272 ***
## weight_c      1.3091  3 62.849 0.2793257
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Analysis of variance on fixed effects

Previous model: let's compare weight vs. weight_c

```
##                   F Df Df.res    Pr(>F)
## (Intercept)  1.2837  1 66.293 0.2612908
## parity      10.0323  2 64.057 0.0001621 ***
## weight       0.5601  1 66.035 0.4568590
## ---
```

```r
lsmeans(lmm,"parity")       # LSM
```

```
##  parity   lsmean       SE    df lower.CL upper.CL
##  1      21.36782 1.312626 22.06 18.64606 24.08958
##  2      25.81548 1.140621 14.04 23.45037 28.18058
##  3      27.73300 1.144417 14.17 25.36003 30.10598
##
## Results are averaged over the levels of: weight_c
## Degrees-of-freedom method: satterthwaite
## Confidence level used: 0.95
```

Least square means: **calculated after averaged the weight classes as they were equal** (i.e., the weight_c effect is set to zero)

Coefficients of the model

```
coef(lmm)                    # See the coefficients of you model
```

```
## $herd
##   (Intercept)   parity1   parity2 weight_c1 weight_c2 weight_c3
## 1    29.58749 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 2    29.03496 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 3    30.36409 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 4    29.74215 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 5    31.91834 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 6    26.85939 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
##
## attr(,"class")
## [1] "coef.mer"
```

Parity3 & weight_c4 have no solutions because they have been set to zero (`contrasts`)

✓ Solutions for the fixed effects (**BLUE**), also reported in the **`summary`** of the model:

```
## Fixed effects:
##              Estimate Std. Error t value
## (Intercept)    29.584      1.614  18.326
## parity1        -6.365      1.583  -4.021
## parity2        -1.918      1.185  -1.618
## weight_c1      -2.968      2.196  -1.351
## weight_c2      -3.347      1.978  -1.692
## weight_c3      -1.091      1.659  -0.658
```

✓ Solutions for the random effect (**BLUP**), each expressed as (intercept-BLUP)

```
             Estimate     ## $herd                    (Intercept)
(Intercept)    29.584     ##   (Intercept)          1    29.58749
               29.584  +  ## 1  0.003092077         2    29.03496
               29.584     ## 2 -0.549440320    =    3    30.36409
               29.584     ## 3  0.779687456         4    29.74215
               29.584     ## 4  0.157743112         5    31.91834
               29.584     ## 5  2.333933433         6    26.85939
                          ## 6 -2.725015758
```

Coefficients of the model: contrasts

```
coef(lmm)                    # See the coefficients of you model
```

```
## $herd
##   (Intercept)   parity1   parity2 weight_c1 weight_c2 weight_c3
## 1    29.58749 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 2    29.03496 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 3    30.36409 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 4    29.74215 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 5    31.91834 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
## 6    26.85939 -6.365184 -1.917528 -2.968023 -3.346611 -1.090962
##
## attr(,"class")
## [1] "coef.mer"
```

```
(Intercept)   parity1   parity2 weight_c1 weight_c2 weight_c3
```

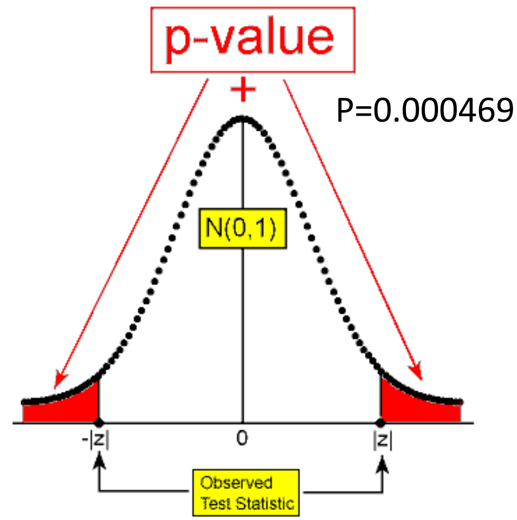
- ✓ This is the order in which fixed & random effects are considered for calculating the contrasts: random effects are included in the first row (see previous slide) and the levels of the two effects, excluding the one set to zero, are in the subsequent rows

Contrast 1: parity 1 vs. parity 2

```
(Intercept)   parity1   parity2 weight_c1 weight_c2 weight_c3
```

```r
library(multcomp)
K <- matrix(c(0, 1, -1, 0, 0, 0), 1)   # Contrast Parity1 vs Parity2
t <- glht(lmm, linfct = K)             # fit a general linear hypothesis test
summary(t)                             # See the results
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = milk ~ parity + weight_c + (1 | herd), data
##     REML = TRUE)
##
## Linear Hypotheses:
##          Estimate Std. Error z value Pr(>|z|)
## 1 == 0    -4.448      1.272  -3.498 0.000469 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```



(Parity1-parity2)/standard error

- ✓ After created a vector **K** with values 1 and -1 for parity 1 & parity 2, & 0 elsewhere, a general linear hypothesis test (`glht`) called **t** is done on **K**
- ✓ Z-test: (parity1 – parity2) is different from 0 at a confidence level P<0.001

Contrast 1: parity 2 vs. parity 3

```
     (Intercept)   parity1   parity2 weight_c1 weight_c2 weight_c3
```

```
K <- matrix(c(0, 0, 1, 0, 0, 0), 1)     # Contrast Parity2 vs Parity3
t <- glht(lmm, linfct = K)               # fit a general linear hypothesis test
summary(t)                               # See the results
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = milk ~ parity + weight_c + (1 | herd), data = cows,
##     REML = TRUE)
##
## Linear Hypotheses:
##          Estimate Std. Error z value Pr(>|z|)
## 1 == 0    -1.918      1.185  -1.618    0.106
## (Adjusted p values reported -- single-step method)
```

✓ (0, 0, 1, 0, 0, 0): since **parity3** has been set to zero before running the model, only **parity2** is written as 1 and tested against zero

```
contrasts(cows$parity)<-contr.SAS   # Change the reference grid to SAS
```

```
## attr(,"contrasts")$parity
##   1 2
## 1 1 0
## 2 0 1
## 3 0 0
##
```

Reference grid for solutions and contrasts of the model
Solutions of the model obtained by setting at zero the last level of the fixed effect parity (parity3)
See the further command to extract the solutions of the fixed effects (BLUE)

Let's go back to see these parts!

Contrast 1: parity 1+2 vs. parity 3

```
(Intercept)   parity1   parity2 weight_c1 weight_c2 weight_c3
```

```
K <- matrix(c(0, 1, 1, 0, 0, 0), 1)     # Contrast Parity1+Parity2 vs Parity3
t <- glht(lmm, linfct = K)              # fit a general linear hypothesis test
summary(t)                              # See the results
```

```
##
##   Simultaneous Tests for General Linear Hypotheses
##
## Fit: lmer(formula = milk ~ parity + weight_c + (1 | herd), data = cows,
##     REML = TRUE)
##
## Linear Hypotheses:
##        Estimate Std. Error z value Pr(>|z|)
## 1 == 0   -8.283      2.491  -3.326 0.000882 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## (Adjusted p values reported -- single-step method)
```

- ✓ (0, 1, 1, 0, 0, 0): the sum of parity1 and parity2, both written as 1, is tested against **parity3** (set to zero before running the model)
- ✓ Using R, contrasts in mixed models are a bit more complicated to be written than in linear models
- ✓ …Let's try to write some contrasts also for the weight_c effect!

Appendix

Model fitting statistics: overview

“fitting a model”

Model fitting statistics: overview

- Model fitting statistics indicate the **goodness-of-fit** of the model regarding to its covariance structure
- AIC, AICC e BIC are all derived from the first statistics, that is
    - **-2 Log (likelihood function L)**
    - **Lower is the value** of -2Log(L), AIC, AICc, BIC, **better is the fitting** of the model (the model should be preferred)
- These statistics allow a comparison among models in terms of goodness of fit, but do not **provide tests for the null hypothesis** → **Likelihood Ratio Test**, LRT
- Other statistics used for model fitting are based on the analysis of the residuals of the model (a «residual» is the difference between the observation – phenotypic record – and its predicted value obtained by running the target model)
- **Analysis of residuals** can use a number of statistics such as:
    - Coefficient of Determination (R^2)
    - MSEP, MAD, PSB (**lower value, better fitting**)
    - correlation phenotypes/predicted values, correlation phenotypes/residuals
    - Bootstrapping or random sampling and correlation of residuals

Akaike Information Criterion (AIC)

(Akaike 1974)

$$AIC = -2\ln L(\hat{\theta} \mid x) + 2k$$

- ✓ where $-2\ln L(\hat{\theta} \mid x)$ is the logarithm of the likelihood function reached at convergence and k is the number of independently adjusted parameters within the model
- ✓ It allow to correct the likelihood of the model for the number of parameters estimated
- ✓ AIC includes the penalty because typically increasing the number of parameters in the model almost always improve the goodness of fit

Corrected AIC (AICc)

(Hurvich & Tsai 1989)

$$\mathbf{AICc} = \mathbf{AIC} + \frac{2(\mathbf{k}+1)(\mathbf{k}+2)}{\mathbf{n}-\mathbf{k}-2}$$

- ✓ where AIC is the Akaike Information Criterion and k is the number of independently adjusted parameters within the model and n the sample size
- ✓ AICc allow to adjust AIC with a correction for finite sample sizes
- ✓ The formula depends upon the statistical model. The formula above assumes that the model is univariate, linear, and has normally-distributed residuals

Bayesian Information Criterion (BIC)

(BIC; Schwartz 1978)

$$BIC = -2\ln L(\hat{\theta} \mid x) + k\ln(n)$$

- ✓ where $-2\ln L(\hat{\theta} \mid x)$ is the logarithm of the likelihood function reached at convergence and k is the number of independently adjusted parameters within the model

- ✓ BIC modifies the AIC increasing the incidence of additional parameters within a model through the introduction of the sample size (n) effect

- ✓ This criterion assigns a greater penalty to models with many parameters (useful for comparing and deciding between different covariance structures)

- ✓ Depending on the model and on the variance structure, it should be preferred alternatively AIC (or AICc) or BIC

Likelihood Ratio Test (LRT)

(Felsenstein (1981)

LRT = 2(logL model with more parameters - logL «null model»)

- ✓ logL is the logarithm of the likelihood function
- ✓ LRT allows to **compare the likelihood of the model as respect to a “null model”** assuming uncorrelated and normally distributed residuals ($\mathbf{R}=\sigma_e^2 \cdot \mathbf{I}$)
- ✓ LRT may also used to test the significance of a covariance estimate by comparing the model with a **null model that does not include the covariance**
- ✓ LRT statistic approximately follows a **chi-square distribution**. To determine if the difference in likelihood scores among the two models is statistically significant, degrees of freedom are necessary. In the LRT, degrees of freedom are equal to the number of additional parameters in the more complex model. → Value compared with critical values of chi-square distribution
- ✓ Only models with a **hierarchy** (i.e. increasing number of parameters, but not different parameters among models) may be compared
- ✓ If the model with a different covariance structure is significantly better than the null model, than the tests of fixed effects may be valued