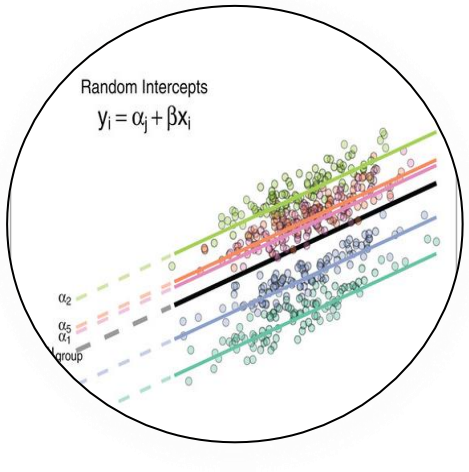
PhD course Mixed Linear Models

Session 2: A gentle introduction to Mixed Models / Variance components models

Gerrit Gort
Biometris, Wageningen University
(slides largely by Bas Engel)

June 10-11-12, 2025



WAGENINGEN UR
For quality of life

Mixed Models

Mixed models is about the analysis of ***dependent*** data.

Overview

- Building a mixed model for dependent data
- Issues about fixed and random effects
- Balanced and unbalanced data
  ANOVA method versus REML
- Summary

First quick impression - 1

Familiar structure from linear model, e.g. analysis of variance:

y = sum of effects + random error term e

main effects and interactions

$\mu + \alpha_i + \beta_j + \alpha\beta_{ij}$

or

regression $\beta_0 + \beta_1 x$

or

…

Normal distribution
Variance independent of the mean

Independent terms, hence independent data

Change to model for dependent data

First quick impression - 2

y = sum of effects + random error term

We will stick to the assumptions of

- linearity
- equal variance
- normality.

But we will relax the assumption of independence.

So, with mixed models you will be able to analyse dependent data.

Dependence between data

Data may be dependent because:

- animals are in the same cage or pen
- pens are on the same farm
- plants are in the same tray or greenhouse
- chemicals are from the same batch
- individuals are from the same family
- data are collected by the same analyst
- individuals are repeatedly measured in time
- …

Start with fictitious data

Animal	A	B
1	31.85	32.03
2	27.31	26.84
3	31.09	29.99
4	31.77	31.47
5	34.62	34.24
6	31.28	31.02
7	36.00	35.50
8	32.39	32.24
9	34.35	34.48
10	27.56	27.14
11	29.05	29.14
12	31.98	31.49

means
(standard deviations):

31.60 versus 31.30
(2.67) (2.72)

How to test for a difference between columns?

t-test !

but which t-test?

Two sample t - test

Sample	Size	Mean	Standard deviation	Standard error of mean
1	12	31.60	2.671	0.7711
2	12	31.30	2.723	0.7859

Difference of means: 0.306
Standard error of difference: 1.101

Test of null hypothesis that population means are the same

Test statistic t = 0.306 / 1.101 = 0.28 with df = (12–1)+(12–1) = 22

P-value = 0.784, so no significant difference found

These are paired observations (on 12 individuals)

Animal	Feed ration A	Feed ration B	Difference
1	31.85	32.03	-0.18
2	27.31	26.84	0.47
.	.	.	.
.	.	.	.
12	31.98	31.49	0.49

mean difference = 0.31

sd = 0.35

Paired t-test !

Paired t - test

One-sample t-test on differences within animals:

Sample	Size	Mean	Standard deviation	Standard error of mean
Difference	12	0.3058	0.3512	0.1014

Test of null hypothesis that mean of differences is equal to 0

Test statistic t = 0.3058 / 0.1014 = 3.02 with df = (12-1) = 11

Probability = 0.012, so we have found a significant difference after all

First conclusions

- It can matter a great deal whether you account for dependence between data or not.
- Here, the treatment effect (0.3058) remains the same, whether we account for dependence or not.
- But se of treatment effect changes considerably (1.101 or 0.3512).
- Consequently, the confidence interval and the P-value of the significance test change considerably too.
- Hence, conclusions and further action would have been quite different when dependence would have been ignored.

What if observations are lost?

Animal	Ration	
	A	B
1	x	x
2	x	x
…	…	…
6	x	x
7	x	
8	x	
9	x	
10		x
11		x
12		x

x = observation is present

This part can be handled by the paired t-test

This part can be handled by the two-sample t-test

But how do we combine these two tests?

Answer: build a model (mixed model) for inspiration.

First, return to simpler situation with complete data.

Start with balanced data

- We start with balanced data, i.e. an equal number of repeats per combination of experimental factors.
- This will lead us to the so-called ANOVA method.

- For unbalanced data the ANOVA does not work out well.
- For unbalanced data, another method has been developed: so-called restricted maximum likelihood or REML.

- For balanced data REML coincides with the ANOVA method.
- So REML is generally applicable, but the ANOVA method is not.

- Despite of this, the ANOVA method is of interest: it is more “transparent” and offers inspiration for e.g. power calculations, while REML is something of a “black box”.

The mixed model – introduction of extra random effects

$$\underline{y}_1 = \mu + \underline{u} + \underline{e}_1$$
$$\underline{y}_2 = \mu + f + \underline{u} + \underline{e}_2$$

← two observations on the same unit (the first animal here) share a common random effect (here a random animal effect).

This common random effect makes them alike = dependent

μ = population mean for ration A (the reference)

f = difference between population means of rations B and A (treatment effect)

u_j = random animal effect $\sim N(0, \sigma_u^2)$

e_{ij} = random (residual) error term $\sim N(0, \sigma^2)$

two variance components: σ_u^2 and σ^2

"Fixed" (μ,f) and "random" (u_j, e_{ij}) effects in one model → mixed model

Extra random effects for dependence - 1

Create dependence by giving observations something in common.

Here, observations on the same animal share random effect for that animal.

Each animal has it's own random effect: u_j

So, random effects u_1, u_2, ... , u_{12} for 12 animals.

Extra random effects for dependence - 2

Each animal has two observations, each with its own error term: $e_{1,j}$, $e_{2,j}$

There are two sets of random effects in the model:

12 animal effects u_1, u_2, ... , u_{12}

24 error terms $e_{1,1}$, $e_{2,1}$, ..., $e_{1,12}$, $e_{2,12}$

Random effects u and e are assumed to be independently normally distributed with zero means and separate variances σ_u^2 and σ^2.

Extra random effects for dependence - 3

31.85	32.03
27.31	26.84
31.09	29.99
31.77	31.47
34.62	34.24
31.28	31.02
36.00	35.50
32.39	32.24
34.35	34.48
27.56	27.14
29.05	29.14
31.98	31.49

These two observations share a positive random animal effect: both are relatively high compared to the respective means of the feed rations.

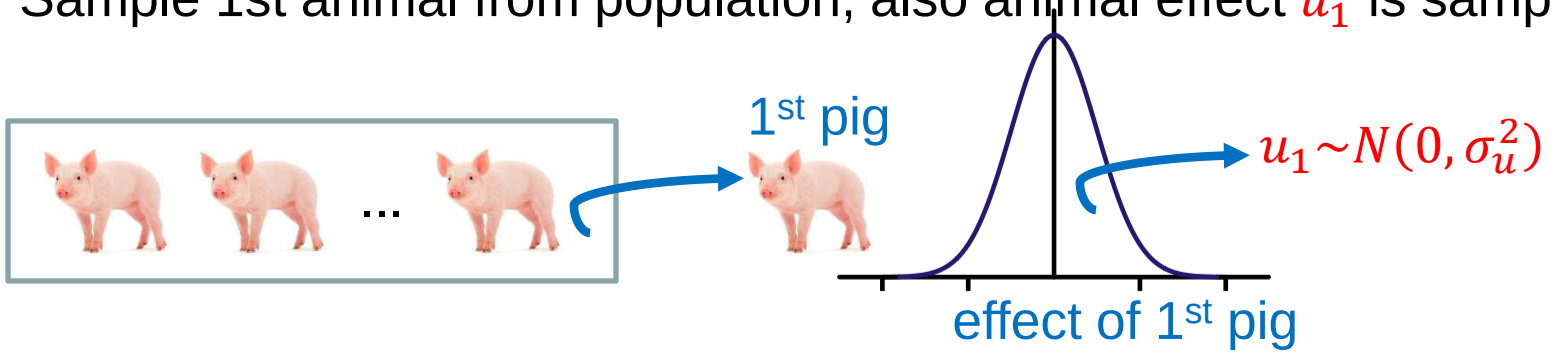
This animal is a fast grower under both feed rations.

These two observations share a negative random animal effect: both are relatively low compared to the respective means of the feed rations.

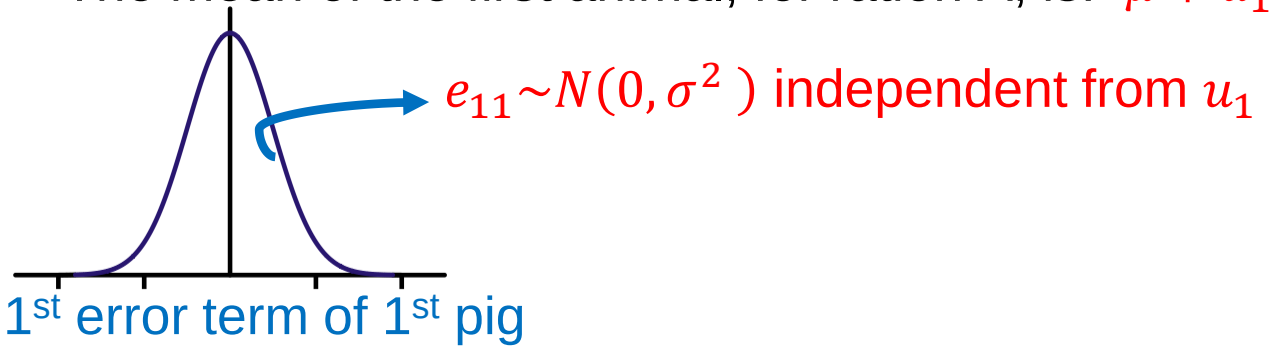
This animal is a slow grower under both feed rations.

How the data are generated (in the model) - 1

- Population mean of all animals with ration A is: μ.
- Sample 1st animal from population, also animal effect u_1 is sampled:



- The mean of the first animal, for ration A, is: $\mu + u_1$.



- 1st observation on 1st animal for ration A is: $y_{11} = \mu + u_1 + e_{11}$.

How the data are generated (in the model) - 2

1st pig

- Population mean of all animals with ration B is: $\mu + f$
- Mean 1st animal, for ration B, is: $\mu + f + u_1$
- 2nd observation, 1st animal for ration is: $y_{21} = \mu + f + u_1 + e_{21}$

$e_{21} \sim N(0, \sigma^2)$, independent from u_1 and e_{11}

- Sample 2nd animal, gives:

$u_2 \sim N(0, \sigma_u^2)$, $e_{12}, e_{22} \sim N(0, \sigma^2)$,

$y_{12} = \mu + u_2 + e_{12}$, $y_{22} = \mu + f + u_2 + e_{22}$, ... etc.

2nd pig

... etc.

The formal model for the feed ration data

$$y_{ij} = \mu + f_i + u_j + e_{ij} \quad i = 1, 2, \quad j = 1 \ldots 12,$$

$u_j \sim N(0, \sigma_u^2)$, independent,

$e_{ik} \sim N(0, \sigma^2)$, independent,

all u_j's and e_{ij}'s mutually independent.

With 1st feed ration as the reference (cornerstone representation): $f_1 = 0, f_2 = f$ (= feed ration effect)

Exercise 1: paired or independent t-test or both?

- In this exercise you repeat the statistical analysis on feed rations as shown in the lecture.
- Open R-script exercise1.r in RStudio.
- Do a paired samples t-test comparing the two feed rations from animals 1-6. How is the dataset organized? Conclusion from the t-test?
- Do an independent samples t-test comparing the two feed rations from animals 7-12. How is the dataset organized? Conclusion from the t-test?
- Fit the mixed model using function `lmer` in package `lme4`. This is like combing the separate paired t-test and independent samples t-test into one single procedure.
- Write down the mixed model in scalar form.
- Give the estimates of the fixed effect parameters and of the variance components.

Several sets of random effects

In general there may be several sets of random effects

e.g. for education data:

random effects for schools,
random effects for classes within schools,
random effects for pupils within classes,
error terms,

4 sets with 4 components of variance

e.g. for animal welfare data:

random effects for test groups,
random effects for animals within test groups,
error terms.

3 sets with 3 components of variance

Fixed or random effects?

To distinguish between fixed and random effects, the following question often helps.

If you would repeat the experiment or study, would you use the same e.g. animals or feed rations?

Fixed or random effects in the example

The answer is no for the animals; ideally we would like to see as many animals as we can.

Animals in the experiment are representative for a population of animals.

Animal effects are random.

Fixed or random effects in the example

The answer is yes for the feed rations, because we have no interest in other feed rations.

Ideally we would like to have as many measurements as possible with these two feed rations only.

Feed ration effects are fixed.

Question: fixed or random effects?

You can have several sets of fixed and random effects in a model.

Identify fixed and random effects in the examples below.

Two instruments for measuring fat thickness of pig carcasses are compared.
Measurements collected in a large number of slaughterhouses.
Instruments handled by different operators.
Measurements collected for females (gilts) and males (boars).

Salt intake from families.
Interview with two persons, a brother and a sister, per family.
Many families from each of three social classes.
Observations collected twice, 2nd time after information and brochure was provided after 1st time.

Measures for dependence

Simple, but popular, measures for dependence:

Covariance: $Cov(y,z) = E\big(y - E(y)\big)\big(z - E(z)\big)$

Correlation: $Corr(y,z) = \frac{Cov(y,z)}{\sqrt{Var(y)}\sqrt{Var(z)}}$

A correlation is a scaled covariance.

Correlation is scale invariant, covariance is not.

["E" stands for "expected value" according to the probability distribution]

Covariance & correlation

Look at observations y_{11} and y_{21} on the same (1st) animal:

$$Cov(y_{11}, y_{21}) =$$

$$Cov(\mu + u_1 + e_{11}, \mu + u_1 + e_{21}) = Cov(u_1 + e_{11}, u_1 + e_{21}) =$$

$$Cov(u_1, u_1) + Cov(u_1, e_{21}) + Cov(e_{11}, u_1) + Cov(e_{11}, e_{21}) =$$

$$Cov(u_1, u_1) = \sigma_u^2$$

$$Var(y_{11}) = Var(\mu + u_1 + e_{11}) =$$

$$Var(u_1 + e_{11}) = Var(u_1) + Var(e_{11}) = \sigma_u^2 + \sigma^2$$

$$Corr(y_{11}, y_{21}) = \frac{Cov(y_{11}, y_{21})}{\sqrt{Var(y_{11})}\ \sqrt{Var(y_{21})}} = \frac{\sigma_u^2}{\sigma_u^2 + \sigma^2}$$

Focus of interest, components or fixed effects

- Sometimes components of variance are of interest.

  E.g. quantify variation in measurements in an assay due to laboratories, analysts within laboratories, batches of material used in the assay, and residual error.

- Often interest is in the fixed effects.

  E.g. compare treatments, like feed rations in the example.

Focus of interest, random effects (BLUP)

Usually random effects $\underline{u}$ are just a way to introduce structured dependence between data.
But random effects may be of primary interest.

Random effects can be predicted: so-called Best Linear Unbiased Predictions or BLUPs.

BLUPs have been in use for a long time to quantify genetic merit of animals, e.g. the breeding value of a bull for milk production.

$$\underline{y} = \mu + f + \underline{u} + \underline{e}$$

- $\underline{y}$: 305 day milk production of a cow
- f: herd effect
- $\underline{u}$: sire random effect
- $\underline{e}$: environmental variation + dam effect + Mendelian sampling

BLUPs are so-called shrinkage estimators and as such are of interest in prediction settings with a high number of explanatory variables.

There are links with ridge-regression, penalized regression, splines and Bayesian statistics.

Notation - indices

$$\underline{y}_{ijk} = \mu + f_i + \underline{u}_{ij} + \underline{e}_{ijk}$$

i = index for social classes, ...

j = index for random effects, e.g. families within classes

k = index for repeats, e.g. people within the same family ...

response y may be some food pattern score

Usual side conditions on fixed effects, such as:

cornerstone representation: $f_1 = 0$, i.e. level 1 is the reference, or
sum to zero condition: $f_1 + f_2 + \ ... = 0$, overall mean is the reference

Notation – indices – example side conditions

$$\underline{y}_{ijk} = \mu + f_i + \underline{u}_{ij} + \underline{e}_{ijk}$$

Say $i = 1, 2, 3$ social classes with means 5, 1, and 6.

cornerstone representation: $f_1 = 0$, i.e. social class 1 is the reference

$$\mu = 5,\ f_1 = 0,\ f_2 = -4,\ f_3 = 1$$

sum-to-zero representation: overall mean of classes is the reference

$$\mu = 4,\ f_1 = 1,\ f_2 = -3,\ f_3 = 2$$

Notation – indices – example feed rations

$$y_{ij} = \mu + f_i + u_j + e_{ij}$$

Back to two feed rations, $i = 1, 2;\ j = 1 \ldots 12$.

Cornerstone representation, say true (unknown) means: 29.5 and 29.2,

so, $\mu = 29.5, f_1 = 0, f_2 = -0.3$

1st animal, say u_1= 2.4,

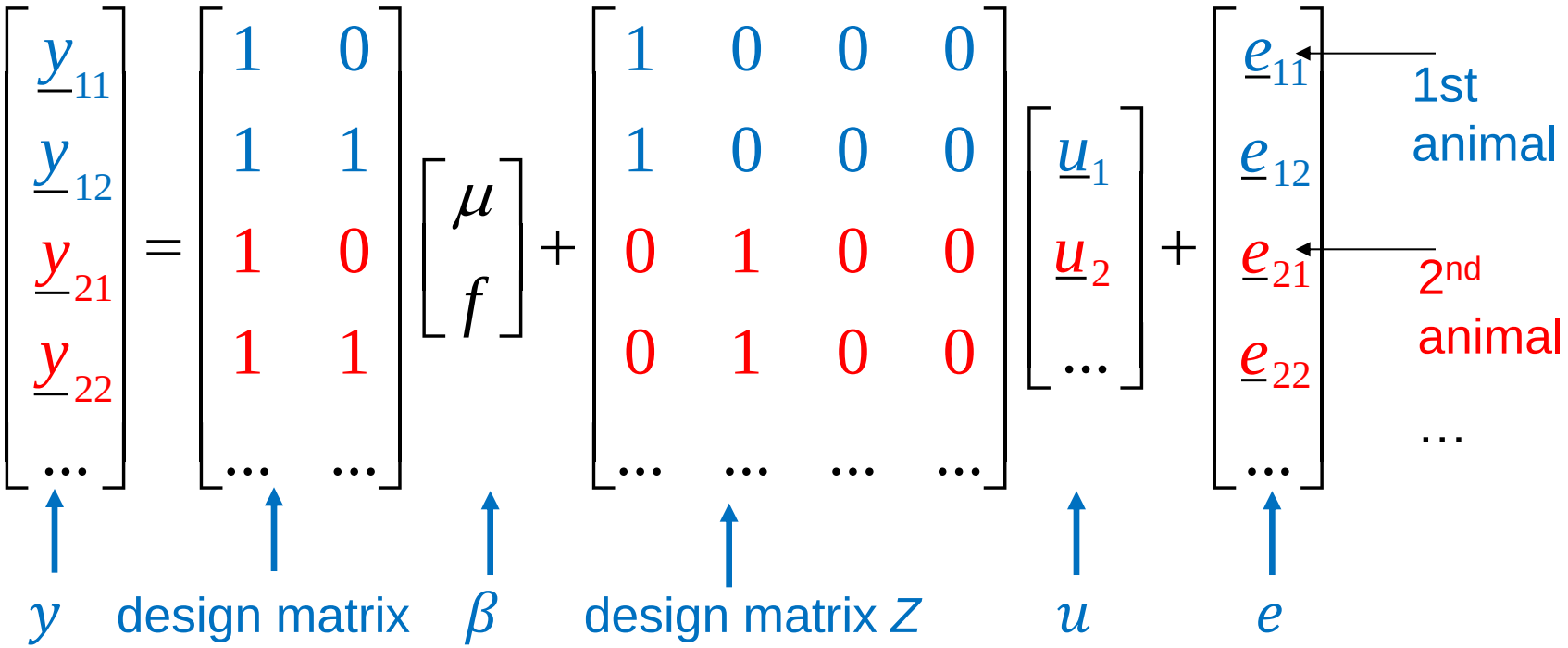
individual means 29.5 + 2.4 = 31.9 and 29.2 + 2.4 = 31.6

1st animal, say $e_{11} = -0.05$ and $e_{21} = 0.43$

observations $y_{11} = 31.9 - 0.05 = 31.85$ and $y_{21} = 31.6 + 0.43 = 32.03$

Notation – vectors & matrices

$$y = X\beta + Zu + e$$



Extending the model - 1

Suppose that we have males and females:

Gender	Animal		Feed ration	
males	1		A	B
	…		…	…
	6		A	B
	7	(1)	A	B
females	…	…	…	…
	12	(6)	A	B

Convenient to renumber animals within sexes (just a detail)

Extending the model - 2

Female:

$$\underline{y}_{11} = \mu + \quad \underline{u}_1 + \underline{e}_{11}$$

$$\underline{y}_{21} = \mu + f + \underline{u}_1 + \underline{e}_{21}$$

random effect 1st animal , say a female

random effect 2nd animal, say a male

Male:

$$\underline{y}_{12} = \mu + s \quad \underline{u}_2 + \underline{e}_{12}$$

$$\underline{y}_{22} = \mu + s + f + sf + \underline{u}_2 + \underline{e}_{22}$$

main effect for sex

main effect feed ration

interaction between sex and feed ration

More formally

$$y_{ijk} = \mu + f_i + s_j + fs_{ij} + \underline{u}_{jk} + \underline{e}_{ijk}$$

$i = 1, 2$ for feed rations, e. g. $f_1 = 0, f_2 = f$,

$j = 1, 2$ for sex, e. g. $s_1 = 0, s_2 = s$,

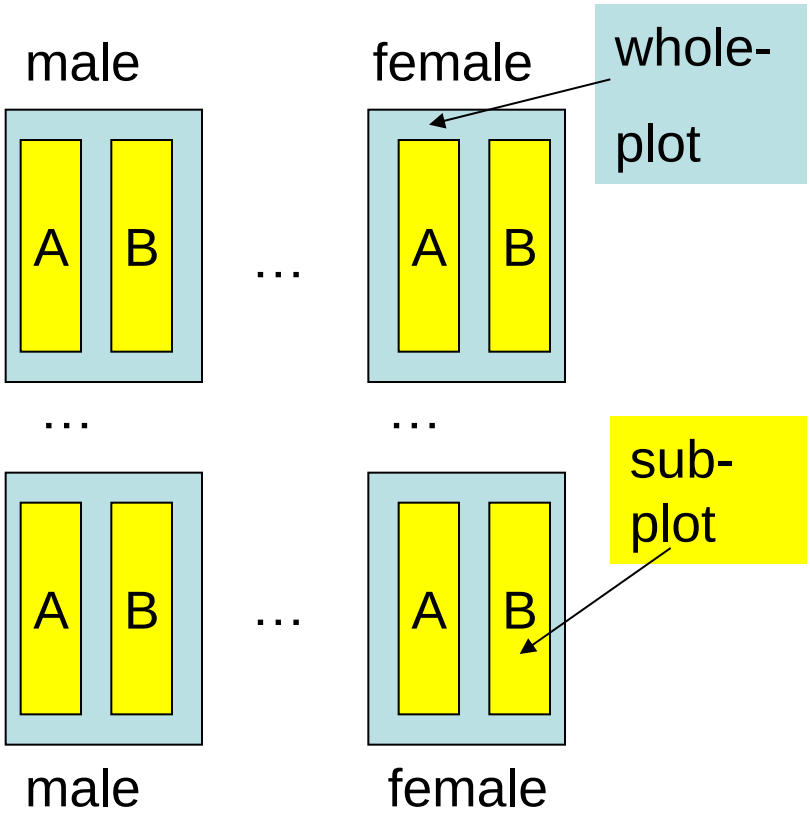
$fs_{11} = fs_{12} = fs_{21} = 0, fs_{22} = fs$ fixed part

$k = 1 \ldots 6$ for animal within gender.

		gender	
		female	male
ration	A	μ	$\mu+s$
	B	$\mu+f$	$\mu+s+f+fs$

A split plot model

This is a split-plot model



Terminology from field experiments:

whole-plots = here: animals
sub-plots = here: repeats within animals

whole-plot treatment = here: gender (male or female)

sub-plot treatment = here: feed ration (ration A or B)

whole-plot variance $= \sigma_u^2$

sub-plot variance $= \sigma^2$

Summary: gender & feed ration

Provide population means, in terms of the parameters of the model, for combinations of gender and feed ration.

Provide population variances, for combinations of gender and feed ration.

Provide population correlations between two observations on the same animal, for combinations of gender and feed ration.

$$\underline{y}_{111} = \mu + \underline{u}_{11} + \underline{e}_{111}$$

$$\underline{y}_{211} = \mu + f + \underline{u}_{11} + \underline{e}_{211}$$

e.g. 1st female, rations A, B

$$\underline{y}_{121} = \mu + s + \underline{u}_{21} + \underline{e}_{121}$$

$$\underline{y}_{221} = \mu + s + f + sf + \underline{u}_{21} + \underline{e}_{221}$$

e.g. 1st male, rations A, B

Estimation & testing
ANOVA method for balanced data
illustrated by split-plot for feed ration data

Balanced data for factorial experiment: equal numbers of observations for combinations of factors (treatments)

Overview:

- Sums of squares (SS) and mean squares (MS) and expected mean square (EMS) as inspiration for estimation and testing
- Estimation of variance components
- F-tests for fixed effects
- Pairwise comparisons by t-tests and Satterthwaite's approximation
- Fixed & random effects re-visited

Estimation & testing
ANOVA method for balanced data
illustrated by split-plot for feed ration data

First steps towards estimation (of variance components) and testing (of fixed effects).

- ANOVA table
- sums of squares (*SS*)
- mean squares (*MS*)
- expected mean square (*EMS*)

ANOVA method

ANOVA method based on sums of squares

Source of variation	Degrees of freedom (df)	Sums of squares (SS)	MS = SS / df	Expected MS E(MS)
Sex	J -1	SSsex	SSsex / (J -1)	$Q_{sex} + I\sigma_u^2 + \sigma^2$
Animal	J(K-1)	SSanimal	…	$I\sigma_u^2 + \sigma^2$
Feed ration	I-1	SSration	…	$Q_{ration} + \sigma^2$
Interaction	(J - 1)(I-1)	SSInteraction	…	$Q_{interaction} + \sigma^2$
Residual	J(K- 1)(I - 1)	SSresidual	…	σ^2
Total	IJK-1	SStot	SStot / (IJK-1)	-

$i = 1 \ldots I$, here $I = 2$ (rations)

$j = 1 \ldots J$, here $J = 2$ (sexes)

$k = 1 \ldots K$, here $K = 6$ (animals per sex, numbered (nested) within sexes)

ANOVA table, some details of expected means squares - 1

Source of variation	Degrees of freedom (df)	Sums of squares (SS)	MS = SS / df	Expected MS E(MS)
Sex	J -1	SSsex	SSsex / (J -1)	$Q_{sex} + I\sigma_u^2 + \sigma^2$
Animal	J(K-1)	SSanimal	…	$I\sigma_u^2 + \sigma^2$
Feed ration	I-1	SSration	…	$Q_{ration} + \sigma^2$
Interaction	(J -1)(I-1)	SSInteraction	…	$Q_{interaction} + \sigma^2$
Residual	J(K-1)(I -1)	SSresidual	…	σ^2
Total	IJK-1	SStot	SStot / (IJK-1)	-

Note that there are two kind of terms in the expected mean squares:

quadratic forms $Q_{...}$ for fixed effects

variance components $\sigma^2_{...}$ for random effects.

When random effects do not contribute to the variation in the model, their associated component of variance is zero.

When fixed effects do not contribute to the population means in the model, their associated quadratic form is zero.

ANOVA table, some details of expected means squares - 2

The coefficients of the components of variance are the number of times that a single random effect is represented in the dataset.

So, the coefficient of σ_u^2 is 2 because there are two observations per animal (per random effect u)

The coefficient of σ^2 is 1 because with each error term e there corresponds only one observation in the dataset.

ANOVA table for the feed ration data in R

```
lmmanova <- aov(y~Ration+Gender+Ration:Gender+Error(Animal))
```

```
Error: Animal
          Df Sum Sq Mean Sq F value  Pr(>F)
Gender     1  83.37   83.37   10.97 0.00785 **
Residuals 10  75.98    7.60
```

This line is about animals

```
Error: Within
              Df Sum Sq Mean Sq F value Pr(>F)
Ration         1 0.5612  0.5612   8.757 0.0143 *
Ration:Gender  1 0.0376  0.0376   0.587 0.4614
Residuals     10 0.6408  0.0641
```

Interaction not significant (P =0.46), main effects for sex (P = 0.008) and feed ration (P = 0.014) significant (i.e. P < 0.05), details will follow.

Balanced lay-out - estimation of components of variance

Estimation of components inspired by expected mean squares of mean squares for random effects:

$$E(MSanimal) = 2\sigma_u^2 + \sigma^2$$
$$E(MSE) = \sigma^2$$

Replace expected mean squares by their outcomes, and solve the equations:

$$MSanimal = 2\sigma_u^2 + \sigma^2$$
$$MSE = \sigma^2$$

Estimated components for the feed ration data

MSanimal = 7.60
MSE = 0.0641

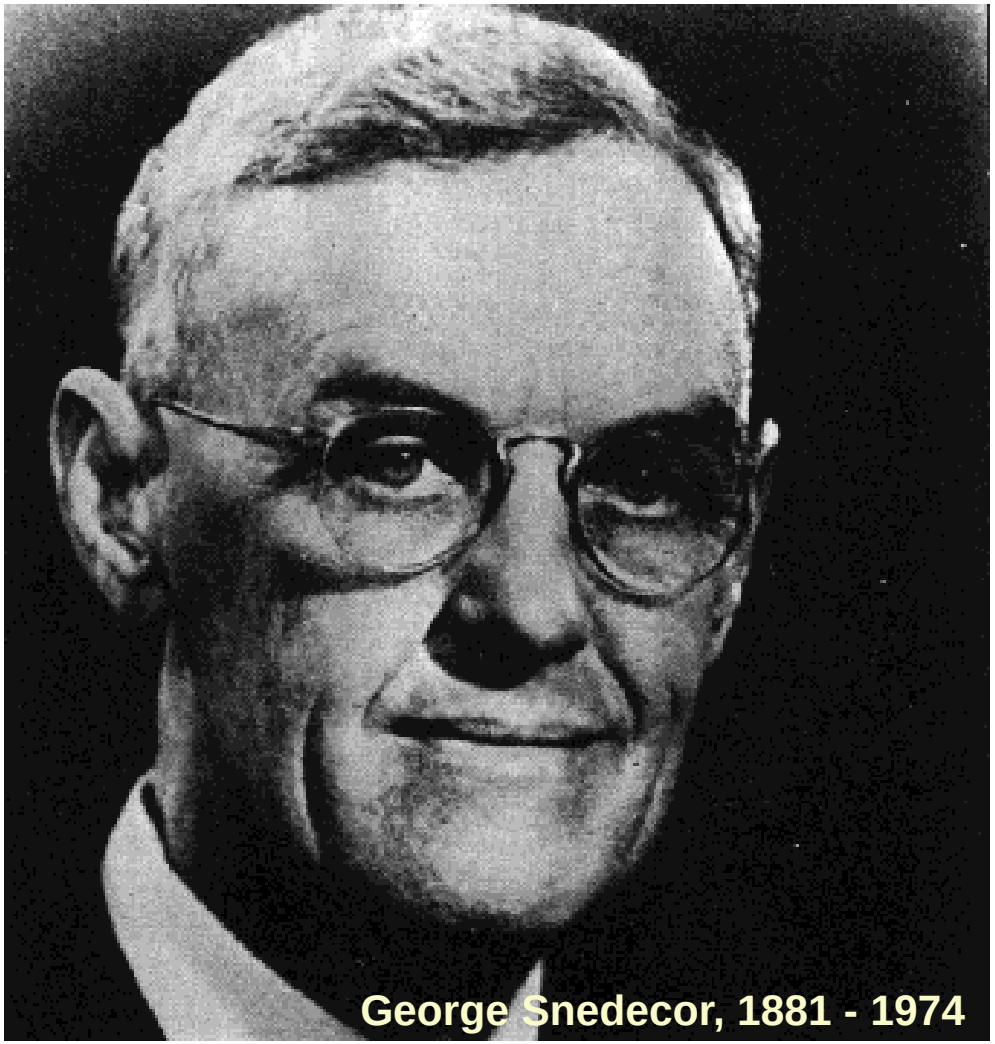
$$E(MSanimal) = 2\sigma_u^2 + \sigma^2$$
$$E(MSresidual) = \sigma^2$$

Replace expected values on the left by observed values:

$$7.60 = 2\sigma_u^2 + \sigma^2$$
$$0.0641 = \sigma^2$$

estimate $\hat{\sigma}^2 = 0.064$,
estimate $\hat{\sigma}_u^2 = (7.60 - 0.0641) / 2 = 3.77$

Snedecor's *F* or Fisher & Snedecor's F-test


George Snedecor, 1881 - 1974

Numerator and denominator of F-test for balanced lay-out - 1

Test statistic F-test looks like:

F = MSnumerator / MSdenominator

MSnumerator corresponds to effects of interest,
e.g. *MSinteraction* for null hypothesis H_0: there is no interaction.

H_0: there is no interaction is equivalent with $Q_{interaction}$ = 0

Numerator and denominator of F-test for balanced lay-out - 2

We want *E*(*MSnumerator*) = *E*(*MSdenominator*) under H_0,
e.g. *F* close to 1 when there is no interaction.

We want to reject H_0 for large values of *F*, e.g. large *F* for marked interaction.

The expected mean squares in the ANOVA table are used to motivate the choice of the mean square in the denominator of *F*.

Exercise F-test for interaction

- *F = MSnumerator / MSdenominator*
- *MSnumerator* corresponds to effects of interest
- Equal E(*MSnumerator*) and E(*MSdenominator*) under H_0
- Reject H_0 for large *F*.

How should *F* look like for testing interaction?

Source of variation	Degrees of freedom (df)	Sums of squares (SS)	MS = SS / df	Expected MS E(MS)
Sex	J -1	SSsex	SSsex / (J -1)	$Q_{sex} + I\sigma_u^2 + \sigma^2$
Animal	J(K-1)	SSanimal	…	$I\sigma_u^2 + \sigma^2$
Feed ration	I-1	SSration	…	$Q_{ration} + \sigma^2$
Interaction	(J - 1)(I-1)	SSInteraction	…	$Q_{interaction} + \sigma^2$
Residual	J(K- 1)(I - 1)	SSresidual	…	σ^2
Total	IJK-1	SStot	SStot / (IJK-1)	-

F-test for interaction, answer - 1

In numerator: $MS_{interaction}$.

$E(MS_{interaction}) = \sigma^2 + Q_{interaction}$

Under H_0: no interaction, we have $Q_{interaction} = 0$.

For denominator we need another MS with expected value σ^2
This is $MS_{residual}$.

So, the test statistic is:

$$F = MS_{interaction}/MS_{residual}$$

Distribution to use for P-value: F distribution, df1 = 1, df2 = 10.

F-Test for interaction, answer - 2

H_0: no interaction, equivalent with $Q_{interaction} = 0$.

F statistic is ratio of $MS_{interaction}$ and $MS_{residual} = MSE$:

$$F_{interaction} = \frac{MS_{interaction}}{MSE}$$

$\approx Q_{interaction} + \sigma^2$

$\approx \sigma^2$

Under H_0, expected values of numerator and denominator are the same.

Large value of *F* statistic suggests $Q_{interaction} > 0$, leads us to reject H_0.

ANOVA table for the feed ration data in R

```
lmmanova<-aov(y ~ Ration+Gender+Ration:Gender + Error(Animal))
```

```
Error: Animal
          Df Sum Sq Mean Sq F value  Pr(>F)
Gender     1  83.37   83.37   10.97 0.00785
Residuals 10  75.98    7.60
```

this is about the animals

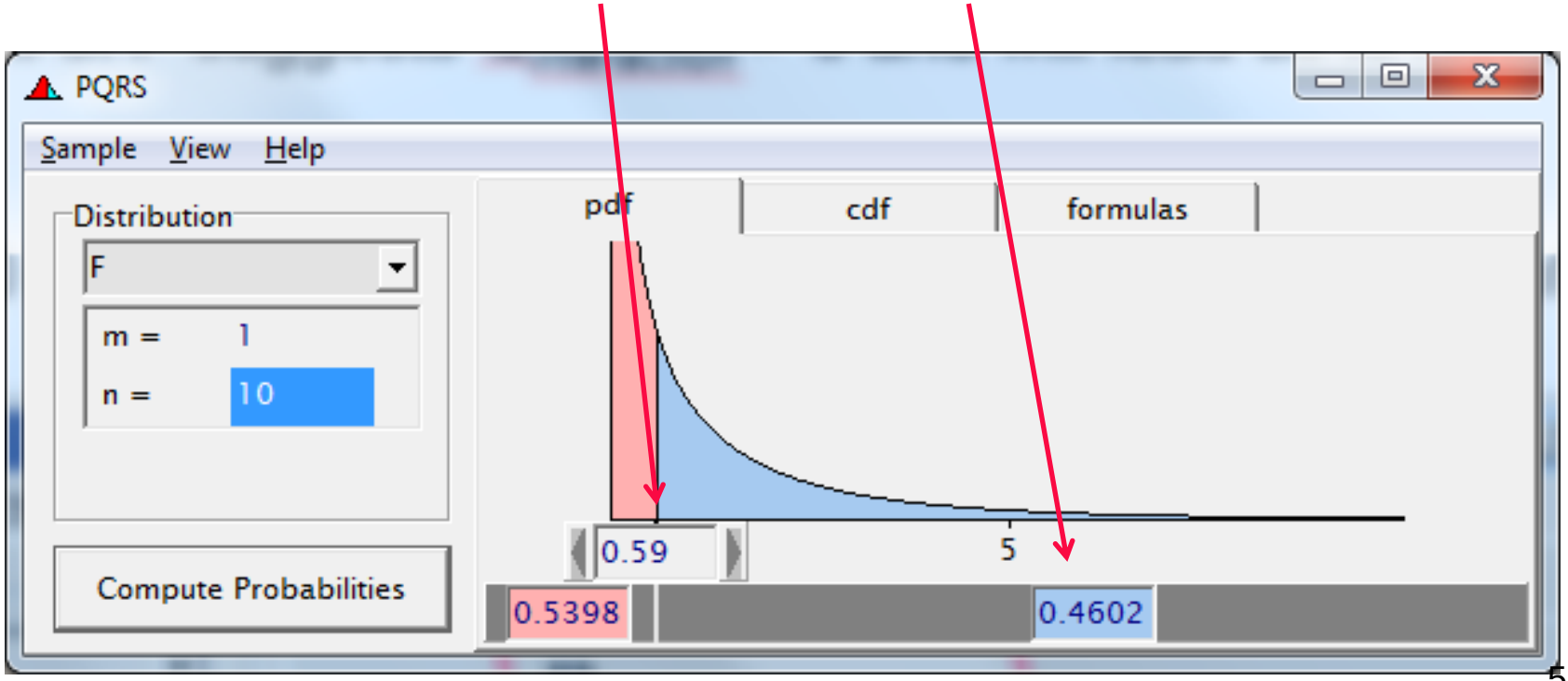
```
Error: Within
              Df Sum Sq Mean Sq F value Pr(>F)
Ration         1 0.5612  0.5612   8.757 0.0143
Ration:Gender  1 0.0376  0.0376   0.587 0.4614
Residuals     10 0.6408  0.0641
```

this is about the residual error terms

Balanced lay-out - test for interaction, P-value

Large values of *F* suggests $Q_{interaction} > 0$ and will lead us to reject H_0.

F = 0.0376 / 0.0641 = 0.59, P-value = 0.46 (area to the right of 0.59).



F-test for main effect feed rations

H_0: no main effect feed rations, equivalent with $Q_{ration} = 0$.

In numerator of F statistic we use MS_{ration} .

For denominator look for *MS* with same expectation under H_0 as MS_{ration}.

This is: $MS_{residual} = MSE$

So: $F = MS_{ration} / MS_{residual}$.

Distribution to use for P-value: F distribution, df1 = 1, df2 = 10.

ANOVA table for the feed ration data in R

```
lmmanova <- aov(y~Ration+Gender+Ration:Gender + Error(Animal))
```

```
Error: Animal
          Df Sum Sq Mean Sq F value  Pr(>F)
Gender     1  83.37   83.37   10.97 0.00785
Residuals 10  75.98    7.60
```

```
Error: Within
              Df Sum Sq Mean Sq F value Pr(>F)
Ration         1 0.5612  0.5612   8.757 0.0143
Ration:Gender  1 0.0376  0.0376   0.587 0.4614
Residuals     10 0.6408  0.0641
```

Exercise F-test for main effect sexes

- $MS_{numerator}$ is MS for main effect sexes
- Under null hypothesis H_0: no main effects sexes, $Q_{sex} = 0$
- $E(MS_{numerator}) = E(MS_{denominator})$ under H_0
- Reject H_0 for large F.

How should F look like for testing main effects for sexes?

Source of variation	Degrees of freedom (df)	Sums of squares (SS)	MS = SS / df	Expected MS E(MS)
Sex	J -1	SSsex	SSsex / (J -1)	$Q_{sex} + I\sigma_u^2 + \sigma^2$
Animal	J(K-1)	SSanimal	…	$I\sigma_u^2 + \sigma^2$
Feed ration	I-1	SSration	…	$Q_{ration} + \sigma^2$
Interaction	(J - 1)(I-1)	SSInteraction	…	$Q_{interaction} + \sigma^2$
Residual	J(K- 1)(I - 1)	SSresidual	…	σ^2
Total	IJK-1	SStot	SStot / (IJK-1)	-

F-test for main effect sexes, answer

H_0: no main effect gender, equivalent with $Q_{sex} = 0$.

$MS_{numerator} = MS_{sex}.$

For denominator look for MS with same expectation under H_0 as MS_{sex}.

This is $MS_{animals}$, and not $MS_{residual}$!

So, F statistic looks like:

$$F_{gender} = \frac{MS_{sex}}{MS_{animals}}$$

$MS_{sex} \approx Q_{Sex} + 2\sigma_u^2 + \sigma^2$

$MS_{animals} \approx 2\sigma_u^2 + \sigma^2$

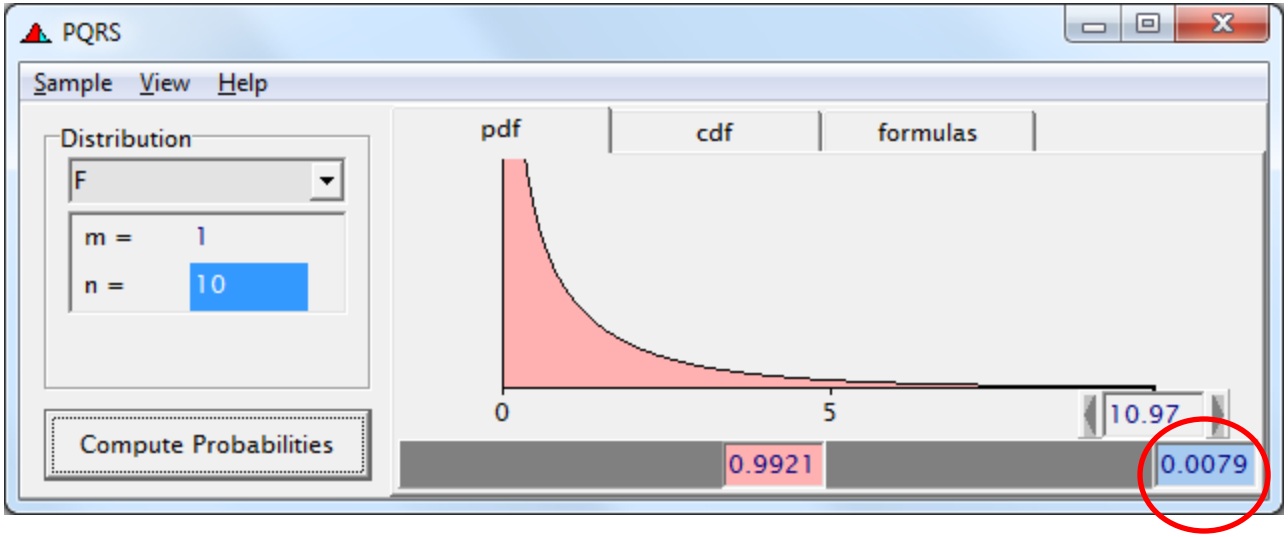
ANOVA table for the feed ration data in R

```
lmmanova <- aov(y ~ Ration + Gender + Ration:Gender + Error(Animal))
```

```
Error: Animal
           Df Sum Sq Mean Sq F value  Pr(>F)
Gender      1  83.37   83.37   10.97 0.00785
Residuals  10  75.98    7.60
```

```
Error: Within
              Df Sum Sq Mean Sq F value Pr(>F)
Ration         1 0.5612  0.5612   8.757 0.0143
Ration:Gender  1 0.0376  0.0376   0.587 0.4614
Residuals     10 0.6408  0.0641
```

P-value main effects gender



F-distribution with df1 = 1 and df2 = 10
Reject for large outcomes of F.
Outcome is F = 10.97.
P-value is area to the right of outcome 10.97, which is 0.0079.

Usual follow-up: pairwise comparisons

When interaction is significant: make pairwise comparisons between combinations of ration and gender.

When interaction is not significant: look at main effects.

When main effects for e.g. rations are significant: make pairwise comparisons between means of rations.

Here trivial since there are only two rations. But imagine an experiment with three littermates, each randomly assigned to one of three feed rations.

Different accuracies in a split-plot analysis – balanced lay-out - 1

sed = standard error of difference (between means)

For sexes:

$$sed^2 = Var(\bar{s}_1 - \bar{s}_2) = Var(\bar{s}_1) + Var(\bar{s}_2)$$

because sexes are attached to different animals (obviously)

$$Var(\bar{s}) = Var\left(\frac{2u_1 + \ldots + 2u_6 + e_1 + \ldots + e_{12}}{12}\right) =$$

$$= \frac{1}{144}(4 * 6 * \sigma_u^2 + 12 * \sigma^2) = \frac{2\sigma_u^2 + \sigma^2}{12} =$$

$$\approx \frac{MSanimal}{12} = \frac{7.59841}{12} = 0.6332$$

$$\text{So}: sed = \sqrt{2 * 0.6332} = 1.1253$$

Notation:

$\bar{S}_1$ = mean of sex 1, say females

Different accuracies in a split-plot analysis – balanced lay-out - 2

```
> emmeans(lmmanova2, pairwise ~ Gender, adjust="none")

$emmeans
 Gender  emmean       SE df lower.CL upper.CL
 1       33.3150 0.795739 10 31.54199 35.08801
 2       29.5875 0.795739 10 27.81449 31.36051

Results are averaged over the levels of: Ration
Degrees-of-freedom method: satterthwaite
Confidence level used: 0.95

$contrasts
 contrast estimate       SE df t.ratio p.value
 1 - 2      3.7275 1.125345 10   3.312  0.0078

Results are averaged over the levels of: Ration
```

the sed for sexes

Different accuracies in a split-plot analysis – balanced lay-out - 3

Because rations can be compared within each animal

For rations:

$$sed^2 = Var(\bar{r}_1 - \bar{r}_2) = Var(\overline{(y - y')}_{\text{within the same animal}}) =$$

$$Var(\overline{(e - e')}_{\text{within the same animal}}) =$$

$$= Var(\bar{e}) + Var(\bar{e}') = 2\frac{\sigma^2}{12} = \frac{\sigma^2}{6} =$$

$$\approx \frac{MSresidual}{6} = \frac{0.06408}{6} = 0.01068$$

So, the random animal effect *u* cancels out !!!

$$\text{So}: sed = \sqrt{0.01068} = 0.1033$$

Question sed's

For comparing sexes: sed = 1.1253

For comparing feed rations: sed = 0.1033

Why is the sed for feed rations smaller than for sexes?

The two-level structure of a split-plot analysis

You have seen that:

- there are two levels of random variation, often corresponding to two levels of randomization: whole plot and sub plot errors
- there are two levels for testing: “against” the whole plot error (in the denominator) or “against” the sub plot error (in the denominator)
- there are two levels for standard errors when comparing means, either within (relatively small se) or between whole plots (relatively large se)

Fixed or random or not at all? - 1

Suppose that you enter animals as fixed and not as random

→ analysis can only employ differences within animals

→ no information about main effects for sexes because animals and sexes are confounded

Fixed or random or not at all? – 2
Fixed animal effects in the feed ration example

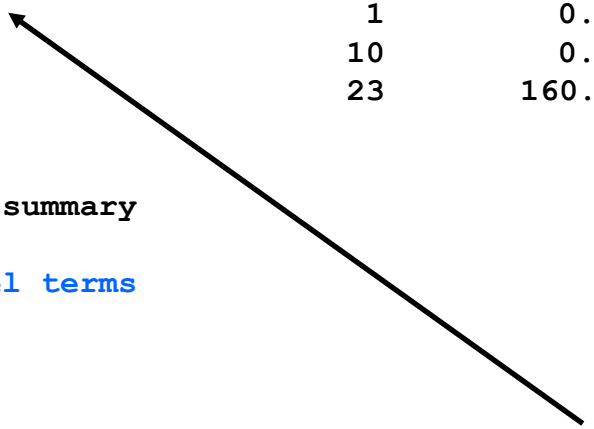
Analysis of variance

Variate: y

Source of variation	df	SS	MS	F-statistic	P-value
Animal	11	159.34961	14.48633	226.05	<.001
Ration	1	0.56120	0.56120	8.76	0.014
Ration.Sex	1	0.03760	0.03760	0.59	0.461
Residual	10	0.64084	0.06408		
Total	23	160.58926			

Information summary

Aliased model terms
Sex



No information about main effects for the sexes, but inference about interaction and main effects for feed rations is OK.

Fixed or random or not at all? - 3

Suppose that you drop animals from the model altogether (and thus ignore the dependence structure)

→ for main effects ration and interaction between ration and sex, P-value will be too high

for main effects sex, P-value will be too low

P-values may be wrong and misleading!

Fixed or random or not at all? – 4
Ignoring animal effects in the feed ration example

compare P-values

Analysis of variance

Variate: y

Source of variation	df	SS	MS	F-stat.	P-value	
Sex	1	83.366	83.366	21.76	<.001	was:**0.0008**
Ration	1	0.561	0.561	0.15	0.706	was:**0.014**
Ration.Sex	1	0.038	0.038	0.01	0.922	was:**0.461**
Residual	20	76.625	3.831			
Total	23	160.59				

Summary so far

- Important to recognize and model dependence structures in data.
- Otherwise P-values may be to high or too low.
- Represent dependence structure by sets of extra random effects.
- Representative example is the split-plot model.
- Different accuracies, depending on whether treatments are applied to whole plots or to sub plots.
- For balanced data there is the ANOVA method.
- This method uses Expected mean squares (EMS) to motivate estimation of components of variance and F-tests for fixed effects

Exercise 2 Split-plot analysis for feed ration data

- Open R-script exercise2.r. In this R-script you can reproduce the analysis for the feed ration data as a split-plot analysis.
- Make the data available.
- To fit the mixed model the ANOVA method can be used (i.e. R function `aov()`) because data are balanced. Using this function, reproduce the ANOVA table as shown on slide 57.
- Do a number of “wrong” analyses using function `lm()` and realize what is wrong:
  - ignoring animal effects completely
  - introducing fixed animal effects
  - nesting animals within sex.
- Study the different accuracies (sed’s) in split-plot analysis.

Question Ants

Introduction of the problem.

Discuss in groups how the model should look like.

When you are fast workers, think about how the F-tests should look like.

Unbalanced data

Overview:

- ML & REML
- An example with several sets of random effects
- Variance-covariance matrices
- More about REML
- Estimation of fixed effects & prediction of random effects
- Mixed model equations (MMEs)
- Odds and ends

Unbalanced data – intro - 1

Animals may die, plots may flood, instruments may fail,
data may be from an observational study, …

So, no equal numbers per combination of levels of factors anymore.

In the feed ration data, there may be some animals that have missing observations for some feed rations.

Throughout we will assume data are completely missing at random.

Unbalanced data – intro 2

Nice properties from balanced data that lead to F-distributions, i.e. independence between mean squares, distributions multiples of chi-square distributions, do not hold anymore.

ANOVA approach is abandoned.

Estimation now by restricted (or residual) maximum likelihood (REML)

No exact F-tests but approximate F-tests for fixed effects.

For a balanced lay-out, ANOVA method and REML yield same results, and exact and approximate F-tests for fixed effects are the same.

Intermezzo - maximum likelihood (ML) estimation

ML is a general method for parameter estimation.

There is an extensive theory about ML.

We will only touch upon it briefly with a simple example.

Simple example to illustrate ML estimation

Estimate probability *p* for an individual to be diseased.

Probability *p* is prevalence of disease in population of individuals.

Random sample of 10 individuals: 0 0 1 1 1 1 0 1 1 1

(1 = diseased, 0 = healthy).

Model: individuals are independent trials, each with probability *p* of success (= individual is diseased).

Probability for observed data:

$$(1-p)\,(1-p)\;\; p\;\; p\;\; p\;\; p\;\; (1-p)\;\; p\,p\;\; p\;\; = p^7\,(1-p)^3$$

The likelihood

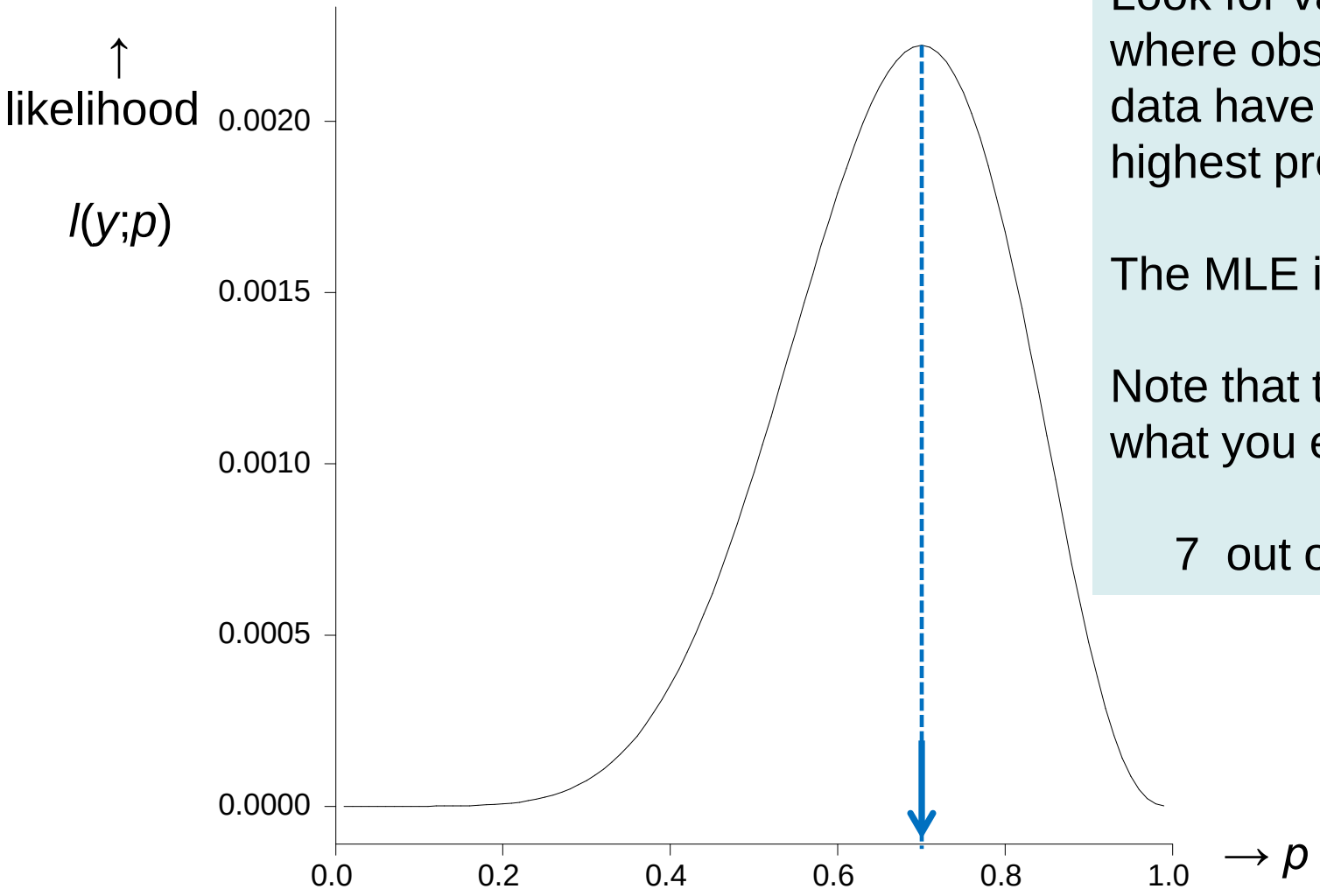
0 0 1 1 1 1 0 1 1 1 ← observed data

Probability for the observed data, as a function of the unknown *p*, is called the likelihood:

$$l((y;p) = p^7 (1-p)^3$$

For any value for *p*, the likelihood tells you how likely that value is to be the true value of *p*.

The maximum likelihood estimator (MLE)



Look for value for *p* where observed data have the highest probability.

The MLE is: 0.7.

Note that this is what you expect:

7 out of 10.

Simple, every day example of ML

Guess which match is played:

m = 1: Ajax - PSV

m = 2: Ajax - FC Vlagtwedde

What is the value of parameter *m*?

Data is the score: 25 – 0.

The loglikelihood

The likelihood l is a product.

For practical and theoretical reasons a sum is preferred over a product.

Therefore, the (natural) logarithm of the likelihood is taken, and the log likelihood $L = \log(l)$ is used.

$L(y;p) = log(l(y;p)) = log(p^7 (1-p)^3) = 7\log(p) + 3\log(1-p)$

Maximum likelihood theory

- ML estimation also works in more complicated situations
- Generally by numerical optimisation of the log likelihood

- There is an extensive theory about ML
- Lots of theoretical results, such as:

MLE approximately normally distributed, around true parameter value, with approximate standard error derived from second order derivatives of log likelihood function

ML and REML - 1

ML is not very good in estimating variances.

This is because ML basically estimates the parameter of interest as if the other parameters were known.

Afterwards the other parameters are replaced by their MLEs.

No correction is made for over adjustment to the data.

Consequently variances tend to be underestimated by ML.

ML and REML - 2

For example: in regression with *p* explanatory variables,
ML estimate for error variance is:

$\frac{SSE}{n} = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2$ and not $MSE = \frac{1}{n-p-1}\sum_i (y_i - \hat{y}_i)^2$

ML does not take estimation of intercept and slopes into account.

ML estimator for variance adjusts too much to the data.

No correction for loss of $(p + 1)$ degrees of freedom.

ML and REML, worst case scenario

Because ML does not always work out well for mixed models, ML is modified into REML.

Example : pairs $(y_{i1}, y_{i2}) \sim N(\mu_i, \sigma^2)$, independent, $i = 1...n$.

$$\text{ML yields}: \hat{\sigma}^2_{ML} = \frac{1}{2n}\sum_{i=1}^{n}\sum_{j=1}^{2}(y_{ij} - \bar{y}_i)^2 \approx \frac{1}{2n} n\sigma^2 = \sigma^2/2$$

So, ML under estimates by 50%!

Problems with many *μ's,* ML does not account for "loss of degrees of freedom"

Example (continued):

$$\text{REML yields}: \hat{\sigma}^2_{REML} = \frac{1}{2n}\sum_{i=1}^{n}(y_{i1} - y_{i2})^2 \approx \sigma^2$$

REML uses 'error' or 'residual' contrasts and is (nearly) unbiased

REML, more formally

REML accounts for loss of degrees of freedom due to fixed effects, while ML does not.

REML is ML applied to error contrasts Ky, with matrix K chosen such that

$$E(Ky) = 0.$$

This way, from the start fixed effects are removed.

REML log likelihood for Ky only depends on unknown components of variance.

Matrix K for the example with paired data

For simplicity, look at $n = 3$ pairs $(y_{11}, y_{12}), (y_{21}, y_{22}), (y_{31}, y_{32})$.

Suppose first two observations from pair 1, next two observations from pair two, last two observations from pair 3.

Matrix K may look like this:

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

This will give you three differences: $y_{11} - y_{12}, y_{21} - y_{22}, y_{31} - y_{32}$

So, effectively we move from 6 observations $y_1 \ldots y_6$ to 3 error contrasts $y_{11} - y_{12}, y_{21} - y_{22}, y_{31} - y_{32}$ as new observations.

We correct for loss of 3 degrees of freedom
(for estimation of means μ_1, μ_2, μ_3 by pair means $\bar{y}_1., \bar{y}_2., \bar{y}_3.$).

Some details of REML

- Log likelihood of REML is unique up to a constant (depending on choice of *K*).

- Kernel of the REML log likelihood (the part that involves the components of variance) is always the same.

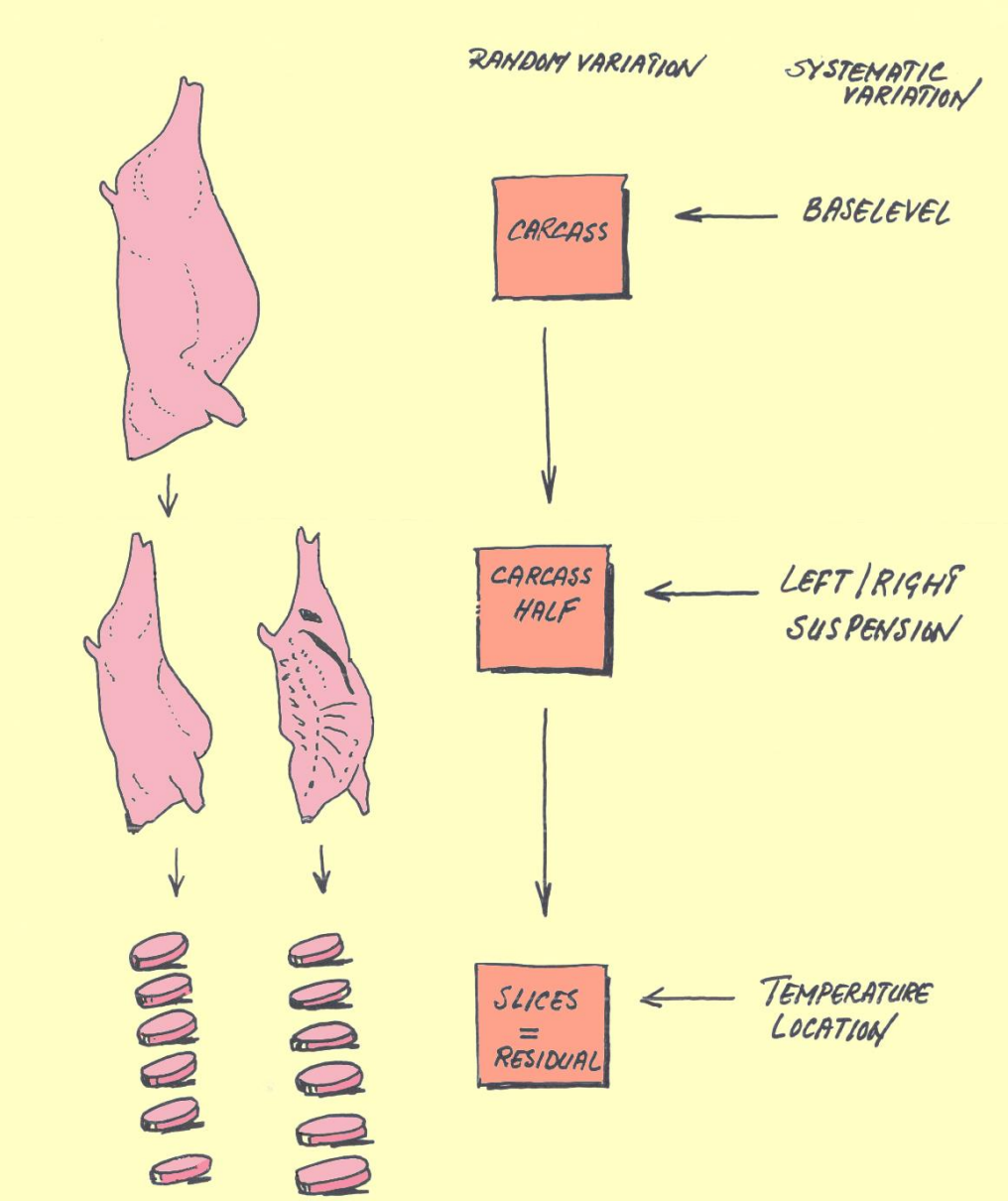
- REML estimates for components of variance do not depend upon choice of *K*.

- There are iterative algorithms for REML, e.g.:

  - EM algorithm, (sometimes too) slow but dependable,
  - Fisher scoring, fast, but less stable,
  - Average Information algorithm, fast and stable.

Example with several sets of random effects

suspension of pig carcasses

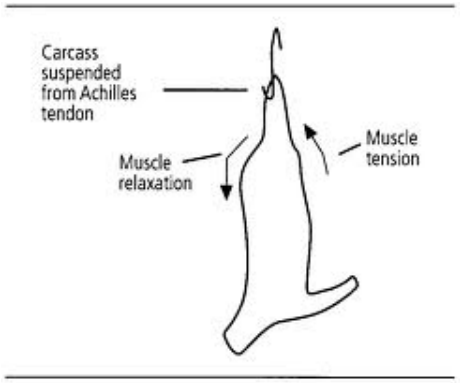


Carcass halves suspended from leg or pelvis

Slices of muscle from carcass halves stored at different temperatures plus control

Slices are from different locations in the muscle

Suspension of pig carcasses – fixed & random effects



Balanced lay-out (variable = sheer force):

Base level	1
Carcass	5, sum to 6 carcasses
Left/right half (systematic)	1
Suspension	1
Carcass half	4, sum to 12 carcass halves
Temperature	5
Location	5
Suspension x Temperature	5
Suspension x Location	5
Slices = residual	40
Total	72 = 6 * 2 * 6

Carcasses

halves

slices

Suspension of pig carcasses - design

In this Latin square: control treatment and 5 temperatures

Locations on muscle

Animals	1	2	3	4	5	6
1	1	2	3	4	5	6
2	2	3	4	5	6	1
3	3	4	5	6	1	2
4	4	5	6	1	2	3
5	5	6	1	2	3	4
6	6	1	2	3	4	5

Not 6 * 6 * 6 = 216 combinations, but only 6 * 6 = 36 combinations in the scheme, therefore not all interactions can be estimated

For some variables, e.g. variable cooking loss, there are missing data

→ unbalanced data, so REML has to be used

Suspension of pig carcasses – ANOVA table

Balanced layout, ANOVA method

Analysis of variance

Variate: SF

Source of variation	df	SS	MS	F-stat.	P-value
Carcass stratum	5	5001.0	1000.2	7.08	
Carcass.Half stratum					
Lr	1	55.7	55.7	0.39	0.564
Susp	1	2724.5	2724.5	19.28	0.012
Residual	4	565.1	141.3	0.45	
Carcass.Half.Slice stratum					
Temp	5	22020.1	4404.0	13.89	<.001
Loc	5	3997.8	799.6	2.52	0.045
Susp.Temp	5	2603.7	520.7	1.64	0.171
Susp.Loc	5	1141.3	228.3	0.72	0.612
Residual	40	12682.4	317.1		
Total	71	50791.4			

Significant main effects

interactions are not significant

Suspension of pig carcasses – follow-up

Tables of means output from GenStat

```
Temp    RAW     T55    T60    T65    T70    T80
        106.5   96.8   68.2   62.4   67.8   62.8
```

Standard errors of differences of means

```
Table   Lr      Susp    Temp    Loc
rep.     36      36      12      12
d.f.      4       4      40      40
s.e.d.  2.80    2.80    7.27    7.27

Table   Susp    Susp
        Temp    Loc
rep.     6       6
s.e.d.  9.79    9.79
d.f.    43.96   43.96
Except when comparing means with the same level(s) of
Susp    10.28   10.28
d.f.    40      40
```

LSD ≈ 2 * 7.27 = 14.5,

so, (raw, T55) versus (T60…T80)

Suspension of pig carcasses - REML

Unbalanced lay-out (Genstat output), main effects model for response cooking loss, REML

Estimated variance components

Random term	component	s.e.
Carcass	0.167	0.506
Carcass.Half	0.277	0.645

Residual variance model

Term	Factor	Model(order)	Parameter	Estimate	s.e.
Carcass.Half.Slice		Identity	Sigma2	3.069	0.695

Tests for fixed effects

(approximate) F-tests from Kenward & Roger

Sequentially adding terms to fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Lr	0.59	1	0.59	4.0	0.485
Susp	4.22	1	4.22	4.0	0.109
Temp	1400.49	4	350.12	39.0	<0.001
Loc	6.03	5	1.21	40.5	0.323

Sequential = type I SS

Dropping individual terms from full fixed model

Fixed term	Wald statistic	n.d.f.	F statistic	d.d.f.	F pr
Lr	0.59	1	0.59	4.0	0.485
Susp	4.22	1	4.22	4.0	0.109
Temp	1400.49	4	350.12	39.0	<0.001
Loc	6.03	5	1.21	40.5	0.323

Dropping = type II SS

Message: denominator degrees of freedom for approximate F-tests are calculated using algebraic derivatives ignoring fixed/boundary/singular variance parameters.

Software in R - 1

For an overview of libraries and routines for linear mixed models in R and in some other packages, see e.g.:

https://bbolker.github.io/mixedmodels-misc/glmmFAQ.html

Software in R - 2

library nlme
function lme for REML

library lme4
function lmer for REML

library pbkrtest
function krmodcomp for approximate F-test

library car
function Anova for approximate F-test (also type III)

library glmmTMB
function glmmTMB for REML, (generalized) linear mixed models

More about approximate F-tests later on.

There is more...

Technical details – estimation of fixed effects

Fixed effects are estimated by generalized (≈ weighted) least squares:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}y, \quad V = Cov(y)$$

Unknown components of variance in *V* replaced by estimated values from REML.

Same estimates for fixed effects can be solved from MMEs (next slide).

Use of MMEs more appealing because much smaller matrices have to be inverted for classic data ($n > p$).
Not for high dimensional data ($n < p$).

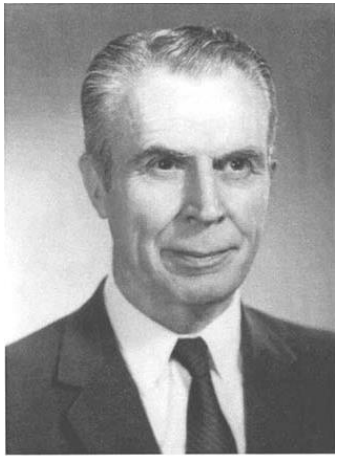
Technical details – Mixed Model Equations (MMEs)

Estimates for fixed effects (β) (BLUE = Best Linear Unbiased Estimates) and predictions for random effects (u) (BLUP = Best Linear Unbiased Predictions) can be solved from the *mixed model equations*:

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + \hat{\sigma}^2/\hat{\sigma}_u^2 I \end{pmatrix} \begin{pmatrix} \beta \\ u \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix}$$

This is how the MMEs look like for a split-plot

MMEs resemble the *normal equations* for the ordinary linear model. Except for part circled in red involving the components of variance.



C.R. Henderson

Technical details – n > p and p > n

The MMEs can reduce the size of matrices to be inverted in so-called n > p situations.

For high dimensional data, so-called p > n situations, other expressions for the calculations are needed.

An example is GBLUP in genetics.

GBLUP – a brief digression

Predict a phenotype from SNPs.

Coding a SNP by x = 0, 1, 2, in the model we have:

$$\underline{y} = \mu + \underline{\beta}_1 x_1 + \underline{\beta}_2 x_2 + \cdots + \underline{e}$$

SNP effects $\beta_1, \beta_2,$ ...are assumed to be random (because there are so many) and BLUPs for the SNP effects $\underline{\beta}_1$, $\underline{\beta}_2$, ... can be derived.

BLUPS are so-called shrinkage estimators.

There is a connection with

- a posterior mean in Bayesian analysis,
- a technique called ridge-regression,
- use of a quadratic penalty function.

Fixed or random effects, revisited

Classical point of view:

- random effects refer to a population, fixed effects do not
- ask the question: should the experiment be repeated, would the same animals be used?
- answer is negative, use of other animals is more appealing, hence animal effects are random (from a population)
- would the same feed rations be used?
- answer is positive, so feed ration effects are fixed (not from a population)

Fixed or random effects, revisited – a broader view

Sometimes a broader view is needed:

- recovery of inter-block information (next slide),
- accounting for higher order interactions by random effects,
- accounting for large number of treatments by random effects,

  e.g. GBLUP in genetics for genome covered by many SNPs
- ...

Broadly speaking, random effects and correlation can be used to model similarity between data, with a relatively small number of parameters.

Fixed or random effects - recovery of inter block information

- In a balanced complete block design, all information about differences between treatments can be derived from contrasts between observations within blocks.

- In an unbalanced block design (perhaps observations were lost), there is information within but also between blocks.

- In that case information between blocks (inter block information) can be recovered when block effects are assumed to be random rather than fixed effects.

  Compare with main effects for sexes in feed ration example: possible to estimate main effects for sexes for random animal effects, but not for fixed animal effects, because all information about sexes is between animals.

Question – dependence in time

25 patients are assumed to represent a random sample from a target population of patients, e.g. patients with a blood pressure disorder.

Patients are repeatedly measured, say at 10 time points per person, i.e. 10 monthly visits of patients to a clinic.

- Suppose that a split-plot model is used to model repeated measurements in time taken on the same person.
- Can you think of any disadvantages of the use of the split-plot model with random effects for patients?
- For dependence in time often more intricate models are required.

Summary, continued

- For balanced data there is the ANOVA method.
- For unbalanced data there is the REML method.
- ANOVA is transparent and can be used for e.g. sample size calculation. REML is more of a black box.

- Exact F-tests for fixed effects for balanced data
- Approximate F-tests (Kenward & Roger) for fixed effects for unbalanced data.

- For balanced data the ANOVA and REML yield the same results (with the possible exception of a negative component estimate in some software, e.g. R).