Statistics with Spa RR ows

Lecture 13

Julia Schroeder

Julia.schroeder@imperial.ac.uk

Outline

- Repeatability
- Pitfalls

Repeatability

- How consistent something is within a group, compared to the whole sample

Repeatability

- How consistent something is within a group, compared to the whole sample
- Sounds familiar?

Repeatability

- How consistent something is within a group, compared to the whole sample
- Sounds familiar?

$$R = \frac{\sigma_A^2}{\sigma_T^2}$$

Repeatability

- How consistent something is within a group, compared to the whole sample
- Sounds familiar?
- It’s the ratio of among-group variance over the total variance

$$R = \frac{\sigma_A^2}{\sigma_T^2}$$

Repeatability

- How consistent something is within a group, compared to the whole sample
- Sounds familiar?
- It’s the ratio of among-group variance over the total variance

$$R = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_W^2}$$

- It’s got lots of interesting uses in biology.

Observer repeatability

- Measuring tarsus consistently is not easy
- Some people measure it 3 times and take the mean
- Are observers consistent in their measures?

Individual behaviour - personality

- Do birds always behave the same way?
- Different from others?

Ecology

- If you calculate diversity indices, are the one in one region more alike than those between regions?

Ecology

- If you calculate diversity indices, are the one in one region more alike than those between regions?

- Is earthworm abundance consistent between day in the same plots?

Repeatability calculated from ANOVA output

$$R = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_W^2} \quad \sigma_A^2 = \frac{MS_A - MS_W}{n_0}$$

$$\sigma_W^2 = MS_W$$

Repeatability calculated from ANOVA output

$$R = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_W^2} \quad \sigma_A^2 = \frac{MS_A - MS_W}{n_0}$$

$$\sigma_W^2 = MS_W$$

Repeatability calculated from ANOVA output

- N_0 not sample size
- Among-group variance difficult to compute (we've had trouble with SS before, remember!)

Repeatability calculated from ANOVA output

- N_0 not sample size
- Among-group variance difficult to compute (we've had trouble with SS before, remember!)
- Balanced dataset

a_i =1,1,2,1,1,1

b_i =3,3,4,3,3,3

c_i =5,5,4,5,5,1

Repeatability calculated from ANOVA output

- N_0 not sample size
- Among-group variance difficult to compute (we've had trouble with SS before, remember!)

- Unbalanced dataset
- Heterogeneous dataset

$a_i = 1,1,2,1,1,1$

$b_i = 3,3,4,3,3,3,4,2,5,5,7$

$c_i = 5,5$

Repeatability calculated from ANOVA output

- N_0 not sample size
- Among-group variance difficult to compute (we've had trouble with SS before, remember!)

$a_i = 1,1,2,1,1,1$

$b_i = 3,3,4,3,3,3,4,2,5,5,7$

$c_i = 5,5$

- Unbalanced dataset
- Heterogeneous dataset

- In ecology we have those often. In exact sciences less so (medicine, any planned experiments)

Repeatability calculated from ANOVA output

- N_0 not sample size
- Among-group variance difficult to compute (we've had trouble with SS before, remember!)

$$n_0 = [\frac{1}{a\text{-}1}[\sum_{i=1}^{a} n_i - (\frac{\sum_{i=1}^{a} n_i^2}{\sum_{i=1}^{a} n_i})]$$

a = number of groups

n_i = sample size in each group i

LMMs

- Linear mixed models

- Combine linear models and variance analysis

Nested data structure

- Repeated measures
- Offspring in families
- …

LMMs

$$y_{i,j} = b_0 + b_1 x_{i,j} + \alpha_j + \varepsilon_{i,}$$

LMMs

$$y_{i,j} = b_0 + b_1 x_{i,j} + \alpha_j + \varepsilon_{i,}$$

Linear model bit that we know

Estimates FIXED intercept, covariates and factors

LMMs

$$y_{i,j} = b_0 + b_1 x_{i,j} + \alpha_j + \varepsilon_i,$$

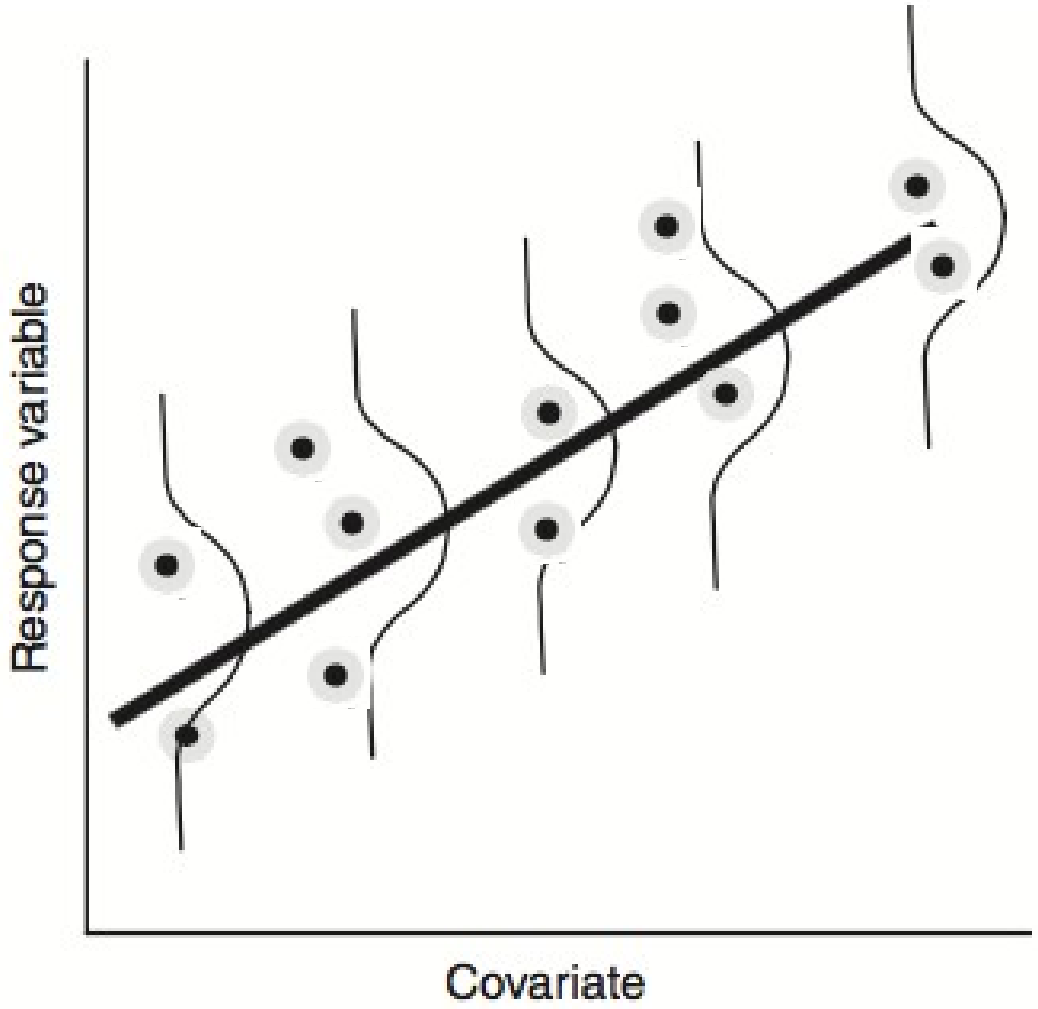
Random factor for a group j (i.e. BirdID)
Estimate variance component AMONG BIRDS

LMMs

$$y_{i,j} = b_0 + b_1 x_{i,j} + \alpha_j + \varepsilon_i,$$

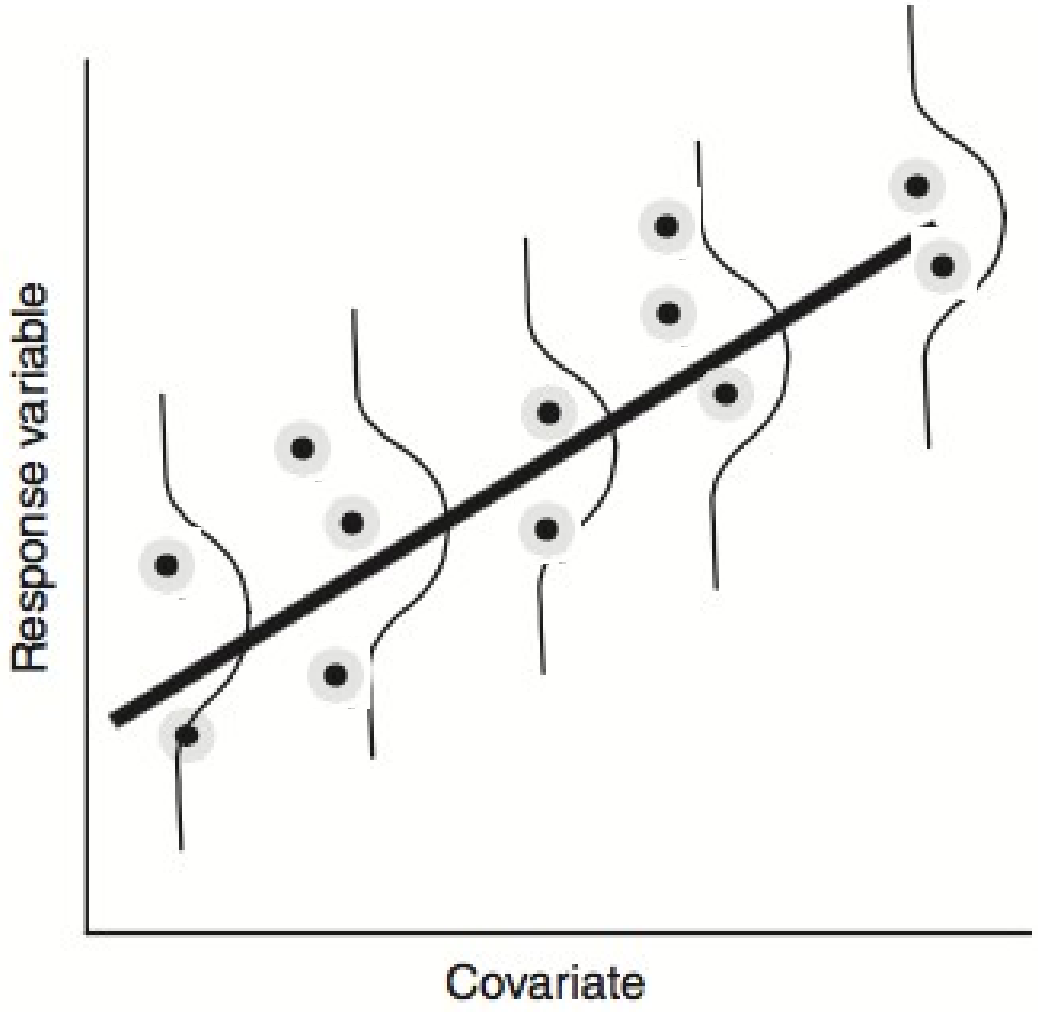
Residual variance

LMMs



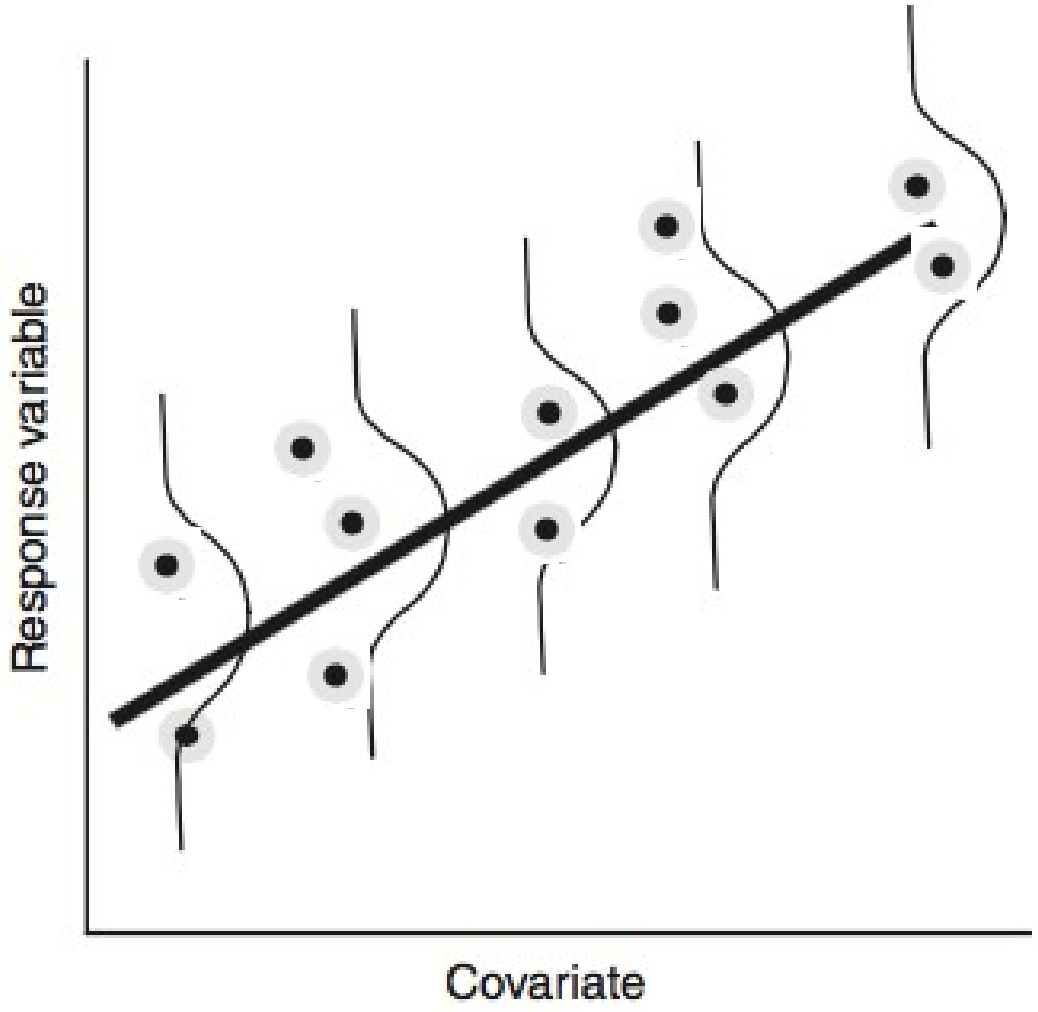
LMMs

- Estimate variance components and fixed parameter estimates simultaneously



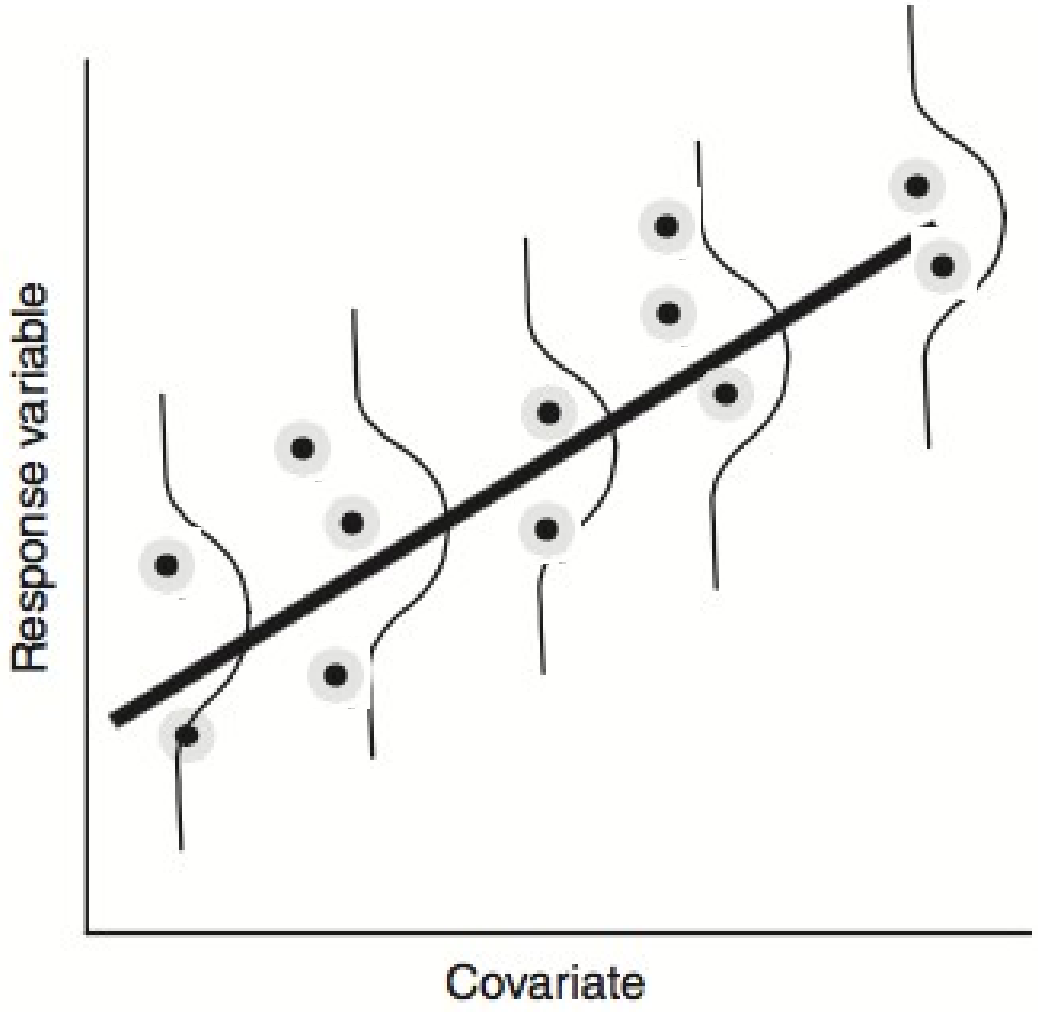
LMMs

- Estimate variance components and fixed parameter estimates simultaneously
- Complicated but very useful



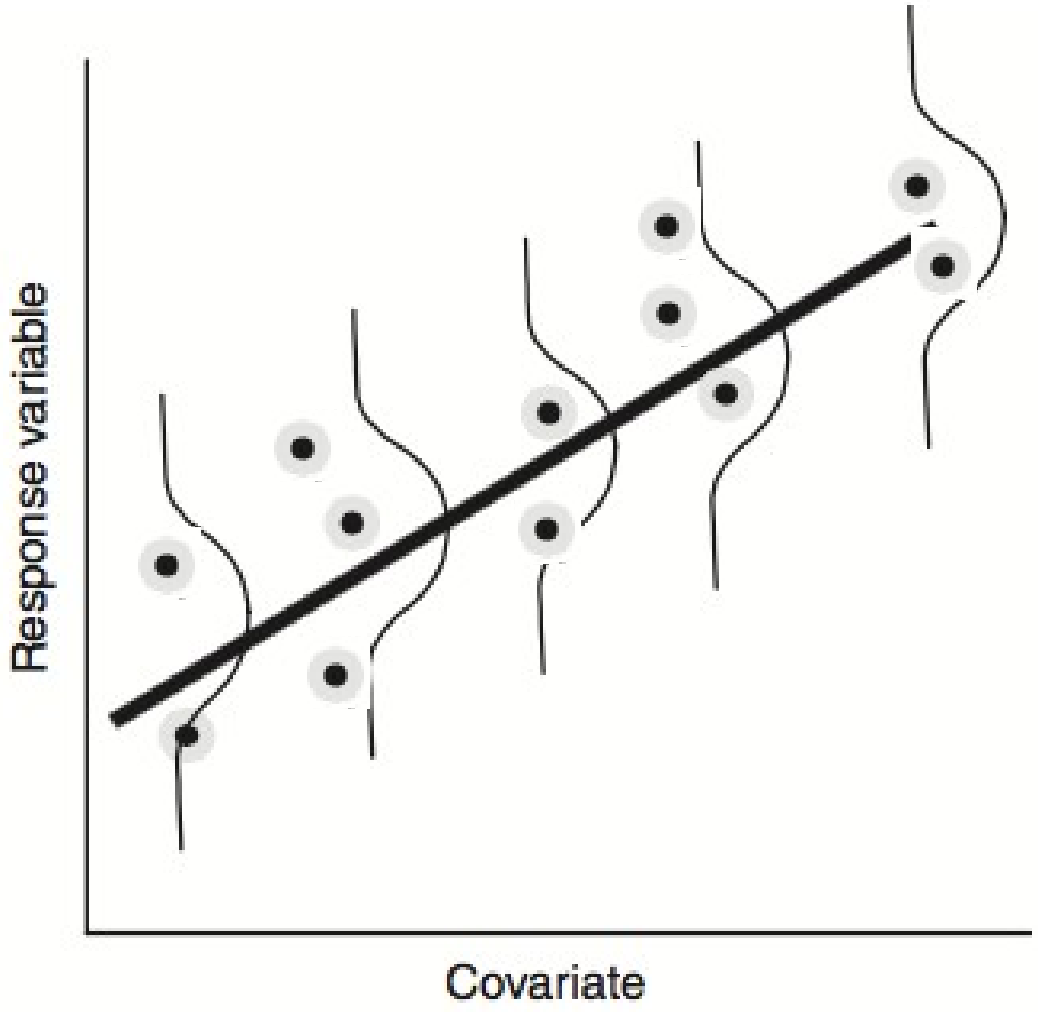
LMMs

- Estimate variance components and fixed parameter estimates simultaneously
- Complicated but very useful
- Better than ANOVA



LMMs

- Estimate variance components and fixed parameter estimates simultaneously
- Complicated but very useful
- Better than ANOVA



Linear mixed models:

- Can deal with heterogenous datasets (unbalanced groups)
- Estimates variance components and fixed effects *at the same time*

Comparing ANOVA vs Linear mixed model

- multiple students measuring the tarus of one bird

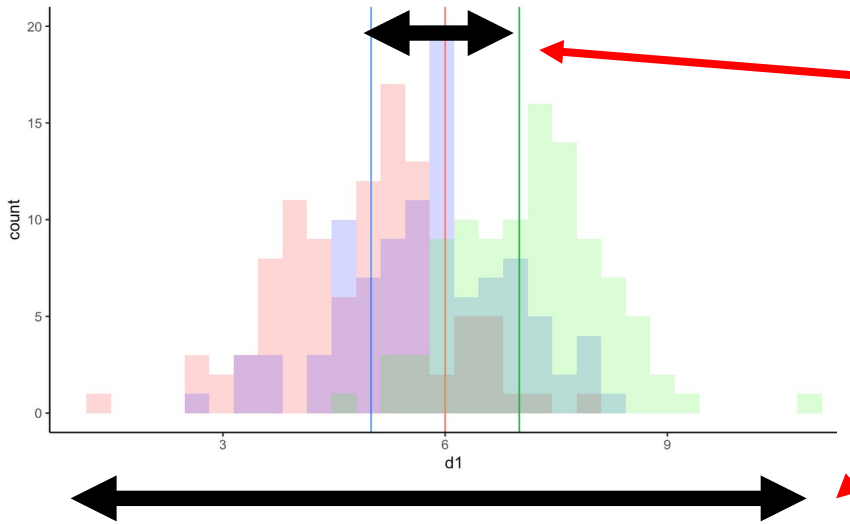
- → observer repeatability describes how much variation is explained by different observers
- Measurement of methodological precision

- R = variance explained by student ID/total variance

ANOVA *vs.* LMM

```
> lm1<-lm(Tarsus~StudentID, data=b)
> anova(lm1)
Analysis of Variance Table

Response: Tarsus
          Df Sum Sq Mean Sq F value    Pr(>F)
StudentID 79 561.95  7.1133  4.2004 2.095e-09 ***
Residuals 71 120.24  1.6935
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

?

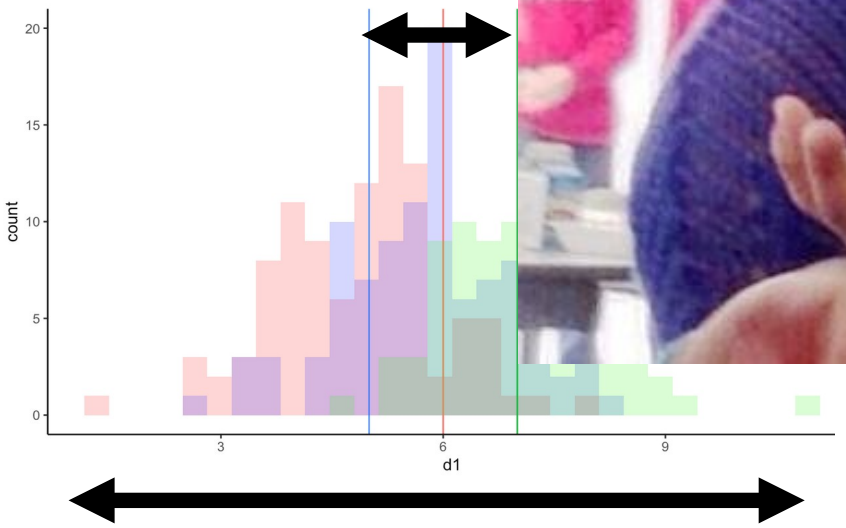


Sum
or
Var(Tarsus)

```
> lmm1<-lmer(Tarsus~1+(1|StudentID), data=b)
> summary(lmm1)
Linear mixed model fit by REML ['lmerMod']
Formula: Tarsus ~ 1 + (1 | StudentID)
   Data: b

REML criterion at convergence: 620.7

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.1168 -0.3591  0.0201  0.4431  3.3629

Random effects:
 Groups    Name        Variance Std.Dev.
 StudentID (Intercept) 2.963    1.721
 Residual              1.684    1.298
Number of obs: 151, groups:  StudentID, 80

Fixed effects:
            Estimate Std. Error t value
(Intercept)  18.7101     0.2228   83.99
```

ANOVA

```
> lm1<-lm(Tarsus~StudentID, data=b)
> anova(lm1)
Analysis of Variance Table

Response: Tarsus
          Df Sum Sq Mean Sq F value
StudentID 79 561.95  7.1133  4.2004
Residuals 71 120.24  1.6935
---
Signif. codes:  0 ‘***’ 0.001 ‘**’
```

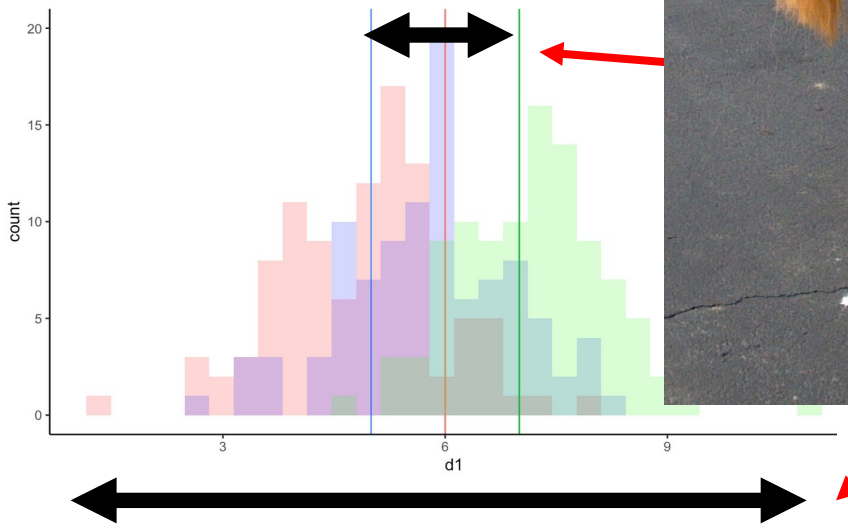


$$\sigma^2_A$$

$$S_W$$

LMM

- Repeatability: Var_A
- 2.96/(2.96+1.68) =
- 64% of variance in
  is explained by diff
  student measuring



Var(Tarsus)

```
ID), data=b)

'lmerMod']
ntID)

20.7

    Max
 3.3629

Std.Dev.
1.721
1.298
dentID, 80

t value
  83.99
```

Learning aim

- Repeatability is intra-class correlation coefficient
- Ratio of how much variance is explained by groups
- N_0 is horrible
- Linear mixed models are better at estimating variance components than ANOVA