

Preliminaries

```
> library(lme4)
> library(DAAG)
```

The final two sentences of Exercise 1 are challenging! Exercises 1 & 2 should be asterisked.

Exercise 1

Repeat the calculations of Subsection 2.3.5, but omitting results from two vines at random. Here is code that will handle the calculation:

```
n.omit <- 2
take <- rep(TRUE, 48)
take[sample(1:48,2)] <- FALSE
kiwishade.lmer <- lmer(yield ~ shade + (1|block) + (1|block:plot),
                      data = kiwishade,subset=take)
vcov <- show(VarCorr(kiwishade.lmer))
gps <- vcov[, "Groups"]
print(vcov[gps=="block:plot", "Variance"])
print(vcov[gps=="Residual", "Variance"])
```

Repeat this calculation five times, for each of `n.omit = 2, 4, 6, 8, 10, 12` and `14`. Plot (i) the plot component of variance and (ii) the vine component of variance, against number of points omitted. Based on these results, for what value of `n.omit` does the loss of vines begin to compromise results? Which of the two components of variance estimates is more damaged by the loss of observations? Comment on why this is to be expected.

For convenience, we place the central part of the calculation in a function. On slow machines, the code may take a minute or two to run.

```
> trashvine <- function(n.omit=2)
+ {
+   k <- k+1
+   n[k] <- n.omit
+   take <- rep(T, 48)
+   take[sample(1:48, n.omit)] <- F
+   kiwishade$take <- take
+   kiwishade.lmer <- lmer(yield ~ shade + (1 | block) + (1|block:plot),
+                         data = kiwishade, subset=take)
+   varv <- as.numeric(attr(VarCorr(kiwishade.lmer), "sc")^2)
+   varp <- as.numeric(VarCorr(kiwishade.lmer)$`block:plot`)
+   c(varp, varv)
+ }
> varp <- numeric(35)
> varv <- numeric(35)
> n <- numeric(35)
> k <- 0
> for(n.omit in c( 2, 4, 6, 8, 10, 12, 14))
+ for(i in 1:5){
+   k <- k+1
```

```

+   vec2 <- trashvine(n.omit=n.omit)
+   n[k] <- n.omit
+   varp[k] <- vec2[1]
+   varv[k] <- vec2[2]
+ }

```

We plot the results:

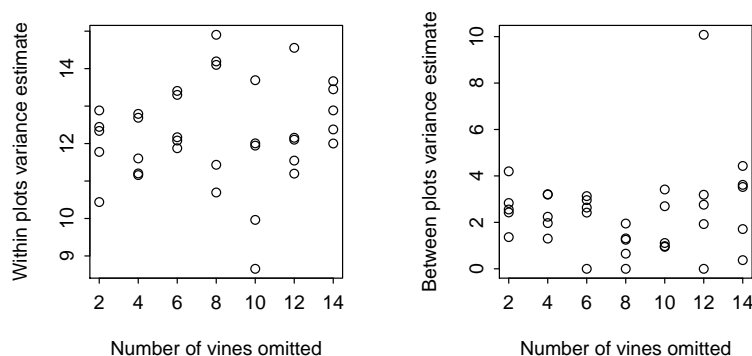


Figure 1: Within, and between plots variance estimates, as functions of the number of vines that were omitted at random

As the number of vines that are omitted increases, the variance estimates can be expected to show greater variability. The effect should be most evident on the between plot variance. Inaccuracy in estimates of the between plot variance arise both from inaccuracy in the within plot sums of squares and from loss of information at the between plot level.

At best it is possible only to give an approximate d.f. for the between plot estimate of variance (some plots lose more vines than others), which complicates any evaluation that relies on degree of freedom considerations.

Exercise 2

Repeat the previous exercise, but now omitting 1, 2, 3, 4 complete plots at random.

```

> trashplot <- function(n.omit=2)
+ {
+   k <- k+1
+   n[k] <- n.omit
+   plotlev <- levels(kiwishade$plot)
+   use.lev <- sample(plotlev, length(plotlev)-n.omit)
+   kiwishade$take <- kiwishade$plot %in% use.lev
+   kiwishade.lmer <- lmer(yield ~ shade + (1 | block) + (1|block:plot),
+                         data = kiwishade, subset=take)
+   varv <- as.numeric(attr(VarCorr(kiwishade.lmer), "sc")^2)
+   varp <- as.numeric(VarCorr(kiwishade.lmer)$`block:plot`)
+   c(varp, varv)
+ }
> varp <- numeric(20)

```

```

> varv <- numeric(20)
> n <- numeric(20)
> k <- 0
> for(n.omit in 1:4)
+ for(i in 1:5){
+   k <- k+1
+   vec2 <- trashplot(n.omit=n.omit)
+   n[k] <- n.omit
+   varp[k] <- vec2[1]
+   varv[k] <- vec2[2]
+ }

```

Again, we plot the results:

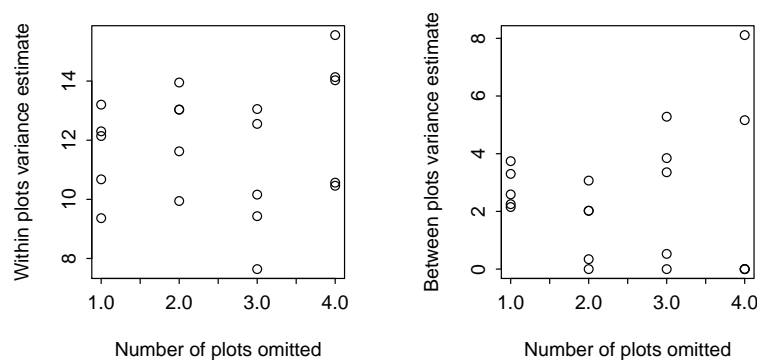


Figure 2: Within, and between plots variance estimates, as functions of the number of whole plots (each consisting of four vines) that were omitted at random.

Omission of a whole plot loses 3 d.f. out of 36 for estimation of within plot effects, and 1 degree of freedom out of 11 for the estimation of between plot effects, i.e., a slightly greater relative loss. The effect on precision will be most obvious where the d.f. are already smallest, i.e., for the between plot variance. The loss of information on complete plots is inherently for serious, for the estimation of the between plot variance, than the loss of partial information (albeit on a greater number of plots) as will often happen in Exercise 1.

Exercise 3

The data set **Gun** (*MEMSS* package) reports on the numbers of rounds fired per minute, by each of nine teams of gunners, each tested twice using each of two methods. In the nine teams, three were made of men with slight build, three with average, and three with heavy build. Is there a detectable difference, in number of rounds fired, between build type or between firing methods? For improving the precision of results, which would be better – to double the number of teams, or to double the number of occasions (from 2 to 4) on which each team tests each method?

It probably does not make much sense to look for overall differences in **Method**; this depends on **Physique**. We therefore nest **Method** within **Physique**.

```
> library(MEMSS)
> Gun.lmer <- lmer(rounds~Physique/Method +(1|Team), data=Gun)
> summary(Gun.lmer)
```

```
Linear mixed model fit by REML
Formula: rounds ~ Physique/Method + (1 | Team)
Data: Gun
    AIC   BIC logLik deviance REMLdev
143.0 155.6 -63.48   133.5   127.0
Random effects:
    Groups      Name      Variance Std.Dev.
Team      (Intercept) 1.0909   1.0444
Residual                2.1797   1.4764
Number of obs: 36, groups: Team, 9
```

```
Fixed effects:
              Estimate Std. Error t value
(Intercept)      23.5889     0.4922   47.92
Physique.L        -0.9664     0.8525   -1.13
Physique.Q         0.1905     0.8525    0.22
PhysiqueSlight:MethodM2 -8.4500     0.8524   -9.91
PhysiqueAverage:MethodM2 -8.1000     0.8524   -9.50
PhysiqueHeavy:MethodM2  -8.9833     0.8524  -10.54
```

```
Correlation of Fixed Effects:
              (Intr) Phys.L Phys.Q PS:MM2 PA:MM2
Physique.L    0.000
Physique.Q    0.000  0.000
PhysqSl:MM2   -0.289  0.353 -0.204
PhysqAv:MM2   -0.289  0.000  0.408  0.000
PhysqHv:MM2   -0.289 -0.353 -0.204  0.000  0.000
```

A good way to proceed is to determine the fitted values, and present these in an interaction plot:

```
> Gun.hat <- fitted(Gun.lmer)
> interaction.plot(Gun$Physique, Gun$Method, Gun.hat)
```

Differences between methods, for each of the three physiques, are strongly attested. These can be estimated within teams, allowing 24 degrees of freedom for each of these comparisons.

Clear patterns of change with **Physique** seem apparent in the plot. There are however too few degrees of freedom for this effect to appear statistically significant. Note however that the parameters that are given are for the lowest level of **Method**, i.e., for M1. Making M2 the baseline shows the effect as closer to the conventional 5% significance level.

The component of variance at the between teams level is of the same order of magnitude as the within teams component. Its contribution to the variance of team means (1.044^2) is much greater than the contribution of the within team component ($1.476^2/4$; there are 4 results per team). If comparison between physiques is the concern; it will be much more effective to double the number of teams; compare $(1.044^2 + 1.476^2/4)/2$ ($=0.82$) with $1.044^2 + 1.476^2/8$ ($=1.36$).

Exercise 4

*The data set `ergoStool` (*MEMSS* package) has data on the amount of effort needed to get up from a stool, for each of nine individuals who each tried four different types of stool. Analyse the data both using `aov()` and using `lme()`, and reconcile the two sets of output. Was there any clear winner among the types of stool, if the aim is to keep effort to a minimum?

For analysis of variance, specify

```
> aov(effort~Type+Error(Subject), data=ergoStool)
```

Call:

```
aov(formula = effort ~ Type + Error(Subject), data = ergoStool)
```

Grand Mean: 10.25

Stratum 1: Subject

Terms:

	Residuals
Sum of Squares	66.5
Deg. of Freedom	8

Residual standard error: 2.883141

Stratum 2: Within

Terms:

	Type	Residuals
Sum of Squares	81.19444	29.05556
Deg. of Freedom	3	24

Residual standard error: 1.100295

Estimated effects may be unbalanced

For testing the **Type** effect for statistical significance, refer $(81.19/3)/(29.06/24)$ ($=22.35$) with the $F_{3,24}$ distribution. The effect is highly significant.

This is about as far as it is possible to go with analysis of variance calculations. When `Error()` is specified in the `aov` model, R has no mechanism for extracting estimates. (There are mildly tortuous ways to extract the information, which will not be further discussed here.)

For use of `lmer`, specify

```
> summary(lmer(effort~Type + (1|Subject), data=ergoStool))
```

Linear mixed model fit by REML

Formula: effort ~ Type + (1 | Subject)

Data: ergoStool

AIC	BIC	logLik	deviance	REMLdev
133.1	142.6	-60.57	122.1	121.1

Random effects:

Groups	Name	Variance	Std.Dev.
Subject	(Intercept)	1.7755	1.3325
Residual		1.2106	1.1003

Number of obs: 36, groups: Subject, 9

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	8.5556	0.5760	14.854
TypeT2	3.8889	0.5187	7.498
TypeT3	2.2222	0.5187	4.284
TypeT4	0.6667	0.5187	1.285

Correlation of Fixed Effects:

	(Intr)	TypeT2	TypeT3
TypeT2	-0.450		
TypeT3	-0.450	0.500	
TypeT4	-0.450	0.500	0.500

Observe that 1.100295^2 (Residual StdDev) is very nearly equal to $29.06/24$ obtained from the analysis of variance calculation.

Also the Stratum 1 mean square of $66.5/8$ ($=8.3125$) from the analysis of variance output is very nearly equal to $1.3325^2 + 1.100295^2/4$ ($= 2.078$) from the `lme` output.

Exercise 5*

In the data set `MathAchieve` (`MEMSS` package), the factors `Minority` (levels `yes` and `no`) and `sex`, and the variable `SES` (socio-economic status) are clearly fixed effects. Discuss how the decision whether to treat `School` as a fixed or as a random effect might depend on the purpose of the study? Carry out an analysis that treats `School` as a random effect. Are differences between schools greater than can be explained by within school variation?

School should be treated as a random effect if the intention is to generalize results to other comparable schools. If the intention is to apply them to other pupils or classes within those same schools, it should be taken as a fixed effect.

For the analysis of these data, both `SES` and `MEANSES` should be included in the model. Then the coefficient of `MEANSES` will measure between school effects, while the coefficient of `SES` will measure within school effects.

```
> library(MEMSS)
> MathAch.lmer <- lmer(MathAch ~ Minority*Sex*(MEANSES+SES) + (1|School),
+                       data=MathAchieve)
```

```
> options(width=90)
> MathAch.lmer
```

Linear mixed model fit by REML

Formula: `MathAch ~ Minority * Sex * (MEANSES + SES) + (1 | School)`

Data: `MathAchieve`

AIC BIC logLik deviance REMLdev

46344 46441 -23158 46308 46316

Random effects:

Groups Name Variance Std.Dev.

School (Intercept) 2.5118 1.5849

Residual 35.7895 5.9824

Number of obs: 7185, groups: School, 160

Fixed effects:

	Estimate	Std. Error	t value
(Intercept)	12.7992	0.1791	71.45
MinorityYes	-2.6055	0.2791	-9.33
SexMale	1.2773	0.1862	6.86
MEANSES	2.2365	0.5039	4.44

SES	2.5085	0.1853	13.54
MinorityYes:SexMale	-0.4623	0.3757	-1.23
MinorityYes:MEANSES	1.4385	0.6837	2.10
MinorityYes:SES	-1.1007	0.3188	-3.45
SexMale:MEANSES	0.5741	0.5740	1.00
SexMale:SES	-0.5166	0.2643	-1.95
MinorityYes:SexMale:MEANSES	-0.7132	0.9034	-0.79
MinorityYes:SexMale:SES	0.1103	0.4683	0.24

Correlation of Fixed Effects:

	(Intr)	MnrtyY	SexMal	MEANSE	SES	MnY:SM	MY:MEA	MY:SES	SM:MEA	SM:SES	MY:SM:M
MinorityYes	-0.346										
SexMale	-0.481	0.268									
MEANSES	-0.095	0.066	0.054								
SES	-0.017	0.031	0.007	-0.355							
MnrtyYs:SxM	0.207	-0.671	-0.433	-0.030	-0.010						
MnY:MEANSES	0.091	0.161	-0.043	-0.510	0.271	-0.142					
MnrtyYs:SES	0.008	0.117	-0.012	0.211	-0.584	-0.089	-0.446				
SxM:MEANSES	0.044	-0.035	-0.141	-0.540	0.315	0.092	0.366	-0.181			
SexMale:SES	0.010	-0.017	-0.081	0.252	-0.703	0.045	-0.194	0.409	-0.430		
MY:SM:MEANS	-0.033	-0.140	0.096	0.316	-0.205	0.120	-0.651	0.332	-0.576	0.280	
MnrY:SM:SES	-0.011	-0.076	0.056	-0.140	0.397	0.122	0.300	-0.678	0.241	-0.567	-0.473

```
> options(width=68)
```

The between school component of variance (1.585^2) is 2.51, compared with a within school component that equals 35.79. To get confidence intervals (strictly Bayesian credible intervals) for these variance estimates, specify:

```
> MathAch.mcmc <- mcmcscamp(MathAch.lmer, n=10000)
> HPDinterval(VarCorr(MathAch.mcmc, type="varcov"))
```

```
      lower      upper
[1,]  1.638323  2.966406
[2,] 34.700599 37.072491
attr(,"Probability")
[1] 0.95
```

The 95% confidence interval for the between school component of variance extended, in my calculation, from 1.64 to 3.0. The confidence interval excludes 0.

The number of results for school varies between 14 and 67. Thus, the relative contribution to class means is 5.51 and a number that is at most $5.982429^2/14 = 2.56$.