Models with Random Effects

- Levels are a random sample
- Variability between levels is of interest
- Nested vs. crossed factors

One Random Factor

Serum measurements of blood samples.

Model:

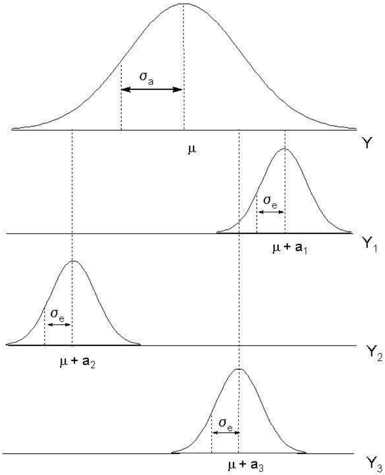
$$Y_{ij} = \mu + a_i + \epsilon_{ij}, \qquad i = 1, \ldots, I; j = 1, \ldots, J$$

a_i random effect of sample i, $a_i \sim \mathcal{N}(0, \sigma_a^2)$,
ϵ_{ij} error of jth measurement of sample i, $\epsilon_i \sim \mathcal{N}(0, \sigma_e^2)$, a_i and ϵ_{ij} are all independent.

$$Var(Y_{ij}) = Var(a_i + \epsilon_{ij}) = \sigma_a^2 + \sigma_e^2, \quad Cov(Y_{ij}, Y_{ij'}) = \sigma_a^2$$

The variance of Y_{ij} consists of two components. Such models are also called variance components models.

Illustration



Anova Table

$$H_0 : \sigma_a^2 = 0 \qquad H_A : \sigma_a^2 > 0$$

Source	SS	df	MS=SS/df
Sample	$SS_a = \sum\sum(y_{i.} - y_{..})^2$	$I - 1$	MS_a
Residual	$SS_{res} = \sum\sum(y_{ij} - y_{i.})^2$	$N - I$	MS_{res}
Total	$SS_{tot} = \sum\sum(y_{ij} - y_{..})^2$	$N - 1$	

$$E(MS_{res}) = \sigma_e^2, \qquad E(MS_a) = J\sigma_a^2 + \sigma_e^2$$

Can use $F = MS_a/MS_{res}$ to test H_0.

Parameter estimations

$$\hat{\sigma}_e^2 = MS_{res}$$

$$\hat{\sigma}_a^2 = (MS_a - MS_{res})/J \quad \text{can be negative!}$$

$$\hat{\mu} = y_{..} \quad \text{with } Var(\hat{\mu}) = \frac{1}{I}(\sigma_a^2 + \sigma_e^2/J)$$

Either Maximum Likelihood estimators or $\hat{\sigma}_a^2 \geq 0$

Variability between Laboratories

$$Y_{ijk} = \mu + a_i + b_j + \epsilon_{ijk}$$

a_i random effect of lab i, $a_i \sim \mathcal{N}(0, \sigma_a^2)$,
b_j random effect of sample j, $b_j \sim \mathcal{N}(0, \sigma_b^2)$,
ϵ_{ijk} measurement error, $\epsilon_{ijk} \sim \mathcal{N}(0, \sigma_e^2)$,
all random variables are independent.

Source	df	E(MS)	F
Lab	$I-1$	$\sigma_e^2 + JK\sigma_a^2$	MS_a/MS_{res}
Sample	$J-1$	$\sigma_e^2 + IK\sigma_b^2$	MS_b/MS_{res}
Residual	$\ll diff \gg$	σ_e^2	
Total	$IJK-1$		

Estimation of Variance Components

$$\hat{\sigma}_e^2 = MS_{res}$$
$$\hat{\sigma}_a^2 = (MS_a - MS_{res})/JK$$
$$\hat{\sigma}_b^2 = (MS_b - MS_{res})/IK$$

Model with Interaction Lab:Sample

Source	E(MS)	H_0	F
Lab	$\sigma_e^2 + JK\sigma_a^2 + K\sigma_{ab}^2$	$\sigma_a^2 = 0$	MS_a/MS_{ab}
Sample	$\sigma_e^2 + IK\sigma_b^2 + K\sigma_{ab}^2$	$\sigma_b^2 = 0$	MS_b/MS_{ab}
Lab : Sample	$\sigma_e^2 + K\sigma_{ab}^2$	$\sigma_{ab}^2 = 0$	MS_{ab}/MS_{res}
Residual	σ_e^2		

$H_0 : \sigma_a^2 = 0$ Test statistic: $F = MS_a/MS_{ab}$

$H_0 : \sigma_a^2 = \sigma_{ab}^2 = 0$ Test statistic: $F = MS_a/MS_{res}$

Crossed factors

Factors A and B are called crossed if every level of B occurs with every level of A. A factorial design involves crossed factors.

	Factor A			
Factor B	1	2	3	4
1	xx	xx	xx	xx
2	xx	xx	xx	xx
3	xx	xx	xx	xx

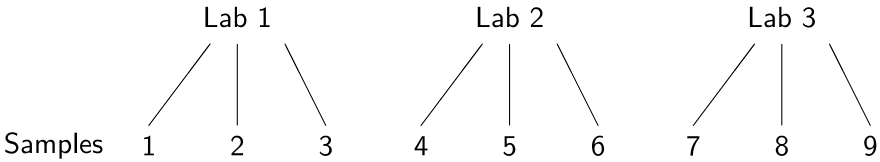
Nested factors

Factors A and B are called nested if there are different levels of B within each level of A. B is nested within A in the following layout.

A		1			2			3			4	
B	1	2	3	4	5	6	7	8	9	10	11	12
	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx	xx

Designs with nested factors are called nested designs or hierarchical designs.

Nested Designs



The factor Sample is nested within Lab.

Model for a two-stage nested design:

$$Y_{ijk} = \mu + a_i + b_{j(i)} + \epsilon_{k(ij)}, \qquad i = 1, .., I; j = 1, .., J; k = 1, .., K$$

The subscript $j(i)$ indicates that the jth level of factor B is nested within the ith level of factor A.

Question: Is an interaction term important?

Anova table

Decomposition of sum of squares:

$$y_{ijk} - y_{...} = y_{i..} - y_{...} \quad + \quad y_{ij.} - y_{i..} \quad + \quad y_{ijk} - y_{ij.}$$

$$SS_{tot} = SS_A + SS_{B(A)} + SS_{res}.$$

Source	df	E(MS)
Lab	$I-1$	$\sigma_e^2 + K\sigma_b^2 + JK\sigma_a^2$
Sample	$I(J-1)$	$\sigma_e^2 + K\sigma_b^2$
Residual	"diff"	σ_e^2
Total	$IJK-1$	

Moisture Content of Cowpea

Effect of milling on moisture content. 3 samples of 100g from 5 batches were milled. From each sample 10g are measured three times.

	sample								
batch		1			2			3	
1	9.3	9.2	8.8	8.6	8.7	9.9	8.9	8.7	8.5
2	8.0	8.2	9.2	9.7	9.4	8.2	9.3	9.5	9.4
3	11.0	10.7	9.9	9.3	13.9	9.2	9.2	10.9	9.7
4	10.1	10.2	9.9	8.6	9.4	8.3	8.3	9.9	9.5
5	12.0	9.3	10.8	12.2	9.6	11.7	11.4	9.8	12.4

Anova Table

```
> mod1=aov(moisture~batch + sample%in%batch)
> summary(mod1)
              Df Sum Sq Mean Sq F value    Pr(>F)
batch          4 30.928  7.7320  7.0390 0.0004027 ***
batch:sample  10  5.911  0.5911  0.5381 0.8491520
Residuals     30 32.953  1.0984
```

$$\begin{aligned}
\hat{\sigma}_e^2 &= 1.0984 && MS_{res} \\
\hat{\sigma}_s^2 &= (0.5911 - 1.0984)/3 = 0 && (MS_s - MS_{res})/K \\
\hat{\sigma}_b^2 &= (7.732 - 1.0984)/9 = 0.737 && (MS_b - MS_s)/JK
\end{aligned}$$

Linear mixed-effects model fit

```
> library(nlme)
> summary(lme(moisture~1,random=~1|batch/sample))
Random effects:
 Formula: ~1 | batch
        (Intercept)
StdDev:   0.8666916
 Formula: ~1 | sample %in% batch
         (Intercept)  Residual
StdDev: 3.783493e-05 0.9857034

Number of Observations: 45
Number of Groups: batch sample %in% batch
                      5                15
```