JOHNS HOPKINS BLOOMBERG SCHOOL of PUBLIC HEALTH

Nested ANOVA: Example

- 3 cages
- 4 mosquitoes within each cage
- 2 independent measurements per mosquito

Cage I				Cage II				Cage III			
1	2	3	4	1	2	3	4	1	2	3	4
58.5	77.8	84.0	70.1	69.8	56.0	50.7	63.8	56.6	77.8	69.9	62.1
59.5	80.9	83.6	68.3	69.8	54.5	49.3	65.8	57.5	79.2	69.2	64.5

Nested ANOVA: models

$$Y_{ijk} = \mu + \alpha_i + \beta_{ij} + \epsilon_{ijk}$$

μ = overall mean

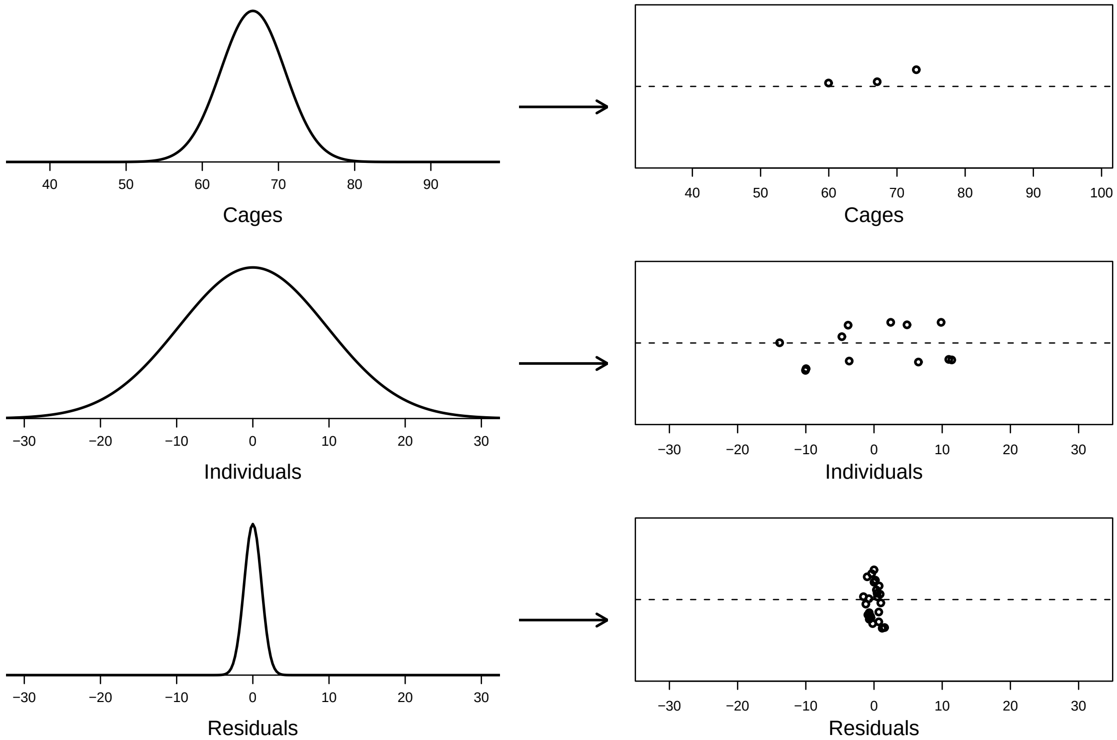
α_i = "effect" for ith cage

β_{ij} = "effect" for jth mosquito within ith cage

ϵ_{ijk} = random error

Random effects model	Mixed effects model
$\alpha_i \sim \text{Normal}(0, \sigma_A^2)$	α_i fixed; $\sum \alpha_i = 0$
$\beta_{ij} \sim \text{Normal}(0, \sigma_{B\|A}^2)$	$\beta_{ij} \sim \text{Normal}(0, \sigma_{B\|A}^2)$
$\epsilon_{ijk} \sim \text{Normal}(0, \sigma^2)$	$\epsilon_{ijk} \sim \text{Normal}(0, \sigma^2)$

The model



Example: sample means

	Cage I				Cage II				Cage III			
	1	2	3	4	1	2	3	4	1	2	3	4
	58.5	77.8	84.0	70.1	69.8	56.0	50.7	63.8	56.6	77.8	69.9	62.1
	59.5	80.9	83.6	68.3	69.8	54.5	49.3	65.8	57.5	79.2	69.2	64.5
$\bar{Y}_{ij.}$	59.00	79.35	83.80	69.20	69.80	55.25	50.00	64.80	57.05	78.50	69.55	63.30
$\bar{Y}_{i..}$	72.84				59.96				67.10			
$\bar{Y}_{...}$					66.63							

Calculations (equal sample sizes)

Source	Sum of squares	df
among groups	$SS_{among} = bn \sum_i (\bar{Y}_{i..} - \bar{Y}_{...})^2$	$a - 1$
subgroups within groups	$SS_{subgr} = n \sum_i \sum_j (\bar{Y}_{ij.} - \bar{Y}_{i..})^2$	$a(b - 1)$
within subgroups	$SS_{within} = \sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{ij.})^2$	$a b (n - 1)$
TOTAL	$\sum_i \sum_j \sum_k (Y_{ijk} - \bar{Y}_{...})^2$	$a b n - 1$

ANOVA table

SS	df	MS	F	expected MS
SS_{among}	$a - 1$	$\frac{SS_{among}}{a - 1}$	$\frac{MS_{among}}{MS_{subgr}}$	$\sigma^2 + n\,\sigma^2_{B\|A} + n\,b\,\sigma^2_A$
SS_{subgr}	$a(b - 1)$	$\frac{SS_{subgr}}{a(b - 1)}$	$\frac{MS_{subgr}}{MS_{within}}$	$\sigma^2 + n\,\sigma^2_{B\|A}$
SS_{within}	$a b (n - 1)$	$\frac{SS_{within}}{ab(n - 1)}$		σ^2
SS_{total}	$a b n - 1$			

Example

source	df	SS	MS	F	P-value
among groups	2	665.68	332.84	1.74	0.23
among subgroups within groups	9	1720.68	191.19	146.88	< 0.001
within subgroups	12	15.62	1.30		
TOTAL	23	2401.97			

```
aov.out <- aov(length ~ cage / individual, data=mosq)
summary(aov.out)
```

Variance components

Within subgroups (error; between measurements on each female)

$s^2 = MS_{within} = 1.30$ $\qquad s = \sqrt{1.30} = 1.14$

Among subgroups within groups (among females within cages)

$$s^2_{B|A} = \frac{MS_{subgr} - MS_{within}}{n} = \frac{191.19 - 1.30}{2} = 94.94 \qquad s_{B|A} = \sqrt{94.94} = 9.74$$

Among groups (among cages)

$$s^2_{A} = \frac{MS_{among} - MS_{subgr}}{nb} = \frac{332.84 - 191.19}{8} = 17.71 \qquad s_{A} = \sqrt{17.71} = 4.21$$

Variance components (2)

$s^2 + s^2_{B|A} + s^2_A$ = 1.30 + 94.94 + 17.71 = 113.95.

$$s^2 \quad \text{represents} \quad \frac{1.30}{113.95} = 1.1\%$$

$$s^2_{B|A} \quad \text{represents} \quad \frac{94.94}{113.95} = 83.3\%$$

$$s^2_A \quad \text{represents} \quad \frac{17.71}{113.95} = 15.6\%$$

Note:

$\text{var}(Y) = \sigma^2 + \sigma^2_{B|A} + \sigma^2_A$

$\text{var}(Y \mid A) = \sigma^2 + \sigma^2_{B|A}$

$\text{var}(Y \mid A, B) = \sigma^2$

Mosquito averages

	I-1	I-2	I-3	I-4	II-1	II-2	II-3	II-4	III-1	III-2	III-3	III-4
	58.5	77.8	84.0	70.1	69.8	56.0	50.7	63.8	56.6	77.8	69.9	62.1
	59.5	80.9	83.6	68.3	69.8	54.5	49.3	65.8	57.5	79.2	69.2	64.5
ave	59.0	79.4	83.8	69.2	69.8	55.2	50.0	64.8	57.0	78.5	69.6	63.3

ANOVA table

source	**df**	**SS**	**MS**	**F**	**P-value**
between	2	332.8	166.4	1.74	0.23
within	9	860.3	95.6		

```
aov.out <- aov(avelen ~ cage, data=mosq2)
summary(aov.out)
```

Ignoring cages

I-1	I-2	I-3	I-4	II-1	II-2	II-3	II-4	III-1	III-2	III-3	III-4
58.5	77.8	84.0	70.1	69.8	56.0	50.7	63.8	56.6	77.8	69.9	62.1
59.5	80.9	83.6	68.3	69.8	54.5	49.3	65.8	57.5	79.2	69.2	64.5

ANOVA table

source	df	SS	MS	F	P-value
between	11	2386.4	216.9	166.7	< 0.001
within	12	15.6	1.3		

```
mosq$ind2 <- factor(paste(mosq$cage,mosq$individual, sep=":"))
aov.out <- aov(length ~ ind2, data=mosq)
summary(aov.out)
```

Ignoring individual mosquitoes

Cage I	Cage II	Cage III
58.5	69.8	56.6
59.5	69.8	57.5
77.8	56.0	77.8
80.9	54.5	79.2
84.0	50.7	69.9
83.6	49.3	69.2
70.1	63.8	62.1
68.3	65.8	64.5

ANOVA table

source	df	SS	MS	F	P-value
between	2	665.7	332.8	4.03	0.033
within	21	1736.3	86.7		

This is wrong!

```
aov.out <- aov(length ~ cage, data=mosq)
summary(aov.out)
```

Example: mixed effects

Strain	Jar	Jar means	Strain means	Strain	Jar	Jar means	Strain means
LDD	1	27.000		LC	1	28.500	
	2	27.750			2	26.875	
	3	26.625	27.125		3	27.000	27.458
OL	1	33.375		RH	1	29.500	
	2	38.125			2	30.375	
	3	31.250	34.250		3	28.250	29.375
NH	1	27.500		NKS	1	30.125	
	2	26.625			2	29.625	
	3	28.500	27.452		3	31.750	30.500
RKS	1	31.750		BS	1	27.875	
	2	31.750			2	25.625	
	3	35.250	32.917		3	27.500	37.000

Results

source	**df**	**SS**	**MS**	**F**	**P-value**
among strains	7	1323.42	189.06	8.47	< 0.001
among jars within strains	16	357.25	22.33	0.80	0.68
within jars	168	4663.25	27.76		

Note: 8 strains; 3 jars per strain; 8 flies per jar

The expected mean squares are

$$\sigma^2 + n\,\sigma^2_{B|A} + n\,b\,\frac{\sum \alpha^2}{a-1}$$

$$\sigma^2 + n\,\sigma^2_{B|A}$$

$$\sigma^2$$

Higher-level nested ANOVA models

You can have as many levels as you like. For example, here is a three-level nested mixed ANOVA model:

$$Y_{ijkl} = \mu + \alpha_i + B_{ij} + C_{ijk} + \epsilon_{ijkl}$$

Assumptions: $B_{ij} \sim N(0,\sigma^2_{B|A})$, $C_{ijk} \sim N(0,\sigma^2_{C|B})$, $\epsilon_{ijkl} \sim N(0,\sigma^2)$.

Calculations

Source	**Sum of squares**	**df**
among groups	$SS_{among} = b\,c\,n \sum_i (\bar{Y}_{i\cdots} - \bar{Y}_{\cdots\cdot})^2$	a – 1
among subgroups	$SS_{subgr} = c\,n \sum_i \sum_j (\bar{Y}_{ij\cdot\cdot} - \bar{Y}_{i\cdots})^2$	a (b – 1)
among subsubgroups	$SS_{subsubgr} = n \sum_i \sum_j \sum_k (\bar{Y}_{ijk\cdot} - \bar{Y}_{ij\cdot\cdot})^2$	a b (c – 1)
within subsubgroups	$SS_{subsubgr} = \sum_i \sum_j \sum_k \sum_l (Y_{ijkl} - \bar{Y}_{ijk\cdot})^2$	a b c (n – 1)

ANOVA table

SS	MS	F	expected MS
SS_{among}	$\frac{bcn \sum_a (\bar{Y}_A - \bar{Y})^2}{a - 1}$	$\frac{MS_{among}}{MS_{subgr}}$	$\sigma^2 + n\sigma^2_{C \subset B} + nc\sigma^2_{B \subset A} + ncb\frac{\sum \alpha^2}{a - 1}$
SS_{subgr}	$\frac{cn \sum_a \sum_b (\bar{Y}_B - \bar{Y}_A)^2}{a(b - 1)}$	$\frac{MS_{subgr}}{MS_{subsubgr}}$	$\sigma^2 + n\sigma^2_{C \subset B} + nc\sigma^2_{B \subset A}$
$SS_{subsubgr}$	$\frac{n \sum_a \sum_b \sum_c (\bar{Y}_C - \bar{Y}_B)^2}{ab(c - 1)}$	$\frac{MS_{subsubgr}}{MS_{within}}$	$\sigma^2 + n\sigma^2_{C \subset B}$
SS_{within}	$\frac{\sum_a \sum_b \sum_c \sum_n (Y - \bar{Y}_C)^2}{abc(n - 1)}$		σ^2

Unequal sample size

It is best to design your experiments such that you have equal sample sizes in each cell. However, once in a while this is not possible.

In the case of unequal sample sizes, the calculations become really painful (though the R function aov() does all of the calculations for you).

Even worse, the F tests for the upper levels in the ANOVA table no longer have a clear null distribution.

We'll ignore the details...seek advice if you are in such a situation.