

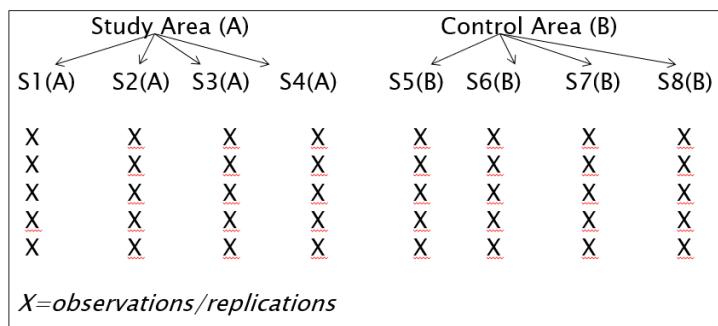
# Experiments with Nested Factors

Norberto E. Milla, Jr.

2025-11-28

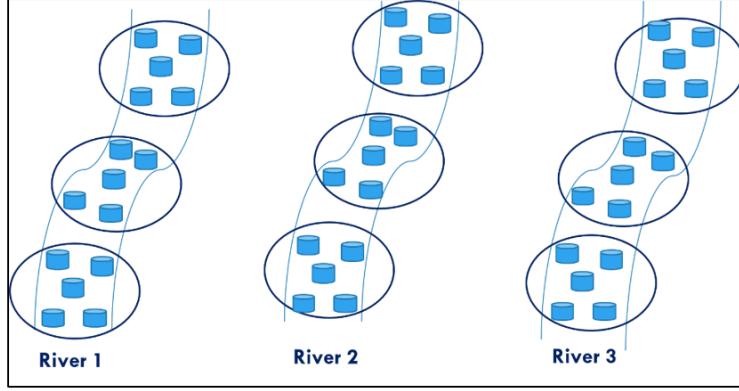
## Nested (hierarchical) factors

- Recall that in a factorial experiment, the levels of factor A are “crossed” with the levels of factor B
- In some two-factor experiments the level of one factor , say B, is not “cross” or “cross classified” with the other factor, say A, but is “NESTED” with it.
- Nesting may occur in the treatment structure or in the design structure
- Nesting in the design structure: two sizes of experimental units with smaller experimental units nested within larger experimental unit (SPLIT-PLOT)
- Nesting in the treatment structure: levels of factor B are not identical for different levels of factor A
- Sampling occurs in two or more stages, say, (1) random selection of 2 study areas in a large forest reserve, (2) random selection of sites within each selected study area
- There is no link from any sites on one area to any sites on another area.



## Example:

- A researcher is interested in determining the biodiversity of river systems in a province
- The first stage of sampling involves selection of three (3) river systems from all rivers in the province
- The second stage involves identification of sampling sites in each of the three (3) selected rivers



## Linear model

$$Y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)}$$

- where:

- $Y_{ijk}$  = observed response for the  $k^{th}$  unit in the  $j^{th}$  level of B within the  $i^{th}$  level of A,  $i = 1, 2, \dots, a$ ;  $j = 1, 2, \dots, b$ ;  $k = 1, 2, \dots, r$
- $\mu$  = general (overall) mean
- $\alpha_i$  = effect of the  $i^{th}$  level of A,
- $\beta_{j(i)}$  = effect of the  $j^{th}$  level of B within the  $i^{th}$  level of A
- $\epsilon_{k(ij)}$  = random error associated with the  $k^{th}$  unit in the  $j^{th}$  level of B within the  $i^{th}$  level of A

## ANOVA table

- there are 3 sources of variation (SoV)
- no INTERACTION effect term

SoV	df	SS	MS	EMS (Fixed)	EMS (Random)
A	$a - 1$	SSA	MSA	$\sigma_\epsilon^2 + br \sum_{i=1}^a \frac{\alpha_i^2}{a-1}$	$\sigma_\epsilon^2 + r\sigma_{\beta(\alpha)}^2 + br\sigma_\alpha^2$
B(A)	$a(b - 1)$	SSB(A)	MSB(A)	$\sigma_\epsilon^2 + r \sum_{j=1}^b \frac{\beta_j^2}{b-1}$	$\sigma_\epsilon^2 + r\sigma_{\beta(\alpha)}^2$
Error	$ab(r - 1)$	SSE	MSE	$\sigma_\epsilon^2$	$\sigma_\epsilon^2$
TOTAL	$n - 1$	SST			

## Sums of squares: Computing formulas

$$\begin{aligned}
 SST &= \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^r Y_{ijk}^2 - \frac{Y_{...}^2}{n}, n = abr \\
 SSA &= \frac{1}{br} \sum_{i=1}^a Y_{i..}^2 - \frac{Y_{...}^2}{n} \\
 SSB(A) &= \frac{1}{r} \sum_{i=1}^a \sum_{j=1}^b Y_{ij.}^2 - \frac{Y_{...}^2}{n} - SSA \\
 SSE &= SST - SSA - SSB(A)
 \end{aligned}$$

## Test of hypothesis (Fixed Model)

### Effect of A

- $H_0 : \alpha_i = 0, \forall i$
- $H_1 : \alpha_i \neq 0, \exists i$
- Test statistic:

### Effect of B

- $H_0 : \beta_{j(i)} = 0, \forall j$
- $H_1 : \beta_{j(i)} \neq 0, \exists j$
- Test statistic:

## Test of hypothesis (Random Model)

### Effect of A

- $H_0 : \sigma_\alpha^2 = 0$
- $H_1 : \sigma_\alpha^2 > 0$
- Test statistic:

### Effect of B

- $H_0 : \sigma_{\beta(\alpha)}^2 = 0$
- $H_1 : \sigma_{\beta(\alpha)}^2 > 0$
- Test statistic:

## An example

A study was conducted to assess the diversity of an orchid species in forest reserves located in three municipalities in Leyte. In each municipality, four (4) sites were randomly identified and three (3) 10m x 10m quadrats were laid out in each selected site. Diversity is recorded per quadrat. The data is shown below.

M <sub>1</sub>				M <sub>2</sub>				M <sub>3</sub>				→	Mun.
1	2	3	4	5	6	7	8	9	10	11	12	→	Sites
10	12	8	13	11	13	9	10	13	14	7	10		
14	8	10	12	14	11	10	9	10	13	9	7		
9	10	12	11	8	9	8	8	16	12	5	4		

Quad.

## Nested factors: an example

```

#Import/Load the data
biod <- read.csv("biodiversity.csv")

#Converts integer codes into nominal (factor) codes
biod$mun <- as.factor(biod$Mun)
biod$site <- as.factor(biod$Site)

#Runs nested ANOVA and store the results in m1
m1 <- aov(Diversity ~ mun/site, data= biod)

#Displays the ANOVA table
knitr::kable(anova(m1))

```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
mun	2	4.50	2.25	0.500000	0.6127098
mun:site	9	128.25	14.25	3.166667	0.0115577
Residuals	24	108.00	4.50	NA	NA

- TASK: Verify the sums of squares in the ANOVA table
- Needs to correct the *F ratio* and *p-value* for *mun*

–  $F = \frac{2.25}{14.25} = 0.16$

– p-value = 0.8545

- INTERPRETATIONS:

- There are no significant differences in biodiversity of the orchid species among the three municipalities.
- There are significant differences in biodiversity of the orchid species among sites within municipality(ies).

## Post hoc analysis: sites within municipality(ies)

- To see which sites differ we do post hoc analysis using the code chunk below

```
emmeans::emmeans(m1, pairwise ~ site | mun, adjust = "bonferroni")
```

```

## $emmeans
## mun = 1:
##   site emmean   SE df lower.CL upper.CL
##   1      11 1.22 24     8.47   13.53
##   2      10 1.22 24     7.47   12.53
##   3      10 1.22 24     7.47   12.53
##   4      12 1.22 24     9.47   14.53
##
## mun = 2:
##   site emmean   SE df lower.CL upper.CL
##   1      11 1.22 24     8.47   13.53
##   2      11 1.22 24     8.47   13.53

```

```

##   3      9 1.22 24     6.47    11.53
##   4      9 1.22 24     6.47    11.53
##
## mun = 3:
##   site emmean   SE df lower.CL upper.CL
##   1      13 1.22 24     10.47    15.53
##   2      13 1.22 24     10.47    15.53
##   3      7 1.22 24      4.47     9.53
##   4      7 1.22 24      4.47     9.53
##
## Confidence level used: 0.95
##
## $contrasts
## mun = 1:
##   contrast   estimate   SE df t.ratio p.value
##   site1 - site2      1 1.73 24     0.577  1.0000
##   site1 - site3      1 1.73 24     0.577  1.0000
##   site1 - site4     -1 1.73 24    -0.577  1.0000
##   site2 - site3      0 1.73 24     0.000  1.0000
##   site2 - site4     -2 1.73 24    -1.155  1.0000
##   site3 - site4     -2 1.73 24    -1.155  1.0000
##
## mun = 2:
##   contrast   estimate   SE df t.ratio p.value
##   site1 - site2      0 1.73 24     0.000  1.0000
##   site1 - site3      2 1.73 24     1.155  1.0000
##   site1 - site4      2 1.73 24     1.155  1.0000
##   site2 - site3      2 1.73 24     1.155  1.0000
##   site2 - site4      2 1.73 24     1.155  1.0000
##   site3 - site4      0 1.73 24     0.000  1.0000
##
## mun = 3:
##   contrast   estimate   SE df t.ratio p.value
##   site1 - site2      0 1.73 24     0.000  1.0000
##   site1 - site3      6 1.73 24     3.464  0.0121
##   site1 - site4      6 1.73 24     3.464  0.0121
##   site2 - site3      6 1.73 24     3.464  0.0121
##   site2 - site4      6 1.73 24     3.464  0.0121
##   site3 - site4      0 1.73 24     0.000  1.0000
##
## P value adjustment: bonferroni method for 6 tests

```

#### Summary table of means with letter designations

Municipality	Site	Mean
1	1	11
	2	10
	3	10
	4	12
2	1	11
	2	11
	3	9

Municipality	Site	Mean
3	4	9
	1	13 <sup>a</sup>
	2	13 <sup>a</sup>
	3	7 <sup>b</sup>
	4	7 <sup>b</sup>