

# Statistics for the Sciences

## Linear Models and Nested Designs

Xuemao Zhang  
East Stroudsburg University

January 18, 2025

# Outline

- Linear Models
- Linear Mixed Models
- Nested Designs
- Example
- Lab

# Linear Models

- A **linear model** or **general linear model** relates a dependent variable and one or more independent variables such that the relationship is modeled as a **linear combination** of the independent variables which are numerical or factors (categorical).
- Linear models get their name because they are **linear in the model parameters**. The general form of a linear model is given by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where  $\mathbf{Y}$  is the vector of dimension  $n \times 1$  observed responses,  $\mathbf{X}$  is the design matrix of  $n \times (p+1)$  fixed constants,  $\boldsymbol{\beta}$  is the vector of  $(p+1) \times 1$  parameters to be estimated (unknown), and  $\boldsymbol{\varepsilon}$  is the vector of  $n \times 1$  random errors.

- ▶  $p+1$  is used for including the intercept
- ▶ So MLR models are linear models

# Linear Models

- Is the model for one-factor experimental design a linear model?
- For example, consider a fixed-effect model

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, j = 1, 2, \dots, n_i, i = 1, 2, \dots, k$$

with  $\sum_i \tau_i = 0$

- For simplicity let  $n_i = 4, i = 1, 2, \dots, k$  and  $k = 3$

$$Y_{ij} = \mu + \tau_i + \varepsilon_{ij}, j = 1, 2, 3, 4, i = 1, 2, 3$$

with  $\tau_1 + \tau_2 + \tau_3 = 0$

# Linear Models

$$Y_{11} = \mu + \tau_1 + \varepsilon_{11}$$

$$Y_{12} = \mu + \tau_1 + \varepsilon_{12}$$

$$Y_{13} = \mu + \tau_1 + \varepsilon_{13}$$

$$Y_{14} = \mu + \tau_1 + \varepsilon_{14}$$

$$Y_{21} = \mu + \tau_2 + \varepsilon_{21}$$

$$Y_{22} = \mu + \tau_2 + \varepsilon_{22}$$

$$Y_{23} = \mu + \tau_2 + \varepsilon_{23}$$

$$Y_{24} = \mu + \tau_2 + \varepsilon_{24}$$

$$Y_{31} = \mu + \tau_3 + \varepsilon_{31}$$

$$Y_{32} = \mu + \tau_3 + \varepsilon_{32}$$

$$Y_{33} = \mu + \tau_3 + \varepsilon_{33}$$

$$Y_{34} = \mu + \tau_3 + \varepsilon_{34}$$

# Linear Models

- Due to  $\tau_1 + \tau_2 + \tau_3 = 0$ , we only need three parameters:  $\mu$ ,  $\tau_1$  and  $\tau_2$ .

$$Y_{11} = \mu + \tau_1 + \varepsilon_{11}$$

$$Y_{12} = \mu + \tau_1 + \varepsilon_{12}$$

$$Y_{13} = \mu + \tau_1 + \varepsilon_{13}$$

$$Y_{14} = \mu + \tau_1 + \varepsilon_{14}$$

$$Y_{21} = \mu + \tau_2 + \varepsilon_{21}$$

$$Y_{22} = \mu + \tau_2 + \varepsilon_{22}$$

$$Y_{23} = \mu + \tau_2 + \varepsilon_{23}$$

$$Y_{24} = \mu + \tau_2 + \varepsilon_{24}$$

$$Y_{31} = \mu - \tau_1 - \tau_2 + \varepsilon_{31}$$

$$Y_{32} = \mu - \tau_1 - \tau_2 + \varepsilon_{32}$$

$$Y_{33} = \mu - \tau_1 - \tau_2 + \varepsilon_{33}$$

$$Y_{34} = \mu - \tau_1 - \tau_2 + \varepsilon_{34}$$

# Linear Models

- In matrix notation

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \end{bmatrix}$$

# Linear Models

- Is the model for two-factor block design a linear model?
- Consider the statistical model (both effects are fixed):

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, j = 1, 2, \dots, b, i = 1, 2, \dots, a,$$

with  $\sum_i \tau_i = 0$  and  $\sum_j \beta_j = 0$

- For simplicity, consider  $a = 3$  and  $b = 4$ :

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, j = 1, 2, 3, 4, i = 1, 2, 3,$$

with  $\tau_1 + \tau_2 + \tau_3 = 0$  and  $\beta_1 + \beta_2 + \beta_3 + \beta_4 = 0$ .



# Linear Models

- Again, there are only six parameters:  $\mu, \tau_1, \tau_2, \beta_1, \beta_2$  and  $\beta_3$ .

$$Y_{11} = \mu + \tau_1 + \beta_1 + \varepsilon_{11}$$

$$Y_{12} = \mu + \tau_1 + \beta_2 + \varepsilon_{12}$$

$$Y_{13} = \mu + \tau_1 + \beta_3 + \varepsilon_{13}$$

$$Y_{14} = \mu + \tau_1 - (\beta_1 + \beta_2 + \beta_3) + \varepsilon_{14}$$

$$Y_{21} = \mu + \tau_2 + \beta_1 + \varepsilon_{21}$$

$$Y_{22} = \mu + \tau_2 + \beta_2 + \varepsilon_{22}$$

$$Y_{23} = \mu + \tau_2 + \beta_3 + \varepsilon_{23}$$

$$Y_{24} = \mu + \tau_2 - (\beta_1 + \beta_2 + \beta_3) + \varepsilon_{24}$$

$$Y_{31} = \mu - (\tau_1 + \tau_2) + \beta_1 + \varepsilon_{31}$$

$$Y_{32} = \mu - (\tau_1 + \tau_2) + \beta_2 + \varepsilon_{32}$$

$$Y_{33} = \mu - (\tau_1 + \tau_2) + \beta_3 + \varepsilon_{33}$$

$$Y_{34} = \mu - (\tau_1 + \tau_2) - (\beta_1 + \beta_2 + \beta_3) + \varepsilon_{34}$$

# Linear Models

- In matrix notation,

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & -1 & -1 & -1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & -1 & -1 & -1 \\ 1 & -1 & -1 & 1 & 0 & 0 \\ 1 & -1 & -1 & 0 & 1 & 0 \\ 1 & -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \end{bmatrix}$$

# Linear Mixed Models

- **Linear Mixed Models** are models which are linear in their parameters and incorporate both fixed and random effects.
  - ▶ Response is numerical
  - ▶ Fixed effects can be numerical or categorical factors
  - ▶ Random effects are categorical factors with factor levels drawn from a larger population of factor levels.
- The matrix notation for a mixed model is highly similar to that for a fixed effects(systematic) model.
- In matrix notation,a linear mixed model can be represented as

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}$$

- ▶ Systematic part  $\mathbf{X}\boldsymbol{\beta}$ :  $\boldsymbol{\beta}$  are the parameters for the fixed effects
- ▶ Random part  $\mathbf{Z}\mathbf{b}$ :  $\mathbf{b}$  is the random effects and  $\mathbf{Z}$  is the design matrix for the random effects
- ▶  $\boldsymbol{\varepsilon}$  is the experimental random error vector
- ▶ In general,  $\mathbf{b}$  and  $\boldsymbol{\varepsilon}$  are assumed to be independent normal vectors.

# Linear Mixed Models

- Consider the mixed model for the randomized block design (treatment effects fixed, block effects random). For simplicity, consider  $a = 3$  and  $b = 4$  again.

$$Y_{ij} = \mu + \tau_i + \beta_j + \varepsilon_{ij}, j = 1, 2, 3, 4, i = 1, 2, 3,$$

with  $\tau_1 + \tau_2 + \tau_3 = 0$  and  $\beta_j \sim N(0, \sigma_b^2), j = 1, 2, 3, 4$ .

- In matrix notation

$$\begin{bmatrix} Y_{11} \\ Y_{12} \\ Y_{13} \\ Y_{14} \\ Y_{21} \\ Y_{22} \\ Y_{23} \\ Y_{24} \\ Y_{31} \\ Y_{32} \\ Y_{33} \\ Y_{34} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \\ 1 & -1 & -1 \end{bmatrix} \begin{bmatrix} \mu \\ \tau_1 \\ \tau_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{13} \\ \varepsilon_{14} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{23} \\ \varepsilon_{24} \\ \varepsilon_{31} \\ \varepsilon_{32} \\ \varepsilon_{33} \\ \varepsilon_{34} \end{bmatrix}$$

# Linear Models

- Estimation Methods
  - ▶ Method of Moments
  - ▶ Maximum likelihood(ML)
  - ▶ Restricted maximum likelihood(REML)
- Maximum likelihood (ML) is obtained from the likelihood functions for the distribution of  $\mathbf{Y}|\mathbf{b}$  and  $\mathbf{b}$
- Restricted maximum likelihood(REML): REML is based on the maximum likelihood method but, instead of maximizing the likelihood function of the original data, it maximizes the likelihood function of the residuals obtained by removing the fixed effects from the original response variables, which are assumed to be known.

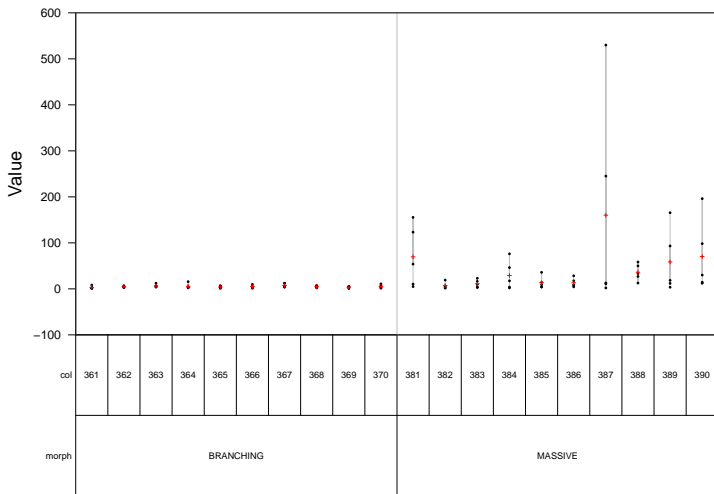
# Nested Designs

- In certain multifactor experiments, the levels of one factor (e.g., factor B) are similar but not identical for different levels of another factor (e.g., A). Such an arrangement is called a **nested, or hierarchical**, design, with the levels of factor B nested under the levels of factor A.
- Example (`lesions.csv`): Lozada-Misa et al. (2015) compared two different morphology types of the coral genus *Porites* for lesions caused by white syndrome disease. They collected ten random colonies of the branching *P. cylindrica* and ten colonies of the massive *Porites* spp., and five random lesions per colony were photographed and measured.
  - ▶ Morphology type was a fixed factor,
  - ▶ with colony a random factor nested within morphology type, and
  - ▶ size of individual lesions as the observations (responses).

```
##           morph col    size
## 1 BRANCHING 361 1.8912
## 2 BRANCHING 361 8.2259
## 3 BRANCHING 361 3.0047
## 4 BRANCHING 361 2.0495
## 5 BRANCHING 361 0.5167
## 6 BRANCHING 362 6.8058
## 7 BRANCHING 362 5.6315
## 8 BRANCHING 362 2.7840
```

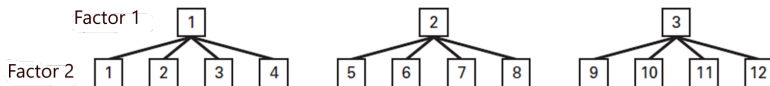
# Nested Designs

- Colony nested within morphology type



# Nested Designs

- This is a **two-stage nested design**, with Colony nested under morphology types.



- In the example, the ten random colonies under BRANCHING have no connection with the ten random colonies under MASSIVE.
- The number of stages can be more than two.



# Nested Designs

- There could be three possible models
  - ▶ Random effects model
  - ▶ Fixed effects model
  - ▶ Mixed Effects Models
    - ★ Treatment effects fixed and nested factor random
- Random effects model:

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}, k = 1, 2, \dots, n, j = 1, 2, \dots, b, i = 1, 2, \dots, a,$$

- ▶ where the  $j(i)$  indicates that the factor corresponding to  $j$  (factor B) is nested in the factor corresponding to  $i$  (factor A),
- ▶ The random errors  $\varepsilon_{(ij)k} \text{ iid } \sim N(0, \sigma^2)$  are nested in factor A-B combinations.
- ▶  $\tau_i \text{ iid } \sim N(0, \sigma_\tau^2)$
- ▶  $\beta_{j(i)} \text{ iid } \sim N(0, \sigma_\beta^2)$

# Nested Designs

- ANOVA Table for the Random effects model

Analysis of Variance Table for the Two-Stage Nested Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
<i>A</i>	$bn \sum (\bar{y}_{i..} - \bar{y}_{...})^2$	$a - 1$	$MS_A$
<i>B</i> within <i>A</i>	$n \sum \sum (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$a(b - 1)$	$MS_{B(A)}$
Error	$\sum \sum \sum (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n - 1)$	$MS_E$
Total	$\sum \sum \sum (y_{ijk} - \bar{y}_{...})^2$	$abn - 1$	

- $E(MS_A) = \sigma^2 + bn\sigma_\tau^2 + n\sigma_\beta^2$
- $E(MS_{B(A)}) = \sigma^2 + n\sigma_\beta^2$
- $E(MS_E) = \sigma^2$

# Nested Designs

- Fixed effects model:

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}, k = 1, 2, \dots, n, j = 1, 2, \dots, b, i = 1, 2, \dots, a,$$

- ▶ where the  $j(i)$  indicates that the factor corresponding to  $j$  (factor B) is nested in the factor corresponding to  $i$  (factor A),
- ▶ The random errors  $\varepsilon_{(ij)k} \text{ iid } \sim N(0, \sigma^2)$  are nested in factor A-B combinations.
- ▶  $\sum_i \tau_i = 0$
- ▶  $\sum_j \beta_{j(i)} = 0$  for all  $i = 1, 2, \dots, a$

# Nested Designs

- ANOVA Table for the Fixed effects model

Analysis of Variance Table for the Two-Stage Nested Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
<i>A</i>	$bn \sum (\bar{y}_{i..} - \bar{y}_{...})^2$	$a - 1$	$MS_A$
<i>B</i> within <i>A</i>	$n \sum \sum (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$a(b - 1)$	$MS_{B(A)}$
Error	$\sum \sum \sum (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n - 1)$	$MS_E$
Total	$\sum \sum \sum (y_{ijk} - \bar{y}_{...})^2$	$abn - 1$	

- $E(MS_A) = \sigma^2 + bn \frac{\sum_i \tau_i^2}{a - 1}$
- $E(MS_{B(A)}) = \sigma^2 + n \frac{\sum_i \sum_j \beta_{j(i)}^2}{a(b - 1)}$
- $E(MS_E) = \sigma^2$

# Nested Designs

- In practice we often encounter models which contain both random and fixed effects. We call them **mixed effects models**.
- Mixed Effects Model:

$$Y_{ijk} = \mu + \tau_i + \beta_{j(i)} + \varepsilon_{(ij)k}, k = 1, 2, \dots, n, j = 1, 2, \dots, b, i = 1, 2, \dots, a,$$

- ▶ where the  $j(i)$  indicates that the factor corresponding to  $j$  (factor B) is nested in the factor corresponding to  $i$  (factor A),
- ▶ The random errors  $\varepsilon_{(ij)k} \text{ iid } \sim N(0, \sigma^2)$  are nested in factor A-B combinations.
- ▶  $\tau_i$  is fixed:  $\sum_{i=1}^a \tau_i = 0$ .
- ▶  $\beta_{j(i)} \text{ iid } \sim N(0, \sigma_\beta^2)$

# Nested Designs

- ANOVA Table for the Mixed effects model

Analysis of Variance Table for the Two-Stage Nested Design

Source of Variation	Sum of Squares	Degrees of Freedom	Mean Square
<i>A</i>	$bn \sum (\bar{y}_{i..} - \bar{y}_{...})^2$	$a - 1$	$MS_A$
<i>B</i> within <i>A</i>	$n \sum \sum (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$a(b - 1)$	$MS_{B(A)}$
Error	$\sum \sum \sum (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n - 1)$	$MS_E$
Total	$\sum \sum \sum (y_{ijk} - \bar{y}_{...})^2$	$abn - 1$	

- $E(MS_A) = \sigma^2 + bn \frac{\sum_i \tau_i^2}{a - 1} + n\sigma_\beta^2$
- $E(MS_{B(A)}) = \sigma^2 + n\sigma_\beta^2$
- $E(MS_E) = \sigma^2$

# Nested Designs

- The first important hypothesis testing problem in nested designs is to test

$H_0$  : The levels of B have the same effect on the response within every given level of A.

Under this null hypothesis, the model is reduced to

$$Y_{ijk} = \mu + \tau_i + \varepsilon_{ijk}, k = 1, 2, \dots, n, j = 1, 2, \dots, b, i = 1, 2, \dots, a,$$

- Another important hypothesis testing problem is to test if the  $a$  levels of A (main factor; fixed effects) averaged over the levels of B have a same effect. That is, we check if the main effects are equal.

# Example

- Calculate summary stats by groups

##	col	morph	N	size	sd	se	ci
## 1	361	BRANCHING	5	3.13760	2.979725	1.3325737	3.699818
## 2	362	BRANCHING	5	4.41892	1.768847	0.7910522	2.196313
## 3	363	BRANCHING	5	6.39866	3.451261	1.5434508	4.285306
## 4	364	BRANCHING	5	5.77804	5.594773	2.5020583	6.946828
## 5	365	BRANCHING	5	4.06686	2.398199	1.0725071	2.977757
## 6	366	BRANCHING	5	5.01400	3.148717	1.4081489	3.909648
## 7	367	BRANCHING	5	7.37804	4.299565	1.9228237	5.338614
## 8	368	BRANCHING	5	4.82584	2.134307	0.9544910	2.650092
## 9	369	BRANCHING	5	3.59112	1.931891	0.8639679	2.398759
## 10	370	BRANCHING	5	5.40918	3.791108	1.6954351	4.707282
## 11	381	MASSIVE	5	69.53184	67.443147	30.1614921	83.741727
## 12	382	MASSIVE	5	7.01802	7.054432	3.1548380	8.759234
## 13	383	MASSIVE	5	11.10566	8.543465	3.8207537	10.608113
## 14	384	MASSIVE	5	29.10582	31.698167	14.1758511	39.358472
## 15	385	MASSIVE	5	13.52838	13.171769	5.8905943	16.354912
## 16	386	MASSIVE	5	13.28264	9.758733	4.3642380	12.117067
## 17	387	MASSIVE	5	160.17892	230.708834	103.1761270	286.462853
## 18	388	MASSIVE	5	36.23336	18.281438	8.1757075	22.699403
## 19	389	MASSIVE	5	58.51856	69.802447	31.2166031	86.671185
## 20	390	MASSIVE	5	70.19178	78.626406	35.1627978	97.627578





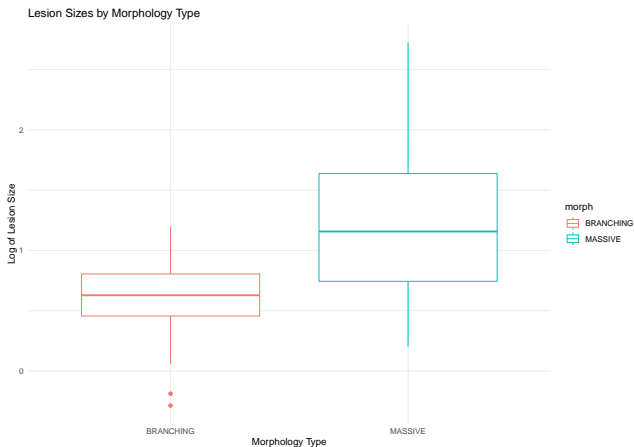
# Nested Designs

- Consider log10 transformation of the response variable size

##		morph	col	size	logsize
## 1	BRANCHING	361	1.8912	0.2767375	
## 2	BRANCHING	361	8.2259	0.9151834	
## 3	BRANCHING	361	3.0047	0.4778011	
## 4	BRANCHING	361	2.0495	0.3116479	
## 5	BRANCHING	361	0.5167	-0.2867615	
## 6	BRANCHING	362	6.8058	0.8328792	
## 7	BRANCHING	362	5.6315	0.7506241	
## 8	BRANCHING	362	2.7840	0.4446692	
## 9	BRANCHING	362	4.0470	0.6071332	
## 10	BRANCHING	362	2.8263	0.4512183	

# Example

- Check the box plots again



# Example

- ANOVA table

```
## # A tibble: 3 x 7
##   stratum term          df sumsq meansq statistic    p.value
##   <chr>   <chr>      <dbl> <dbl>  <dbl>      <dbl>    <dbl>
## 1 col     morph         1  9.82   9.82       32.3  0.0000217
## 2 col     Residuals     18  5.47   0.304      NA      NA
## 3 Within  Residuals     80 17.1   0.214      NA      NA
```

# Example

- REML-estimation with hypothesis testing

```
## Linear mixed model fit by REML. t-tests use Kenward-Roger's method [
## lmerModLmerTest]
## Formula: logsize ~ morph + (1 | col)
## Data: lesions
##
## REML criterion at convergence: 141
##
## Scaled residuals:
##      Min       1Q   Median       3Q      Max
## -2.23033 -0.54021 -0.00755  0.49227  3.04868
##
## Random effects:
##  Groups   Name                Variance Std.Dev.
##  col      (Intercept)  0.01809   0.1345
##  Residual                    0.21368   0.4623
## Number of obs: 100, groups:  col, 20
##
## Fixed effects:
##              Estimate Std. Error      df t value Pr(>|t|)
## (Intercept)  0.60557    0.07799  18.00000   7.765 3.74e-07 ***
## morphMASSIVE 0.62674    0.11029  18.00000   5.683 2.17e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

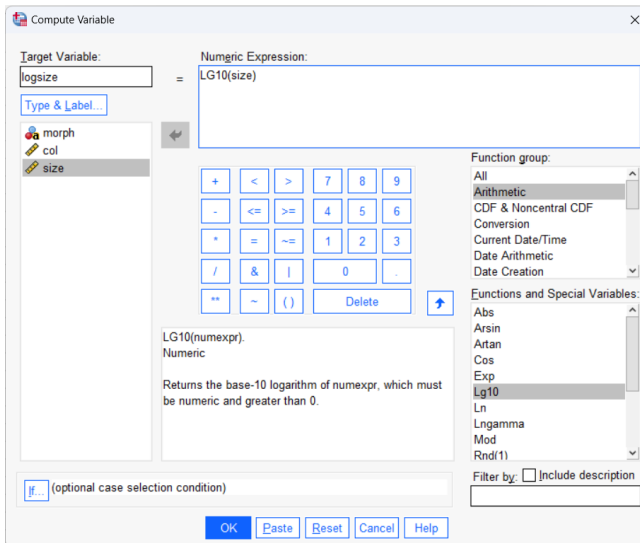
# Example

- Variance component for random-effects

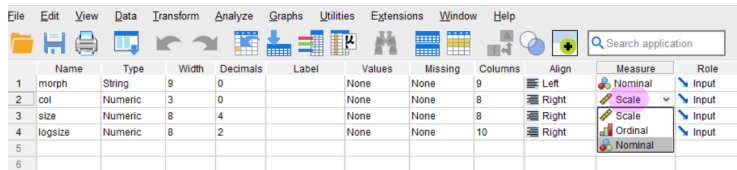
```
##
##
## Result Variance Component Analysis:
## -----
##
##   Name      DF      SS      MS      VC      %Total      SD      CV[%]
## 1 total      4.555803
## 2 morph      1      9.82009  9.82009  0.19032  45.090258  0.436256  47.473953
## 3 morph:col  18      5.473944  0.304108  0.018085  4.284774  0.134482  14.634504
## 4 error      80      17.094469  0.213681  0.213681  50.624968  0.462256  50.303293
##
## Mean: 0.918938 (N = 100)
##
## Experimental Design: balanced | Method: ANOVA
```

# Lab

- After importing the data `lesions.csv`, add a new variable `logsize` which is  $\log_{10}$  transformation of `size`.



- Convert measure of col to Nominal



The screenshot shows the Minitab software interface. The menu bar includes File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Extensions, Window, and Help. Below the menu bar is a toolbar with various icons. The main workspace displays a table with the following data:

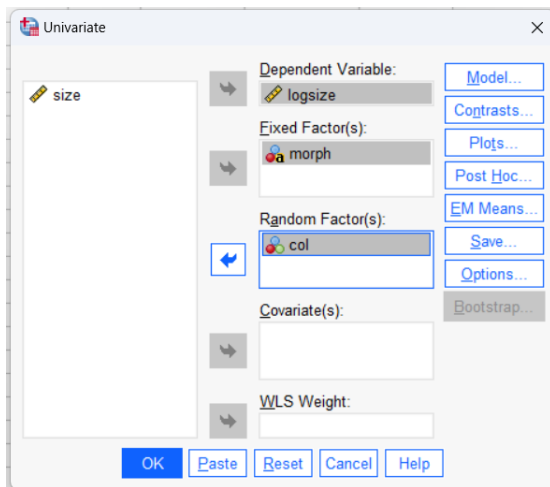
	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	morph	String	9	0		None	None	9	Left	Nominal	Input
2	col	Numeric	3	0		None	None	8	Right	Scale	Input
3	size	Numeric	8	4		None	None	8	Right	Scale	Input
4	logsize	Numeric	8	2		None	None	10	Right	Ordinal	Input
5										Nominal	
6											

The 'Measure' dropdown menu for the 'col' variable is open, showing the following options: Nominal, Scale, Ordinal, and Nominal. The 'Scale' option is currently selected.



# Lab

- Click on Analyze → General Linear Model → Univariate...



**Univariate**

Dependent Variable: **size**

Fixed Factor(s): **morph**

Random Factor(s): **col**

Covariate(s):

WLS Weight:

OK Paste Reset Cancel

**Univariate: Model**

Specify Model

☐ Full factorial ☒ Build terms ☐ Build custom terms

Factors & Covariates:

- morph**
- col**

Model:

- morph**

Build Term(s)

Type:

Main effects ▾

Build Term:

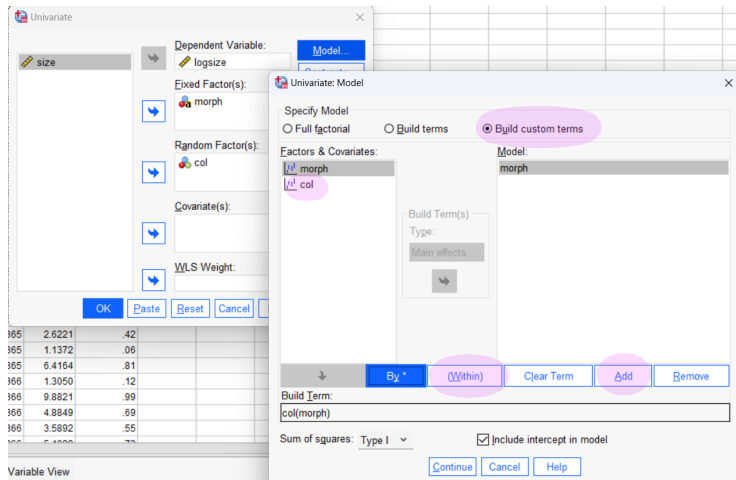
Sum of squares: **Type I**

☒ Include intercept in model

Continue Cancel Help

365	2.6221	.42
365	1.1372	.06
365	6.4164	.81
366	1.3050	.12
366	9.8821	.99
366	4.8849	.69
366	3.5892	.55
366	5.4888	.73

Variable View



The image shows two overlapping SPSS dialog boxes. The background box is the 'Univariate' dialog, and the foreground box is the 'Univariate: Model' sub-dialog.

**Univariate Dialog:**

- Dependent Variable:** size
- Fixed Factor(s):** logsize
- Random Factor(s):** col
- WLS Weight:** (empty)

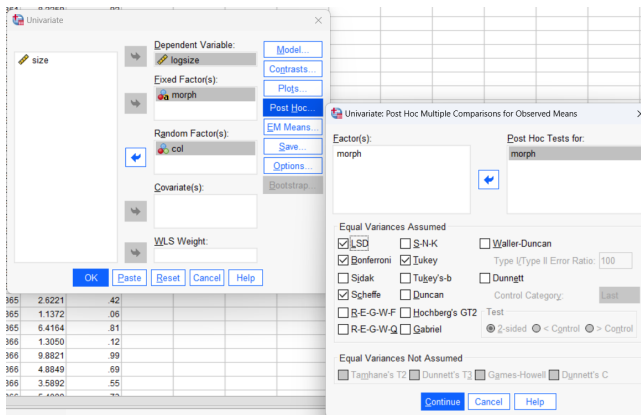
**Univariate: Model Dialog:**

- Specify Model:** ☒ Build custom terms
- Factors & Covariates:** morph, col
- Model:** morph, col(morph)
- Build Term(s):** Type: Main effects
- Sum of squares:** Type I
- ☒ Include intercept in model

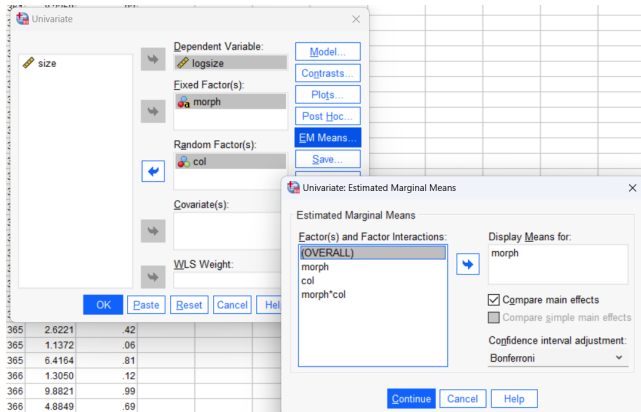
Buttons at the bottom of the 'Univariate: Model' dialog include 'Continue', 'Cancel', and 'Help'.

# Lab

- Set up for multiple comparisons. It is not useful here.
  - ▶ It will be useful if there are more than two levels.



- Compare the means of the fixed effects



- ANOVA table

*Tests of Between-Subjects Effects*

Dependent Variable: logsize

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	84.445	1	84.445	277.680	<.001
	Error	5.474	18	.304 <sup>a</sup>		
morph	Hypothesis	9.820	1	9.820	32.291	<.001
	Error	5.474	18	.304 <sup>a</sup>		
col(morph)	Hypothesis	5.474	18	.304	1.423	.144
	Error	17.094	80	.214 <sup>b</sup>		

a. MS(col(morph))

b. MS(Error)

## Estimated Marginal Means

morph

*Estimates*

Dependent Variable: logsize

morph	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
BRANCHING	.606 <sup>a</sup>	.065	.475	.736
MASSIVE	1.232 <sup>a</sup>	.065	1.102	1.362

a. Based on modified population marginal mean.

*Pairwise Comparisons*

Dependent Variable: logsize

(I) morph	(J) morph	Mean Difference (I-J)	Std. Error	Sig. <sup>d</sup>	95% Confidence Interval for Difference <sup>d</sup>	
					Lower Bound	Upper Bound
BRANCHING	MASSIVE	-.627 <sup>**b,c</sup>	.092	<.001	-.811	-.443
MASSIVE	BRANCHING	.627 <sup>**b,c</sup>	.092	<.001	.443	.811

<sup>d</sup>Based on estimated marginal means



# License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](#).