

Statistics for the Sciences

ANOVA for Complete Block Design

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Complete Block Design
 - ▶ Fixed Effects Model
 - ▶ Mixed Effects Models
 - ▶ Friedman Rank-Sum Test
- Lab

Block Design

- The Blocking Principle
 - ▶ Blocking is a technique for dealing with nuisance factors
 - ▶ A nuisance factor is a factor that probably has some effect on the response, but it's of no interest to the experimenter. . . however, the variability it transmits to the response needs to be minimized
- Typical nuisance factors include batches of raw material, operators, pieces of test equipment, time (shifts, days, etc.), different experimental units

Block Design

Randomized Complete Block Design

- There are two factors: factor of interest and nuisance factor.
- A block is a specific level of the nuisance factor
 - ▶ The design uses blocks such that units in each block are relatively similar or homogeneous, with one unit within each block randomly assigned to each treatment (all runs within a block are randomized).
 - ▶ If the design involves a treatments within each of b blocks, then the total number of observations is ab .
 - ▶ Variability between blocks can be large, variability within a block should be relatively small
- The purpose of blocking is to isolate the block-to-block variability that might hide/affect the effect of the treatments.
- The design must be balanced: the number of units in each block must be equal to the number of treatments.

2-Way ANOVA for Block Design

- Example (Can plants grow on Mars and the Moon `wamelink1.csv`): Wamelink et al. (2014) compared the growth of 14 species of plant grown in soil collected from near the Rhine River with simulant regolith representing the soil conditions on the moon and on Mars.
 - ▶ We will focus on total biomass of *Sedum reflexum* (`totbiomass`) as the response variable.
 - ▶ Twenty spatial blocks were established in a glasshouse and for this species, there were 3 small pots (each with five seeds) in each block, representing the three soils types ($a = 3$, $b=20$).
 - ▶ Both Soil type block are considered fixed-effect factors.

```
##    block  soil  ngerm  nleaves  nflowers  nalive  totbiomass  abovebiomass
## 1      1  Moon      3        2          0        2          2.3          2.3
## 2      1 Earth      5        4          0        4          8.3          4.5
## 3      1  Mars      4        4          0        4         106.7         90.7
## 4      2  Moon      4        4          0        4          3.0          2.4
## 5      2  Mars      5        5          0        5          72.5         65.4
##    belowbiomass
## 1              0.1
## 2              3.8
## 3             16.0
## 4              0.6
```

2-Way ANOVA for Block Design

- Suppose that there are a treatments (factor levels) and b blocks
- Let y_{ij} be the response for the i -th treatment applied to the j -th block.
 $i = 1, 2, \dots, a, j = 1, 2, \dots, b$
- The data will be look like this

Table 1: Data from Block Design

Blocks	Treatments				
	1	2	\dots	a	
1	y_{11}	y_{21}	\dots	y_{a1}	$\bar{y}_{\cdot 1}$
2	y_{12}	y_{22}	\dots	y_{a2}	$\bar{y}_{\cdot 2}$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
b	y_{1b}	y_{2b}	\dots	y_{ab}	$\bar{y}_{\cdot b}$
	$\bar{y}_{1\cdot}$	$\bar{y}_{2\cdot}$	\dots	$\bar{y}_{a\cdot}$	

2-Way ANOVA for Block Design

- Statistical model (both effects are fixed):

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

where $\varepsilon_{ij} \sim N(0, \sigma^2)$ are independent normal random errors with common variance σ^2 and

- Side conditions $\sum_i \tau_i = 0$ and $\sum_j \beta_j = 0$.
- Model fit and factor effects (if both are fixed-effect)
 - ▶ Overall mean: $\hat{\mu} = \bar{y}_{..}$
 - ▶ Treatment: $\hat{\tau}_i = \bar{y}_{i.} - \bar{y}_{..}$, $i = 1, \dots, a$
 - ▶ Block: $\hat{\beta}_j = \bar{y}_{.j} - \bar{y}_{..}$, $j = 1, \dots, b$
- The relevant null hypothesis for comparing the a treatment means

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_a$$

2-Way ANOVA for Block Design

- Partition of the total variation:

$$SS_{total} = SS_T + SS_B + SS_E$$

That is,

$$\begin{aligned} \sum \sum (y_{ij} - \bar{y}_{..})^2 &= \sum \sum (\bar{y}_{i.} - \bar{y}_{..})^2 + \sum \sum (\bar{y}_{.j} - \bar{y}_{..})^2 \\ &\quad + \sum \sum (y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..})^2, \end{aligned}$$

where $y_{i.} = \sum_{j=1}^b y_{ij}, i = 1, \dots, a; y_{.j} = \sum_{i=1}^a y_{ij}, j = 1, \dots, b$.

- The corresponding degrees of freedom for the sums of squares are

$$ab - 1 = (a - 1) + (b - 1) + (a - 1)(b - 1).$$

- Therefore, ratios of sums of squares to their degrees of freedom result in mean squares and the ratio of the mean square for treatments to the error mean square is an F statistic that can be used to test the hypothesis of equal treatment means.

2-Way ANOVA for Block Design

- 2-way ANOVA Table

Table 2: 2-way ANOVA Table for Block Design

Source	df	SS	MS	F
Treatments	$a - 1$	SS_T	$SS_T / (a - 1)$	MS_T / MSE
Blocks	$b - 1$	SS_B	$SS_B / (b - 1)$	MS_B / MSE
Error	$(a - 1)(b - 1)$	SSE	$SSE / (a - 1)(b - 1)$	
Total	$ab - 1$	SS_{total}		

- Larger values of the F-test statistic (resulting in smaller p-values), so the ANOVA test is right-tailed.
- If the b block means are not significantly different, the b blocks are homogeneous and the block design is not necessary for future same study.

2-Way ANOVA for Block Design

- Example

```
## Analysis of Variance Table
##
## Response: totbiomass
##           Df Sum Sq Mean Sq F value    Pr(>F)
## soil        2  52443 26221.7  75.2483 6.097e-14 ***
## block       19   3776   198.8   0.5704  0.9043
## Residuals   38  13242   348.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

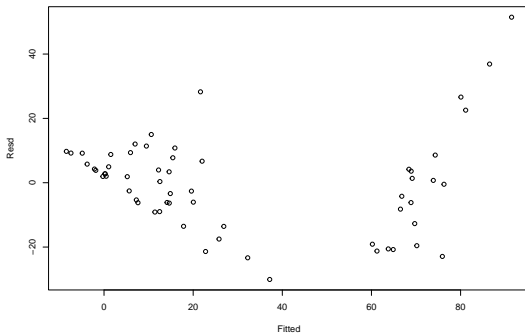
2-Way ANOVA for Block Design

- Example

```
##  
## Call:  
## lm(formula = totbiomass ~ soil + block, data = wamelink)  
##  
## Residuals:  
##      Min      1Q  Median      3Q      Max   
## -30.073  -8.403   1.620   7.967  51.457   
##  
## Coefficients:  
##              Estimate Std. Error t value Pr(>|t|)      
## (Intercept)   25.807     11.304   2.283  0.0281 *      
## soilMars      54.270      5.903   9.193 3.35e-11 ***   
## soilMoon     -14.390      5.903  -2.438  0.0196 *      
## block2       -11.167     15.242  -0.733  0.4683        
## block3        -9.900     15.242  -0.650  0.5199        
## block4       -10.933     15.242  -0.717  0.4776        
## block5       -13.567     15.242  -0.890  0.3790        
## block6         6.433     15.242   0.422  0.6753      
```

2-Way ANOVA for Block Design

- Model adequacy check: We check the fitted values and residuals to see if the three model assumptions are satisfied. These are to be done in a similar way as those in one-way ANOVA.



2-Way ANOVA for Block Design

- Now we use Levene test to further check the constant variance assumptions. That is, we are testing

$$H_0 : \sigma_{Earth} = \sigma_{Mars} = \sigma_{Moon}$$

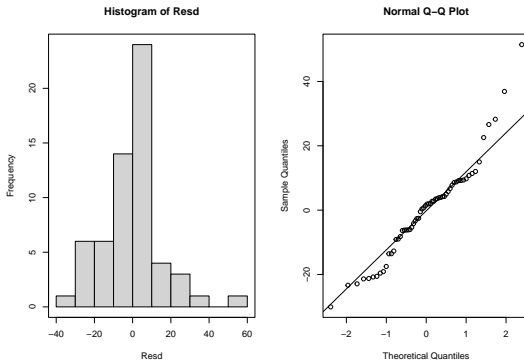
$$H_0 : \sigma_1 = \sigma_2 = \cdots = \sigma_{20}$$

```
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value Pr(>F)
## group  2   4.5259 0.01499 *
##           57
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value Pr(>F)
## group 19   0.5919   0.89
##           40
```

2-Way ANOVA for Block Design

- Normality assumption can be checked using histogram and QQ plot of the residuals.



2-Way ANOVA for Block Design

- Normality test using shapiro test:

```
##  
## Shapiro-Wilk normality test  
##  
## data:  Resd  
## W = 0.95006, p-value = 0.01568
```

Mixed Effects Models

- In practice we often encounter models which contain both random and fixed effects. We call them **mixed effects models**.
- In the Wamelink example,
 - ▶ Soil type is fixed-effect factor
 - ▶ The block is better treated as random since each block has its own (random) deviation.

Mixed Effects Models

- Consider this Mixed Effects Model:

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

- ▶ where $\varepsilon_{ij} \sim N(0, \sigma^2)$ are independent normal random errors with common variance σ^2 .
- ▶ τ_i (for machine) is fixed: $\sum_{i=1}^a \tau_i = 0$.
- ▶ $\beta_j \text{ iid } \sim N(0, \sigma_\beta^2)$ (for block)

Mixed Effects Models

```
##
##
## ANOVA-Type Estimation of Mixed Model:
## -----
##
## [Fixed Effects]
##
##      int soilEarth  soilMars  soilMoon
##      3.25      14.39      68.66      0.00
##
##
## [Variance Components]
##
##      Name  DF  SS           MS           VC           %Total  SD           CV
## 1 total  57
## 2 block  19 3776.286667  198.75193  0*           0*           0*           0*
## 3 error  38 13241.809333  348.468667  348.468667  100          18.667315  60
##
## Mean: 30.93333 (N = 60)
##
```

Mixed Effects Models

- ANOVA table (using the fixed-effects model)

```
##              Df Sum Sq Mean Sq F value    Pr(>F)
## soil           2  52443    26222   75.25 6.1e-14 ***
## block         19   3776     199    0.57  0.904
## Residuals     38  13242     348
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

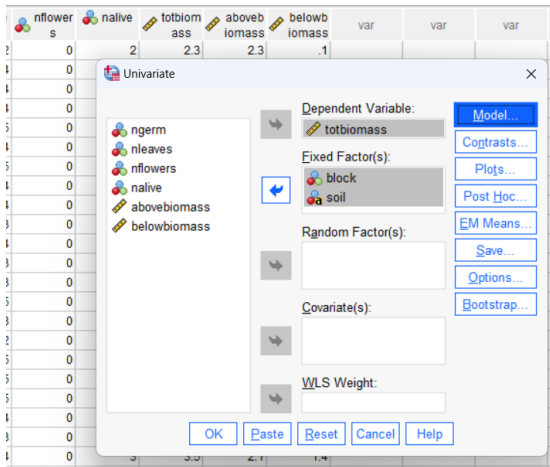
Friedman Rank-Sum Test

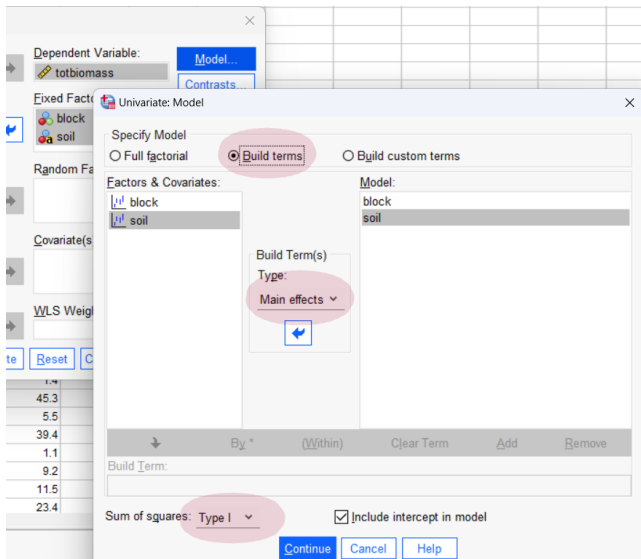
- The Friedman rank-sum test is a nonparametric procedure that can be used to compare more than two population medians in a randomized complete block design. The procedure involves ranking each row(block) together, then considering the values of ranks by treatments(columns).

```
##  
## Friedman rank sum test  
##  
## data:  totbiomass and soil and block  
## Friedman chi-squared = 40, df = 2, p-value = 2.061e-09
```

Lab

- For data `wamalink1.csv`, suppose both Soil type block are fixed-effect factors.
- Click on Analyze → General Linear Model → Univariate.





Tests of Between-Subjects Effects

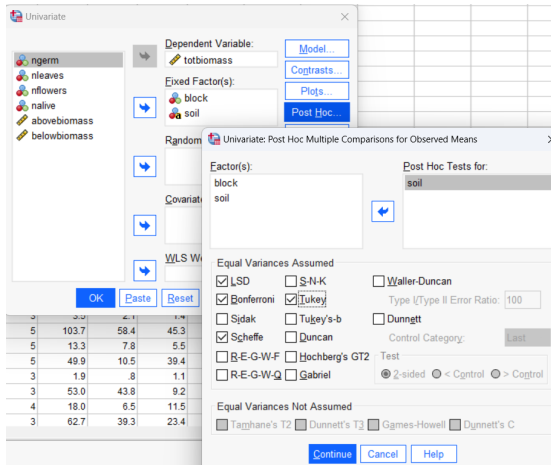
Dependent Variable: totbiomass

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	56219.624 ^a	21	2677.125	7.683	<.001
Intercept	57412.267	1	57412.267	164.756	<.001
block	3776.287	19	198.752	.570	.904
soil	52443.337	2	26221.669	75.248	<.001
Error	13241.809	38	348.469		
Total	126873.700	60			
Corrected Total	69461.433	59			

a. R Squared = .809 (Adjusted R Squared = .704)

Lab

- The p-value for testing homogeneity of block effects is large.
- The p-value for testing difference among three treatment effects is small.
 - ▶ Then we can conduct multiple comparisons to check which is (are) different from the others.



- Next, we consider the mixed-effect model that block is random. We'll see the anova table

Tests of Between-Subjects Effects

Dependent Variable: totbiomass

Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Intercept	Hypothesis	57412.267	1	57412.267	288.864	<.001
	Error	3776.287	19	198.752 ^a		
soil	Hypothesis	52443.337	2	26221.669	75.248	<.001
	Error	13241.809	38	348.469 ^b		
block	Hypothesis	3776.287	19	198.752	.570	.904
	Error	13241.809	38	348.469 ^b		

a. MS(block)

b. MS(Error)

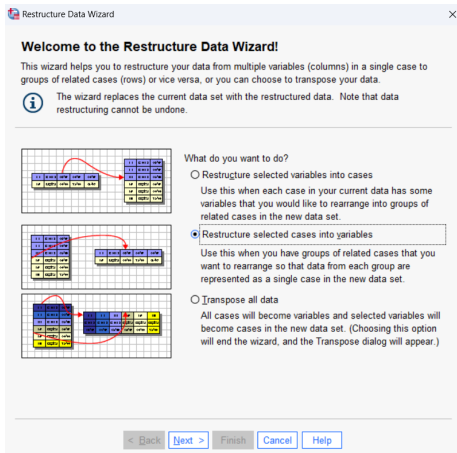
Lab

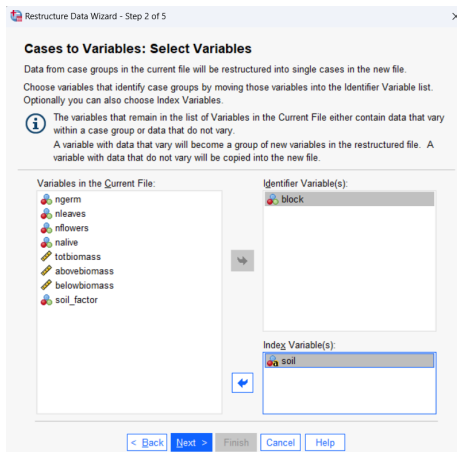
- Before conducting Friedman Rank-Sum Test, we need to reconstruct the data from long format to wide format.

##	block	soil_Moon	soil_Earth	soil_Mars
## 1	1	2.3	8.3	106.7
## 2	2	3.0	8.3	72.5
## 3	3	10.3	26.7	50.6
## 4	4	2.5	11.5	70.5
## 5	5	2.1	16.2	58.3
## 6	6	4.3	8.9	123.4
## 7	7	1.3	15.3	41.1
## 8	8	3.5	13.3	103.7
## 9	9	1.9	49.9	53.0
## 10	10	3.0	18.0	62.7

Lab

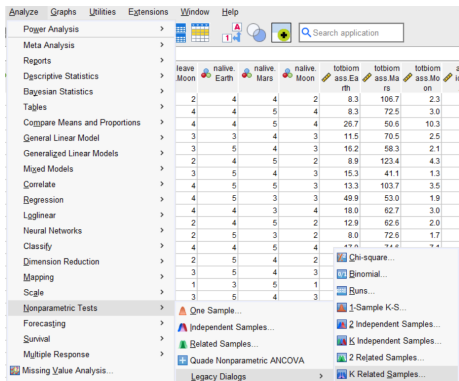
- Let's see how the wide format data can be reconstructed using SPSS
- Data → Restructure...





- Remaining setting unchanged

- Then Click on Analyze → Nonparametric Tests → Legacy Dialogs → K Related Samples.



The screenshot shows the Minitab software interface. A dialog box titled "Tests for Several Related Samples" is open. The "Test Variables" list contains three variables: "totbiomass.Earth", "totbiomass.Mars", and "totbiomass.Moon". The "Test Type" section has "Friedman" selected. The background data table has columns for "block", "nflower.s", "ngerm.Earth", "ngerm.Mars", and three "totbiomass" columns.

block	nflower.s	ngerm.Earth	ngerm.Mars	totbiomass.Earth	totbiomass.Mars	totbiomass.Moon
1	1	0	5			
2	2	0	4			
3	3	0	4			
4	4	0	3			
5	5	0	5			
6	6	0	5			
7	7	0	5			
8	8	0	5			
9	9	0	5			
10	10	0	4			
11	11	0	4			
12	12	0	5			
13	13	0	5			
14	14	0	5			
15	15	0	5	4	3	5
16	16	0	3	5	1	3
17	17	0	5	5	3	5
18	18	0	4	4	1	4
19	19	0	4	5	5	4
20	20	0	3	5	5	3

License



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