Statistics for the Sciences

ANOVA for Factorial Design

Xuemao Zhang
East Stroudsburg University

January 18, 2025

Outline

- Factorial Design
- Lab

Factorial Design

Two-factor Factorial Design

- A two-way classification which involves two factors, both of which are of interest to the experimenter.
  - There are a levels of factor A and b levels of factor B — the experiment is replicated n times at each factor-level combination.
  - The effect of a factor is defined to be the change in mean response produced by a change in the level of the factor (generally when the factor is changed from low to high)
- The replications in the design allow the experimenter to investigate the **interaction** between factors A and B.

Factorial Design

Interaction

- There is an interaction between two factors if the effect of one of the factors changes for different levels of the other factor.
- **Interaction** describes the effect of one factor on the behavior of the other. If there is no interaction, the two factors behave independently.

Factorial Design

- Example (`linton.csv`): Linton et al. (2009) studied the effects of the insecticide pyriproxyfen on ovarian development in an endemic Christmas Island land crab, *Geocarcoidea natalis*. The insecticide was proposed as a means of controlling numbers of an introduced ant species that was viewed as a major threat, and it is an endocrine disruptor. The experiment was designed to test whether the insecticide might pose risks to the crabs, which have a hormone similar to the one targeted in insects, and consisted of feeding crabs a mixture of leaf litter and a bait. Half of the baits contained the insecticide, and the other half were controls (bait type factor). The baits were supplied at three rates, with two levels corresponding to levels used in field applications (2 kg ha^{-1} and 4 kg ha^{-1}), with the third rate being ad libitum feeding (bait dosage factor).
  - The experimental units in this case were large plastic tubs, each containing a single female crab, and there were 7 crabs for each combination of factors. The response variable was the dry mass of the ovaries of each crab.
- The data are categorized with two factors:
  - `type`: Control, Experimental
  - `dosage`: 2, 4, and `ad lib`

Factorial Design

Factorial Design

- The subcategories are called **cells**, and the response variable was the dry mass of the ovaries of each crab. The first 15 rows of data

```
##       type dosage drymass nitrogen
## 1  Control      2   0.524     3620
## 2  Control      2   0.535     4030
## 3  Control      2   1.094     6530
## 4  Control      2   0.525     3938
## 5  Control      2   0.707     4312
## 6  Control      2   0.551     3740
## 7  Control      2   0.489     3860
## 8  Control      4   0.461     4329
## 9  Control      4   0.584     5108
## 10 Control      4   0.715     5877
## 11 Control      4   0.206     1231
## 12 Control      4   0.885     6583
## 13 Control      4   0.835     6342
## 14 Control      4   2.475    11933
## 15 Control ad lib   0.619     4232
```

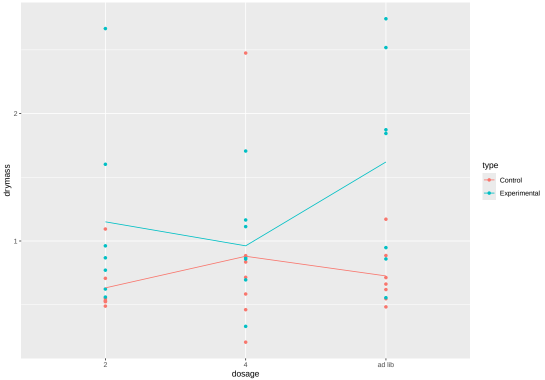
Factorial Design

- Cell means

```
##           type dosage   drymass
## 1      Control      2 0.6321429
## 2 Experimental      2 1.1504286
## 3      Control      4 0.8801429
## 4 Experimental      4 0.9621429
## 5      Control ad lib 0.7258571
## 6 Experimental ad lib 1.6201429
```

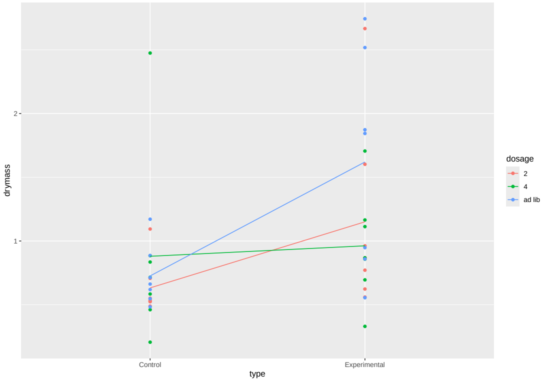
Factorial Design

- Let's construct an interaction graph based on the cell meams.



Factorial Design

- Let's construct an interaction graph based on the cell meams.



2-Way ANOVA for Factorial Design

Interaction

- An interaction effect is suggested if the line segments are far from being parallel.
- No interaction effect is suggested if the line segments are approximately parallel.
- For the linton data, it appears there is an interaction effect:

2-Way ANOVA for Factorial Design

- Let y_{ijk} be the k-th replication at the i-th level of A and the j-th level of B. $i = 1, 2, \ldots, a, j = 1, 2, \ldots, b, k = 1, 2, \ldots, r$.
- The data will be look like this

Table 1: Data from a Factorial Design

	Factor A			
Factor B	1	2	$\cdots$	a
1	$y_{111}, \ldots, y_{11r}$	$y_{211}, \ldots, y_{21r}$	$\cdots$	$y_{a11}, \ldots, y_{a1r}$
2	$y_{121}, \ldots, y_{12r}$	$y_{221}, \ldots, y_{22r}$	$\cdots$	$y_{a21}, \ldots, y_{a2r}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
b	$y_{1b1}, \ldots, y_{1br}$	$y_{2b1}, \ldots, y_{2br}$	$\cdots$	$y_{ab1}, \ldots, y_{abr}$

2-Way ANOVA for Factorial Design

- A factorial design becomes blocking design when there is no replicate in each cell.
- Random effects model:

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

  - where $\varepsilon_{ijk} \sim N(0, \sigma^2)$ are independent normal random errors with common variance σ^2.
  - Both τ_i or β_j are random and are independent of the random errors:
  - τ_i *iid* $\sim N(0, \sigma_\tau^2)$
  - β_j *iid* $\sim N(0, \sigma_\beta^2)$
  - $(\tau\beta)_{ij}$ *iid* $\sim N(0, \sigma_{\tau\beta}^2)$

2-Way ANOVA for Factorial Design

- For the above random-effects model, to test the treatment effects for both factors

$$
\begin{aligned}
H_0 :& \sigma_\tau^2 = 0 \\
H_a :& \sigma_\tau^2 > 0
\end{aligned}
$$

and

$$
\begin{aligned}
H_0 :& \sigma_\beta^2 = 0 \\
H_a :& \sigma_\beta^2 > 0
\end{aligned}
$$

2-Way ANOVA for Factorial Design

- Fixed-effects model

$$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, k = 1, 2, \ldots, n, j = 1, 2, \ldots, b, i = 1, 2, \ldots, a,$$

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

- τ_i is fixed: $\sum_{i=1}^{a} \tau_i = 0$.
- β_j is fixed: $\sum_{j=1}^{b} \beta_j = 0$.
- $(\tau\beta)_{ij}$ is fixed: $\sum_{i=1}^{a} (\tau\beta)_{ij} = \sum_{j=1}^{b} (\tau\beta)_{ij} = 0$

2-Way ANOVA for Factorial Design

- Mixed Effects Model

  $$Y_{ijk} = \mu + \tau_i + \beta_j + (\tau\beta)_{ij} + \varepsilon_{ijk}, k = 1, 2, \ldots, n, j = 1, 2, \ldots, b, i = 1, 2, \ldots, a,$$

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

2-Way ANOVA for Factorial Design

- The SS_{total} is divided into 4 parts:
  - SSA (sum of squares for factor A): measures the variation among the means for factor A
  - SSB (sum of squares for factor B): measures the variation among the means for factor B
  - SS(AB) (sum of squares for interaction): measures the variation among the ab combinations of factor levels
  - SSE (sum of squares for error): measures experimental error

2-Way ANOVA for Factorial Design

- Partition of the total variation:

$$SS_{total} = SSA + SSB + SS(AB) + SS_E$$

That is,

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

- Factor effects (for **fixed-effect models**)
  - A: $\hat{\tau}_i = \bar{y}_{i..} - \bar{y}_{...}$, $i = 1, \dots, a$
  - B: $\hat{\beta}_j = \bar{y}_{.j.} - \bar{y}_{...}$, $j = 1, \dots, b$
  - Interaction: $\widehat{(\tau\beta)}_{ij} = \bar{y}_{ij.} - \bar{y}_{i..} - \bar{y}_{.j.} + \bar{y}_{...}$, $i = 1, \dots, a$, $j = 1, \dots, b$

2-Way ANOVA for Factorial Design

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

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

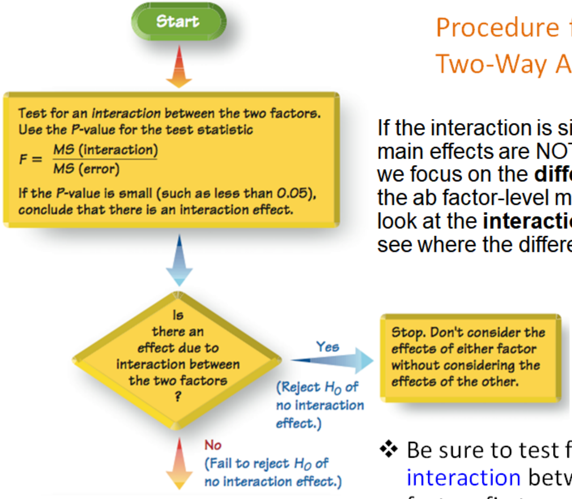
- 2-way ANOVA Table

Table 2: 2-way ANOVA Table for Factorial Design

Source	df	SS	MS	F
Trt A	$a-1$	SSA	$SSA/(a-1)$	MSA/MSE
Trt B	$b-1$	SSB	$SSB/(b-1)$	MSB/MSE
Interaction	$(a-1)(b-1)$	$SS(AB)$	$SS(AB)/(a-1)(b-1)$	$MS(AB)/MSE$
Error	$ab(r-1)$	SSE	$SSE/ab(r-1)$	
Total	$abr-1$	SS_{total}		

- Again, all ANOVA F-tests are right-tailed.
- Procedure for Two-Way ANOVA, see the next two slides

2-Way ANOVA for Factorial Design

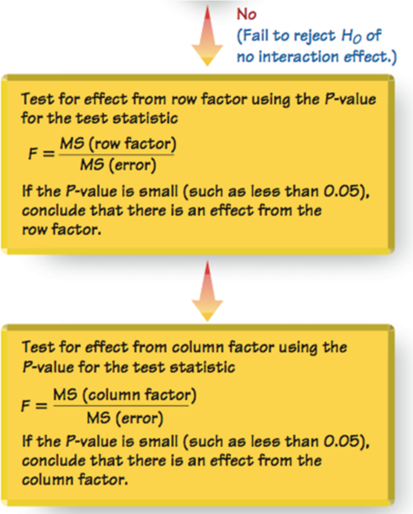


Procedure for Two-Way ANOVA

If the interaction is significant, the main effects are NOT tested, and we focus on the **differences** in the ab factor-level means. We look at the **interaction plot** to see where the differences lie

❖ Be sure to test for an interaction between the two factors first.

2-Way ANOVA for Factorial Design



Procedure for Two-Way ANOVA

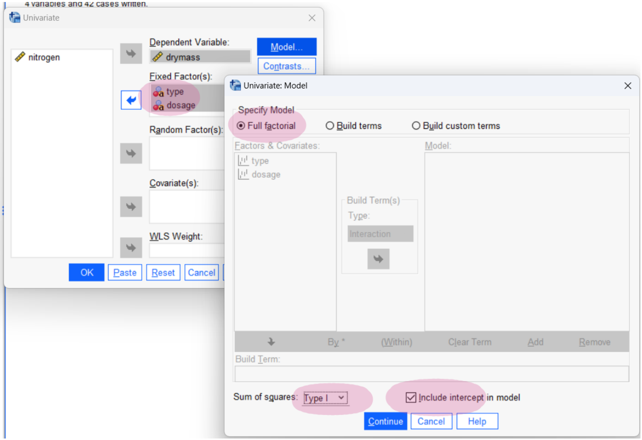
If the interaction effect is not significant, the main effects A and B can be individually tested using F = **MSA/MSE** and F = **MSB/MSE**, respectively.

2-Way ANOVA for Factorial Design

```
## Analysis of Variance Table
##
## Response: drymass
##              Df  Sum Sq Mean Sq F value  Pr(>F)
## type          1  2.6060 2.60603  7.3602 0.01016 *
## dosage        2  0.6705 0.33527  0.9469 0.39739
## type:dosage   2  1.1568 0.57839  1.6336 0.20937
## Residuals    36 12.7465 0.35407
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

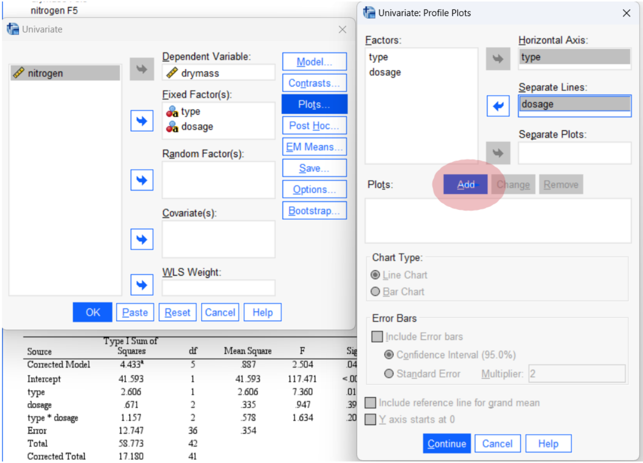
Lab

- For the `linton` data, consider both factors as fixed-effects
  - bait `type`
  - bait `dosage`
- Click on `Analyze` → `General Linear Model` → `Univariate`.



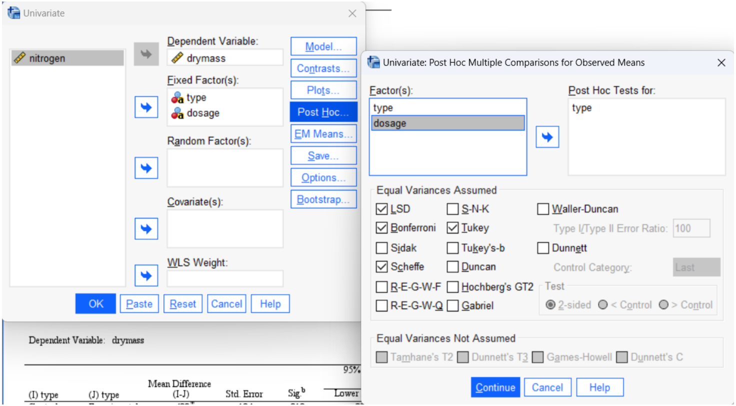
Lab

- We need to add an interaction plot



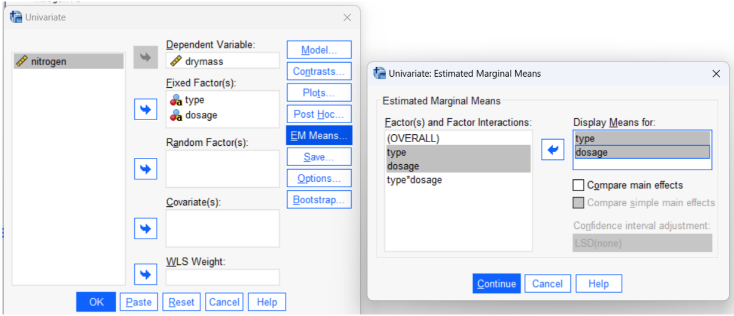
Lab

- Multiple comparisons



Lab

- treatment means



Lab

- The SPSS sytax for the above operations

```
DATASET ACTIVATE DataSet1.
UNIANOVA drymass BY type dosage
  /METHOD=SSTYPE(1)
  /INTERCEPT=INCLUDE
  /POSTHOC=type(TUKEY SCHEFFE LSD BONFERRONI)
  /PLOT=PROFILE(type*dosage) TYPE=LINE ERRORBAR=NO MEANREFERENCE=NO YAXIS=AUTO
  /EMMEANS=TABLES(type)
  /EMMEANS=TABLES(dosage)
  /CRITERIA=ALPHA(0.05)
  /DESIGN=type dosage type*dosage.
```

Lab

Univariate Analysis of Variance

Warnings

Post hoc tests are not performed for type because there are fewer than three groups.

Between-Subjects Factors

		N
type	Control	21
	Experimental	21
dosage	2	14
	4	14
	ad lib	14

Lab

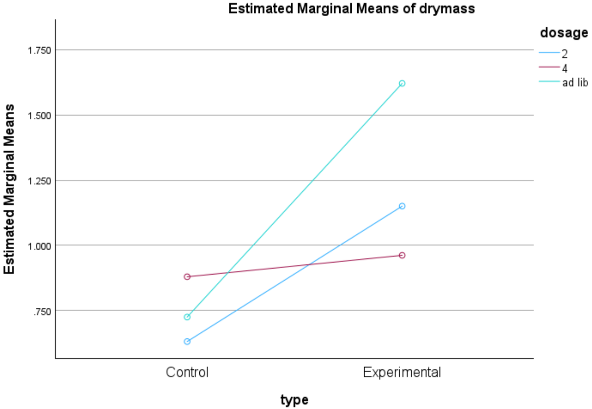
Tests of Between-Subjects Effects

Dependent Variable: drymass

Source	Type I Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.433[a]	5	.887	2.504	.048
Intercept	41.593	1	41.593	117.471	<.001
type	2.606	1	2.606	7.360	.010
dosage	.671	2	.335	.947	.397
type * dosage	1.157	2	.578	1.634	.209
Error	12.747	36	.354		
Total	58.773	42			
Corrected Total	17.180	41			

a. R Squared = .258 (Adjusted R Squared = .155)

Lab



Estimated Marginal Means of drymass
Estimated Marginal Means
1.750
1.500
1.250
1.000
.750
Control
Experimental
type
dosage
2
4
ad lib

License