Course 02403 Introduction to Mathematical Statistics

Lecture 11: Two-way Analysis of Variance, ANOVA

DTU Compute
Technical University of Denmark
2800 Lyngby – Denmark

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Introduction

- Lecture 10: One classification criterion (one-way ANOVA)

$$Y_{ij} = \mu + \alpha_i + \varepsilon_{ij} \quad , \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

- Today: Two classification criteria (two-way ANOVA)

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij} \quad , \quad \varepsilon_{ij} \sim N(0, \sigma^2)$$

- Classification criterion = factor (or categorical variable)
- The first factor is typically called *treatment*, and the second factor *block*

Two-Way Analysis of Variance - Example

- Same data as for one-way ANOVA, but now it is known that the experiment was divided into blocks:

	Group A	Group B	Group C
Block 1	2.8	5.5	5.8
Block 2	3.6	6.3	8.3
Block 3	3.4	6.1	6.9
Block 4	2.3	5.7	6.1

  - Three *treatments* distributed across four *persons*

- *One-way* or *two-way* ANOVA

- *Completely randomized experiment* or *Randomized block experiment*

Two-Way Analysis of Variance - Example

- Same data as for one-way ANOVA, but now it is known that the experiment was divided into blocks:

	Group A	Group B	Group C
Block 1	2.8	5.5	5.8
Block 2	3.6	6.3	8.3
Block 3	3.4	6.1	6.9
Block 4	2.3	5.7	6.1

- Is there a difference (in means) among groups A, B, and C?

- Analysis of Variance (ANOVA) can be used for the analysis if the observations in each cell are assumed to be normally distributed, or if there are sufficiently many observations (CLT).

Example in Python

- Open today's Python notebook in VS Code
  - "Example: Two-way ANOVA"

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Two-Way Analysis of Variance - Model

- The model:

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

  where the errors are independent and identically distributed with

$$\varepsilon_{ij} \sim N(0, \sigma^2).$$

  - μ is the overall mean
  - α_i represents the effect of treatment $i \in \{1, ..., k\}$
  - β_j represents the effect of block $j \in \{1, ..., l\}$
  - There are k treatments and l blocks

- Note: In this course, we only have one observation in each cell (i.e., with the same α and the same β) when performing two-way ANOVA.

Example

Two way ANOVA (the model):

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim \text{ i.i.d. } N(0, \sigma^2), \quad i = 1,2,3, \quad j = 1,2.$$

Some constraint are needed here, e.g. $\alpha_1 = \beta_1 = 0$ or $\sum \alpha_i = \sum \beta_j = 0$. An expanded view of this model is (when $\alpha_1 = \beta_1 = 0$, as in "`smf.ols()`"):

$$\begin{array}{lllll}
y_{11} = \mu & & & & + \varepsilon_{11} \\
y_{12} = \mu & & & +\beta_2 & + \varepsilon_{12} \\
y_{21} = \mu & +\alpha_2 & & & + \varepsilon_{21} \\
y_{22} = \mu & +\alpha_2 & & +\beta_2 & + \varepsilon_{22} \\
y_{31} = \mu & & +\alpha_3 & & + \varepsilon_{31} \\
y_{32} = \mu & & +\alpha_3 & +\beta_2 & + \varepsilon_{32}
\end{array}$$

The exact same in matrix notation:

$$\underbrace{\begin{pmatrix} y_{11} \\ y_{12} \\ y_{21} \\ y_{22} \\ y_{31} \\ y_{32} \end{pmatrix}}_{\mathbf{y}} = \underbrace{\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 \end{pmatrix}}_{\mathbf{X}} \underbrace{\begin{pmatrix} \mu \\ \alpha_2 \\ \alpha_3 \\ \beta_2 \end{pmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{pmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \varepsilon_{31} \\ \varepsilon_{32} \end{pmatrix}}_{\boldsymbol{\varepsilon}}$$

2-way anova as an LM

We can write the design matrix as

$$\boldsymbol{X} = [\mathbf{1} \quad \boldsymbol{X}_{Tr} \quad \boldsymbol{X}_{Bl}].$$

Under the hypothesis that $\alpha_i = 0$ then the design matrix is

$$\boldsymbol{X}_{0,Tr} = [\mathbf{1} \quad \boldsymbol{X}_{Bl}],$$

and under the hypothesis that $\beta_i = 0$ then the design matrix is

$$\boldsymbol{X}_{0,Bl} = [\mathbf{1} \quad \boldsymbol{X}_{Tr}].$$

and under the hypotheis that $\alpha_i = \beta_j = 0$ we have

$$\boldsymbol{X}_0 = \mathbf{1}.$$

We will denote the corresponding projection matrices by $\boldsymbol{H}$, $\boldsymbol{H}_{0,Tr}$, $\boldsymbol{H}_{0,Bl}$ and $\boldsymbol{H}_0$.

Two-Way Analysis of Variance - Estimation

- The parameter estimates $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$ are calculated as (using the constraint $\sum \alpha_i = \sum \beta_j = 0$, i.e. as in Chapter 8):

$$\hat{\mu} = \frac{1}{k \cdot l} \sum_{i=1}^{k} \sum_{j=1}^{l} y_{ij} = \bar{\bar{y}}$$

$$\hat{\alpha}_i = \left(\frac{1}{l} \sum_{j=1}^{l} y_{ij} \right) - \hat{\mu} = \bar{y}_{i.} - \bar{\bar{y}}$$

$$\hat{\beta}_j = \left(\frac{1}{k} \sum_{i=1}^{k} y_{ij} \right) - \hat{\mu} = \bar{y}_{.j} - \bar{\bar{y}}$$

- $\hat{\alpha}_i$ and $\hat{\beta}_j$ describe the estimates of the *marginal* effects of being in a specific treatment group or block.

Example in Python

- Let us estimate these parameters in data from the example:
  - "Example: Estimate parameters μ, α_i, and β_j"

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Two-Way Analysis of Variance - Decomposition and Variance Analysis Table - Theorem 8.20

- With the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2),$$

  the total variation in the data can be decomposed as:

$$SST = SS(Tr) + SS(Bl) + SSE.$$

  or

$$\begin{aligned} \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H}_0)\boldsymbol{Y} =& \boldsymbol{Y}^T(\boldsymbol{H}_{0,Tr} - \boldsymbol{H}_0)\boldsymbol{Y} + \boldsymbol{Y}^T(\boldsymbol{H} - \boldsymbol{H}_{0,Tr})\boldsymbol{Y} + \\ & \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} \\ =& \boldsymbol{Y}^T(\boldsymbol{H}_{0,Bl} - \boldsymbol{H}_0)\boldsymbol{Y} + \boldsymbol{Y}^T(\boldsymbol{H} - \boldsymbol{H}_{0,Bl})\boldsymbol{Y} + \\ & \boldsymbol{Y}^T(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} \end{aligned}$$

- The method is called analysis of variance because the testing involves comparing variances.

Balanced design and Type I/III

Definition (Balanced design)

A design matrix is said to be balanced if the number of observations for any given combination of factors is the same fixed number.

- I all cases we look at there is 1 observation in each cell.

Theorem (Equivalence between Type I and Type III)

For two-way ANOVA with balanced design, the Type I and Type III partitioning of variation is equivalent.

Balanced design and Type I/III - Two way ANOVA

In our case this imply that

$$(\boldsymbol{H}_{0,Tr} - \boldsymbol{H}_0) = (\boldsymbol{H} - \boldsymbol{H}_{0,Bl})$$
$$(\boldsymbol{H}_{0,Bl} - \boldsymbol{H}_0) = (\boldsymbol{H} - \boldsymbol{H}_{0,Tr})$$

or it does not matter in which order we add the hypothesis.

Formulas for Sums of Squares

- Total variation (same as for one-way analysis):

$$SST = \sum_{i=1}^{k} \sum_{j=1}^{l} (y_{ij} - \hat{\mu})^2$$

- *Total Sum of Squares*

Formulas for Sums of Squares

- Variation between treatments/groups (variation explained by *treatments*):

$$SS(Tr) = l \cdot \sum_{i=1}^{k} (\bar{y}_{i.} - \hat{\mu})^2 = l \cdot \sum_{i=1}^{k} \hat{\alpha}_i^2$$

- *Treatment Sum of Squares*
- *"Variation between treatment groups"*

Formulas for Sums of Squares

- Variation between blocks/persons (variation explained by *blocks*):

$$SS(Bl) = k \cdot \sum_{j=1}^{l} (\bar{y}_{\cdot j} - \hat{\mu})^2 = k \cdot \sum_{j=1}^{l} \hat{\beta}_j^2$$

- *Block Sum of Squares*
- *"Variation between blocks"*

Formulas for Sums of Squared Deviations

- Variation of residuals (Variation not explained by the model)

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{l}(y_{ij} - \hat{y}_{ij})^2 = \sum_{i=1}^{k}\sum_{j=1}^{l}(y_{ij} - \hat{\alpha}_i - \hat{\beta}_j - \hat{\mu})^2$$

- *Sum of Squared Errors*

Example in Python

- Open today's Python notebook in VS Code
  - "Example: SST, SS(Tr), SS(Bl) and SSE"

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Two-Way ANOVA - Hypothesis About Different *Treatment Effects* - Theorem 8.22

- The goal is to compare the treatment effects (means α_i) in the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

- The null hypothesis of *no difference/effect among treatments* can be formulated as:

$$\begin{aligned} H_{0,Tr}: &\quad \alpha_i = 0 \quad \text{for all } i \\ H_{1,Tr}: &\quad \alpha_i \neq 0 \quad \text{for at least one } i \end{aligned}$$

- Under $H_{0,Tr}$, the test statistic

$$F_{Tr} = \frac{SS(Tr)/(k-1)}{SSE/((k-1)(l-1))}$$

is F-distributed with $k-1$ and $(k-1)(l-1)$ degrees of freedom.

Two-Way ANOVA - Hypothesis About Different *Block Effects* - Theorem 8.22

- The goal is to compare the block effects (means β_j) in the model

$$Y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \overset{\text{i.i.d.}}{\sim} N(0, \sigma^2).$$

- The null hypothesis of *no difference/effect among blocks* can be formulated as:

$$\begin{aligned} H_{0,Bl}: &\quad \beta_j = 0 \quad \text{for all } j \\ H_{1,Bl}: &\quad \beta_j \neq 0 \quad \text{for at least one } j \end{aligned}$$

- Under $H_{0,Bl}$, the test statistic

$$F_{Bl} = \frac{SS(Bl)/(l-1)}{SSE/((k-1)(l-1))}$$

is F-distributed with $l-1$ and $(k-1)(l-1)$ degrees of freedom.

ANOVA-table

Source of variation	Deg. of freedom	Sums of squares	Mean sum of squares	Test-statistic F	p-value
Treatment	$k-1$	$SS(Tr)$	$MS(Tr) = \frac{SS(Tr)}{k-1}$	$F_{\text{Tr}} = \frac{MS(Tr)}{MSE}$	$P(F > F_{\text{Tr}})$
Block	$l-1$	$SS(Bl)$	$MS(Bl) = \frac{SS(Bl)}{l-1}$	$F_{\text{Bl}} = \frac{MS(Bl)}{MSE}$	$P(F > F_{\text{Bl}})$
Residual	$(k-1)(l-1)$	SSE	$MSE = \frac{SSE}{(k-1)(l-1)}$		
Total	$n-1$	SST			

Example in Python

- Open today's Python notebook in VS Code
  - "Example: F-test and ANOVA table"

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Example in Python

- Open today's Python notebook in VS Code
  - "Example: Model diagnostics"

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Post Hoc Confidence Intervals

- As in one-way ANOVA (use method 8.9 and 8.10), but replace $n-k$ degrees of freedom with $(k-1)(l-1)$ and use MSE from two-way ANOVA.
- Can be done for either treatments or blocks.
- A single pre-planned comparison of the difference between treatment a and b is given by:

$$\bar{y}_a - \bar{y}_b \pm t_{1-\alpha/2}\sqrt{\frac{SSE}{(k-1)(l-1)}\left(\frac{1}{n_a}+\frac{1}{n_b}\right)}$$

  where $t_{1-\alpha/2}$ is from the t-distribution with $(k-1)(l-1)$ DOF.
- If M combinations of pairwise confidence intervals are calculated, use the formula M times but each time with $\alpha_{\text{Bonferroni}} = \alpha/M$.

Post Hoc Pairwise Hypothesis Testing

- For a single *pre-planned* hypothesis test:

$$H_0:\ \alpha_a = \alpha_b,\ \ H_1:\ \alpha_a \neq \alpha_b$$

compute the test statistic as:

$$t_{\text{obs}} = \frac{\bar{y}_a - \bar{y}_b}{\sqrt{MSE\left(\frac{1}{n_a} + \frac{1}{n_b}\right)}}$$

and the p-value as:

$$p = 2P(T > |t_{\text{obs}}|),$$

where T follows a t-distribution with $(k-1)(l-1)$ degrees of freedom.

- If M combinations of pairwise confidence intervals are calculated, use the adjusted significance level: $\alpha_{\text{Bonferroni}} = \alpha/M$.

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons

Generalizations, its is all about $\boldsymbol{X}$

We have studied the model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}; \quad \boldsymbol{\epsilon} \sim N(\mathbf{0},^2 \boldsymbol{I}).$$

The (very) general theory is given in Chapter 9, and specific examples are given in

1. The columns of $\boldsymbol{X}$ given as zero and ones (t-test and ANOVA, Chapter 3 and 8)
2. The columns of $\boldsymbol{X}$ given as real numbers (regression, Chapter 5 and 6)

Further we assumed balanced design in the two-way ANOVA. The set up is fairly easily generalized (meaning that all general formulas for the LM transfer directly), to

- Non-balanced design, i.e. different number of observations in each group (e.g. missing observations, or multiple observations in some cells).
- Multiple (more than 2) factors.
- Mix of regression and factor analysis (different slopes in different groups)
- Interaction effects, i.e. "effect of factor A and B" $\neq$ "effect of factor A + factor B".

Overview

1. Introduction and Example
2. The Model
3. Computation: Variance Decomposition and ANOVA Table
4. Hypothesis Testing (F-test)
5. Model Diagnostics
6. Post Hoc Comparisons
7. The general linear model, generalizatons