Introduction to Statistics Note

2024 Spring Semester

21 CST H3Art

Chapter 13: Two-Way Analysis of Variance

13.1 The Two-Way ANOVA Model

Two-Way Designs（双因子）

In a two-way design, **two factors (independent variables)** are studied in **conjunction（结合）** with **the response (dependent) variable**. There are thus two ways of organizing the data, as shown in a two-way table.

When the dependent variable is **quantitative**, the data are **analyzed with a two-way ANOVA procedure**. A **chi-square test** is used instead if the dependent variable is **categorical**. **（当因变量是定量的时，使用双向方差分析分析数据。如果因变量是分类变量，则使用卡方检验）**

Two-Way ANOVA（双因子方差分析）

- It is **more efficient to study two factors at once（同时研究两个因子更有效）** than separately.
  - A two-way design requires smaller sample sizes per condition than does a series of one-way designs because the samples for all levels of factor $\mathbf{B}$ contribute to sampling for factor $\mathbf{A}$. **（与一系列单向设计相比，双向设计每个条件所需的样本量更小，因为因子$\mathbf{B}$所有级别的样本都有助于因子$\mathbf{A}$的采样。）**
- Including a second factor thought to influence the response variable helps reduce the residual variation in a model of the data.
  - In a **one-way ANOVA** for factor $\mathbf{A}$, any effect of factor $\mathbf{B}$ is assigned to the $\textbf{residual}$ ($\textbf{“error”}$ term). In a **two-way ANOVA**, **both factors** contribute to the $\textbf{“fit”}$ part of the model.
- **Interactions between factors（因子间相互作用）** can be investigated.
  - The two-way ANOVA breaks down the fit part of the model between each of the **main components (the two factors)** and an **interaction effect**. The interaction cannot be tested with a series of one-way ANOVAs. **（交互作用无法被单因子方差分析测试）**

The Two-Way ANOVA Model

- We record a quantitative variable in a two-way design with I **levels of factor $\mathbf{A}$** and J **levels of factor $\mathbf{B}$**.
- There are $I \times J$ combinations of the two sets of factor levels. We'll use (i, j) to denote the population for which **factor $\mathbf{A}$ is at level i** and **factor $\mathbf{B}$ is at level j**.
- We assume that we have $I \times J$ **independent SRSs**, one from each population, and that each population is **Normal**. The **population means may be different** but **all populations have the same σ**.
- Let x_{ijk} represent the k-th observation from population (i, j). The statistical model is:

$$
\begin{aligned}
x_{ijk} &= \mu_{ij} & +\varepsilon_{ijk} \\
&= \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} & +\varepsilon_{ijk}
\end{aligned}
$$

  where μ_{ij} **is the mean of population** (i, j) and ε_{ijk} **is an error term**.
  μ_{ij}, the mean of combination (i, j), **can be broken down into** the **overall mean** μ, **plus** α_i, the effect of level i of Factor $\mathbf{A}$, **plus** β_i, the effect of level j of Factor $\mathbf{B}$, **plus** $(\alpha\beta)_{ij}$, the interaction effect of the two levels

Interaction（相互作用）: Two variables interact if a particular combination of variables leads to results that **would not be anticipated on the basis of the main effects of those variables（根据这些变量的主要影响却无法被预期的的结果）**. An interaction implies that the effect of one variable is different at different levels of another variable.

Main Effects（主要影响）: This is the impact on the response (dependent variable) of varying levels of that factor, **regardless of the other factor** (i.e., pooling together the levels of the other factor). **（只却决于单个解释变量的变化水平）**

- When there is **no clear interaction**, the **main effects are enough** to describe the data. In the **presence of interaction**, the **main effects could mask what is really going on with the data**.

In a two-way design, statistical significance can be found for each factor, for the interaction effect, or for any combination of these:

Neither factor is significant	Neither factor is significant	Only one factor is significant	Both factors are significant
No interaction	Interaction effect is significant	No interaction	With or without significant interaction

13.2 Inference for Two-Way ANOVA

Inference for Two-Way ANOVA（双因子方差分析推断）:

- A one-way ANOVA tests the following model of your data:

$$\text{data}(\text{"Total"}) = \text{fit}(\text{"Groups"}) + \text{residual}(\text{"Error"})$$

  so that the sums of squares and degrees of freedom are:

$$\mathbf{SST} = \mathbf{SSG} + \mathbf{SSE}$$
$$\mathbf{DFT} = \mathbf{DFG} + \mathbf{DFE}$$

- When the **sample sizes are equal（样本大小相同）**, a two-way design breaks down the "fit" part of the model into more specific subcomponents, so that:

$$\mathbf{SST} = \mathbf{SSA} + \mathbf{SSB} + \mathbf{SSAB} + \mathbf{SSE}$$
$$\mathbf{DFT} = \mathbf{DFA} + \mathbf{DFB} + \mathbf{DFAB} + \mathbf{DFE}$$

  where $\mathbf{A}$ and $\mathbf{B}$ are the two **main effects** from each of the two factors, and $\mathbf{AB}$ represents the **interaction** of factors $\mathbf{A}$ and $\mathbf{B}$.

The Two-Way ANOVA Table（双因子方差分析表）

Source of variation	df	Sum of Square $\mathbf{SS}$	Mean Square MS	F	P-value
Factor $\mathbf{A}$	$\mathbf{DFA} = I - 1$	$\mathbf{SSA}$	$\mathbf{MSA} = \mathbf{SSA}/\mathbf{DFA}$	$F_{\mathbf{A}} = \mathbf{MSA}/\mathbf{MSE}$	for $F_{\mathbf{A}}$
Factor $\mathbf{B}$	$\mathbf{DFB} = J - 1$	$\mathbf{SSB}$	$\mathbf{MSB} = \mathbf{SSB}/\mathbf{DFB}$	$F_{\mathbf{B}} = \mathbf{MSB}/\mathbf{MSE}$	for $F_{\mathbf{B}}$
Interaction	$\mathbf{DFAB} = (I-1)(J-1)$	$\mathbf{SSAB}$	$\mathbf{MSAB} = \mathbf{SSAB}/\mathbf{DFAB}$	$F_{\mathbf{AB}} = \mathbf{MSAB}/\mathbf{MSE}$	for $F_{\mathbf{AB}}$

Source of variation	df	Sum of Square $\mathbf{SS}$	Mean Square MS	F	P-value
Error	$\mathbf{DFE} = N - IJ$	$\mathbf{SSE}$	$\mathbf{MSE} = \mathbf{SSE}/\mathbf{DFE}$		
Total	$\mathbf{DFT} = N - 1 = \mathbf{DFA} + \mathbf{DFB} + \mathbf{DFAB} + \mathbf{DFE}$	$\mathbf{SST} = \mathbf{SSA} + \mathbf{SSB} + \mathbf{SSAB} + \mathbf{SSE}$			

- **Main effects（主要效应）**: P-value for factor $\mathbf{A}$, P-value for factor $\mathbf{B}$.
- **Interaction**: P-value for the interaction effect of $\mathbf{A}$ and $\mathbf{B}$.
- **Error**: Represents variability in the measurements within the groups.
- $\mathbf{MSE}$ is an unbiased estimate of the population variance σ^2.

Cautions for Two-Way ANOVA（双因子方差分析注意事项）:

1. Always **construct a plot of factor level means（构建因子水平均值图）**. Examine for possible interaction.
2. **Examine the test for interaction first（首先检验交互作用）**. Presence of a strong interaction may influence the interpretation of the main effects. **（强相互作用的存在可能会影响主要效应的解释）**
3. When an interaction is present, carefully examine the means to properly interpret the data. **（当存在相互作用时，需要仔细检验均值以合理解释数据）**
    - The marginal means do not tell the whole story **（边界均值不能说明所有情况）**
    - An examination of treatment means may provide a better interpretation than the ANOVA F tests for main effects. (A treatment mean is the mean response to a particular combination of factor levels.) **(对措施均值的检验也许会提供比ANOVA F检验更好的对主要效应的解释，因为采取某种措施的均值是特定因素水平组合下的响应均值，其可以更好地解释主要效应)**