Introduction to Statistics Note

2024 Spring Semester

21 CST H3Art

Chapter 12: One-Way Analysis of Variance

12.1 Inference for One-Way Analysis of Variance

Compare any number of means using **analysis of variance(ANOVA/方差分析)**.

The Idea of ANOVA:

- Analysis of variance (ANOVA) is the technique used to **compare several means（比较多个均值）**.
- **One-way ANOVA（单因子方差分析）** is used for situations in which there is **only one factor（单因子）**, or **only one way to classify the populations（单向总体分类）** of interest.
- **Two-way ANOVA（双因子方差分析）** is used to analyze the effect of **two factors（双因子）**.
- Analysis of variance(ANOVA) compares the variation due to specific sources with the variation among individuals who should be similar. In particular, ANOVA tests whether several populations **have the same means（具有相同的均值）** by comparing **how far apart the sample means are（样本均值之间相距/组间均方）** with **how much variation there is within a sample（样本的变异程度/组内均方）**.
- An **overall test（整体测试）** to find any differences among the parameters we want to compare **（找出我们想要比较的参数的差异）**
- A detailed **follow-up analysis（后续分析）** to decide which groups differ and how large the differences are **（确定哪些组存在差异，以及差异的大小）**

Random sampling always produces chance variations. Any "**factor effect（因素效应）**" would thus show up in our data as the **factor-driven differences（因素驱动的差异）** plus **chance variations ("error")（机会变化/偶然差异）**：

$$\text{data} = \text{factor effect} + \text{error}$$

And the one-way ANOVA model analyzes data x_{ij} where **chance variations are normally distributed（偶然差异满足正态分布）** $N(0, \sigma)$, and we take random samples from each of I **different populations（总体大小为I）**：

$$\begin{aligned} x_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \mu + \alpha_i + \varepsilon_{ij} \end{aligned}$$

for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$. The ε_{ij} are assumed to be from a $N(0, \sigma)$ distribution. The **parameters of the model（模型参数）** are the **population means（总体均值）** $\mu_1, \mu_2, \ldots, \mu_I$ and the **common standard deviation（共同标准差）** σ.

μ_i can be broken into the overall mean μ and the effect of level i of the factor. **（μ_i可以分解为总体均值μ和i水平下因子的效应）**

Test the **null hypothesis（零假设）** that there are **no differences among the means（均值之间没有差异）** of the populations, and the **alternative hypothesis（备选假设）** is that there is **some difference, namely not all means are equal, the hypothesis is not one-sided or two-sided. It is "many-sided"（均值之间存在差异，这个假设是多尾的）**.

The above test is called the **analysis of variance F test（ANOVA F检验）**.

The conditions under which we can use ANOVA are:

- Conditions for ANOVA Inference
  - We have I **independent SRSs（有I个独立的简单随机抽样样本）**, one from each population. We **measure the same response variable for each sample（测量每个样本中同样的相应变量）**.
  - The i-th population has a **Normal distribution** with unknown **mean** μ_i. **（均值μ_i满足正态分布）**
  - All the populations **have the same standard deviation** σ, whose **value is unknown**. **（所有总体有同样的方差，但方差值未知）**
- Checking Standard Deviations in ANOVA
  - The results of the ANOVA F test are approximately correct **when the largest sample standard deviation is no more than twice as large as the smallest sample standard deviation（当最大样本标准差不超过最小样本标准差的两倍时）**.

To determine statistical significance, we need a test statistic—the ANOVA F Statistic, it has this form:

$$F = \frac{\textbf{variation among the sample means}}{\textbf{variation among individuals in the same sample}}$$

- F is always **zero or positive（F值大于等于0）**
  - F is **zero** only when **all sample means are the same（所有样本均值相同时F值为0）**.
  - F gets **larger** as **means move further apart（样本均值偏差越大，F值越大）**.
- Large values of F are evidence against H_0 : **equal means**
- The F test is **upper-one-sided（x坐标轴上方的单尾检验）**.

When describing an F distribution, always give the **numerator degrees of freedom first（分子作为F分布的第一个自由度参数）**. $F(df_1, df_2)$ with

- df_1 degrees of freedom in the **numerator（分子）**
- df_2 degrees of freedom in the **denominator（分母）**
- We want to compare the means of I populations. We have an SRS of size n_i from the i-th population, so that the total number of observations in all samples combined is:

$$N = n_1 + n_2 + \cdots + n_I$$

If the null hypothesis is all population means are equal, the ANOVA F statistic has the F distribution with $I-1$ **degrees of freedom in the numerator** and $N-I$ **degrees of freedom in the denominator**.

$$F(I - 1, N - I)$$

The **measures of variation** in the numerator and denominator are **mean squares**:

- Numerator: **Mean Square for Groups** ($\textbf{MSG}$)

$$\textbf{MSG} = \frac{n_1(\bar{x_1} - \bar{x})^2 + n_2(\bar{x_2} - \bar{x})^2 + \cdots + n_I(x_I{}^- - \bar{x})^2}{I - 1}$$

- Denominator: **Mean Square for Error** ($\textbf{MSE}$)

$$\textbf{MSE} = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \cdots + (n_I - 1)s_I^2}{N - I}$$

  - MSE is also called the **pooled sample variance**, written as s_p^2 (s_p is the **pooled standard deviation**)
  - s_p^2 estimates the **common variance（共同方差）** σ^2

The ANOVA Table

Source of variation	Sum of squares SS	df	Mean square MS	F	P-value	F crit
Among or between "groups"	$\sum n_i(\bar{x}_i - \bar{x})^2$	$I-1$	$\mathbf{MSG} = \mathbf{SSG/DFG}$	$\mathbf{MSG/MSE}$	Tail area above F	Value of F for α
Within groups or "error"	$\sum(n_i-1)s_i^2$	$N-I$	$\mathbf{MSE} = \mathbf{SSE/DFE}$			
Total	$\mathbf{SST} = \mathbf{SSG} + \mathbf{SSE}$ $\sum(x_{ij}-\bar{x})^2$	$N-1$				

Coefficient of determination:

$$R^2 = \mathbf{SSG/SST}$$

Pooled standard deviation:

$$\sqrt{\mathbf{MSE}} = s_p$$

The sums of squares represent different sources of variation in the data:

$$\mathbf{SST} = \mathbf{SSG} + \mathbf{SSE}$$

The degrees of freedom mirror the sums of squares:

$$\mathbf{DFT} = \mathbf{DFG} + \mathbf{DFE}$$

$$\mathbf{data(``Total")} = \mathbf{factor\ effect(``Groups")} + \mathbf{error(``Error")}$$

12.2 Comparing the Means

Contrasts（对比） can be used only when there are clear expectations **BEFORE starting an experiment**, and these are reflected in the experimental design. Contrasts are **planned comparisons（有计划的比较）** .

- It is **NOT appropriate to use a contrast** test when that test is suggested only **AFTER the data are collected**. **（在数据收集后进行对比是不合适的）**
- To **test a specific hypothesis that some treatments are different from other treatments（测试特定假设中的一些不同措施）** , we can use a contrast to test for significant differences
- Contrasts are more powerful than multiple comparisons because they are more specific. **（对比相比多重比较更有说服力，因为它更具体）**
- Use a t-**test** on the contrast or calculate a t confidence interval
- Since the hypothesis is **pre-specified（假设是预先指定的）** , the corresponding t-test is typically done in lieu of the multiple sample ANOVA test **（通常会进行相应的t检验来代替多样本方差分析检验）**

A contrast is **a combination of population means（总体均值的组合）** of the form, where the coefficients a_i have sum 0:

$$\psi = \sum a_i\mu_i$$

To test the null hypothesis $H_0 : \psi = 0$ use the t statistic with degrees of freedom $\mathbf{DFE}$ that is associated with s_p. The alternative hypothesis can be **one- or two-sided**:

$$t = \frac{c}{\mathrm{SE}_c}$$

The corresponding sample contrast is:

$$c = \sum a_i \bar{x}_i$$

The standard error of c is:

$$\mathbf{SE}_c = s_p \sqrt{\sum \frac{a_i^2}{n_i}} = \sqrt{\mathbf{MSE} \times \sum \frac{a_i^2}{n_i}}$$

A level C confidence interval for the difference ψ, where t^* is the critical value defining the middle $C\%$ of the t distribution with $\mathbf{DFE}$ degrees of freedom:

$$c \pm t^* \times \mathbf{SE}_c$$

Multiple comparisons（多重比较） should be used when **the pattern of differences between means is unknown beforehand（平均值差异事先未知）**. Such comparisons are **pair-wise（成对比较）** tests of significance.

Planned vs. Post-Hoc Tests（计划测试与事后测试）：

- **Planned tests** are tests formulated **prior to the data collection（计划测试是在数据收集之前制定的测试）**. These are **contrasts** that are of interest to the researchers on top of the overall F-test
- **Post-hoc tests** are tests formulated **after ANOVA has been performed（事后测试是在执行方差分析后制定的测试）**
- After a significant F-test has been obtained, different contrasts that are of interest to the researchers may be formulated. **In most cases, contrasts in the form of pairwise comparisons（大多数情况下以成对比较的形式进行对比）** are formulated
- Some people proceed with testing the different contrasts **without considering when the contrasts were formulated – before or after the data collection – or how many contrasts involved（不考虑对比何时制定或涉及多少对比）**
  - This approach **inflates the overall Type I error（夸大了总体的I类误差）** (i.e., **false positive rate（假阳性率）**) because it only **controls for what is called the testwise Type I error rate（仅控制所谓的 testwise I类错误率）**
  - This means that the **test for each contrast will have an error rate of α（每个对比的测试都会有α的错误率）**
  - However, the error rate for all the different tests taken together (known as familywise Type I error rate) is not known, but can be assumed to be higher than α
  - To **keep the familywise Type I error at most at α（保持在最多α的familywise的I类误差）**, some adjustments need to be adopted to **make the individual tests more conservative（使个体测试更加保守）**

The Bonferroni Procedure（邦费罗尼校正）

The Bonferroni procedure is an example of **a procedure that adjusts for simultaneously performing many tests at the same time（调整同时执行多个测试的过程）**.

The Bonferroni procedure can be done in either of two equivalent ways. In either approach, one **chooses an overall significance level α** and then does a number of pair-wise t-tests. Let k **be the total number of t-tests performed**.

- Use the procedure with t^* being the t-value with degrees of freedom $\mathbf{DFE}$ and **an area of $\alpha/(2k)$ to its right（右侧面积为 $\alpha/(2k)$，即总体显著性水平α除以2×进行t检验的次数k）**
- Compute a P-value for each t-test in the usual way. Conclude that a particular pair of means is significantly different only when k times that P-value is no larger than α **（仅当 k 乘以 P 值不大于 α 时，得出结论，特定的均值对/pair of means，其存在显著的不同）**
- When one does multiple comparisons the chance of committing at least one **type I error** increases with the number of tests done **（当进行多次比较时，犯至少一个I类错误的机会会随着完成的测试次数的增加而增加）**
- To compensate for this tendency, the Bonferroni procedure lowers the working significance level of each test to control the probability of making at least one type I error among all tests performed **（降低了每个测试的显著性水平，以控制在所有执行的测试中至少出现一个I类错误的概率）**
- As a consequence, the more pair-wise comparisons you perform, the more difficult it will be to show statistical significance for each test **（执行的成对比较越多，显示每个测试的统计显着性就越困难）**
- The Bonferroni procedure is **typically used with planned contrasts（通常与计划对比一起使用）**

Scheffé's Method（薛费氏事后比较法）

- Scheffé's method **keeps the Type I error at α for all possible (post-hoc) contrasts（将所有可能的、事后对比的I类误差保持在α）**
- Consequently, **sleuthing or fishing for significant results is allowed（因此，允许探寻或捞取重大成果？）**
- That is, we can try as many contrasts as we want after getting a significant F-test
- However, there's no guarantee that the significant results for the contrasts would be particularly interesting
- Scheffé's method **requires that a different distribution by used（要求使用不同的分布）**
- Scheffé's method **adjust the critical t^* as a function of the F distribution $F(I-1, N-I)$ and the number of groups I**（将临界t^*调整为F分布$F(I-1, N-I)$和组数I的函数）
- Scheffé's method is **the most conservative method（最保守的方法）** that can be used without getting more conservative than necessary
- A related method for post-hoc tests is **Tukey's method**
  - This gives a Type I experimentwise error rate of α for all possible pairwise comparisons
  - Is less conservative than Scheffé's method **(Tukey's method不如Scheffé's method那么保守)**
- As long as we adjust for (1) the number of contrasts or (2) when the contrasts are formulated, the procedure will be more conservative (e.g., more likely to retain the null) than if we do not make any adjustments at all