

# ANOVA

BWB

## 1 ANOVA 简介

ANOVA 是"Analysis of Variance" (分散分析) 的缩写。它是 t-test 的推广与扩展, 用于检验三组及三组以上实验结果是否有显著差异。

## 2 ANOVA 的主要概念

### 2.1 假设

- $H_0: \mu_1 = \mu_2 = \dots = \mu_g$  vs  $H_1$ : 至少有一个  $\mu_i \neq \mu_j$

### 2.2 假设条件

- 观测值在每个样本内部是独立的
- 所有总体具有相同的方差  $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_g^2 = \sigma$
- 每个总体都服从正态分布 (或样本量足够大, 可以应用中心极限定理)

### 2.3 检验统计量

$$T = \frac{\text{Treatment Mean Sq}}{\text{Residual Mean Sq}} \sim F_{g-1, N-g} \text{ under } H_0$$

### 2.4 观测检验统计量

记为  $t_0$

### 2.5 p 值

$$P(T \geq t_0) = P(F_{g-1, N-g} \geq t_0)$$

注意: 总是查看上尾概率。

### 2.6 决策

如果 p 值小于显著性水平  $\alpha$ , 则拒绝原假设, 得出至少有一组的总体均值与其他组显著不同的结论。  
如果 p 值大于  $\alpha$ , 则不拒绝原假设, 认为没有足够的证据表明各组间存在显著差异。

### 3 ANOVA 的数学模型

对于每个观测值  $y_{ij}$  (其中  $j = 1, 2, \dots, n_i$  且  $i = 1, 2, \dots, g$ ), 我们假设:

$$Y_{ij} \sim N(\mu_i, \sigma^2)$$

且所有随机变量都是独立的。

### 4 一些需要记住的 dot notation

- 样本  $i$  的总和:  $\sum_{j=1}^{n_i} y_{ij} = y_{i\bullet}$
- 样本  $i$  的平均值:  $\frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} = \bar{y}_{i\bullet}$
- 所有观测值的总和:  $\sum_{i=1}^g \sum_{j=1}^{n_i} y_{ij} = y_{\bullet\bullet}$
- 所有观测值的总平均:  $\frac{1}{N} \sum_{i=1}^g \sum_{j=1}^{n_i} y_{ij} = \bar{y}_{\bullet\bullet}$
- 其中  $N = n_1 + \dots + n_g$  是观测值总数
- 第  $i$  组的样本方差:  $s_i^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2$

### 5 ANOVA 分解

总平方和 (Total Sum of Squares) 可以分解为:

$$\sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{\bullet\bullet})^2 = \sum_{i=1}^g \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_{i\bullet})^2 + \sum_{i=1}^g n_i (\bar{y}_{i\bullet} - \bar{y}_{\bullet\bullet})^2$$

即: Total SS = Residual SS + Treatment SS

### 6 处理均方和残差均方

#### 6.1 处理均方 (Treatment Mean Square)

$$\text{Treatment Mean Square} = \frac{\sum_{i=1}^g n_i (\bar{Y}_{i\bullet} - \bar{Y}_{\bullet\bullet})^2}{g - 1}$$

#### 6.2 残差均方 (Residual Mean Square)

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^g \sum_{j=1}^{n_i} (Y_{ij} - \bar{Y}_{i\bullet})^2}{N - g} \sim \left( \frac{\sigma^2}{N - g} \right) \chi_{N-g}^2$$

#### 6.3 F 统计量

$$\frac{\text{Treatment Mean Square}}{\text{Residual Mean Square}} \sim F_{g-1, N-g} \text{ under } H_0$$

## 7 R 代码

首先，让我们看一下完整的 R 代码：

```
1 # 加载数据
2 data(PlantGrowth)
3
4 # 查看数据结构
5 str(PlantGrowth)
6
7 # 计算每组的描述性统计
8 tapply(PlantGrowth$weight, PlantGrowth$group, summary)
9
10 # 创建箱线图
11 boxplot(weight ~ group, data = PlantGrowth,
12         main = "Plant Growth in Different Conditions",
13         xlab = "Treatment", ylab = "Weight")
14
15 # 进行方差分析
16 #aov(numerical data ~ categorical data, data = data)
17 anova_result <- aov(weight ~ group, data = PlantGrowth)
18
19 # 查看 ANOVA 结果
20 summary(anova_result)
```

Source	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	2	3.766	1.8832	4.846	0.0159*
Residuals	27	10.492	0.3886		

表 1: ANOVA Table

## 8 HATPC 步骤分析

现在，让我们按照 HATPC 步骤来分析这个 ANOVA。

### 8.1 H - Hypothesis (假设)

原假设 ( $H_0$ ): 所有处理组的植物重量平均值相等。

$$H_0 : \mu_{ctrl} = \mu_{trt1} = \mu_{trt2}$$

备择假设 ( $H_1$ ): 至少有一组的平均值与其他组不同。

$$H_1 : \text{至少有一个 } \mu_i \neq \mu_j$$

其中， $\mu_{ctrl}$ ,  $\mu_{trt1}$ , 和  $\mu_{trt2}$  分别代表对照组、处理 1 组和处理 2 组的总体平均值。

## 8.2 A - Assumptions (假设条件)

1. 独立性: 每个观察值都是独立的。2. 正态性: 每个组内的数据近似服从正态分布。3. 方差齐性: 所有组的总体方差相等。

我们可以通过图形方法 (如 Q-Q 图和残差图) 和统计测试 (如 Shapiro-Wilk 测试和 Levene 测试) 来检验这些假设。

## 8.3 T - Test Statistic (检验统计量)

我们使用 F 统计量作为检验统计量。F 统计量是处理均方与残差均方的比值:

$$F = \frac{\text{Treatment Mean Square}}{\text{Residual Mean Square}}$$

从 ANOVA 表中, 我们可以得到:

$$F = 4.846$$

## 8.4 P - P-value (p 值)

从 ANOVA 表中, 我们可以看到 p 值:

$$p = 0.0159$$

这个 p 值表示在原假设为真的情况下, 观察到当前或更极端结果的概率。

## 8.5 C - Conclusion (结论)

由于 p 值 (0.0159) < 0.05 (假设我们选择 0.05 作为显著性水平), 我们拒绝原假设。

结论: 在 5% 的显著性水平下, 我们有足够的证据认为至少有一种处理方法对植物生长有显著影响。

在进行方差分析 (ANOVA) 后, 如果我们发现存在显著差异, 下一步就是确定具体哪些组之间存在差异。以下是这个过程的关键步骤和概念:

## 9 1. 残差正态性检验

在进行方差分析 (ANOVA) 后, 如果发现存在显著差异, 我们通常需要进行多重比较来确定具体哪些组之间存在差异。以下是三种常用的多重比较方法及其特点。

## 10 1. Bonferroni 方法

**R 代码:**

```
confint(pairs(flicker_em, adjust = "bonferroni"))
test(pairs(flicker_em, adjust = "bonferroni"))
```

**特点:**

- The Bonferroni test should be used when you have a set of planned comparisons you would like to make beforehand.

Bonferroni method 需要我们提前想好对比。

- Bonferroni post-hoc test produces the most narrow confidence intervals, which means it has the greatest ability to detect true difference between the groups of interest.  
Bonferroni method 可以给到最窄的 confidence intervals (越窄越强)。
- Note that the Bonferroni post-hoc test can also be used whether or not the group sample sizes are equal.  
Bonferroni method 对 sample size 没有要求。
- 最简单和保守的方法。
- 通过将 p 值乘以比较次数来调整显著性水平。
- 适用于预先计划好的少量比较。
- 当比较次数增加时, 可能会过于保守, 导致统计检验力降低。

## 11 2. Tukey 方法

**R 代码:**

```
confint(pairs(flicker_em, adjust = "tukey"))
test(pairs(flicker_em, adjust = "tukey"))
```

**特点:**

- 用于所有可能的成对比较。
- 比 Bonferroni 方法的统计检验力更高。
- 适用于样本量相等的情况。
- 控制整个实验的错误率。

## 12 3. Scheffe 方法

**R 代码:**

```
confint(pairs(flicker_em, adjust = "scheffe"))
test(pairs(flicker_em, adjust = "scheffe"))
```

**特点:**

- The Scheffe test should be used when you would like to make all possible contrasts between group means. 并不是只有两两对比, 把所有可能的组合都做了个遍。
- This test allows you to compare more than just two means at once, unlike the Tukey post-hoc test. 与之前的两个 Method 都去做一个区分。
- While the Scheffe test is the most flexible, it is also the most conservative and produces the widest confidence intervals. This means it has the lowest statistical power and the lowest ability to detect true differences between the groups. 是所有校正方法里最弱的, 有着最宽的 confidence interval。
- Note that the Scheffe post-hoc test can be used whether or not the group sample sizes are equal. 对 Sample size 没有提出要求。

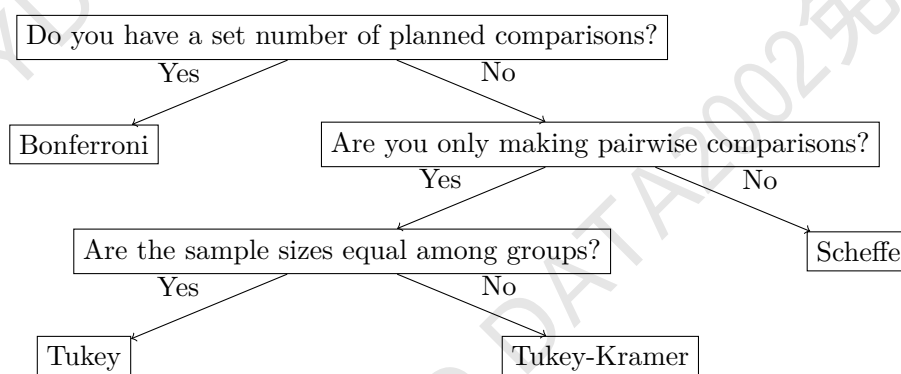
- 最灵活的方法，允许所有可能的对比。
- 可以比较两个以上的均值。
- 统计检验力较低。

## 13 方法比较

1. 保守程度：Scheffe > Bonferroni > Tukey
2. 统计检验力：Bonferroni > Tukey > Scheffe
3. 适用情况：

- Bonferroni：预先计划的少量比较
- Tukey：所有可能的成对比较，样本量相等
- Scheffe：复杂比较，样本量可以不等

## 14 选择合适的后续检验方法



## 15 使用建议

1. 如果只关心少数预先计划的比较，可以使用 Bonferroni 方法。
2. 如果需要进行所有可能的成对比较，且样本量相等，Tukey 方法是一个很好的选择。
3. 如果需要进行复杂的比较（不仅仅是成对比较），或者样本量不等，可以考虑使用 Scheffe 方法。
4. 在实际应用中，可以尝试多种方法并比较结果。如果结果一致，那么可以更有信心地得出结论。
5. 无论使用哪种方法，都要注意解释结果时的谨慎性，特别是在处理边界情况时。
6. 选择方法时，要在研究开始前就决定，以避免选择性报告的问题。

## 16 ANOVA 假设不满足时的替代方法（之后的不怎么考，了解就行）

在某些情况下，ANOVA 的基本假设可能不满足。以下是两种常见情况及其替代方法：

### 16.1 共同方差假设不满足

当各组的方差不相等时，可以使用 Welch 检验并应用 Bonferroni 校正：

```
pairwise.t.test(wolfSAldrin, wolfSDepth,
                p.adjust.method = "bonferroni",
                pool.sd = FALSE)
```

这种方法考虑了各组的不同方差，并使用 Bonferroni 校正来控制总体的 I 类错误率。

## 16.2 正态分布假设不满足

当数据不满足正态分布假设时，可以使用 Kruskal-Wallis 检验：

### 16.2.1 ANOVA 的正态分布假设

在标准 ANOVA 中：

- 每个总体都应该服从正态分布
- 方差应该相等

在这些假设下，原假设简化为：

所有观测值来自同一个正态分布

一个更弱的假设是：

所有观测值来自同一个分布（不一定是正态分布）

### 16.2.2 Kruskal-Wallis 程序

Kruskal-Wallis 检验是 ANOVA 的非参数替代方法：

1. **假设:**  $H_0$ : 响应变量在所有组中分布相同  $H_1$ : 至少有一个组的响应变量系统地高于其他组
  2. **假设条件:** 每个组内的观测值是独立的，组与组之间也是独立的。不同的组遵循相同的分布（仅在位置参数上有所不同）。
  3. **检验统计量:** 类似于对秩次应用 ANOVA。在原假设下，Kruskal-Wallis 检验统计量近似服从自由度为  $g - 1$  的卡方分布，其中  $g$  是组的数量。
  4. **p 值:**  $P(T \geq t_0) = P(\chi_{g-1}^2 \geq t_0)$
  5. **决策:** 如果 p 值小于  $\alpha$ ，我们拒绝原假设，得出至少有一组的总体均值与其他组显著不同的结论。如果 p 值大于  $\alpha$ ，我们不拒绝原假设，认为没有足够的证据表明各组间存在显著差异。
- 在 R 中，可以使用以下函数进行 Kruskal-Wallis 检验：

```
kruskal.test(Time ~ factor(Game), data = rugby)
```

## 17 结论

1. 在进行 ANOVA 之前，应该先检查数据是否满足基本假设。
2. 如果共同方差假设不满足，可以考虑使用 Welch 检验并应用 Bonferroni 校正。
3. 如果正态分布假设不满足，Kruskal-Wallis 检验是一个很好的非参数替代方法。
4. 无论使用哪种方法，都要谨慎解释结果，并考虑到所使用方法的限制和假设。
5. 在实际应用中，可能需要结合多种方法来获得更全面的理解。