# 第 9 章：混合设计方差分析和完全嵌套方差分析

# 引言

方差分析与实验设计密不可分。实验条件的安排以及随之带来的数据结构决定了合适的方差分析的种类。第 6 章的单因素独立样本 ANOVA 分析的是单个因素的多个水平之间的不同独立样本是否有差异，第 7 章的单因素重复测量 ANOVA 分析的是同一组被试在单个因素的不同水平下重复测量得到的多个相关样本之间是否有差异，第 8 章的双因素独立样本 ANOVA 分析的是两个因素的不同水平缔造多个组合条件之间的差异，不同条件使用不同组的被试；而条件之间的差异包括两个因素各自的主效应，也有它们之间的交互效应。实验研究当然不止这 3 种设计，但是，上述 3 种实验设计是其他更加复杂的实验设计的基础。可以说，用上面三种 ANOVA 形式的灵活组合，我们可以应对更多、更复杂的实验及其数据。

本章我们将介绍两种更复杂的方差分析，一种是混合设计方差分析（mixed-design ANOVA），一种是完全嵌套方差分析（nested ANOVA）。因为两者涉及的研究问题有比较大的差异，我们将先后分别对两者进行介绍。

# 混合设计方差分析

## 混合设计方差分析的研究问题

混合设计涉及的因素中不仅有独立样本，也有重复测量的相关样本。因此，它具备了独立样本 ANOVA 和重复测量 ANOVA 的特点。同时，因为有多个因素共同决定因变量，混合设计 ANOVA 在分析上和多因素 ANOVA 相似（例如第 8 章的双因素独立样本 ANOVA）类似，需要考虑交互效应。

我们以最简单的混合设计方差 ANOVA，即双因素混合设计 ANOVA，为例子来介绍。双因素混合设计 ANOVA 只包含一个独立分组因素和一个重复测量因素。研究者想比较两种的学习时间安排对短期和长期的学习效果的影响。一种时间安排是临场抱佛脚学习法（简称佛脚法），就是把学习时间集中安排在一个时间段中；第二种时间安排是细水长流学习法（简称流水法），就是把学习时间分散使用，在一个比较长的时期内多次、小段时间地学习。公平起见，佛脚法和流水法在总学习时间上是一样的，只是时间编排不同。为了量化学习效果，研究者在学习前，学习后即刻，学习完成后第 6 个月三个时间点通过测验的方式考察了被试学习记忆的效果（学习表现为 0-10 分）。学习前测验得到的是被试对学习材料掌握的基线水平，学习后即刻是看学习短期效果（类似复习后立刻参加考试），学习完成后第 6 个月的回访测试是看学习的长期效果（因为这时候学习效果相对稳定）。研究者招募了共 20 名被试，随机分配到 2 种时间安排的条件下，然后跟踪在上述三个时间点测试了他们的学习效果。数据表格（表格 9.1）和误差棒图（图 9.1）如下。

表格 9.1: 两种学习方法（1：佛脚法；2：流水法）对短期和长期学习效果的影响数据集。学习前、后即刻、6 个月后的学习表现分布罗列在三列中；学习表现值越大表现月好。一行数据是一个被试。

| 被试编号 | 学习方法 | 学习前 | 学习后即刻 | 学习后6 个月 |
|---|---|---|---|---|
| 1 | 1 | 1 | 6 | 5 |
| 2 | 1 | 1 | 6 | 3 |
| 3 | 1 | 3 | 8 | 4 |
| 4 | 1 | 5 | 9 | 5 |
| 5 | 1 | 2 | 8 | 4 |
| 6 | 1 | 3 | 5 | 3 |
| 7 | 1 | 4 | 7 | 4 |
| 8 | 1 | 5 | 8 | 7 |
| 9 | 1 | 4 | 8 | 4 |
| 10 | 1 | 4 | 9 | 2 |
| 11 | 2 | 4 | 6 | 5 |
| 12 | 2 | 2 | 7 | 5 |
| 13 | 2 | 3 | 5 | 5 |
| 14 | 2 | 3 | 7 | 7 |
| 15 | 2 | 3 | 6 | 6 |
| 16 | 2 | 5 | 5 | 6 |
| 17 | 2 | 2 | 5 | 4 |
| 18 | 2 | 3 | 5 | 5 |
| 19 | 2 | 4 | 6 | 6 |
| 20 | 2 | 6 | 7 | 6 |
| | | $M_1 = 3.35$ | $M_2 = 6.65$ | $M_3 =4.8$ |
| | | $SD_1 = 1.35$ | $SD_2 =1.35$ | $SD_3 =1.32$ |

观察数据图我们发现两种时间安排都有学习效果，两组被试的学习表现在学习后即刻和 6 个月跟踪回访的表现都好于学习前的基线水平。但是，两组之间又有差异：佛脚组在学习后即刻的平均表现要好于流水组，但是流水组在跟踪回访的平均表现又反过来好于佛教组。但是，无论是学习效果本身，还是组间差异，这些差异是否达到了显著水平，光看各个条件的均值是无法判断的。这里，我们需要使用方差分析来做判断。

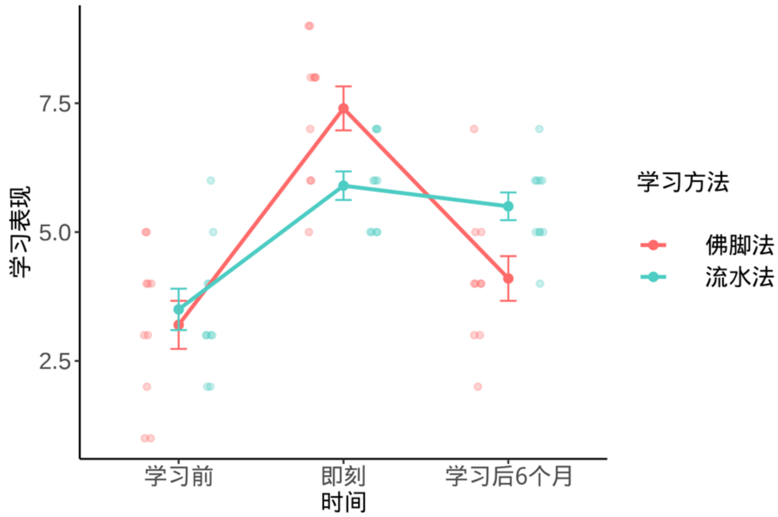


图 9. 1：两种学习方法的被试的长期和短期学习效果。

在学习方法的示例中，学习方法条件就是独立因素(independent factor)或者分组因素(group factor)，而不同测试时间就是重复测量因素(repeated-measure factor)或者相关因素(dependent factor)。因此，这是混合设计方差分析的数据结构。每一个被试只被分配到一个组中，而他/她被多次重复测量，贡献了 3 个数据点（表 9.1）。虽然数据结构类似双因素独立样本 ANOVA（第八章表 8.1）, 但是这里的一行数据来自于同一个被试，而不是来自于不同的被试。因此，我们不能套用第 8 章独立样本 ANOVA 的做法，而是需要使用混合设计 ANOVA 的方法。

# 混合设计方差分析的主要内容

和双因素独立样本 ANOVA 类似，双因素混合设计 ANOVA 也涉及两个主效应和一个交互效应。示例中，学习方法的分组变量是自变量 A，它有 2 个水平。而测量时间点是重复测量变量 B，它有 3 个水平。对于 A 因素的主效应，我们有：

$$H_0: u_{A1} = u_{A2}$$

这里的虚无假设是一个综括假设，即两个条件之间(A1, A2)都没有差别。也就是说两组不同被试之间的表现没有差异，学习方法的不同没有造成学习效果的不同。

相应的，对于 B 因素的主效应，我们有：

$$H_0: u_{B1} = u_{B2} = u_{B3}$$

即 B 的三个水平之间没有差异。也就是说不同时间点测量的学习效果没有差别。

第三个虚无假设是 A 的不同水平对 B 对因变量的影响不会带来差异。也可以说，B 的不同水平对 A 对因变量的影响不会带来差异。这两者说法对应：

$$H_0: u_{ij} = u_{ik}$$

$$H_0: u_{ji} = u_{ki}$$

这 2 种交互效应的表达中，第一个下缀 i 表达 A 的不同水平，第二个下缀 j 表达 B 的不同水平。交互效应的虚无假设就是说无论学习效果的时间效应是什么样子，佛脚组和流水组

没有差别。或者说，无论佛脚组和流水组有什么组间差异，他们的组间差异在不同时间阶段上没有变化。

如果用线性模型表达混合因素 ANOVA 的数据的话：

$$y_{ijk} = u + a_j + \beta_k + \gamma_{jk} + s_i + \varepsilon_{ijk}$$

其中$y_{ijk}$是 A 自变量的第 j 个水平和 B 自变量的第 k 个水平下的第 i 个被试的分数。$u$是总体均值，$a_j$是自变量 A 第 j 个水平的效应，$\beta_k$是自变量 B 第 k 个水平的效应，$\gamma_{jk}$就是相应的交互效应，$s_i$是第 i 个被试的效应。$\varepsilon_{ijk}$就是 ANOVA 模型的残差（residue），即除了两个主效应、交互效应和被试效应以外的额外的、模型不能解释的数据变异性；残差假设服从均值为 0 的高斯分布。

和双因素独立样本 ANOVA 类似，混合设计 ANOVA 也有 3 个效应，即两个主效应和一个交互效应。但是，和双因素独立样本 ANOVA 不同，混合设计 ANOVA 有两个不同的误差项而不是一个。

组间效应，即独立样本因素 A 的效应，是分组变量的不同水平之间的差异，其对应的平方和是：

$$SS_A = \sum n_j(\overline{Y_j} - \overline{\mathrm{G}})^2$$

其中，$n_j$是第 j 组的被试数目，$\overline{Y_j}$是第 j 组的分数的均值。注意，这里的均值是包括了所有的重复测量水平的分数。$\bar{G}$是所有分数的总体平均值。$SS_A$对应的自由度是分组变量的水平数目 a-1。在学习方法的示例中，$n_j$ = 10， a = 2。

组间效应所对应的误差项是被试间的效应，即个体差异。它的平方和可以用下面公式来计算：

$$SS_{S/A} = k\sum(\overline{Y_i} - \overline{Y_j})^2$$

其中，k 是重复测量水平的次数，$\overline{Y_j}$是第 i 个被试的平均分数，而$\overline{Y_j}$是他/她所在组的平均分数。因此，$SS_{S/A}$对应的自由度是 a(s-1), s 是每一组被试的数目。在学习方法的示例中，s = 10, k = 3.

由此，我们可以知道组间效应的 F 检验公式是：

$$F = \frac{MS_A}{MS_{S/A}} = \frac{SS_A/(a-1)}{SS_{S/A}/a(s-1)}$$

F 检验的自由度分别是 a 和 a(s-1)。组间效应的计算方法和单因素独立样本 ANOVA 类似，而且$MS_{S/A}$的计算方法和单因素 ANOVA 的$MS_{within}$一样。

对于组内效应，情况就更复杂一些。混合设计 ANOVA 的组内效应不仅包括重复测量因素 B 的主效应，还包括重复测量和组间效应的交互效应。

$$SS_B = \sum n_k(\overline{Y_{..k}} - \overline{\mathrm{G}})^2$$

$$SS_{AB} = \sum n_{jk}(\overline{Y_{.jk}} - \overline{\mathrm{G}})^2$$

其中，$n_k$是第 k 次测量（重复测量因素 B 的第 k 个水平）下的被试量，$\overline{Y_{..k}}$是第 k 次测量的平均值。相应的，$n_{jk}$是第 j 组被试在第 k 次测量的被试量，$\overline{Y_{.jk}}$是这个条件下的平均值。两个效应对应的自由度分别是 b-1 和 (a–1)×(b-1)。从公式可以看出，两个组内效应

（或者重复测量效应）的平方和和均方的计算方法和两因素独立样本 ANOVA 类似。但是，两个组内效应的误差项和前述的组间效应的误差项不一样，使用的是模型的残差$\varepsilon_{ijk}$，在数值计算上，残差$\varepsilon_{ijk}$的平方和可以表达为：

$$SS_{B*S/A} = \sum (Y_{ijk} - \overline{Y_{.jk}})^2 - k\sum (\overline{Y_{i..}} - \overline{Y_{.j}})^2$$

可以看出，这里的残差是每个条件内（不同 A 和 B 的组合）分数的变异性减去了被试效应，剩下的变异性才是 ANOVA 模型无法解释的残差。它的自由度是 a(b-1)(s-1)。至此，我们可以完成混合设计 ANOVA 的两个组内效应 F 检验:

$$F_A = \frac{MS_A}{MS_{S/A}} = \frac{SS_A/(a-1)}{SS_{S/A}/a(s-1)}$$

$$F_B = \frac{MS_B}{MS_{B*S/A}} = \frac{SS_B/(b-1)}{SS_{B*S/A}/a(b-1)(s-1)}$$

$$F_{AB} = \frac{MS_{AB}}{MS_{B*S/A}} = \frac{SS_{AB}/(a-1)(b-1)}{SS_{B*S/A}/a(b-1)(s-1)}$$

通过以上的计算步骤，我们可以构建两因素独立样本 ANOVA 的最终结果：

表 9.2:学习方法示例的混合设计 ANOVA 结果表

| | *MS* | *MSE* | *df1* | *df2* | *F* | *p* | $\eta^2$ |
|---|---|---|---|---|---|---|---|
| 学习方法 | 0.067 | 2.574 | 1.00 | 18.00 | 0.026 | .874 | .001 |
| 时间 | 55.085 | 0.964 | 1.99 | 35.76 | 57.151 | <.001*** | .760 |
| 学习方法*时间 | 10.789 | 0.964 | 1.99 | 35.76 | 11.193 | <.001*** | .383 |

首先，我们发现交互效应显著，表明佛脚组和流水组的学习效果的时间效应不尽相同。组间主效应不显著，但是时间的主效应显著。后者表明学习总体上来说是有效果的，综括两种学习方法来说。当然，组间的主效应不显著不能解读成为两组没有差别，因为这里关键的交互效应是显著的。获取这部分结果的 R 代码如下：

```
mixed_fit <- MANOVA(mixed_data,
                    dvs = "time1:time3",
                    dvs.pattern = "time(.)",
                    between = "group",
                    within = "time") ###CAUTION###
```

```
## ──────────────────────────────────────────────────────────────────────
##                   MS   MSE   df1    df2      F     p      η²p [90% CI of η²p]  η²G
## ──────────────────────────────────────────────────────────────────────
## group          0.067 2.574 1.000 18.000  0.026  .874      .001 [.000, .094] .001
## time          55.085 0.964 1.987 35.759 57.151 <.001 ***  .760 [.636, .830] .575
## group * time  10.789 0.964 1.987 35.759 11.193 <.001 ***  .383 [.165, .540] .210
## ──────────────────────────────────────────────────────────────────────
```

我们还有几个问题亟待回答：随机分组的两组人在学习前测中是否没有显著差异？学习后即刻和 6 个月后的回测是否有显著差异？和学习前相比，两组被试是否在 6 个月后的学习效果都消失了？这些问题需要用简单主效应和事后检验来回答。

# 混合设计方差分析的简单主效应和事后检验

因为两因素 ANOVA 的交互效应显著，我们需要检验在一个因素的一个水平之下，另一个因素的不同水平是否有显著差异，这需要进一步做简单主效应的检验。注意，如果交互效应不显著，我们是不能做简单主效应的，也没有必要。

混合设计 ANOVA 和独立样本 ANOVA（第 8 章）在简单主效应和事后检验上的计算方法基本一致。我们只显示上述示例的 R 代码和检验结果。

```
# Post hoc
EMMEANS(mixed_fit, effect = "time", by = "group")
```

```
## Multivariate Tests of "time":
## ──────────────────────────────────────────────────────────────────────
##                  Pillai's trace Hypoth. df Error df Exact F     p
## ──────────────────────────────────────────────────────────────────────
## group1: "time"           0.861      2.000   17.000  52.520 <.001 ***
## group2: "time"           0.665      2.000   17.000  16.849 <.001 ***
## ──────────────────────────────────────────────────────────────────────
## Pairwise Comparisons of "time":
## ─────────────────────────────────────────────────────────────────────────────────────
##       Contrast "group" Estimate    S.E. df      t     p     Cohen's d [95% CI of d]
## ─────────────────────────────────────────────────────────────────────────────────────
##  time2 - time1  group1    4.200 (0.422) 18  9.961 <.001 ***  3.118 [ 2.292,  3.945]
##  time3 - time1  group1    0.900 (0.453) 18  1.988  .187      0.668 [-0.219,  1.555]
##  time3 - time2  group1   -3.300 (0.438) 18 -7.538 <.001 *** -2.450 [-3.308, -1.592]
##  time2 - time1  group2    2.400 (0.422) 18  5.692 <.001 ***  1.782 [ 0.956,  2.608]
##  time3 - time1  group2    2.000 (0.453) 18  4.417 <.001 ***  1.485 [ 0.598,  2.372]
##  time3 - time2  group2   -0.400 (0.438) 18 -0.914 1.000     -0.297 [-1.155,  0.561]
## ─────────────────────────────────────────────────────────────────────────────────────
```

```
EMMEANS(mixed_fit, effect = "group", by = "time")
```

```
## Multivariate Tests of "group":
## ──────────────────────────────────────────────────────────────────────
##                  Pillai's trace Hypoth. df Error df Exact F     p
## ──────────────────────────────────────────────────────────────────────
## time1: "group"           0.013      1.000   18.000   0.238  .632
## time2: "group"           0.326      1.000   18.000   8.691  .009 **
## time3: "group"           0.295      1.000   18.000   7.538  .013 *
## ──────────────────────────────────────────────────────────────────────
```

```
## Pairwise Comparisons of "group":
## ─────────────────────────────────────────────────────────────────────────────────
##          Contrast "time" Estimate    S.E. df      t     p     Cohen’s d [95% CI of d]
## ─────────────────────────────────────────────────────────────────────────────────
##  group2 - group1  time1    0.300 (0.616) 18  0.487  .632        0.223 [-0.737,  1.183]
##  group2 - group1  time2   -1.500 (0.509) 18 -2.948  .009 **   -1.114 [-1.907, -0.320]
##  group2 - group1  time3    1.400 (0.510) 18  2.746  .013 *     1.039 [ 0.244,  1.835]
## ─────────────────────────────────────────────────────────────────────────────────
```

# 混合设计方差分析的结果报告

混合设计 ANOVA 的分析步骤和双因素独立样本 ANOVA 完全一样，也是先获得综括 ANOVA 的结果，然后根据交互效应是否显著决定是否进行下一步的简单主效应和事后检验。根据上面不同学习方法的跟踪研究的范例数据分析，我们可以对其 ANOVA 分析结果进行解读：

我们通过 2（学习方法）×3（时间点）的双因素混合设计 ANOVA 分析了学习效果的数据，其中学习方法是组间变量，而时间点是组内变量。方差分析的交互效应显著（$F(2，36) = 11.193$，$p < .001$），表明两种学习方法对学习效果的变化的影响不同。学习方法的主效应不显著（$F(1,18) = 0.026$，$p = .874$），而时间点的主效应显著（$F(2,36) = 57.151$, $p < .001$）。简单主效应及其事后检验表明表明佛脚组学习结束后即刻学习水平显著好于学习前（$p < .001$），但是它在 6 个月后的回测中的表现和学习后即刻有显著差异（$p < .001$）,而和学习前不再有显著差异（$p = .187$）。这表明佛脚组的学习效果只是学习后即刻有提升，6 个月后学习效果就几乎消失。但是，流水组学习结束后即刻和 6 个月后回测的学习效果都显著好于学习前（$p < .001$）。这表明流水组的学习效果不仅在学习后即刻有提升，而且 6 个月后还是保持了一定的水平。另外，简单主效应也表明两组在学习前学习效果没有差异（$p$=.632），学习后即刻佛脚组的表现好于流水组（$p = .009$）；但是，这种优势在 6 个月后的回测中翻转了过来：这时候流水组反而好于佛脚组（$p = .013$）。

# 混合设计方差分析的前提假设和补充说明

混合设计方差 ANOVA 同时有组间因素和组内因素，所以其前提假设包括了双因素独立样本 ANOVA 和重复测量 ANOVA 的前提假设。包括：1）每一个组间因素的水平的数据要服从高斯假设，2）方差同质性假设：不仅组间因素的不同水平之间方差同质，而且所有条件（即组间因素和组内因素的组合）之间的方差同质。3）球形假设：组内因素不同水平之间的差异分数的方差同质，而且在每一个组间因素的水平下都要满足球形假设。对方差同质性假设和球形假设，我们可以分布使用 Levene’s F 检验和 Mauchly’s 球形检验来测试数据是否满足假设 2 和 3，并且在使用 bruceR 包中的 MANOVA 函数进行方差分析时，会直接进行 Levene’s F 检验和 Mauchly’s 球形检验并生成结果。见 R 代码示例。

```r
# check normality
mixed_data %>%
  group_by(group) %>%
  shapiro_test(time1)
```

```
## # A tibble: 2 × 4
##   group variable statistic     p
##   <dbl> <chr>        <dbl> <dbl>
## 1     1 time1        0.901 0.225
## 2     2 time1        0.903 0.238
```

```r
mixed_data %>%
  group_by(group) %>%
  shapiro_test(time2)
```

```r
mixed_data %>%
  group_by(group) %>%
  shapiro_test(time3)
```

```r
mixed_fit <- MANOVA(mixed_data,
                    dvs = "time1:time3",
                    dvs.pattern = "time(.)",
                    between = "group",
                    within = "time") ###CAUTION###
```

```
## Levene’s Test for Homogeneity of Variance:
## ────────────────────────────────────────────────────
##              Levene’s F df1 df2     p
## ────────────────────────────────────────────────────
## DV: time1        0.371   1  18  .550
## DV: time2        2.584   1  18  .125
## DV: time3        0.536   1  18  .473
## ────────────────────────────────────────────────────
##
## Mauchly’s Test of Sphericity:
## ───────────────────────────────────────────────
##                Mauchly's W     p
## ───────────────────────────────────────────────
## time                0.9933  .944
## group * time        0.9933  .944
## ───────────────────────────────────────────────
```

本章讲解了最简单的混合设计方差分析，即一个组间因素，一个组内因素的两因素的混合设计方差分析。当然，针对研究问题，我们可以使用更复杂的设计，例如两个组间因素和一个组内因素。假设学习效果受学习方法影响的研究中，研究者不仅对学习时间安排感兴趣，还对被试者的年龄感兴趣。因此，研究者引入了小学生、初中生和高中生三个

年龄组。这样，上述的示例就变成了 2（时间安排组）×3（年龄组）×3（时间点）的三因素混合设计。其中，时间安排和年龄是组间因素（或随机分组因素），而时间点还是组内因素（或重复测量因素）。我们用下面的表格（表 9.3）来总结 4 种常见的混合设计 ANOVA 的情况：

表 9.3: 四种常见的混合设计 ANOVA。

| | | | **Sources of Variability** | | |
|---|---|---|---|---|---|
| 实验设计 | Randomized-Groups IVs | Repeated-Measures IVs | *Randomized Groups* | *Repeated Measures* | *Error Terms* |
| Two-way mixed | A | B | A | 7 | S/A |
| | | | | B, A×B | B×S/A |
| Three-way mixed | A, B | C | A, B, A×B | | S/AB |
| | | | | C, A×C, B×C, A×B×C | C×S/AB |
| | A | B, C | A | | S/A |
| | | | | B, A×B | B×S/A |
| | | | | C, A×C | C×S/A |
| | | | | B×C, A×B×C | B×C×S/A |
| Four-way mixed | A, B | C, D | A, B, A×B | | 4 |
| | | | | C, A×C, B×C, A×B×C | 5 |
| | | | | D, A×D, B×D, A×B×D | 7 |
| | | | | C×D, A×C×D, B×C×D, A×B×C×D | 4 |

# 完全嵌套方差分析

## 完全嵌套方差分析的研究问题

我们用一个例子来说明完全嵌套 ANOVA 的研究问题和数据结构。假设我们研究初中生使用的记忆方法对隔天记忆效果的影响。研究者使用了无意义的字符作为记忆材料，被试隔天能回忆的字符数目可以作为记忆效果的衡量，而且该测量只进行了一次。学习方法有三种，联想法、谐音法和记忆宫殿法。为了让被试更有代表性，研究者选取了 12 个不同的中学的学生，每个中学抽取了 5 个被试。12 个中学被随机分配到 3 种记忆方法分组中。这样实验设计对应的数据结构如表 9.4 所示。

表 9.4: 来自不同初中的初中生使用不同记忆方法在隔天记忆测试中的表现，每个中学 5 个被试

| 记忆方法 | 中学 | 记忆表现水平 | | | | |
|---|---|---|---|---|---|---|
| 1 | 1 | 8 | 7 | 7 | 4 | 5 |
| | 2 | 6 | 4 | 3 | 5 | 5 |
| | 3 | 6 | 7 | 6 | 8 | 7 |
| | 4 | 8 | 7 | 9 | 9 | 8 |
| 2 | 5 | 8 | 7 | 6 | 9 | 9 |
| | 6 | 6 | 5 | 4 | 9 | 6 |
| | 7 | 4 | 7 | 6 | 7 | 4 |
| | 8 | 6 | 3 | 6 | 6 | 4 |
| 3 | 9 | 5 | 2 | 3 | 4 | 4 |
| | 10 | 5 | 6 | 5 | 5 | 7 |
| | 11 | 8 | 5 | 8 | 7 | 6 |
| | 12 | 5 | 7 | 3 | 4 | 6 |

研究者意识到这些中学有些是民办中学，有些是普通公办中学，有些是重点公办中学，有些是外来务工人员的专门中学。而不同中学的同学在记忆测验中的表现可能会有差异，因为背景知识可以帮助字符记忆，而不同中学的同学在背景知识上可能有差别。重要的是，研究者关心的问题是记忆方法是否会对记忆表现造成差别，而不是中学带来的记忆表现的差异。很明显，我们需要量化和控制无关变量（被试从属的中学）的影响。

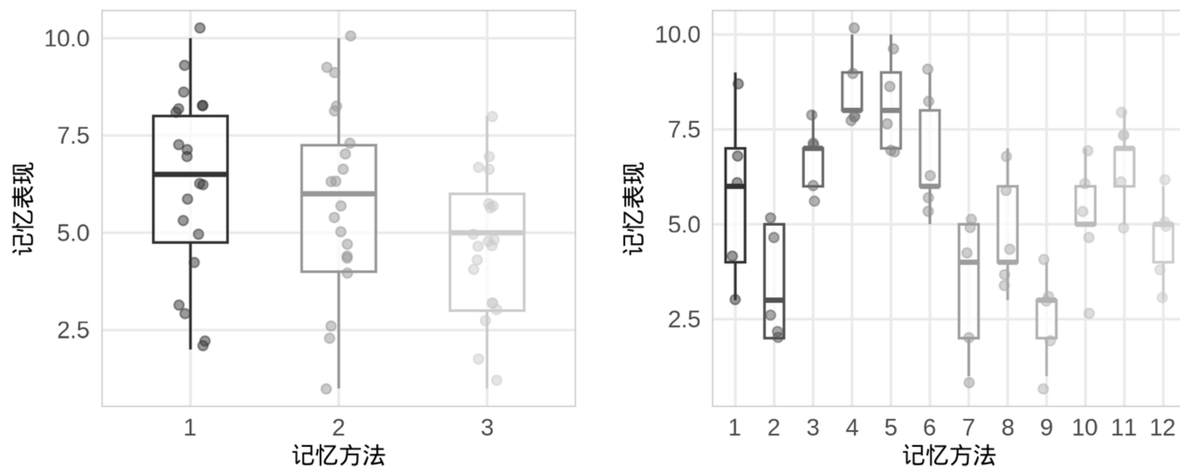


图 9.2: 三种记忆方法对应的隔天记忆表现。（左）每个记忆方法的箱图。（右） 每个学校的记忆表现的箱图。

完全嵌套方差分析是该问题适用的方法。图 9.2 显示了本例的数据结果和大致的数据情况。从表 9.4 中可以看出，学校这个分组变量是嵌套在主因素记忆方法中的，而且这个分组变量在每一种记忆方法内部是不重复的。因此，这样的数据结果是无法使用记忆方法和学校作为两个随机分组变量的两因素独立样本 ANOVA 的。独立样本 ANOVA 要求的数据结构是两个分组变量是完全交叉（fully factorial）。如果这里的学校和记忆方法完全交叉，那么学校的分组变量的数值应该在不同记忆方法的水平下都是是 1~4 的重复，而不是目前的 1~12（见表格 9.4 的第一列）。在本例中，一个分组变量（学校）完全嵌套在另一个分组变量（记忆方法）中，而不是“交叉”在后者中。更重要的是，嵌套变量是一个混淆变量（confounding variable）或者额外变量（nuisance variable），而不是研究者感兴趣的实验变量或者被操控的变量。

很不凑巧的是，学校这个混淆变量对数据的影响比较大：从图 9.2B 可以看出，不同学校的学生的记忆表现是有比较大的差异的，这些差异需要被量化和控制，这样才能保证我们能有效地衡量记忆方法的效应。在这种情况下，我们应该使用完全嵌套 ANOVA 而不是双因素 ANOVA 来分析。

# 完全嵌套方差分析的主要内容

完全嵌套 ANOVA 涉及分组变量 A 的主效应和一个混淆变量 B 的无关效应。在记忆研究的示例中，记忆方法的分组变量是自变量 A，它有 3 个水平；而被试所属的中学是无关变量 B，它有 12 个水平。对于 A 因素的主效应，我们有：

$$H_0: u_{A1} = u_{A2} = u_{A3}$$

这是完全嵌套方差分析的唯一的虚无假设，即 3 个条件之间(A1, A2，A3)都没有差别。也就是说记忆方法对记忆效果没有影响。

用线性模型表达完全嵌套 ANOVA 的数据：

$$y_{ijk} = u + a_j + \beta_{jk} + \varepsilon_{ijk}$$

其中$y_{ijk}$是 A 自变量的第 j 个水平和 B 自变量的第 k 个水平下的第 i 个被试的分数。$u$是总体均值，$a_j$是自变量 A 第 j 个水平的效应，$\beta_{jk}$是自变量 B 第 k 个水平的效应而它是嵌套在 A 的第 j 个水平下的，$\varepsilon_{ijk}$就是 ANOVA 模型的残差，残差假设服从高斯分布。

在我们的示例中，A 有 a = 3 个水平，B 有 b = 12 个水平，每个 B 的水平下我们有 n = 5 个分数（被试）。完全嵌套 ANOVA 在这里只分析 A 变量是否显著，而混淆变量 B 的影响被量化，而且从残差中剔除。

组间效应，即独立样本因素 A 的效应对应的平方和是：

$$SS_A = \sum nb_j(\overline{y_{.J.}} - \bar{G})^2$$

其中，$\overline{y_{.J.}}$是第 j 组的分数的均值，$b_j$是第 j 组内部嵌套的组数，n 是每组被试数目，$\bar{G}$是所有分数的总体平均值。$SS_A$对应的自由度是分组变量的水平数目 a-1。可以看出来，组间效应的计算和单因素独立样本 ANOVA 没有差异。用我们的示例来说明，这里 n = 5, $b_j = 4$。

被嵌套的分组变量 B 的效应所对应的平方和是：

$$SS_B = n\sum_j\sum_k(\overline{y_{.Jk}} - \overline{Y_{.J.}})^2$$

其中，$\overline{y_{.Jk}}$是第 j 组内部嵌套的 B 的 k 个水平的平均值。可以看出$SS_B$是每一个 A 的水平下 B 的不同水平的均值的变异性的表达。它的自由度是 a(b-1)。

最后的误差项$\varepsilon_{ijk}$是两个分组变量 A 和 B 不能解释的残差，它实质上是每个条件下的独立取样之间的变异性，其平方和是：

$$SS_{within} = \sum_j\sum_k\sum_i(y_{ijk} - \overline{y_{.Jk}})^2$$

$SS_{within}$的自由度是 ab(n-1)。基于上面平方和的计算，我们可以最终获得完全嵌套 ANOVA 分析的结果，包含 A 的主效应和 B 的嵌套因素的效应：

表 9.5: 完全嵌套方差分析 ANOVA 结果表

| | SS | *df* | MS | *F* | *p* |
|---|---|---|---|---|---|
| Method | 21.23 | 2 | 10.62 | 4.90 | .0116 |
| Method*school | 179.50 | 9 | 19.94 | 9.21 | .000 |
| Error | 104.00 | 48 | 2.17 | | |

我们来思考一下完全嵌套 ANOVA 和传统独立样本 ANOVA 的差别。和单因素独立样本 ANOVA 相比，完全嵌套 ANOVA 引入了被嵌套的因素 B。和双因素独立样本 ANOVA 相比，完全嵌套 ANOVA 虽然也有 A 和 B 两个随机分组因素，但是两个因素并不交互（或者是假设不交互），而且 B 因素在 A 因素的不同水平下不重复。

在我们的示例中，如果我们忽略被嵌套的因素 B，那么结果就会完全不同。如果研究者没有意识到学校是一个影响结果的重要因素，那么原始数据就是 3 个条件而每个条件有被试 20 人。如果使用单因素独立样本 ANOVA 来分析上述的数据，我们会发现主效应不显著（$F(2,57) = 2.135$, $p = 0.128$），从而做出记忆方法没有差异的错误判断。

## 完全嵌套方差分析的事后检验

完全嵌套 ANOVA 不存在嵌套和被嵌套变量之前的交互项，也没有简单主效应的分析。因为被嵌套的变量 B 通常是混淆变量或者无关变量，因此也没有对其进行进一步事后检验的必要。例如，示例中，学校变量是显著的，但是 12 所学校之间那些两两比较显然不是研究的目标，而且在例子中的样本也不足，即使做了两两比较也无法获得可靠的结论。

但是，被嵌套变量 A 是我们的研究兴趣所在，它可以进一步做事后检验。比如，这里的三种记忆方法，联想法、谐音法和记忆宫殿法的相对差别有研究意义。完全嵌套 ANOVA 的事后检验类似于独立样本 ANOVA 的事后检验，可以使用的方法包括 Tukey, Bonferroni 校正、Scheffe 等。以 Tukey HSD 检验为例，我们发现：记忆宫殿法显著好于谐音法（$p = .011$）。

完全嵌套方差分析的范例 R 代码如下：

```
head(nestedData)
```

```
## # A tibble: 6 × 3
##   method school performance
##    <dbl>  <dbl>       <dbl>
## 1      1      1           6
## 2      1      1           9
## 3      1      1           7
## 4      1      1           3
## 5      1      1           4
## 6      1      2           2
```

```
# note nested anova can use error type I and III.
# here type I and III does not make a difference, since no orders
aov_model = aov(performance ~ factor(method)/factor(school),
                data = nestedData,summation="III")
summary(aov_model)
```

```
##                               Df Sum Sq Mean Sq F value   Pr(>F)
## factor(method)                 2  21.23  10.617   4.900   0.0116 *
## factor(method):factor(school)  9 179.50  19.944   9.205 6.15e-08 ***
## Residuals                     48 104.00   2.167
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

## 完全嵌套方差分析的结果报告

我们随机选择了 12 所中学，每所学校从初中三年级随机抽取了 5 名同学参加了记忆实验。因为学校本身可能对记忆效绩产生影响，我们把 12 所学校随机分配到了三种记忆条件中，每种记忆条件的 20 名同学记忆的隔天的记忆效果分数用完全嵌套 ANOVA 分析，其中学校嵌套在记忆方法中。ANOVA 结果表明记忆方法的主效应显著（$F(2,48) = 4.311$，$p = .019$）,而且学校的效应也显著（$F(9,48) = 43.953$，$p < .001$）。Tukey HSD 检验表明记忆宫殿法显著好于谐音法（$p = .011$），而联想法和谐音法、记忆宫殿法之间没有显著差异（$p = 0.734$; $p = 0.072$）。因此，虽然学校的确对记忆表现有比较大的贡献，记忆方法还是能显著地影响记忆表现，而记忆宫殿法在隔天记忆表现中效果最好。

# 完全嵌套方差分析的前提假设和补充说明

完全嵌套 ANOVA 中，被嵌套的因素 B 是独立取样因素，或称为随机分组因素。示例中，学校这个因素是随机取样的，而且相互独立。因此它是“完全”嵌套到记忆方法中。在这种情况下，独立样本 ANOVA 的前提假设都需要满足，包括高斯假设、独立性假设、方差同质性假设（针对嵌套因素 A）和随机取样假设（B 是随机取样）。

但是，嵌套 ANOVA 还有别的形式。本章只讲述了 A 和 B 都是独立取样因素的完全嵌套设计。从某种意义上来说，第 7 章的单因素重复测量 ANOVA 也是嵌套 ANOVA，因为重复测量因素是嵌套在被试因素里的。这种情况下，方差分析需要满足额外的球形假设，即不同的重复测量水平之前的差异必须满足方差同质性。

重复测量因素作为嵌套 ANOVA 的因素还可以有别的形式。假设我们想知道三种文化人群的智商差异，从每种文化中我们随机抽取了 30 个被试。但是，智商测量只做一次的话可能效度有限，很多情况下智商量表得分受熟悉效应的影响。如果智商量表中的题目的样式是常见的话，人的表现会更好。而在某些文化背景中，智商题目类似的问题是常见的，而在某些文化背景下，智商类的问题是罕见的。如果只有一次测量，仅仅因为熟悉效应，某些文化下的被试可能会表现得更好。为了防止熟悉效应，研究者要求每一名被试参加智商测验 3 次。在这种实验设计中，测试时间点是重复测量因素，虽然不是本研究的主要兴趣问题。重要的是，测试时间点（3 次重复测试）是嵌套在文化条件中的，我们假设两者之间并没有交互作用（即假设不同文化对重复测量的“学习”效应没有不同的影响）。这种情况下，合适的方差分析还是嵌套 ANOVA，只是这次被嵌套的因素是重复测量因素。因此，还需要额外满足球形假设。