

基于 R 语言与 SPSS

的混合模型方差分析过程及可视化效果对比分析

王超

沈阳体育学院 辽宁省 沈阳市 110102

1. 引言

1.1 研究背景

混合效应模型 (Mixed-Effects Model, MEM) 是统计学中的一种重要方法, 能够同时处理固定效应和随机效应, 从而在多种研究设计中具有广泛的应用。尤其在生物医学、教育学和心理学等领域中, 混合效应模型被广泛应用于分析具有重复测量、嵌套结构或多层次的数据。通过结合固定效应和随机效应的分析, 混合效应模型提供了比传统的方差分析 (ANOVA) 更为精确的估计, 能够有效应对复杂的统计模型需求。

方差分析 (ANOVA) 是一种用于比较不同组别间差异显著性的经典统计方法。然而, 在涉及随机效应的情形下, 传统的 ANOVA 可能会失去准确性。混合效应模型通过考虑数据中的随机变异, 能够在复杂的设计情境下提供更为稳定和准确的分析结果。

R 语言作为一种开源的统计分析工具, 以其强大的功能、灵活性和丰富的扩展包在数据分析领域占据主导地位。通过 R 的 lme4、nlme 等包, 可以构建和分析混合效应模型。而 SPSS 作为一款图形化统计分析软件, 其简便易用的界面使得非统计专业人员可以高效地完成统计分析, 但其灵活性和深度定制能力相比 R 语言有所欠缺。近年来, 随着数据分析需求的多样化, 选择 R 语言还是 SPSS 进行混合模型分析成为了一个重要课题, 本文的研究旨在对这两种工具的使用过程、可视化效果以及性能进行对比分析。

1.2 研究目的

本研究旨在对比 R 语言与 SPSS 在混合效应模型方差分析中的操作流程与可视化效果, 评估两者在数据处理、模型构建、分析深度和可视化展示方面的异同。具体而言, 本研究将:

- 比较 R 语言与 SPSS 在混合效应模型方差分析中的操作流程, 分析数据导入、预处理、模型构建与检验的差异;
- 比较两者可视化呈现的效果;
- 对比两者在用户友好性、灵活性和分析深度方面的差异。

2. 研究方法

2.1 数据集

ToothGrowth 源于 R 语言内置数据集, 包含了不同剂量的维生素 C 对小白鼠牙齿生长的影响。数据集中包括 3 个变量 (表 1)。

表 1 ToothGrowth 数据集说明

变量	说明	数据类型		值 (值标签)
		R 语言	SPSS	
supp	组间变量	factor	名义	1-VC; 2-OJ
dose	组内变量	factor	有序	0.5; 1.0; 2.0
len	因变量	num	度量	

2.2 R 语言分析过程

2.2.1 数据描述 (R Markdown 格式脚本)

```
```{r}
data(ToothGrowth)
library(dplyr)
ToothGrowth$dose <- factor(ToothGrowth$dose)
stats <- ToothGrowth %>%
 group_by(supp, dose) %>%
 summarise(n=n(), mean=mean(len), sd=sd(len),
 ci = qt(0.975, df = n - 1) * sd / sqrt(n))

stats
```
```

2.2.2 混合模型方差分析 (脚本为 R Markdown 格式)

(1) 正态性检验及 QQ 图绘制

```
```{r}
fit <- aov(len ~ supp * dose, data = ToothGrowth)
shapiro.test(fit$residuals)
```
```

(2) 方差齐性检验

```
```{r}
library(car)
leveneTest(len ~ supp, data = ToothGrowth)
leveneTest(len ~ dose, data = ToothGrowth)
leveneTest(len ~ supp * dose, data = ToothGrowth)
```
```

(3) 方差分析

```
```{r}
summary(fit)
```
```

(4) 简单效应分析

```
```{r}
library(emmeans)
emmeans(aov_model, pairwise ~ dose | supp)
emmeans(aov_model, pairwise ~ supp | dose)
```
```

(5) 可视化效果

```
```{r}
library(ggplot2)
library(ggpubr)
pd <- position_dodge(0.2)
p1 <- ggplot(stats, aes(x = dose, y = mean, group=supp, color=supp, linetype=supp)) +
 geom_point(size = 2, position=pd) + geom_line(position=pd) +
 geom_errorbar(aes(ymin = mean - ci, ymax = mean + ci),
 width = .1, position=pd) + theme_classic() +
 theme(legend.position = "top") + theme(legend.direction = "horizontal") +
```
```

```

scale_color_manual(values=c("#e87d72", "#53b74c")) + labs(x = NULL, y=NULL) +
scale_y_continuous(breaks = seq(5,30,5),limits = c(5,30))
p2 <- ggplot(stats,
aes(x = supp, y = mean, group=dose,color=dose,linetype=dose)) +
geom_point(size = 2,position=pd) + geom_line(position=pd) +
geom_errorbar(aes(ymin = mean - ci, ymax = mean + ci),
width = .1,position=pd) +
theme_classic() + theme(legend.position = "top") +
theme(legend.direction = "horizontal") +
scale_color_manual(values=c("#e87d72", "#53b74c", "#6c9df8")) +
labs(x = NULL, y=NULL) + scale_y_continuous(breaks = seq(5,30,5),limits = c(5,30))
p3 <- ggplot(ToothGrowth) +
aes(x = supp, y = len, fill = supp) + geom_boxplot() +
scale_fill_hue(direction = 1) + theme_classic() +
theme(legend.position = "top") + theme(legend.direction = "horizontal") +
labs(x = NULL, y = NULL) + scale_y_continuous(breaks = seq(5,30,5),limits = c(5,30))
p4 <- ggplot(ToothGrowth) +
aes(x = dose, y = len, fill = dose) + geom_boxplot() +
scale_fill_hue(direction = 1) + theme_classic() +
theme(legend.position = "top") + theme(legend.direction = "horizontal") +
labs(x = NULL,y = NULL) + scale_y_continuous(breaks = seq(5,30,5),limits = c(5,30))
ggarrange(p2,p4,p3,p1,nrow = 2,ncol = 2)
...

```

2.3 SPSS 分析过程

2.3.1 数据描述 (SPSS Syntax 格式脚本)

* 定制表.

CTABLES

```

/VLABELS VARIABLES=supp len dose DISPLAY=LABEL
/TABLE supp > len [COUNT F40.0, MEAN, STDDEV] BY dose
/CATEGORIES VARIABLES=supp ORDER=A KEY=VALUE EMPTY=INCLUDE
/CATEGORIES VARIABLES=dose ORDER=A KEY=VALUE EMPTY=EXCLUDE
/CRITERIA CILEVEL=95.

```

2.2.2 混合模型方差分析 (SPSS Syntax 格式脚本)

UNIANOVA len BY supp dose

/METHOD=SSTYPE(3)

/INTERCEPT=INCLUDE

/SAVE=RESID

/PLOT=PROFILE(supp*dose dose*supp) TYPE=LINE ERRORBAR=CI

MEANREFERENCE=NO YAXIS=AUTO

/EMMEANS=TABLES(dose*supp) COMPARE(dose) ADJ(BONFERRONI)

/EMMEANS=TABLES(dose*supp) COMPARE(supp) ADJ(BONFERRONI)

/PRINT HOMOGENEITY

/CRITERIA=ALPHA(.05)

/DESIGN=supp dose dose*supp.

```
EXAMINE VARIABLES=RES_1
/PLOT BOXPLOT NPLOT
/COMPARE GROUPS
/STATISTICS DESCRIPTIVES
/CINTERVAL 95
/MISSING LISTWISE
/NOTOTAL.
```

3. 结果与分析

3.1 R 语言分析结果

3.1.1 数据描述

表 2 数据描述（R 语言）输出结果

| supp | dose | n | mean | sd | ci |
|--------|--------|-------|-------|-------|-------|
| <fctr> | <fctr> | <int> | <dbl> | <dbl> | <dbl> |
| OJ | 0.5 | 10 | 13.23 | 4.46 | 3.19 |
| OJ | 1 | 10 | 22.70 | 3.91 | 2.80 |
| OJ | 2 | 10 | 26.06 | 2.66 | 1.90 |
| VC | 0.5 | 10 | 7.98 | 2.75 | 1.96 |
| VC | 1 | 10 | 16.77 | 2.52 | 1.80 |
| VC | 2 | 10 | 26.14 | 4.80 | 3.43 |

3.1.2 混合模型方差分析

(1) 正态性检验

表 3 模型残差正态性检验（R 语言）输出结果

| w | | p |
|---|-------|-------|
| Model fit <- aov(len ~ supp * dose, data = ToothGrowth) | | |
| fit\$residuals | 0.985 | 0.669 |

(2) 方差齐性检验

表 4 模型方差齐性检验（R 语言）输出结果

| Data | df | F | p |
|--------------------|----|--------|--------|
| len by supp | 1 | 1.2136 | 0.2752 |
| len by dose | 2 | 0.6457 | 0.5281 |
| len by supp * dose | 5 | 1.7086 | 0.1484 |

(3) 方差分析

表 5 模型方差分析（R 语言）输出结果

| | df | Sum Sq | Mean Sq | F | p | Sig. |
|-----------|----|--------|---------|--------|--------|------|
| supp | 1 | 205.4 | 205.4 | 15.572 | <0.001 | *** |
| dose | 2 | 2426.4 | 1213.2 | 92.000 | <0.001 | *** |
| supp:dose | 2 | 108.3 | 54.2 | 4.107 | 0.0219 | * |
| Residuals | 54 | 712.1 | 13.2 | | | |

Note:***:0.001;**0.01*;0.05.

(4)简单效应分析

表 6 模型简单效应分析（R 语言）输出结果

| Contrasts | Estimate | SE | df | t.ratio | p |
|-----------------|----------|------|----|---------|--------|
| supp = OJ | | | | | |
| dose0.5 - dose1 | -9.47 | 1.62 | 54 | -5.831 | <0.001 |
| dose0.5 - dose2 | -12.83 | 1.62 | 54 | -7.9 | <0.001 |
| dose1 - dose2 | -3.36 | 1.62 | 54 | -2.069 | 0.1060 |
| supp = VC | | | | | |
| dose0.5 - dose1 | -8.79 | 1.62 | 54 | -5.413 | <0.001 |
| dose0.5 - dose2 | -18.16 | 1.62 | 54 | -11.182 | <0.001 |
| dose1 - dose2 | -9.37 | 1.62 | 54 | -5.77 | <0.001 |
| dose = 0.5 | | | | | |
| OJ - VC | 5.25 | 1.62 | 54 | 3.233 | 0.0021 |
| dose = 1.0 | | | | | |
| OJ - VC | 5.93 | 1.62 | 54 | 3.651 | <0.001 |
| dose = 2.0 | | | | | |
| OJ - VC | -0.08 | 1.62 | 54 | -0.049 | 0.9609 |

(5) 可视化效果

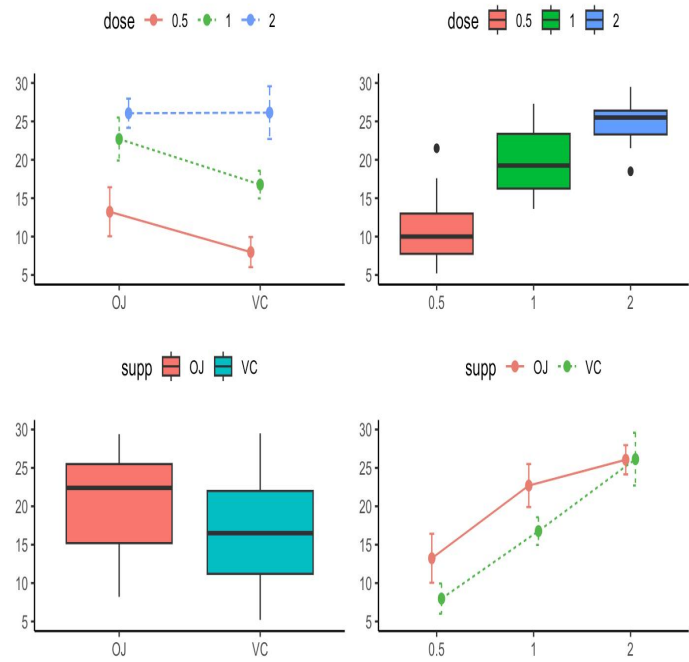


图 1 混合模型方差分析结果组合图（R 语言）

3.2 SPSS 分析结果

3.2.1 数据描述

表 7 数据描述（SPSS）输出结果

| | dose=0.5 | | | dose=1 | | | dose=2 | | |
|----|----------|-------|------|--------|-------|------|--------|-------|------|
| | N | M | SD | N | M | SD | N | M | SD |
| VC | 10 | 8.00 | 2.70 | 10 | 16.80 | 2.50 | 10 | 26.10 | 4.80 |
| OJ | 10 | 13.20 | 4.50 | 10 | 22.70 | 3.90 | 10 | 26.10 | 2.70 |

3.2.2 混合模型方差分析

(1) 正态性检验

表 8 模型残差正态性检验 (SPSS) 输出结果

| | 柯尔莫戈洛夫-斯米诺夫 | | | 夏皮洛-威尔克 | | |
|---------|-------------|-----|-------|---------|-----|-------|
| | 统计 | 自由度 | 显著性 | 统计 | 自由度 | 显著性 |
| Len 的残差 | 0.068 | 60 | 0.200 | 0.985 | 60 | 0.669 |

(2) 方差齐性检验

表 9 模型方差齐性检验 (SPSS) 输出结果

| | 莱文统计 | df1 | df2 | p |
|-----|-------|-----|-----|-------|
| 平均数 | 1.940 | 5 | 54 | 0.103 |
| 中位数 | 1.709 | 5 | 54 | 0.148 |

注：设计：截距 + supp + dos + supp * dose

(3) 方差分析

表 10 模型方差分析 (SPSS) 输出结果

| 变异来源 | III 类平方和 | 自由度 | 均方 | F | 显著性 |
|-----------|----------|-----|----------|--------|--------|
| supp | 205.350 | 1 | 205.350 | 15.572 | <0.001 |
| dose | 2426.434 | 2 | 1213.217 | 92.000 | <0.001 |
| supp*dose | 108.319 | 2 | 54.159 | 4.107 | 0.022 |
| Residuals | 712.106 | 54 | 13.187 | | |
| 合计 | 3452.209 | 59 | | | |

(4)简单效应分析

表 11 模型简单效应分析 (SPSS) 输出结果

| | 平均值差值 | SE | p |
|-----------------|---------|-------|--------|
| supp = OJ | | | |
| dose0.5 - dose1 | -9.470 | 1.624 | <0.001 |
| dose0.5 - dose2 | -12.830 | 1.624 | <0.001 |
| dose1 - dose2 | -3.360 | 1.624 | 0.130 |
| supp = VC | | | |
| dose0.5 - dose1 | -8.790 | 1.624 | <0.001 |
| dose0.5 - dose2 | -18.160 | 1.624 | <0.001 |
| dose1 - dose2 | -9.370 | 1.624 | <0.001 |
| dose = 0.5 | | | |
| VC - OJ | -5.250 | 1.624 | 0.002 |
| dose = 1.0 | | | |
| VC - OJ | -5.930 | 1.624 | <0.001 |
| dose = 2.0 | | | |
| VC - OJ | -0.080 | 1.624 | 0.961 |

(5) 可视化效果

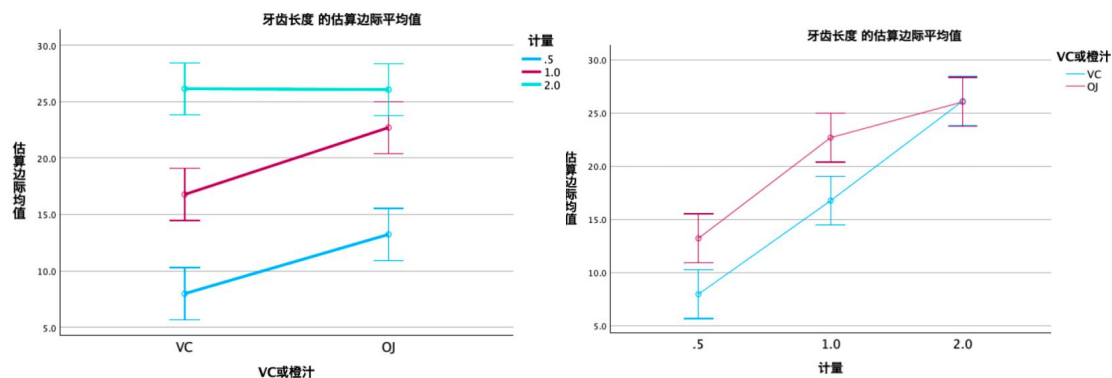


图 2 混合模型方差分析结果图 (SPSS)

4. 讨论

4.1 操作过程对比分析

R 语言在分析过程中提供了更多的灵活性和可扩展性，适合复杂的分析需求。而 SPSS 在界面友好性和操作简便性上具有优势，适合初学者或需要快速得出结果的情况。

4.2 分析深度对比分析

R 语言可以轻松结合各种扩展包进行进一步分析，适合复杂的统计模型和多层次分析。SPSS 适合常规的方差分析和模型构建，但在处理复杂数据结构时稍显局限。

4.3 可视化效果对比分析

R 语言提供了更强的可视化定制性，图形美观且适合高质量展示。SPSS 的可视化虽然清晰，但缺乏高度的定制和复杂性。

5. 结论

本研究通过对比 R 语言和 SPSS 在混合模型方差分析中的应用，发现 R 语言在灵活性、定制性和可扩展性方面具有明显优势，但随着技术的不断发展，R 语言和 SPSS 可能会进一步融合其优点，R 语言的易用性可能会有所提升，而 SPSS 也可能会增强其灵活性和定制化功能。因此，了解和掌握两者的优势，选择合适的工具，将有助于提高混合模型方差分析效率和结果的准确性。

参考文献：

- [1]Robert I. Kabacoff.R 语言实践 (第 3 版) [M].王韬, 译. 北京: 中国邮电出版社, 2023.
- [2]Pinheiro J C, Bates D M. 混合效应模型的应用[M].纽约: Springer, 2000.