

使用rstatix包进行独立样本方差分析

Alex / 2024-04-10 / free_learner@163.com / AlexBrain.cn

本文介绍在R环境下使用rstatix包进行独立样本方差分析（ANOVA）的基本方法。

一、背景

在前面的博客中，我总结了在R中进行独立样本单因素方差分析和双因素方差分析的基本方法，这里使用 rstatix 包重复这些分析。使用 rstatix 包的好处是，对于各种不同类型的ANOVA可以使用同样的调用方式，更简洁更方便。

二、独立样本单因素方差分析

由于在前面的博客中已经介绍过测试数据集的情况，这里不再重复。

```
> ## Check data
> data("PlantGrowth")
> str(PlantGrowth)
'data.frame':   30 obs. of  2 variables:
 $ weight: num  4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...
 $ group : Factor w/ 3 levels "ctrl","trt1",...: 1 1 1 1 1 1 1 1 1 1 ...

> ## ANOVA test
> lm_mod <- lm(weight ~ group, data = PlantGrowth)
> anova_test(lm_mod)
ANOVA Table (type II tests)

  Effect DFn DFd    F    p p<.05  ges
1  group    2   27 4.846 0.016    * 0.264

> ## Post-hoc tests
> emmeans_test(PlantGrowth, weight ~ group, p.adjust.method = 'fdr')
# A tibble: 3 × 9
  term  .y.  group1 group2  df statistic    p  p.adj p.adj.signif
* <chr> <chr> <chr> <chr> <dbl> <dbl> <dbl> <dbl> <chr>
1 group weight ctrl  trt1    27    1.33 0.194  0.194  ns
2 group weight ctrl  trt2    27   -1.77 0.0877 0.132  ns
3 group weight trt1  trt2    27   -3.10 0.00446 0.0134 *
```

在 rstatix 包中， anova_test 用于进行ANOVA检验，除了上面代码展示的调用方式，还可以使用下面的方式实现同样的目的：

```
anova_test(PlantGrowth, weight ~ group)
anova_test(PlantGrowth, dv=weight, between=group)
```

`emmeans_test` 用于事后检验，这个函数是对 `emmeans` 包的二次封装，由于我并不喜欢用管道（`%>%`）操作，所以我觉得直接用 `emmeans` 包会更方便。

三、独立样本双因素方差分析

```
> ## Check data
> data("ToothGrowth")
> ToothGrowth$dose <- as.factor(ToothGrowth$dose)
> str(ToothGrowth)
'data.frame': 60 obs. of 3 variables:
 $ len : num 4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
 $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 ...
 $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 ...
> ## Make data unbalanced
> ToothGrowth <- ToothGrowth[-c(1,42),]
> table(ToothGrowth$supp, ToothGrowth$dose)

      0.5  1  2
OJ    10  9 10
VC     9 10 10

> ## ANOVA test
> lm_mod <- lm(len ~ supp + dose + supp:dose, data = ToothGrowth)
> anova_test(lm_mod)
ANOVA Table (type II tests)

      Effect DFn DFd      F      p p<.05 ges
1      supp   1  52 13.039 6.87e-04 * 0.200
2      dose   2  52 85.059 4.03e-17 * 0.766
3 supp:dose   2  52  3.696 3.20e-02 * 0.124
```

```
> ## Simple effect analysis if significant interaction effect
> ToothGrowth %>% group_by(dose) %>% emmeans_test(len ~ supp, p.adjust.method = "fdr", model =
lm_mod)
```

```
# A tibble: 3 × 10
```

	dose	term	.y.	group1	group2	df	statistic	p	p.adj	p.adj.signif
*	<fct>	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
1	0.5	supp	len	OJ	VC	52	2.87	0.00586	0.00586	**
2	1	supp	len	OJ	VC	52	3.49	0.000998	0.000998	***
3	2	supp	len	OJ	VC	52	-0.0489	0.961	0.961	ns

```
> ## Post-hoc tests if non-significant interaction effect
```

```
> emmeans_test(ToothGrowth, len ~ dose, p.adjust.method = "fdr", model = lm_mod)
```

NOTE: Results may be misleading due to involvement in interactions

```
# A tibble: 3 × 9
```

	term	.y.	group1	group2	df	statistic	p	p.adj	p.adj.signif
*	<chr>	<chr>	<chr>	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>
1	dose	len	0.5	1	52	-7.48	8.59e-10	1.29e-9	****
2	dose	len	0.5	2	52	-13.0	5.18e-18	1.56e-17	****
3	dose	len	1	2	52	-5.46	1.37e-6	1.37e-6	****