

第8章 方差分析

1. 请简述方差分析和回归分析的差别。

【参考答案】方差分析主要研究预测变量对响应变量(试验结果)影响程度的定性关系,即响应变量均值差别的显著性检验,从而剔除对结果影响较小的变量,提高试验的效率和精度。

回归分析是研究预测变量对响应变量(试验结果)的定量关系,得出相应的数学模型。在回归分析中,需要将各变量对结果的影响进行方差分析,以剔除影响不大的变量,提高回归分析的有效性。

2. 请以年级和性别作为单因素,采用单因素方差法分析其对其他体质健康数据的影响,并对结果进行解释。

【参考答案】以年级 grade 为单因素,判断其对身高是否有显著影响。注意,本章示例数据集 g56 中, grade 和 gender 均已转换为 factor。

(1) 先查看 g56 数据集的基本信息。

```
> str(g56)
'data.frame': 120 obs. of 12 variables:
 $ snum : chr "y01" "y02" "y03" "y04" ...
 $ grade : Factor w/ 2 levels "5","6": 2 2 2 2 2 2 2 2 2 2 ...
 $ gender: Factor w/ 2 levels "1","2": 1 1 1 1 1 1 1 1 1 1 ...
 $ height: int 171 145 149 155 140 148 153 145 152 171 ...
 $ weight: num 72 29 45 55 35 41 34 40 43 54 ...
 $ m50 : num 8.9 8.6 9.4 8.3 8.3 8.7 8.4 8.2 8.5 7.6 ...
 $ sitFB : int 7 -4 1 4 9 -3 6 2 8 4 ...
 $ jp : int 53 77 73 50 70 26 109 56 137 37 ...
 $ sitUPs: int 28 24 27 25 45 28 25 36 31 48 ...
 $ lungCp: int 4380 1458 2660 1468 2581 2601 2274 1775 2710 3030 ...
 $ m50x8 : num 2.07 2.12 2.05 2.02 2.15 2.08 1.88 1.77 1.85 1.48 ...
```

(2) 以 grade 作为预测变量, height 作为响应变量进行方差分析。

```
> height.aov2 <- aov(height~grade, data = g56)
> summary(height.aov2)
          Df Sum Sq Mean Sq F value    Pr(>F)
grade      1   2001  2000.8    33.09 7.06e-08 ***
Residuals 118   7135    60.5
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

上述分析结果中 Pr 值为 7.06e-08, 远小于 0.05, 表明否定原假设, 换言之, 年级对身高有显著影响。

(3) 对其他体质指标数据分析方法与上述步骤类似。

3. 请以年级和性别作为预测变量, 采用双因素方差法分析其对其他体质健康数据的影响, 并对结果进行解释。

【参考答案】(1) 这里我们将以数据集集中的 sitFB 变量(坐位体前屈, 单位厘米)为响应变量, grade 和 gender 为预测变量, 进行双因素方差分析。注意, 本章示例数据集 g56 中, grade 和 gender 均已转换为 factor。代码如下:

```
> sitFB.both.aov <- aov(sitFB ~ gender + grade ,data = g56)
> summary(sitFB.both.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender	1	1415	1414.5	40.693	3.7e-09 ***
grade	1	10	9.6	0.277	0.6
Residuals	117	4067	34.8		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

结果显示 gender(性别)对 sitFB 有显著影响(Pr 值远小于0.05), 而 grade 对 sitFB 则无显著影响。该分析结论说明小学 5 至 6 年级, 性别对坐位体前屈体质指标有显著影响, 同时这份观测数据尚不能充分地说明 grade(年级)会显著影响坐位体前屈成绩。

(2)分析年级和性别双因素对其他体质指标影响与上述步骤类似。

4. 在第 3 题基础上加入交互效应, 对比结果的差异, 并对结果进行解释。

【参考答案】(1) 本题将进一步分析 grade 和 gender 双因素对 sitFB(坐位体前屈)体质指标的影响是否存在交互效应。代码如下:

```
> sitFB.both.aov2 <- aov(sitFB ~ gender * grade ,data = g56)
> summary(sitFB.both.aov2)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender	1	1415	1414.5	41.334	2.97e-09 ***
grade	1	10	9.6	0.281	0.5967
gender:grade	1	97	97.2	2.840	0.0946 .
Residuals	116	3970	34.2		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

在上述分析结果中, gender 和 grade 交互效应对应项的 Pr 值为 0.0946, 这表明在这个年龄段, 年龄和性别的交互效应对 sitFB(坐位体前屈)指标一定的影响, 但不是很显著。

(2)再选取 m50(50 米跑)指标作为响应变量, 以 grade 和 gender 预测变量进行双因素分析, 并考察是否存在交互效应。代码如下:

```
> m50.both.aov <- aov(m50~gender * grade, data = g56)
> summary(m50.both.aov)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
gender	1	18.49	18.487	25.694	1.53e-06 ***
grade	1	0.09	0.093	0.129	0.7199
gender:grade	1	4.28	4.279	5.947	0.0163 *
Residuals	116	83.46	0.720		

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

在上述分析结果中, gender 和 grade 交互效应对应项的 Pr 值为 0.0163, 这表明在这个年龄段, 年龄和性别的交互效应对 m50(50 米跑)指标显著影响。这也从另一个角度表明在这个年龄段, 随着年龄的增长, 性别不同导致体质健康数据的差异会逐渐显现。

(3)分析年级和性别双因素对其他体质指标影响与上述步骤类似。