方差分析

1. 单因素方差分析

方差分析中的假设检验：

假设有k个样本，原假设H0：样本均数都相等，H1：样本均数不全相等。

如果经过计算，p<0.05.推翻原假设，说明样本来自不同的正态总体，说明处理造成均值的差异有统计学意义。否则，承认原假设，样本来自相同总体，处理间无差异。

R中的anova()函数用来做方差分析，使用方法为：

例：利用四种不同配方的材料A1、A2、A3、A4生产出来的元件，测得寿命如下：



问：四种不同配方下元件的使用寿命有无显著性差异。

X A

>lamp<-data.frame(

X=c(1600, 1610, 1650, 1680, 1700, 1700, 1780, 1500, 1640,

1400, 1700, 1750, 1640, 1550, 1600, 1620, 1640, 1600,

1740, 1800, 1510, 1520, 1530, 1570, 1640, 1600),

A= c(rep(1,7),rep(2,5), rep(3,8), rep(4,6))

)

> lamp

1 1600 1

2 1610 1

3 1650 1

4 1680 1

5 1700 1

6 1700 1

7 1780 1

8 1500 2

9 1640 2

10 1400 2

11 1700 2

12 1750 2

13 1640 3

14 1550 3

15 1600 3

16 1620 3

17 1640 3

18 1600 3

19 1740 3

20 1800 3

21 1510 4

22 1520 4

23 1530 4

24 1570 4

25 1640 4

26 1600 4

> anova(lm(X~factor(A),data=lamp))

Analysis of Variance Table

Response: X

Df Sum Sq Mean Sq F value Pr(>F)

factor(A) 3 49212 16404.1 2.1659 0.1208

Residuals 22 166622 7573.7

df表示自由度，Sum Sq是平方和，Mean Sq表示均方，F value为F值，Pr(>F)为p值。

Factor(A)是因素，Residuals是残差即误差。

从p值来看，0.121>0.05，没有充分理由说明H0不正确，接受原假设，说明四种材料生产出的元件的平均寿命无显著性差异。

1. 双因素方差分析

例1：使用4种燃料、3种推进器做火箭射程试验，每一种组合情况做一次试验，则得火箭射程列在表中，试分析各种燃料A与各种推进器B对火箭射程有无显著影响。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | A1 | A2 | A3 | A4 |
| B1 | 582 | 491 | 601 | 758 |
| B2 | 562 | 541 | 709 | 582 |
| B3 | 653 | 516 | 392 | 487 |

用anova（）函数进行双因素方差分析：

>rocket<-data.frame(

Y=c(582, 491, 601, 758, 562, 541,

709, 582, 653, 516, 392, 487),

A=c(1,2,3,4,1,2,3,4,1,2,3,4),

B=c(1,1,1,1,2,2,2,2,3,3,3,3)

)

> rocket

Y A B

1 582 1 1

2 491 2 1

3 601 3 1

4 758 4 1

5 562 1 2

6 541 2 2

7 709 3 2

8 582 4 2

9 653 1 3

10 516 2 3

11 392 3 3

12 487 4 3

> anova(lm(Y~factor(A)+factor(B), data=rocket))

Analysis of Variance Table

Response: Y

Df Sum Sq Mean Sq F value Pr(>F)

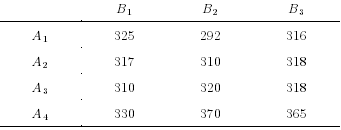
factor(A) 3 15759 5253 0.4306 0.7387

factor(B) 2 22385 11192 0.9174 0.4491

Residuals 6 73198 12200

由此可见，A和B的p值均大于0.05，说明各种燃料A对火箭射程无显著影响，各种推进器B对火箭射程也无显著影响。

例2：在一个农业试验中，考虑四种不同的种子品种A1、A2、A3、A4和三种不同的施肥方法B1、B2、B3，得到产出数据如下：



试分析种子与施肥对产出有无显著影响。

>agriculture<-data.frame(

Y=c(325, 292, 316, 317, 310, 318,

310, 320, 318, 330, 370, 365),

A=c(1,1,1,2,2,2,3,3,3,4,4,4),

B=c(1,2,3,1,2,3,1,2,3,1,2,3)

)

> agriculture

Y A B

1 325 1 1

2 292 1 2

3 316 1 3

4 317 2 1

5 310 2 2

6 318 2 3

7 310 3 1

8 320 3 2

9 318 3 3

10 330 4 1

11 370 4 2

12 365 4 3

> anova(lm(Y~factor(A)+factor(B),data=agriculture))

Analysis of Variance Table

Response: Y

Df Sum Sq Mean Sq F value Pr(>F)

factor(A) 3 3824.2 1274.75 5.2262 0.04126 \*

factor(B) 2 162.5 81.25 0.3331 0.72915

Residuals 6 1463.5 243.92

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

从p值来看，A因素（不同品种）对产量有显著影响，而没有充分理由说明B因素（施肥方法）对产量有显著影响。