

生物实验数据的两因素方差分析*

刘加妹^① 彭景樾^{②**}

(^①中国农业大学理学院 北京 100083; ^②中国科学院动物研究所生殖生物学国家重点实验室 北京 100080)

摘要: 结合实例详细地叙述了生物实验数据的两因素方差分析计算方法,具有较强的实用意义。并介绍了如何分析计算结果。

关键词: 两因素方差分析;实验数据

中图分类号: Q-332 **文献标识码:** A **文章编号:** 0250-3263(2003)02-51-05

Two-factor Analysis of Variance on Biological Experimental Data

LIU Jia-Mei^① PENG Jing-Pian^②

(^① *Mathematics Division, China Agricultural University, Beijing 100083;*

^② *Institute of Zoology, Chinese Academy of Sciences, Beijing 100080, China)*

Abstract: This paper described how to use Two-Factor analysis of variance on biological experimental data, including computational formulae, examples and how to analyze the computational results.

Key words: Two-factors analysis of variance; Experimental data

在“生物实验数据的单因素方差分析”一文中^[1],作者介绍了实验数据的单因素方差分析及如何推断实验数据在组与组之间是否存在显著性差异。在许多生物学实验中影响实验的因素往往不是单一的,有时可能是两个或多个因素,这样的实验数据就不宜用单因素方差分析,而应采用两因素(或多因素)方差分析来处理实验结果。本文将结合实验实例重点介绍实验数据的两因素方差分析(two-factors analysis of variance)。

1 实验设计

在上文中^[1],作者介绍了研究鸟的鸣叫与内分泌关系的实验,实验目的是要了解鸟的鸣叫对雌鸟生殖激素 LH 的分泌是否产生影响,该实验选择性固定了鸟的年龄,而声音刺激因素有 4 个水平,实验数据经单因素方差分析处理,结果表明雌鸟的 Female nest coo 对 LH 的分泌有显著影响。接下来令人感兴趣的问题是,如果雌鸟的年龄不同,声音刺激亦不同,这两种因素是如何影响雌鸟 LH 分泌的呢?为此,设计实验如下^[2]。

以雌鸟 Ring dove (*Streptopelia risoria*) 为实验对象,随

机选取年龄为 130~150 d 的雌鸟 16 只,随机分为 4 组,每组均接受声音刺激分别为对照(不施加任何声音刺激);施以 Female nest coo(雌鸟的一种鸣叫,以下简称为 FNC)声音刺激;施用 Male nest coo 声音刺激(雄鸟的一种鸣叫,以下简称为 MNC);施以 White noise(WN)声音刺激(由计算机模拟合成的一种 0~11 kHz 的声音)。声音刺激强度均为(75±2)dB SPL,刺激时间亦相同。250~280 d、360~390 d 的雌鸟各 16 只,分组与声音刺激的安排同上,详细的实验描述见 Cheng 和 Peng^[2]等报道。根据第一次实验结果分析可知^[1],鸟在接受声音刺激后 40 min 左右 LH 分泌水平变化最为明显,所以本次实验每只鸟接受声音刺激后(40±2)min,分别收集垂体门脉血样,测定血样中 LH 含量,整理得其数据见表 2。在本实验中,声音刺激是一个因素,雌鸟的年龄则是另一个因素,这种实验所得到的数据是两因素的,称

* 中国科学院知识创新工程领域前沿项目资助(KSCX3-10Z-07);

** 通讯作者;

第一作者介绍 刘加妹,女,副教授;研究方向:生物数学。

收稿日期:2002-05-20,修回日期:2002-11-10

为两因素多组群实验。在两因素实验中,由于因素水平的改变造成因素效应的改变称为该因素的主效应。在上述实验中,鸟年龄因素水平的改变造成因素效应的改变则称为年龄因素的主效应;声音刺激因素水平的变化而引起因素效应的改变称为声音因素的主效应。有时某一因素在另一因素的不同水平上产生的效应不同,即两因素间存在交互作用,例如声音刺激因素在年龄因素的不同水平上产生的效应是否有差异。当两因素之间存在交互作用时,了解两因素之间的交互作用比仅了解某一因素的主效应更有意义。实验数据只有经过严格的统计分析之后,才能判断两因素之间是否存在交互作用。

2 整理数据

在两因素实验中,设定 A 因素有 a 水平, B 因素有 b 水平,那么每一次重复就有 $a \times b$ 次实验,设实验的重复次数为 n 次,则实验总数据的数目为 $a \times b \times n$,用 X_{ijk} 表示 A 因素第 i 水平, B 因素第 j 水平在第 k 次重复时所得的实验数据,经排列形成下列的数据表 1,实例的数据排列见表 2。如本实验中 $X_{124} = 34.6$ 表示 A 因素为 130~150 d, B 因素为 FNC 的第 4 次重复实验的观测值。 $X_{232} = 21.8$ 表示 A 因素为 250~280 d, B 因素为 MNC 的第 2 次重复实验的观测值(表 1)。

3 根据公式计算检验统计量

表 1 提供了数据整理形式和部分计算公式,表 2 提供了实验数据和具体计算过程。

在表 2 中声音刺激因素 B 有 4 个水平($b = 4$),分别为对照、FNC、MNC 和 WN;年龄因素 A 有 3 个水平($a = 3$),即鸟龄分别为 130~150 d、250~280 d、360~390 d。每一次重复都包括 $a \times b = 3 \times 4$ 次实验。实验重复 4 次,共获得了 $a \times b \times n = 3 \times 4 \times 4 = 48$ 个实验数据。

表 3 给出方差分析总结表,包括平方和(SS)、自由度(df)、均值平方(MS)和 F 比率。表 4 是方差分析实例总结表。

4 结果解释

根据表 4,年龄因素 A ,当分子自由度为 $a - 1 = 2$,分母自由度为 $a \times b(n - 1) = 36$ 时,查 F 分布表(一般统计类书均提供 F 表^[3])得标准值 $F_{2,36,0.05} = 3.32$, $F_{2,36,0.01} = 5.39$, $F > F_{0.05}$, $F < F_{0.01}$, 计算结果表明,声音刺激引起雌鸟生殖激素 LH 分泌由于雌鸟年龄的不

同而有差异,且在 $\alpha = 0.05$ 水平上有显著差异(表 4 中“*”表示);声音因素 B ,当分子自由度为 $a - 1 = 3$,分母自由度为 $a \times b(n - 1) = 36$ 时, $F_{3,36,0.01} = 4.51$, $F >> F_{0.01}$,表明不同的声音对雌鸟 LH 分泌量的影响在 $\alpha = 0.01$ 水平上有极显著差异,用“**”表示; AB 交互作用,当分子自由度为 $(3 - 1)(4 - 1) = 6$,分母自由度为 $a \times b(n - 1) = 36$ 时, $F_{6,36,0.01} = 3.47$, $F > F_{0.01}$,表明在 A 和 B 两因素之间存在交互作用,交互作用在 $\alpha = 0.01$ 水平上显著。从生物学意义解释上述结果:雌鸟生殖激素 LH 分泌量的变化不仅与 Nest coo 声音刺激及雌鸟自身的年龄有关,而且与两者的交互作用也有关。

表 1 两因素方差分析的数据排列格式及相关计算公式^[3,4]

因素 A $i = 1, 2, 3, \dots, a$	因素 B $j = 1, 2, 3, \dots, b$				总计
	B_1	B_2	...	B_b	
A_1	X_{111}	X_{121}	...	X_{1b1}	$X_{1..}$
	X_{112}	X_{122}	...	X_{1b2}	
	
	X_{11n}	X_{12n}	...	X_{1bn}	
A_2	X_{211}	X_{221}	...	X_{2b1}	$X_{2..}$
	X_{212}	X_{222}	...	X_{2b2}	
	
	X_{21n}	X_{22n}	...	X_{2bn}	
⋮	⋮	⋮	⋮	⋮	⋮
A_a	X_{a11}	X_{a21}	...	X_{ab1}	$X_{a..}$
	X_{a12}	X_{a22}	...	X_{ab2}	
	
	X_{a1n}	X_{a2n}	...	X_{abn}	
总计	$X_{.1}$	$X_{.2}$...	$X_{.b}$	$X_{...}$

$$X_{i..} = \sum_{j=1}^b \sum_{k=1}^n X_{ijk}, \quad \bar{X}_{i..} = \frac{X_{i..}}{bn}, \quad i = 1, 2, 3, \dots, a$$

$$X_{.j.} = \sum_{i=1}^a \sum_{k=1}^n X_{ijk}, \quad \bar{X}_{.j.} = \frac{X_{.j.}}{an}, \quad j = 1, 2, 3, \dots, b$$

$$X_{ij.} = \sum_{k=1}^n X_{ijk}, \quad \bar{X}_{ij.} = \frac{X_{ij.}}{n}, \quad i = 1, 2, 3, \dots, a; j = 1, 2, 3, \dots, b$$

$$X_{...} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}, \quad \bar{X}_{...} = \frac{X_{...}}{abn}$$

$X_{i..}$ 表示 A 因素第 i 水平的所有实验数据之和;
 $X_{.j.}$ 表示 B 因素第 j 水平的所有实验数据之和;
 $X_{ij.}$ 表示 A 因素第 i 水平和 B 因素第 j 水平的所有实验数据之和;
 $X_{...}$ 表示所有实验数据的总和

表 2 两因素(固定型)实验数据整理以及方差分析计算^[3]

年龄因素 A (i = 1, 2, 3)	声音刺激因素 B (j = 1, 2, 3, 4)			
	对照	FNC	MNC	WN
130 ~ 150 d	12.4	39.7	18.2	13.2
	11.5	43.7	18.5	12.6
	9.5	38.5	17.7	12.3
	10.7	34.6	20.4	13.2
250 ~ 280 d	13.2	28.2	22.5	12.4
	12.8	35.8	21.8	14.3
	14.1	30.2	20.6	11.7
	11.5	37.8	24.2	13.5
360 ~ 390 d	11.2	26.7	19.1	13.1
	12.6	32.2	24.3	12.6
	13.9	25.4	22.5	10.7
	13.5	27.5	20.7	11.8
$X_{11\cdot} = X_{111} + X_{112} + X_{113} + X_{114} = 12.4 + 11.5 + 9.5 + 10.7 = 44.1$				$\bar{X}_{11\cdot} = 12.25$
$X_{12\cdot} = X_{121} + X_{122} + X_{123} + X_{124} = 39.7 + 43.7 + 38.5 + 34.6 = 156.5$				$\bar{X}_{12\cdot} = 39.125$
$X_{13\cdot} = X_{131} + X_{132} + X_{133} + X_{134} = 18.2 + 18.5 + 17.7 + 20.4 = 74.8$				$\bar{X}_{13\cdot} = 18.7$
$X_{14\cdot} = X_{141} + X_{142} + X_{143} + X_{144} = 13.2 + 12.6 + 12.3 + 13.2 = 51.3$				$\bar{X}_{14\cdot} = 12.825$
$X_{21\cdot} = X_{211} + X_{212} + X_{213} + X_{214} = 13.2 + 12.8 + 14.1 + 11.5 = 51.6$				$\bar{X}_{21\cdot} = 12.9$
$X_{22\cdot} = X_{221} + X_{222} + X_{223} + X_{224} = 28.2 + 35.8 + 30.2 + 37.8 = 132.0$				$\bar{X}_{22\cdot} = 33.0$
$X_{23\cdot} = X_{231} + X_{232} + X_{233} + X_{234} = 22.5 + 21.8 + 20.6 + 24.2 = 89.1$				$\bar{X}_{23\cdot} = 22.275$
$X_{24\cdot} = X_{241} + X_{242} + X_{243} + X_{244} = 12.4 + 14.3 + 11.7 + 13.5 = 51.9$				$\bar{X}_{24\cdot} = 12.975$
$X_{31\cdot} = X_{311} + X_{312} + X_{313} + X_{314} = 11.2 + 12.6 + 13.9 + 13.5 = 51.2$				$\bar{X}_{31\cdot} = 12.8$
$X_{32\cdot} = X_{321} + X_{322} + X_{323} + X_{324} = 26.7 + 32.2 + 25.4 + 27.5 = 111.8$				$\bar{X}_{32\cdot} = 27.95$
$X_{33\cdot} = X_{331} + X_{332} + X_{333} + X_{334} = 19.1 + 24.3 + 22.5 + 20.7 = 86.6$				$\bar{X}_{33\cdot} = 21.65$
$X_{34\cdot} = X_{341} + X_{342} + X_{343} + X_{344} = 13.1 + 12.6 + 10.7 + 11.8 = 48.2$				$\bar{X}_{34\cdot} = 12.05$
$X_{1\cdot\cdot} = X_{11\cdot} + X_{12\cdot} + X_{13\cdot} + X_{14\cdot} = 44.1 + 156.5 + 74.8 + 51.3 = 326.7$				$\bar{X}_{1\cdot\cdot} = 20.419$
$X_{2\cdot\cdot} = X_{21\cdot} + X_{22\cdot} + X_{23\cdot} + X_{24\cdot} = 51.6 + 132.0 + 89.1 + 51.9 = 324.6$				$\bar{X}_{2\cdot\cdot} = 20.288$
$X_{3\cdot\cdot} = X_{31\cdot} + X_{32\cdot} + X_{33\cdot} + X_{34\cdot} = 51.2 + 111.8 + 86.6 + 48.2 = 297.8$				$\bar{X}_{3\cdot\cdot} = 18.613$
$X_{\cdot 1\cdot} = X_{11\cdot} + X_{21\cdot} + X_{31\cdot} = 44.1 + 51.6 + 51.2 = 146.9$				$\bar{X}_{\cdot 1\cdot} = 12.242$
$X_{\cdot 2\cdot} = X_{12\cdot} + X_{22\cdot} + X_{32\cdot} = 156.5 + 132 + 111.8 = 400.3$				$\bar{X}_{\cdot 2\cdot} = 33.358$
$X_{\cdot 3\cdot} = X_{13\cdot} + X_{23\cdot} + X_{33\cdot} = 74.8 + 89.1 + 86.6 = 250.5$				$\bar{X}_{\cdot 3\cdot} = 20.875$
$X_{\cdot 4\cdot} = X_{14\cdot} + X_{24\cdot} + X_{34\cdot} = 51.3 + 51.9 + 48.2 = 151.4$				$\bar{X}_{\cdot 4\cdot} = 12.617$
所有实验数据总和 $X_{\cdot\cdot\cdot} = 949.1$				$\bar{X}_{\cdot\cdot\cdot} = 949.1/48 = 19.773$
$X_{1\cdot\cdot}^2 = 326.7^2 = 106\ 732.89$				$X_{2\cdot\cdot}^2 = 324.6^2 = 105\ 365.16$
$X_{3\cdot\cdot}^2 = 297.8^2 = 88\ 684.84$				
$X_{\cdot 1\cdot}^2 = 146.9^2 = 21\ 579.61$				$X_{\cdot 2\cdot}^2 = 400.3^2 = 160\ 240.09$
$X_{\cdot 3\cdot}^2 = 250.5^2 = 62\ 750.25$				$X_{\cdot 4\cdot}^2 = 151.4^2 = 22\ 921.96$
$X_{\cdot\cdot\cdot}^2 = 949.1^2 = 900\ 790.81$				

5 讨论

在表 1 中,由于 A 因素和 B 因素可以是固定型(或选择型)因素,也可以是随机型因素,还可是混合型因素(即 A 选择, B 随机),因此有三种不同的统计模型。这三种模型在计算时虽无太大的差异,但在实验设计、检验方法和对实验结果的解释则有极大的不同。在生

物学实验中,各因素一般都是选择型的。本文实例中两因素均为固定型因素,因此统计模型选用固定模型。

关于重复实验,有一点值得注意:表 1 和表 2 有可能给人造成一种误解即重复实验是重复测定几次数据,其实这种理解和设计是错误的。这里所指的重复实验一定是每一次实验都从头到尾完整地做一次得到 $a \times b$ 个实验结果,重复实验做 n 次则得到 $a \times b \times n$

个实验数据。如本实例中年龄因素 A 有三个水平, 重复实验是指每一次重复都用三种不同年龄的雌鸟共 12 只进行 4 种不同声音刺激得到的实验数据为 $a \times b = 3$

$\times 4 = 12$ 次, 该实验重复 4 次, 共获得了 48 个实验数据。

表 3 两因素方差分析总结表

变差来源	平方和(SS)	自由度(df)	均方(MS)	F
A 因素	SS_A	$a - 1$	MS_A	$\frac{MS_A}{MS_e}$
B 因素	SS_B	$b - 1$	MS_B	$\frac{MS_B}{MS_e}$
AB 交互作用	SS_{AB}	$(a - 1)(b - 1)$	MS_{AB}	$\frac{MS_{AB}}{MS_e}$
误差	SS_e	$ab(n - 1)$	MS_e	
总和	SS_T	$abn - 1$		

$$SS_A = bn \sum_{i=1}^a (\bar{X}_{i..} - \bar{X}_{...})^2 = \frac{1}{bn} \sum_{i=1}^a X_{i..}^2 - \frac{X_{..}^2}{abn}$$

$$= \frac{1}{4 \times 4} \sum_{i=1}^3 X_{i..}^2 - \frac{X^2}{3 \times 4 \times 4} = \frac{1}{16} \times (106\ 732.89 + 105\ 365.16 + 88\ 684.84) - 18\ 766.475 = 32.456$$

$$SS_B = an \sum_{j=1}^b (\bar{X}_{.j.} - \bar{X}_{...})^2 = \frac{1}{an} \sum_{j=1}^b X_{.j.}^2 - \frac{X_{...}^2}{abn}$$

$$= \frac{1}{3 \times 4} \sum_{j=1}^4 X_{.j.}^2 - 18\ 766.475 = \frac{1}{12} \times (21\ 579.61 + 160\ 240.09 + 62\ 750.25 + 22\ 921.96) - 18\ 766.475$$

$$= 3\ 524.518$$

$$SS_{AB} = n \sum_{i=1}^a \sum_{j=1}^b (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2 = 4 \times \sum_{i=1}^3 \sum_{j=1}^4 (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2 = 258.122$$

$$SS_T = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n X_{ijk}^2 - \frac{X_{...}^2}{abn} = \sum_{i=1}^3 \sum_{j=1}^4 \sum_{k=1}^4 X_{ijk}^2 - 18\ 766.475 = 22\ 758.13 - 18\ 766.475 = 3\ 991.655$$

$$SS_e = SS_T - SS_A - SS_B - SS_{AB} = 3\ 991.655 - 32.456 - 3\ 524.518 - 258.122 = 176.559$$

$$1. \sum_{i=1}^3 \sum_{j=1}^4 (\bar{X}_{ij.} - \bar{X}_{i..} - \bar{X}_{.j.} + \bar{X}_{...})^2 = \sum_{j=1}^4 (\bar{X}_{1j.} - \bar{X}_{1..} - \bar{X}_{.j.} + \bar{X}_{...})^2 + \sum_{j=1}^4 (\bar{X}_{2j.} - \bar{X}_{2..} - \bar{X}_{.j.} + \bar{X}_{...})^2 + \sum_{j=1}^4 (\bar{X}_{3j.} - \bar{X}_{3..} - \bar{X}_{.j.} + \bar{X}_{...})^2$$

$$2. \sum_{j=1}^4 (\bar{X}_{1j.} - \bar{X}_{1..} - \bar{X}_{.j.} + \bar{X}_{...})^2 = (\bar{X}_{11.} - \bar{X}_{1..} - \bar{X}_{.1.} + \bar{X}_{...})^2 + (\bar{X}_{12.} - \bar{X}_{1..} - \bar{X}_{.2.} + \bar{X}_{...})^2 + (\bar{X}_{13.} - \bar{X}_{1..} - \bar{X}_{.3.} + \bar{X}_{...})^2 + (\bar{X}_{14.} - \bar{X}_{1..} - \bar{X}_{.4.} + \bar{X}_{...})^2$$

表 4 实验数据方差分析总结表

变差来源	平方和(SS)	自由度(df)	均方(MS)	F
年龄因素 A	32.456 (SS_A)	3 - 1	$MS_A = 32.456/2 = 16.23$	$\frac{MS_A}{MS_e} = 3.33^*$
声音因素 B	3 524.518 (SS_B)	4 - 1	$MS_B = 3\ 524.518/3 = 1\ 174.84$	$\frac{MS_B}{MS_e} = 239.76^{**}$
AB 交互作用	258.122 (SS_{AB})	(3 - 1)(4 - 1)	$MS_{AB} = 258.122/6 = 43.02$	$\frac{MS_{AB}}{MS_e} = 8.78^{**}$
误差	176.559 (SS_e)	3 × 4(4 - 1)	$MS_e = 176.559/36 = 4.90$	
总和	3991.655 (SS_T)	3 × 4 × 4 × 1		

表 4 中, 判定 A 和 B 两因素是否存在交互作用, 计算得知 $F = 8.78$, 而 $F_{0.36, 0.01} = 3.47$, $F > F_{0.01}$, 依此判定 A 和 B 两因素存在交互作用。从生物学意义解释该实验结果时, 不能只考虑声音刺激引起雌鸟生殖激素 LH

分泌变化, 也不能只考虑不同年龄的雌鸟接受同一声音刺激 LH 分泌变化, 而必须考虑这两种因素对生殖激素 LH 分泌产生的交互效应。如果计算结果显示: 年龄因素 A 与声音因素 B 之间不存在显著交互作用, 在此

情况下,可以不考虑年龄因素,把各声音水平下的数据作为单因素多组群的数据来处理,具体的统计推断计算、分析和解释结果详见“生物学实验数据的单因素方差分析”一文^[1]

本文介绍的两因素方差分析,它适合对有重复两因素组群实验数据进行方差分析。此外,了解了本文介绍的两因素方差分析的基本原理和一些基本概念,可选用合适的统计软件进行分析处理,如 SPSS,^[5]其操作过程简述如下:首先建立数据文件,打开该数据文件,单击主菜单“Statistics”,展开下拉菜单。在下拉菜单中选择“ANOVA”弹出小菜单,在小菜单上寻找“General Factorial”,单击“General Factorial”即可得到一个对话框(通用多因素方差分析对话框)。将 Value 调入“Dependent Variable:”下的矩形框。将 A、B 调入“Factor(s)”下的矩形框,并定义范围。单击通用多因素方差分析对话框下方的“Model”钮,随即弹出一个对话框,在该对话框上方“Specify Model”中有两项即 Full factorial 和 Custom。单击 Custom 激活这个对话框。在“Build Term

(s)”框下有一个可供选择的下拉菜单,把数据文件中的 A、B 和 $A * B$ 调入 Model 下的矩形框中,然后击“Continue”钮,回到通用多因素方差分析对话框。单击该对话框中的 OK 钮即得到输出的结果。

参 考 文 献

- [1] 刘加妹,彭景榭. 生物实验数据的单因素方差分析. 动物学杂志, 2001, 36(6): 34 ~ 37.
- [2] Cheng M F, Peng J P, Patricia Johnson. Hypothalamic neurons preferentially respond to female nest coo stimulation: demonstration of direct acoustic stimulation of luteinizing hormone release. *The Journal of Neuroscience USA*, 1998, 18 (14): 5 477 ~ 5 489
- [3] Mike Derocco. *Applied Statistics for the Behavioral Sciences, Third Edition*. Boston(USA), 1994. 316 ~ 348
- [4] 杜荣寿. 生物统计学. 北京: 高等教育出版社, 施普林格出版社, 2001. 135 ~ 150
- [5] 马斌荣主编. SPSS for Windows 在医学科研统计中的应用. 北京: 科学出版社, 2000. 55 ~ 77