

· 研究原著 ·

文章编号 1000-2790(2008)03-0266-04

医学研究方差分析模型中效应量的探讨

李运明¹, 袁天峰², 陈长生¹, 曹文君¹, 赵东涛³, 潘峰¹, 李晨⁴ (第四军医大学¹预防医学系卫生统计学教研室,²学报编辑部,³生物医学工程系数学教研室, 陕西 西安 710033, ⁴四川大学华西公共卫生学院预防医学系 2004 级 3 班, 四川成都 610041)

Exploration of effect size for ANOVA in medical research

LI Yun-Ming¹, YUAN Tian-Feng², CHEN Chang-Sheng¹, CAO Wen-Jun¹, ZHAO Dong-Tao³, PAN Feng¹, LI Chen⁴

¹Department of Health Statistics, School of Preventive Medicine,

²Editorial Department, ³Department of Mathematics, School of Biomedical Engineering, Fourth Military Medical University, Xi'an 710033, China, ⁴Class 3, Grade 2004, Faculty of Preventive Medicine, West China School of Public Health, Sichuan University, Chengdu 610041, China

【Abstract】 AIM: To explore the standardized mean difference, the common effect size for ANOVA in medical research. **METHODS:** Formulas of standardized mean difference in different trial designs were provided. **RESULTS:** The standardized mean difference of the same intervention was comparable between different design studies. **CONCLUSION:** The research papers in biomedicine need provide the index of effect size in the future, and researchers should calculate and interpret the standardized mean difference correctly to avoid and reduce the misapplication of effect size.

【Keywords】 ANOVA; effect size; standardized mean difference; hypothesis test

【摘要】目的 探讨医学研究中方差分析常用的效应量——标准均数差的计算方法。方法 针对不同的实验设计类型, 给出标准均数差的计算方法。结果 不同设计的研究间, 相同干预的标准均数差具有可比性。结论 生物医学论文报道效应量是未来的发展趋势, 研究者应正确计算和解释标准均数差, 避免和减少效应量的误用。

【关键词】 方差分析; 效应量; 标准均数差; 假设检验

【中图分类号】 R196 **【文献标识码】** A

0 引言

效应量(effect size)是一类用来描述处理效应的

收稿日期 2007-09-21; 接受日期 2007-10-06

基金项目: 国家自然科学基金(39900126); 陕西省科技计划项目(2003F11)

通讯作者: 陈长生. Tel (029) 84774853 Email zhencs@fmmu.edu.cn

作者简介: 李运明. 硕士生(导师陈长生). Tel (029) 84772180

Email lee3082@sina.com.cn

统计量。在 20 世纪 60 年代, 生物统计学家(Cohen, 1965; Hays, 1963)就强调效应量的应用, 认为效应量是假设检验的补充^[1]。然而医学领域的绝大多数的研究者在报道结果时, 往往仅提供假设检验的 *P* 值^[2-3]。1996 年美国心理学会(APA)的统计推断机构 TFSI 建议报道研究结果时应同时提供处理效应的方向、大小及其的可信区间^[4]。1998 年 Wilkinson 和 TFSI 建议对于主要结果必须报道效应量, 即报道 *P* 值时同时应报道效应量^[5]。2001 年美国心理学会(APA)科研论文发表手册上规定: 论文的结果部分必须报道效应量^[6]。至今已有 24 种心理学、医学期刊要求研究者投稿时报道效应量^[7]。国内教科书对 Meta 分析所涉及的效应量作了简单介绍, 但对效应量的系统研究很少。依资料类型和研究设计不同, 效应量又有很多种类, 我们主要研究方差分析(ANOVA)模型中常用的一类效应量——标准均数差(standardized mean difference)。

1 材料和方法

1.1 材料 为研究不同的实验设计类型的标准均数差的计算方法, 我们采用了 Bauman 等^[1]人的实验数据(表 1)。该实验采用前后测量设计研究了 66 名四年级学生不同阅读习惯对理解能力的影响。阅读习惯(研究干预)分为: 单纯朗读(TA), 阅读并积极思考(DRTA), 阅读(DRA), 其中 DRA 为对照组。理解能力用错误检测任务(EDT)的得分表示, 干预前后两次测量结果用 EDT1, EDT2 表示。该研究考虑了一个控制因素(即研究前的理解能力): 各组前两列的学生研究前理解能力较低, 后两列理解能力较高。

1.2 方法 在统计分析中, 需要解决均数的对比(contrast)问题, 即一个研究有 *J* 个处理组, 则均数的对比可以表示为:

$$\Psi = c_1\mu_1 + c_2\mu_2 + \dots + c_j\mu_j \quad (1)$$

其中, $c_1 + c_2 + \dots + c_j = 0$ 。 $\Psi = \mu_i - \mu_j$ 是最常见的对比。对比含有量纲, 与反应变量的量纲相同, 不能直接用于不同研究间比较; 而标准均数差无量纲, 可用于不同研究间比较的效应量。按反应变量的不

同,可将标准均数差分为单变量和多变量标准均数差. 不同设计标准均数差计算方法如下:

表1 66名四年级学生接受不同干预后 EDT 得分情况

TA				DRTA				DRA			
EDT1	EDT2										
7	4	11	12	7	7	12	13	4	5	12	8
4	4	14	14	7	5	10	5	6	9	16	10
7	4	13	12	8	13	16	14	9	5	15	9
7	3	9	7	7	8	15	14	8	7	14	12
6	8	12	5	6	7	9	10	9	4	12	5
4	1	13	9	8	10	13	12	10	9	12	8
6	7	13	13	9	8	12	11	8	3	13	12
6	7	11	11	8	10	9	8	11	4	12	4
8	9	12	5	9	12	13	10	8	2	12	8
5	6	11	11	8	7	10	11	7	5	12	6
8	4	14	15	8	8	10	12	9	7	12	5

1.2.1 单变量标准均数差

1.2.1.1 单因素完全随机设计 该设计的处理因素有 J 个水平,实验拟研究的问题可表示为对比(1),其标准均数差为:

$$\delta = \frac{\Psi}{\sigma} \quad (2)$$

总体参数 δ 的估计方法:用样本均数 \bar{x} 估计总体

表2 含有控制因素的多因素设计标准均数差的计算方法

分析目的	对比	标准均数差的计算方法
干预因素 A 的主效应	$\Psi = \frac{1}{2}(\mu_{a1 b1} + \mu_{a1 b2}) - \frac{1}{2}(\mu_{a2 b1} + \mu_{a2 b2})$	准则二 a. 按照干预因素分组,计算各组的标准差 b. 用准则一中的一种方法估计 σ .
干预因素 A 在 b1 水平的单独效应	$\Psi = \mu_{a1 b1} - \mu_{a2 b1}$	同准则二.
因素 A 与 B 的交互作用	$\Psi = (\mu_{a1 b1} - \mu_{a2 b1}) - (\mu_{a1 b2} - \mu_{a2 b2})$	同准则二.
控制因素 B 的主效应	$\Psi = \frac{1}{2}(\mu_{a1 b1} + \mu_{a2 b1}) - \frac{1}{2}(\mu_{a1 b2} + \mu_{a2 b2})$	准则三 a. 按照干预因素及对比中含有的控制因素分组,计算各组的标准差 b. 用准则一中的一种方法估计 σ .
控制因素 B 在 a1 水平的单独效应	$\Psi = \mu_{a1 b1} - \mu_{a1 b2}$	同准则三.

多因素实验研究的对比可能仅含有控制因素,不含有处理因素,如在 $2 \times 2 \times 2$ 析因设计中,对比为:

$$\Psi = \frac{1}{2}(\mu_{b1 c1} + \mu_{b1 c2}) - \frac{1}{2}(\mu_{b2 c1} + \mu_{b2 c2}) \quad (3)$$

其中 A 为处理因素, B, C 为控制因素. 仅含有控制因素对比的标准均数差计算方法 a 按照实验研究的控制因素分组,计算各组的标准差,在对比(3)中,按照因素 B 分组 b 用准则一估计 σ .

1.2.1.3 含有协变量的多因素设计 协方差分析 (ANOCVA)通过建立协变量与反应变量的线性回归关系,对各组的反应变量的均数进行校正后,再进行

均数 μ , σ 可以用准则一中的一种方法进行估计. 准则一 a 设计中的某个处理组的标准差,常用对照组的标准差 b 对比中所有处理组的合并标准差 c 设计中所有处理组的合并标准差.

当对比中包含所有的处理组时, b, c 得到的 σ 估计值相同,并与 ANOVA 分析中误差均方 (MSE) 正的平方根相等. 当所有处理组满足方差齐性条件时, c 法是估计 σ 的最佳方法,当不满足时,用 a 法估计. Hedges 指出按照准则一估计的标准均数差是 δ 的有偏估计,需要乘以系数 $1 - 3/(4df - 1)$ 进行校正,其中 df 为用于估计 σ 的标准差或合并标准差的自由度^[8].

1.2.1.2 多因素设计 该设计的因素可为干预因素 (处理因素)和控制因素 (非研究因素、混杂因素). 当所有因素均为干预因素时,标准均数差的计算与单因素完全随机设计相同. 多因素实验中若含有控制因素,如将控制因素与干预因素不加区别,按照准则一计算标准均数差时,会出现相同干预的效应量在不同实验设计间不可比的问题^[1]. 根据所研究对比的特征,标准均数差的计算方法不同,如以 2×2 析因设计为例,见表 2. 设实验含有:处理因素 A (a1 a2),控制因素 B (b1 b2).

假设检验. ANOCVA 标准均数差的计算方法为:用样本校正均数 \bar{x}_c 估计总体均数 μ ,将协变量作为控制因素,按照准则二来估计 σ .

1.2.1.4 含有重复测量因素的多因素设计 含有重复测量因素的设计可分为:①仅含有 1 个或多个重复测量因素的设计;②含有重复测量因素和观测间因素的设计. 因为重复测量因素为处理因素,所以①中不存在控制因素引起的相同处理的效应量在不同实验设计间不可比的问题,标准均数差的计算方法,与因素为处理因素的设计相同. 含有重复测量因素和观

测间因素的设计计算标准均数差时,将重复测量因素作为处理因素,如观测间因素含有控制因素按照表2中准则二或三计算。

1.2.2 多变量标准均数差 马氏距离在多元方差分析中即是一种多变量标准均数差。马氏距离公式为:

$$D = \sqrt{d'R^{-1}d}$$

其中 d 为单变量标准均数差向量, R 为合并的组内相关矩阵。实际计算中,马氏距离可以由多元检验统计量 Wilks's Λ 计算得到:

$$D = \sqrt{\frac{dk(1-\Lambda)\sum_{i=1}^k c_i^2/n_i}{\Lambda}} \quad (4)$$

其中 k 为处理组数, c_i, n_i 分别为 i 组对比系数和样本量。 df 的计算公式为 $df = \sum n_i - k$ 。

1.2.3 标准均数差的解释 标准均数差的解释准则不多,因为医学研究领域所涉及的内容很广泛,想给出普遍适用的准则,需要冒很大风险。Cohen 建议标准均数差为 0.2 时,效应为小,0.5 为中等,0.8 为大。如果样本满足正态分布,总体间重叠的比例(percent of overlap, OL%)有助于标准均数差的解释。若处理组与对照组的均数差为 0.70,那么可认为处理组 50% 的研究对象反应变量值大于对照组 76% 的研究对象的值(图 1)。

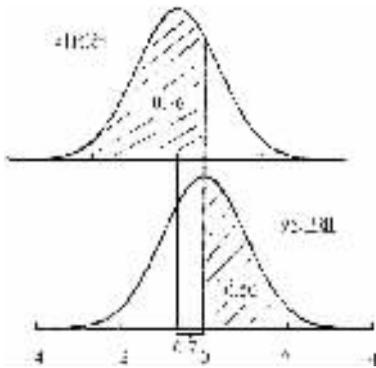


图 1 标准均数差与 OL% 示意图

2 结果

Bauman 等人的研究关心阅读方法 TA 和 DRTA 的平均效应与 DRA 的差别(对比 Ψ_1)以及阅读方法 TA 与 DRTA 的差别(对比 Ψ_2)。

$$\Psi_1 = \frac{1}{2}(\mu_{TA} + \mu_{DRTA}) - \mu_{DRA}, \Psi_2 = \mu_{DRTA} - \mu_{TA}$$

若仅考虑 EDT2 和干预因素(阅读习惯),本例的研究设计为单因素完全随机设计。表 3 为各组的均数和标准差,表 4 为对比 Ψ_1, Ψ_2 的标准均数差。按照 Cohen 准则,两对比均为中等效应。校正后 Ψ_2 的效应量为 0.697,可认为 50% 阅读并积极思考的学生

的 EDT 成绩高于 76% 的单纯朗读的学生成绩。

表 3 各组 EDT1, EDT2 成绩 ($n=22, \bar{x} \pm s$)

变量	TA	DRTA	DRA
EDT1	9.14 ± 3.34	9.73 ± 2.69	10.50 ± 2.97
EDT2	7.77 ± 3.93	9.77 ± 2.72	6.68 ± 2.77

表 4 单因素完全随机设计标准均数差

估计方法	Ψ_1		Ψ_2	
	未校正	校正后	未校正	校正后
$\hat{\delta}_a = \Psi/S_c$	2.09/2.77 = 0.756	0.728	2.00/2.77 = 0.723	0.697
$\hat{\delta}_b = \Psi/S_{pooled}$	2.09/3.19 = 0.656	0.644	2.00/3.38 = 0.592	0.581
$\hat{\delta}_c = \Psi/S_{pooled}$	2.09/3.19 = 0.656	0.648	2.00/3.19 = 0.627	0.620

注: $\hat{\delta}_a, b, c$ 分别是按照准则一中的 a, b, c 法计算。

若将 EDT2 作为研究的反应变量,考虑干预因素 A 和控制因素 B(阅读能力),本例为析因设计。为了便于公式的演算,假设干预因素为两水平(TA, DRTA)本例研究干预因素、控制因素的主效应、单独效应及两因素的交互作用。这些效应的可以用表 2 中相应的对比表示,其标准均数差的计算见表 5。

表 5 多因素设计各组 EDT2 成绩及标准均数差 ($\bar{x} \pm s$)

控制因素 B	干预因素 A	
	TA(a1)	DRTA(a2)
低(b1)	5.18 ± 2.40	8.64 ± 2.38
高(b2)	10.36 ± 3.44	10.91 ± 2.66
合计	7.77 ± 3.93	9.77 ± 2.72
标准均数差		
干预因素 A 的主效应	-2.00/3.38 = -0.592	
干预因素 A 在 b1 水平的单独效应	-3.45/3.38 = -1.022	
因素 A 与 B 的交互作用	-2.91/3.38 = -0.861	
控制因素 B 的主效应	-7.45/2.76 = -2.706	
控制因素 B 在 a1 水平的单独效应	-3.45/2.76 = -1.254	

注:计算标准均数差时,估计 σ 的准则一用 c 法。

若将 EDT2 作为研究的反应变量,考虑干预因素,并将干预前的测量结果 EDT1 作为协变量,本例为含有协变量的单因素设计(协方差设计)。通过协方差分析,各组校正后的均数见表 6。按照校正均数计算对比 Ψ_1, Ψ_2 的标准均数差,见表 6。

将 EDT 作为研究的反应变量,考虑干预因素和重复测量因素,干预前后 EDT 做了两次,重复测量因素有两水平,本例为含有 1 个重复测量因素的两因素设计。不同阅读方式的效应用两次测量的差值表示,

两对比 Ψ_1, Ψ_2 可以表示为:

表6 各组 EDT2 成绩及标准均数差 ($\bar{x} \pm s$)

	TA	DRTA	DRA
EDT2	8.22 ± 3.93	9.81 ± 2.72	6.19 ± 2.77
标准均数差			
Ψ_1	2.83/3.19 = 0.888		
Ψ_2	1.59/3.19 = 0.499		

注: 计算标准均数差时, 估计 σ 的准则一用 c 法。

$$\Psi_1 = \frac{1}{2} (\mu_{EDT2, TA} - \mu_{EDT1, TA}) + \frac{1}{2} (\mu_{EDT2, DRTA} - \mu_{EDT1, DRTA}) - (\mu_{EDT2, DRA} - \mu_{EDT1, DRA}),$$

$$\Psi_2 = (\mu_{EDT2, DRTA} - \mu_{EDT1, DRTA}) - (\mu_{EDT2, TA} - \mu_{EDT1, TA}).$$

根据表 3, 可计算对比 Ψ_1, Ψ_2 的标准均数差分别为 1.018, 0.439。

将 EDT1, EDT2 作为研究的反应变量, 考虑干预因素, 本例为多元单因素完全随机设计。对比 Ψ_1, Ψ_2 中的 μ 为均数向量, 检验统计量 Wilks's Λ , 可以用 SAS/GLM CONTRAST 计算得到^[9]。由公式(4)可计算对比 Ψ_1, Ψ_2 的多元标准均数差 D 分别为 1.228, 0.689。

3 讨论

标准均数差是方差分析模型中常用的一类效应量, 也是目前心理学、医学研究领域和 Meta 分析中最常用到的效应量。本文按照不同的实验设计, 考虑相同干预不同设计间效应量的可比性, 介绍了标准均数差的计算方法, 总结给出了相应的计算准则, 并给出了实例。Meta 分析常遇到研究干预相同、研究设计不同的情况下, 效应量的计算问题。本文介绍的标准均数差的计算方法可以很好的解决这一问题。另外, 本文介绍的标准均数差的计算可适用于两组和多分组的情况, 有些资料和文献上针对两组资料的比较对标准均数差进行介绍。专用于两组比较的标准均数差有 Cohen's d , Glass's Δ , Hedges's g 和 Cohen's f^2 ^[10]。

尽管 APA 和 24 种期刊要求研究者进行假设检验时, 必须报道一种或多种效应量作为其补充, 但是对效应量能否帮助研究者或读者提供有关干预效应有无实际意义的信息, 也有统计学家提出疑问^[11]。Cohen 对标准均数差解释制定的准则, 能否适用医学研究领域, 也存在争议。Cohen 也建议统计学者制定其他的准则来解释标准均数差。目前, 国内的生物医学期刊还未要求报道效应量, 国外对效应量的研究和报道较多, 尤其是在心理测量领域的研究, 并有关于效应量误用的分析报道, 因此我国生物医学论文要求报道效应量是未来的发展趋势。

【参考文献】

【参考文献】

- [1] Olejnik S, Algina J. Measures of effect size for comparative studies: Applications, interpretations, and limitations[J]. Contemp Educ Psychol, 2000, 25(3): 241-286.
- [2] Glaser DN. The controversy of significance testing: Misconceptions and alternatives[J]. Am J Crit Care, 1999, 8(5): 291-296.
- [3] Cohen J. The earth is round ($P < 0.05$) [J]. Am Psychol, 1994, 49(12): 997-1003.
- [4] <http://www.apa.org/science/tfsi.html>.
- [5] Wilkinson L. Task force on statistical inference APA board of scientific affairs. Statistical methods in psychology journals: Guidelines and Explanations[J]. Am Psychol, 1999, 54(8): 594-604.
- [6] American Psychological Association. Publication manual of the American Psychological Association[M]. 5th ed. Washington: American Psychological Association Press, 2001: 1-5.
- [7] <http://www.coe.tamu.edu/bthompson>.
- [8] Hedges LV. Distributional theory for Glass's estimator of effect size and related estimators[J]. J Educ Stat, 1981, 6(6): 107-128.
- [9] 胡良平. 现代统计学与 SAS 应用[M]. 北京: 军事医学科学出版社, 2000: 146-150.
- [10] Rosnow RL. Effect sizes for experimenting psychologists[J]. Can J Exp Psychol, 2003, 57(3): 221-237.

编辑 黄良田