

The Application of ANOVA in Psychological Experiment

Lei Zhang

Education School of Soochow University, Suzhou Jiangsu
Email: 798257682@qq.com

Received: Jun. 4th, 2015; accepted: Jun. 26th, 2015; published: Jun. 29th, 2015

Copyright © 2015 by author and Hans Publishers Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY).
<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

As a basic manner of statistical approach, ANOVA is widely used in all kinds of psychology researches. But there are still some problems when researchers use this particular tool in research, such as the reports being too brief, not being able to eliminate the relevant disturbing variables. This article will discuss these issues of ANOVA in psychological experiments.

Keywords

Experimental Research, ANOVA, Effect Size, Statistical Power

方差分析在实验研究中的应用

张 蕾

苏州大学教育学院, 江苏 苏州
Email: 798257682@qq.com

收稿日期: 2015年6月4日; 录用日期: 2015年6月26日; 发布日期: 2015年6月29日

摘 要

作为一个基本的统计手段, 方差分析在心理学研究领域得到了广泛地应用, 然而在实际应用过程中, 还存在结果报告简单, 无法排除干扰变量等的不足。本篇文章就将会针对这些问题对方差分析在实验研究中的

应用进行讨论。

关键词

实验研究, 方差分析, 效果大小, 统计检验效能

1. 引言

实验研究是研究者运用科学实验的原理和方法, 以一定的理论或者假设为指导, 有目的地控制自变量的变化, 并观察应变量的变化, 得到因果关系, 发现规律的研究方法。

与自然科学一样, 心理学实验研究以追求真理、确认事实为目的, 从实际出发, 尊重事实; 与自然科学不同的是, 许多心理现象难以直接观察, 并且可能受到多个变量的影响, 这就要求心理学研究者们做出更为严谨的实验设计、选择更加科学的分析手段。例如在一项研究中, 研究者想要探寻紧张情绪和认知负荷对前瞻记忆的影响(张晶晶, 张茗, 2011), 如何准确地控制被试的情绪和认知负荷对于实验者是相当重要的, 因为它们无法直观、精确地度量。

心理学研究中, 研究者往往需要考察一个(或者多个)因素的不同水平对实验指标是否有影响, 也就是从质量因子(自变量)的角度探讨各水平对实验结果影响的差异。一般来说, 研究者是可以控制质量因子的, 并能够根据需要取不同的水平。方差分析很好地满足了这一需求, 并且在现有的研究中也得到了大量的应用。张孝义, 贺晓玲, 宋灿, 郭英, 张妍等人在关于创造力(张孝义, 2012)、成就动机(贺晓玲, 2012)、刻板印象(宋灿, 2012)、认知偏向(郭英, 张妍, 2012)等研究中, 都选择方差分析作为数据统计方法。

虽然方差分析在得到了极为广泛的应用, 但是在此应用的过程中, 多数研究者做的还是不过完善, 主要表现在两个方面: 首先, 对于额外变量的统计控制不足; 其次, 研究结果中只报告方差分析的 F 值和 p 值。

2. 协方差分析对于方差分析的补充

方差分析将来自各子总体抽样样本汇总在一起, 先假设他们来自一个总体(即假设没有差异), 然后将这个汇总样本的总变动(SS_T)分解为两个部分。一部分是组内变动(SS_W), 代表着本组内(即某个子总体内); 在多因素分析时则是按多因素进行划分的交互分组内)各案例值关于总平均值的分布离散程度。另一部分是组间变动(SS_B), 代表着各组平均值关于总平均值的分布离散程度。用公式表示为: $SS_T = SS_W + SS_B$ 。

然而在心理学实验研究中, 存在着各种实验误差。这些实验误差对实验结果会产生不同的影响, 常使实验结果不能达到预期的效果。为了克服实验误差的影响, 研究者常采用恰当的实验设计来控制实验误差对实验结果的影响。但有些影响实验结果的因素不易通过实验设计的方法进行控制, 这时汇总样本的总变动(SS_T)就不仅仅包括组内变动(SS_W)和组间变动(SS_B), 它还包括不可控制变量造成的误差(SS_E)。此时: $SS_T = SS_W + SS_B + SS_E$ 。

研究者可借助统计的方法对实验误差进行有效的统计分析方法。然而在现有的心理学实验研究中, 对于这种不可控制变量造成的试验误差的有效规避是极为少见的(高艳慧, 万迪昉, 郭海星, 2012)(王洪利, 王刊良, 李艳, 2011)。

心理学的实验研究中, 常常会出现既有可以控制的变量, 又有不可以控制的变量同时影响实验结果的情况。这时就需要采用协方差分析的统计处理方法, 将可控制变量和不可控制变量(即协变量)综合起来加以考虑。比如, 要研究独生子女与非独生子女儿童的认知能力, 而父亲的文化水平对儿童的认知能力

又有一定的影响, 要消除这一因素对实验结果的影响, 就需要将父亲的文化水平这一因素作为协变量进行协方差分析, 才能得到更准确的结果。由此可见, 协方差分析在心理科学研究中有着广泛的用途(于义良, 李爱玲, 1999)。

协方差分析是建立在方差分析和回归分析基础上的一种统计分析方法, 其基本思路是根据因变量对协变量的回归系数, 从因变量中扣除受协变量影响的部分, 从而正确分析自变量对因变量的关系。

协方差分析的步骤如下: (黄中, 2000)

第一步: 计算各组均值、平方和及协方和。

$$l_{xxi} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{ij} \cdot l_{xxi} = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2; \quad l_{yyi} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij} \cdot l_{yyi} = \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2;$$

$$l_{xyi} = \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(y_{ij} - \bar{y}_i)$$

第二步: 计算公共组内平方和及协方和

$$l_{xxw} = \sum_{i=1}^p l_{xxi}; \quad l_{yyw} = \sum_{i=1}^p l_{yyi}; \quad l_{xyw} = \sum_{i=1}^p l_{xyi}$$

第三步: 计算总平均值、总平方和及总协方和

$$l_{xxT} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} x_{ij} \cdot l_{xxT} = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2;$$

$$l_{yyT} = \frac{1}{n} \sum_{i=1}^p \sum_{j=1}^{n_i} y_{ij} \cdot l_{yyT} = \sum_{i=1}^p \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2;$$

$$l_{xyT} = \sum_{i=1}^p \sum_{j=1}^{n_i} (x_{ij} - \bar{x})(y_{ij} - \bar{y});$$

其中 $n = \sum_{i=1}^p n_i$

最后根据如上计算结果, 列出协方差分析表:

方差来源	修正平方和	修正自由度	均方	F比	临界值	显著性
组内	$Q_w = l_{yyw} - \frac{l_{xyw}^2}{l_{xxw}}$	$n - p - 1$	$V_w = \frac{Q_w}{n - p - 1}$	$F_n = \frac{V_B}{V_w}$	$F(n-1, n-p-1)$	
组间	$Q_B = Q_T - Q_w$	$p - 1$	$V_B = \frac{Q_B}{p - 1}$			
总和	$Q_T = l_{yyT} - \frac{l_{xyT}^2}{l_{xxT}}$	$n - 2$				

当P个总体均值有显著差异时, 就需要对均值排序, 又由于有协变量的影响, 所以需把协变量同时都取在相同的水平上, 这时就有 $y'_i = \bar{y}_i + b_\omega(\bar{x} - \bar{x}_i) = \bar{y}_i - b_\omega(\bar{x}_i - \bar{x})$

其中 $b_\omega = \frac{l_{xyw}}{l_{xxw}}$

通过具体案例, 我们可以更加清晰地看出协方差分析的重要性。

如果我们以某专业的五个班级的第一学期经济致学基础课(微积分)的期末统考成绩(y)和高考数学成绩(x)作为分析的原始数据。

首先研究五个不同专业的期末统考成绩有无显著差异。当不考虑协变量 x 的影响时, 只需作单因素方差分析, 由表 1 可见在显著性水平 $\alpha = 0.01$ 下差异是极其显著的。

按平均成绩由高到低排序为

班别	2	1	5	3	4
平均成绩	80.2	72.5	72.3	70.8	69.3

但是这样得到的结论可能是有问题的。因为四班学生的入学数学成绩在五个教学班中最低(119.4), 而二班学生的入学数学成绩在五个教学班中最高(138.7), 如果入学数学成绩对大学第一学期数学课成绩有影响, 那么入学成绩(即学生的基础)与大学教学的效果就会混杂, 这时四班的平均成绩低就很难判断是由大学教学效果的差异还是由学生基础所致。

学生的基础对学生后续课的学习有直接影响, 这是众人的共识, 因此上述的方差分析方法需要进行修正, 即用协方差分析方法, 对大学教学效果的评价扣除学生入学考试成绩的影响。

按照协方差分析计算步骤, 我们列出协方差分析表如表 2。

根据协方差分析可见, 五个教学班第一学期数学统考成绩有极显著差异, 注意到有协变量的影响, 所以对均值的排序, 需要把协变量同时都取在相同的水平上, 这时就有 $\bar{y}'_i = \bar{y}_i + b_\omega(\bar{x} - \bar{x}_i) = \bar{y}_i - b_\omega(\bar{x}_i - \bar{x})$,

其中 $b_\omega = \frac{l_{xyw}}{l_{xxw}}$ 。

先按 $\bar{y}'_i (i = 1, 2, 3, 4, 5)$ 由大到小的顺序重新排列为

班别	3	4	1	2	5
平均成绩	75.57	74.98	73.09	71.22	70.76

由此我们可以看出, 单纯运用方差分析的结果与运用协方差分析的结果可能大不相同究其原因主要是由于协方差分析排除了协变量对因变量的影响作用, 这种把回归分析和方差分析结合起来运用的方法正是协方差分析的实质和优点所在, 使其所做结论更切合实际一些。

这启示我们在进行研究的过程中要综合考察各种可能影响研究结果的因素, 在进行研究之前尽量通

Table 1. Analysis of variance of grades

表 1. 成绩的方差分析表

方差来源	平方和	自由度	均方	F比	临界值	显著性
班别间	2518.24	4	629.56	51.68	$F_{\alpha=0.01} = 3.34$	**
班别内	2436.36	200	12.18			
总和	4954.60	204				

**在 0.01 水平下显著。

Table 2. Analysis of covariance

表 2. 协方差分析表

方差来源	修正平方和	修正自由度	均方	F比	临界值	显著性
班别间	1209.324	302.33	121.42	48.76	$F_{\alpha=0.01} = 3.34$	**
班别内	496.16	199	2.49			
总和	1705.48	203				

**在 0.01 水平下显著。

过完善的研究设计避免可能造成的影响。

3. 方差分析的统计检验力和效果大小报告不足

现有的大量实验研究中，研究者通常只报告数据分析中方差分析的 F 值以及 p 值，以此说明实验处理是否产生影响，然而仅仅做到这些还是略显草率。例如在我们做相关分析的时候，尽管结果可能会显示相关系数在 $\alpha = 0.01$ 水平上显著，但是其值仅仅为 0.25 的话，我们也很难确定实验结果的可信度，因为相关分析的效果大小是 R^2 (郭志刚, 1999)，而 0.25 的平方仅为 0.0625。因此研究者在报告研究结果是应当连同统计方法的效果大小和统计检验效能一起报告。

效果大小(effect size, ES)是指某个特定总体中的某种特殊的非零的数值。这个数值越大，就表明由研究者所处理的研究现象所造成的效果越大。

目前，学术界对于如何评估方差分析效果大小至少存在两类不同的指标体系：一类以 η^2 作为指标，另一类以粗体小写字母 f 值作指标。两类指标所根据的理论背景不同，被接受的程度也有所不同。其中， η^2 的含义是“实验处理之后各组间平方和在总体平方和中所占的比重”， η^2 的指标值大，反映实验效果大， η^2 的指标值小，则反映实验效果小，因此这一指标所反映的效果大小的内涵最容易让人们所理解(胡竹菁, 戴海琦, 2011)。

美国著名心理统计学家 J. Cohen 指出(Cohen J., 1973)，可以用 η^2 来表示效果大小的指标，其计算公式为：
$$\eta^2 = \frac{SS_w}{SS_T}$$

目前较多的西方学者采用这种方法来估计方差分析的效果大小。

比如有人研究了个人表现的反馈类型对其自尊的影响。让 15 名被试参加一项知识测验，每组各 5 名被试。不管被试在测验中的实际表现如何，对积极反馈组，都告诉他们水平很高；对消极反馈组，都告诉他们表现很差；对控制组，不提供任何反馈信息。最后，让所有的被试都参加一个自尊测验，测验总分为 100 分，得到的分数越高，表明自尊越强。实验结果如表 1 所示，问不同反馈类型的各组被试的自尊水平是否存在显著差异？

对表 3 的数据进行方差分析后可以得到如表 4 所示的方差分析表。

根据表 4，我们可以算出效果大小

$$\eta^2 = SS_B / SS_T = 448 / 878 = 0.51$$

说明在该实验中，总体变异中约有 51% 是来自反馈类型的实验处理。由于 η^2 的结果作效果大小指标的便于理解性，著名的统计软件 SPSS 就采用 η^2 作为该软件“一般线性模型”模块中有关效果大小的指标。将表 1 中的数据输入 SPSS 运行后，SPSS 给出的效果大小值是“ $\eta^2 = 0.51$ ”（注：英文版 SPSS 给出的结果标为“Partial Eta Squared”；中文版 SPSS 18.0 给出的结果标为“偏 eta 方”），与上述用 η^2 作为效果大小的指标计算出的结果完全一样。

如前所述，J. Cohen 认为，当用 η^2 作方差分析效果大小的指标时，在 $\eta^2 = 0.14$ 时就属于大的效果，但 0.51 的效果大小还是远远大于 0.14。因此有的统计学家认为用 η^2 作为实验处理在总变异中所占比重的估计方法一般会高估实验处理的效果，提出另外一个含义与 η^2 大致相同的指标 ω^2 来反映实验处理效果大小，其计算公式为：

$$\omega^2 = \frac{SS_B - (k-1)MS_e}{SS_T + MS_e}$$

还是一表 1 和表 2 的数据为例，将已知数据代入公式后可得：

$$\omega^2 = \frac{SS_B - (k-1)MS_e}{SS_T + MS_e} = \frac{448 - (3-1) \times 28.67}{878 + 28.67} = 0.43$$

虽然 $\omega^2 = 0.43$ 比 $\eta^2 = 0.51$ 的值更小, 但被认为能更准确地反映反馈类型对被试自尊水平实验中的效果。

在统计决策中, 存在范 I 型错误和 II 型错误的可能性。错误地拒绝虚无假设 H_0 为 I 型错误, 我们可以通过 α 检验(p 值)来判断。错误地接受虚无假设 H_0 为 II 型错误, 判断犯 II 型错误的可能性用检验效能来表示。检验效能等于 $1-\beta$ 。我们一般通过采用更低的 α 水平来防止犯 I 型错误, 但防止 II 型错误却没有那么容易, 因为 β 很难求得。方差分析备择假设的期望分布值通常用希腊字母 Φ 表示。当用 η^2 来作为方差分析效果大小的指标时, 可以根据 η^2 值, 各组人数和分组数直接查相应的换算表求统计检验效能 $1-\beta$ 的值。以表 1 的数据为例, 分组数 $k=3$ 的相应换算表有如表 5 所示。

如表 3 所示, 各组人数是 $n=6$, 在表 5 中最接近的值是 $n=10$ 那一行, 效果大小 $\eta^2=0.51$, 虽然其值远远大于 $\eta^2=0.14$, 也只能查最接近的 $\eta^2=0.14$ 那一列, 结果为对本次实验数据进行方差分析的统计检验力 $1-\beta$ 为 51%。

Table 3. The scores of self-esteem under different feedback conditions

表 3. 不同反馈类型条件下被试自尊水平测试得分表

积极反馈组	控制组	消极反馈组
84	71	59
74	75	64
81	73	62
75	74	69
84	69	75
70	82	67

Table 4. ANOVA of self-esteem under different feedback conditions

表 4. 不同反馈类型对被试自尊水平影响的方差分析表

变异来源	平方和	自由度	均方	F
组间	448	2	224	7.814**
组内	430	15	28.67	
总变异	878	17		

**在 0.01 水平下显著。

Table 5. Scores of the power of a statistical test when $k=3$

表 5. 分组数为 $k=3$ 的统计检验力换算表

n	效果大小		
	$\eta^2 = 0.01$	$\eta^2 = 0.06$	$\eta^2 = 0.14$
10	0.07	0.21	0.51
20	0.10	0.43	0.85
30	0.13	0.61	0.96
40	0.16	0.76	0.99
50	0.19	0.85	1.0
100	0.36	0.99	1.0

尽管通过方差检验的 F 值和 p 值, 我们可以看出实验处理能够起到预期的作用, 但是这些还不够, 我们还需要考察其效果大小或者统计检验效能, 探索实验处理能发挥多大程度的作用, 并且还要尽可能剔除其它因素对研究结果的影响。

参考文献 (References)

- 高艳慧, 万迪昉, 郭海星(2012). 基于监管、学习和声誉的联盟稳定性: 实验研究. *管理学报*, 9 期, 1154-2261.
- 郭英, 张妍(2012). 海洛因成瘾者对吸毒线索图片认知偏向的实验研究. *四川大学学报*, 39 期, 124-128.
- 郭志刚(1999). *社会统计分析方法*. 中国人民大学出版社.
- 贺晓玲(2012). 成就动机对反事实思维影响的实验研究. *湖北成人教育学院学报*, 18 期, 23-24.
- 胡竹菁, 戴海琦(2011). 方差分析的统计检验力和效果大小的常用方法比较. *心理学探新*, 13 期, 254-259.
- 黄中(2000). 心理学实验研究中协方差分析的应用. *集宁师专学报*, 22 期, 37-40.
- 宋灿(2012). 大学生职业刻板印象激活的行为效应实验研究. *江苏技术师范学院学报*, 18 期, 106-109.
- 王洪利, 王刊良, 李艳(2011). 考虑观察成本的秘书问题仿真实验研究. *管理学报*, 9 期, 1707-1713.
- 于义良, 李爱玲(1999). 评价教学效率的协方差分析法. *数理统计与管理*, 18 期, 18-21.
- 张晶晶, 张茗(2011). 紧张情绪与认知负荷对前瞻记忆影响的实验研究. *南京工程学院学报*, 11 期, 54-58.
- 张孝义(2012). “留守儿童之家”项目对留守儿童创造力影响的实验研究. *黄山学院学报*, 14 期, 124-127.
- Cohen, J. (1973). Statistical power analysis and research results. *American Educational Research Journal*, 10, 225-229.