

方差分析效果大小报告的新指标

刘铁川 王闪闪

(赣南师范大学教育科学学院, 赣州 341000)

摘要:心理学期刊论文中可重复性不高的现象,原因之一是研究结果的效果普遍较小。并且在报告效果大小的文章中,效果大小指标使用不当。在方差分析中最经常报告的是 η^2 和 η_c^2 ,但是在不同的研究设计中,这些效果大小是无法直接进行比较的。Olejnik和Algina提出了 η_c^2 这个指标,解决了跨不同研究设计可比性的问题,但是国内文献并没有对此进行介绍,并且近五年的主流的几个心理学期刊的结果报告中也没有使用过该指标[1]。我们通过介绍包含操作因素或者测量因素等不同单因素或多因素方差分析设计中 η_c^2 的使用条件,介绍了它的优势,也通过一个实例说明了我们的假设。最后我们期望研究人员在报告效果大小时要考虑到不同的研究设计和研究假设并选择恰当的效果大小指标防止过高估计效果大小。

关键词:可重复性;效果大小; η_c^2 ;可比性

中图分类号:B841.2

文献标识码:A

文章编号:1003-5184(2019)03-0238-06

1 引言

近年来,可重复性(replicability)或重现性(reproducibility)在心理学界引起了广泛重视。可重复性研究指重复实验能够得到与原始结论相同或相似的结果的研究。近几年心理学界的一些知名学者接二连三的学术不端行为的新闻报道引发了心理学界人士对可重复性研究的重视,比如2010年心理学领域的Marc Hauser被指控伪造数据,2011年Daryl Bem被指控选择性筛选数据以及Diederik Stapel的学术造假事件,并且心理学界人士开始反思心理学研究中有多少研究是可靠的、可重复的[2]。2015年,科学开放中心(Open Science Collaboration)从2008年发表在Psychological Science(PSCI)、Journal of Personality and Social Psychology(JPSP)和Journal of Experimental Psychology: Learning, Memory, and Cognition(JEP:LMC)三个杂志的448篇文章中的选取了111(70%)篇文章进行113项重复研究(其中有2篇各包括两个研究),最后完成了100(88%)项重复研究,使用显著性、 p 值、效果大小以及效果大小的元分析进行评估,原始研究中97%的结果是显著的($p < 0.5$),但重复研究中只有36%的结果是显著的,主观评定为重复了原始结果的研究仅为39%[3]。

为了使研究数据结果中的效果达到更大,一些研究人员选择性报告研究结果,一项研究结果发现,大约40%的研究未能完整报告所有实验条件,约70%的研究选择性地报告问卷中的结果变量,报告出来的效果大小是未报告效果大小的两倍,具有统计学显著性的结果是不具有显著性结果的约3倍。

也就是说,相比于报告的效果大小,未报告出来的效果大小较小,差异可能不具有统计学意义[4]。Baker调查了1576位研究人员,其中超过70%的研究人员尝试但未能重复另一位科学家的实验,超过半数的人未能重现他们自己的实验[5]。

对统计推断的误用或误解也是导致可重复性低的原因之一。我们应该开始重新审视心理学的研究方法问题。我们提出假设,收集数据,并检验数据是否与这些假设一致。我们期望永远正确,虽然错误无法完全避免,不过最应该避免的错误就是假阳性,即对零假设的错误拒绝。Simmons, Nelson和Simonsohn揭示了心理学研究假阳性过高的现象及原因,并提出了解决方案[6]。John, Loewenstein和Prelec认为可疑操作(QRPs)是导致心理学研究结果假阳性过高的原因,他们总结了10条可疑操作,如根据显著性检验结果选择性地终止数据采集、选择性报告显著性结果、四舍五入 p 值(例如,0.054报告 p 值小于0.05)、查看结果后决定是否剔除数据、伪造数据来得到统计显著性结果等[7]。然而 p 值表示的统计显著性不代表研究实际显著。国外一些学者最早提出报告效果大小(effect size)来弥补 p 值对统计结果解释的不足。也有学者提出使用置信区间来表示心理学实验的可重复性,区间宽度越小,实验的可重复性越好[1,8]。国内也有学者对此进行了分析和解读,聂丹丹等人也分析了把报告统计结果时尽可能地报告效果大小和置信区间来提高结果的可重复性的必要性(相关概念在后面章节给出)[2,9]。

本研究主要从规范地报告研究结果的角度来分

析心理学论文中可重复性低的这种现象。我们主要介绍的是在报告方差分析研究结果时的一个很重要的指标—— η^2 ，虽然一直不被研究者所重视，但是该指标的结果却能影响我们的研究结论。

2 国内外广义 eta 方的研究现状

从1994年开始，美国心理协会(American Psychological Association, APA)建议在主要结果部分报告效果大小[10]。随后的几年，APA多次重申并规范对效果大小的报告，杂志编辑和学术协会也提出报告效果大小。美国教育研究协会(AERA)实证社会科学研究报告标准指出当报告统计结果时，应包括“变量之间的量化关系的指标(即效果大小)”和“该效果指标的不确定性的说明”，例如置信区间[11]。美国国家教育统计中心(NCES)统计标准规定，当统计结果具有统计学意义时，考虑对效果大小的实质性解释是有用的，为此，观察到的差异可以转化为效果大小，以便解释差异大小[12]。美国心理协会(APA)对效果大小的报告要求力度日益增加。心理学及相关学科的期刊编辑们明确表示要求报告效果大小[12-14]。例如，Educational and Psychological Measurement杂志要求“对具有统计意义的结果需要报告和解释效果大小”[15]；Journal of Applied Psychology杂志指出，如果作者没有报告效果大小以及显著性检验的结果，(编辑)将要求作者提供不报告效果大小的具体理由[16]。同样，Journal of Consulting and Clinical Psychology杂志指出，“报告统计学意义、效果大小以及临床意义时，增强了对心理治疗效果的评价”[17]。Psychological Science杂志的投稿指南要求作者“效果大小应伴随主要结果”。从2014年1月起，心理科学杂志建议使用“新统计量”(效果大小，置信区间和元分析)以避免虚无假设检验相关的问题[18]。此外，报告统一规范(the Consolidated Standards of Reporting Trials)以及非随机设计评估的透明报告(the Transparent Reporting of Evaluations with Nonrandomized Designs)都表明应该报告主要和次要结果的效果大小及置信区间[19,20]。

近年来，国内一些学者也逐渐认识到了效果大小的重要性并介绍了常见统计方法的效果大小指标的使用和计算，心理学学科的几个主流期刊投稿论文的统计报告结果做了要求。

2.1 方差分析效果大小的常见指标

研究人员通过报告效果大小和统计显著性检验来增强其研究结果的呈现。效果大小是一个标准化的指标和估计参数，量化总体之间的差异大小以及自变量和因变量之间的关系强度并且不依赖样

本量。

效果大小种类很多，一般将其分为平均差异和关系强度型指标，基于标准化平均差异，由Cohen的 d ，Hedges的 g 以及Glass的 Δ 表示；当有两个以上水平的独立变量或者是连续变量时，效果大小通常描述的是每个变量所占的方差比例，包括 η^2 (有时也称 R^2)， η_p^2 和 η_c^2 ，相关的 ω^2 ， ω_p^2 和 ω_c^2 ，以及常见的相关指标，比如 r^2 ， R^2 和 R_{adj}^2 。除此之外还有一些不太常见的指标，比如 ε^2 、 f^2 、 φ 和Cramer的 V 等，这里不作展开介绍[21]。本研究主要介绍方差分析中使用的效果大小中指标，包括 η^2 、 η_p^2 和 η_c^2 。 η^2 或者 R^2 ，描述的时观察到的样本量但是可能高估了总体参数， ω^2 或者 R_{adj}^2 ，能够估计抽样总体的变化。因为它们和总体相关，不易受到偶然因素的影响，所以教科书通常推荐使用总体估计值。然而，也因为其值往往比样本统计数据值更小，所以研究人员很少报告总体估计值。

在方差分析结果中报告最多的就是 η^2 和 η_p^2 ，在单因素方差分析中他们的值是相同的，。如，在一个组间设计中，

$$\eta^2 = \frac{SS_{Effect}}{SS_{Total}}$$

SS_{Effect} 是感兴趣的因子的平方和， SS_{Total} 是总的因子平方和。然而在更复杂的设计中，有不正一个因子，这时， η_p^2 可以识别每一个因子的效果大小，

$$\eta_p^2 = \frac{SS_{Effect}}{(SS_{Total} + SS_{Error})}$$

SS_{Error} 是被试内平方和，也称为组内平方和。

2.2 广义 eta 方的提出和使用

Olejnik和Algina整合上述两个指标，为研究者提供了一个一般化形式的效果大小指标，用于包含一个或多个测量因子或操作因子的设计[22]，

$$\eta_c^2 = SS_{Effect} / (\delta * SS_{Effect} + \sum SS_{Measured} + \sum SS_{Subjects/Cov'})$$

$SS_{Subjects/Cov'}$ 包括涉及被试或协变量的平方和，等同于被试间设计的 SS_{Error} ，其他参数含义同上。

如果感兴趣的是操纵因素则 $\delta = 1$ ；如果感兴趣的是测量因素则 $\delta = 0$ 。

η_c^2 与 η^2 、 η_p^2 之间的不同体现在分母上， η^2 分母中包含所有的方差项， η_p^2 和 η_c^2 分母中包含了其中的一部分方差项。一般来说， η_c^2 比 η_p^2 分母中包含的方差项更多。比如重复测量设计时，因为包括被试或被试与测量因素的交互作用，所以 η_c^2 分母更大，所以一般得出的 η_c^2 小于 η_p^2 。

η_p^2 消除了个体差异的其他来源，使其在跨被试

间和被试内设计中不能直接可比。 η_c^2 作为 η^2 和 η_p^2 的替代方案,适用于至少有一个分类独立变量的设计,使得效果大小在不同的设计间可以比较,用于调查来自同一个总体,但是使用不同的分组因子、协变量或附加因子的一个因子或因子间的交互作用。在两种设计中操纵不同的因素,被试间或被试内的操纵因素的设计是不同的,使用分组因子和协变量的设计也是不同的。

下面用几个例子来具体说明效果大小参数及其各部分是如果起作用的,以及在不同的研究设计中如何估计参数。

假设在一个平衡组设计中调查四种过敏药物(因素 D)对智力操作性能的影响的。效果大小即,

$$\sigma_p^2 / (1 \times \sigma_D^2 + \sigma_{s/Cells}^2) = \sigma_D^2 / (\sigma_D^2 + \sigma_{s/Cells}^2)$$

这里, $\sigma_{s/Cells}^2 = \sigma_{Individual-Difference}^2$, 在这个例子中,分母是总的平方和。

但是这个例子并不意味着分母总是包含总方差。现在假设被试根据性别进行分组,那么仍然是两因素平衡设计,效果大小变为

$$\sigma_D^2 / (1 \times \sigma_D^2 + \sigma_G^2 + \sigma_{PG}^2 + \sigma_{s/Cells}^2) = \sigma_D^2 / (\sigma_D^2 + \sigma_G^2 + \sigma_{DG}^2 + \sigma_{s/Cells}^2)$$

这里, $\sigma_{s/Cells}^2$ 是药物和性别组合的标准差,不包括所有由个体差异引起的变异。相反 $\sigma_G^2 + \sigma_{DG}^2 + \sigma_{s/Cells}^2 = \sigma_{Individual-Difference}^2$, 因为性别效应以及性别与因素 D 交互作用反映了个体差异的变化。除了 σ_D^2 一项,上面两个公式的分母是相同的,所以,两个设计中由两个公式定义的参数估计也是可比的。即,尽管两个例子中两个设计是不同的,但是效果大小与用于调查的设计并没有混淆。

2.2.1 广义 eta 方的计算和使用条件

进行方差分析 η_c^2 的计算时,研究中的数据是由两个方差来源引起的:研究中的操作因素和个体差异。个体差异是由于被试稳定、暂时的特性以及实验设置中的不可控特性造成的。例如在一项旨在调查体重训练方案对力量增长的有效性的研究中,治疗方案中个体力量的测量可能因为性别、身体类型和先前活动水平的个体差异而有所不同。它们也可能因为动机水平和个人暂时健康状况的暂时性特征而不同。设备质量或仪器校准等实验设置(如)的差异,也是同一体重训练计划中个人得分差异的一个来源。如果要得到一个在各种研究设计中可比较的效果大小指标,那么效果大小参数要能够识别和调整操纵因素的数量和类型,以及个体差异来源的估计或控制程度上的差异。效果大小参数如下:

$$\frac{\sigma_{Effect}^2}{\delta \times \sigma_{Effect}^2 + \sigma_{Individual-Differences}^2}$$

如果效果只包含操纵因素,那么, $\delta = 1$, 如果效果包含一个或多个测量因素及其交互作用,如性别、性别 \times 测量因素,那么 $\delta = 0$ 。参数 σ_{Effect}^2 定义为分析的方差。比如 J 水平的单因素被试间设计,

$$\sigma_{Effect}^2 = \sum_{j=1}^J (\mu_j - \mu)^2 / J$$

μ_j 是因子的第 j 水平, μ 是均值。 $\sigma_{Individual-Differences}^2$ 是由于测量因素(如性别)、测量因素与其他因素的交互作用以及协变量引起的方差分量之和。需要指出的是,如果 σ_{Effect}^2 是测量因素的主效应,或者是测量因素与其他因素的交互作用的一个方差分量时, σ_{Effect}^2 已经包含在 $\sigma_{Individual-Differences}^2$ 中,设置 $\delta = 0$ 只是为了避免将其计算分母内两次。另一方面,如果 σ_{Effect}^2 是操作因素的主效应,或者仅仅是操作因素之间的交互作用的一个方差分量时, σ_{Effect}^2 则没有包含在 $\sigma_{Individual-Differences}^2$ 中,设置 $\delta = 1$ 将其加入分母内计算,因为操作因素增加了数据的方差。

由此得到,不同设计间可比较的效果大小的公式:

$$G^2 = \frac{SS_{Effect}}{\delta \times SS_{Effect} + \sum_{Meas} SS_{Meas} + \sum_{\kappa} SS_{\kappa}}$$

如果感兴趣的是操纵因素,则 $\delta = 1$; 否则为 0。 $Meas$ 为测量因素的所有的差异来源(比如分组因素或者分组 \times 操作因素的交互作用,不包括被试),并且 SS_{Meas} 是这一类效果的总和。 κ 为所有被试或者协变量的方差来源。 SS_{κ} 是这种差异来源的总和。 SS_{Effect} 是估计因素的效果大小的平方和。

比如在完全随机设计(completely random design, CR)中,所有的因素都是操作因素,对于所有的效应, $\delta = 1$, 并且,

$$\sum_{Meas} SS_{Meas} = 0$$

在多因素完全随机设计中 η_c^2 就等于 η_p^2 。如,一个两因素的设计, A 是操作因素, b 是测量因素(使用小写字母代表测量因素,比如性别)如果感兴趣的是因素 A, 那么 $SS_{Effect} = SS_A + \delta = 1$ 。

$$\sum_{Meas} SS_{Meas} = SS_b + SS_{Ab}$$

并且,

$$G^2 = \frac{SS_A}{SS_A + SS_b + SS_{Ab} + SS_{s/Cells}} = \frac{SS_A}{SS_{Total}}$$

如果感兴趣的是测量因素 $SS_{Effect} = SS_b$, $\delta = 0$ 。

$$\sum_{Meas} SS_{Meas} = SS_b + SS_{Ab}$$

并且,

$$G^2 = \frac{SS_b}{SS_b + SS_{Ab} + SS_{s/Cells}} = \frac{SS_b}{SS_{Total} - SS_A}$$

当两个因素 a 和 b 都是测量因素时,

$$\sum_{Meas} SS_a = SS_b + SS_{ab}$$

η_c^2 就能通过 SS_{ab}/SS_T 计算。

在混合设计中,当被试间设计(组间设计)和被试内设计(组内设计)结合时,通常被称为混合设计。如上所述,重复测量的因素几乎都不是测量因素,但是被试间因素(控制因素)可能是操作因素或测量因素。比如在一个研究中,一个重复测量因素(B),一个被试间因素(A),被试嵌套在因素A水平之内(s/A),如果被试间因素(控制因素)是操作因素,那么, $SS_{Effect} = SS_A \delta = 1$,

$$\sum_{Meas} SS_{Meas} = 0$$

并且, $\sum_{\kappa} SS_{\kappa} = SS_{s/A} + SS_{Bs/A}$

$$G^2 = \frac{SS_A}{SS_A + SS_{s/A} + SS_{Bs/A}} = \frac{SS_A}{SS_T - SS_B - SS_{AB}}$$

这里, SS_T 是每个观察值与总平均的平方差的总和。

Olejnik 和 Algina 提供了包含操纵因素或测量因素的三因素设计公式, Bakeman 强调了这些公式在单因素或多因素重复测量设计中的应用[1,23]。

表 1 A、P、AP、aP 和 PQ 设计的 η_p^2 、 η_c^2 的计算公式

设计	效应	F 比	η_p^2	η_c^2
A	SS_A	$MS_A/MS_{s/A}$	$SS_A/(SS_A + SS_{s/A})$	$SS_A/(SS_A + SS_{s/A})$
P	SS_P	MS_P/MS_{P_s}	$SS_P/(SS_P + SS_{P_s})$	$SS_P/(SS_P + SS_s + SS_{P_s})$
AP	SS_{PA}	$MS_{PA}/MS_{P_{s/A}}$	$SS_{PA}/(SS_{PA} + SS_{P_{s/A}})$	$SS_{PA}/(SS_{PA} + SS_{s/A} + SS_{P_{s/A}})$
aP	SS_{Pa}	$MS_{Pa}/MS_{P_{s/a}}$	$SS_{Pa}/(SS_{Pa} + SS_{P_{s/a}})$	$SS_{Pa}/(SS_{Pa} + SS_a + SS_{s/a} + SS_{P_{s/a}})$
PQ	SS_{PQ}	MS_{PQ}/MS_{PQ_s}	$SS_{PQ}/(SS_{PQ} + SS_{PQ_s})$	$SS_{PQ}/(SS_{PQ} + SS_s + SS_{P_s} + SS_{Q_s} + SS_{PQ_s})$

注:“A”代表的是被试间操作因素,“a”代表被试间测量因素,“P”和“Q”代表被试内因素,“s”代表被试因素。

以上公式中,规定字母(P、Q)表示被试内因素,因其与被试分组的个体差异因素无关,故不会出现小写字母。根据定义,重复测量因素是操作因素。在被试内设计中,被试因素的交互作用作为误差项,不是操作因素,故是小写字母 s。用大写字母和小写字母表示的几个简单的包含重复测量的例子(见表 1)。

表 1 是常用到的重复测量方差分析中各效应大小的计算公式。SPSS 结果中可通过方差分析表直接计算得出 η_c^2 , 也可使用 R 语言中的 ez 包直接进行计算。

3 重复测量方差分析中广义 eta 方应用的一个实例

当设计包括重复测量时, η_c^2 比 η_p^2 更小,正如之前指出的, η_c^2 的分母因为包含了被试平方和以及所

接下来将重点介绍 η_c^2 在重复测量方差分析中的计算公式和使用。

2.2.2 重复测量方差分析中的广义 eta 方

Bakeman 进一步补充了 η_c^2 在重复测量方差分析中的应用[1],

$$\frac{\sigma_{Effect}^2}{\delta \times \sigma_{Effect}^2 + \sigma_{Measured}^2}$$

$\sigma_{Measured}^2$ 是由个体差异产生的方差,如果效应包含一个或多个测量因素则 $\delta = 0$,如果效应只涉及操作因素则 $\delta = 1$ 。所以,效果大小的分母包括涉及测量因素的所有的方差来源,排除只包含操作因素的方差来源。例外情况是感兴趣的效应只涉及操作因素,在这种情况下,相关方差也计入分母中,那么 η_c^2 的计算公式为:

$$\eta_c^2 = \frac{SS_{Effect}}{\delta \times SS_{Effect} + \sum_{Measured} SS_{Measured}}$$

特别是,对于包含重复测量的设计,不只被试平方和,所有的被试 \times 重复测量因素的交互作用也包括在分母中(见表 1)。

有被试与重复测量因素的交互作用,所以,得出的 η_c^2 也会更小。比如,两因素重复测量设计中(PQ 设计),P 效应的 η_p^2 的分母是 $SS_P + SS_{P_s}$, η_c^2 的分母是 $SS_P + SS_s + SS_{P_s} + SS_{Q_s} + SS_{PQ_s}$ (即 $SS_T - SS_Q - SS_{PQ}$)。正如 Olejnik 和 Algina 所说, η_p^2 消除了个体差异的其他来源,使它在被试间设计和被试内设计的研究中不可以直接比较[24]。

举一个具体的例子,《教育与心理统计学》(第三版)P437 一个时间间隔 \times 区组的混合设计中,需检验时间间隔(被试内因素)、区组(被试间因素)是否显著,区组因素(a),与时间间隔即重复测量因素(P),即为 aP 设计。经 SPSS 输出结果整理为面下的方差分析表(表 2)。

表2 方差分析表

来源	平方和	自由度	均方	F 值	p 值	偏 Eta 方
Group(区组)	10046. 7	1	10046. 7	274. 63	0. 00	0. 97
E(Group)	292. 67	8	36. 58			
Time(时间间隔)	3272. 27	2	1636. 13	701. 20	0. 00	0. 99
Time × Group	10. 40	2	5. 20	2. 23	0. 14	0. 22
E(Time)	37. 33	16	2. 33			
总和	13659. 37	29				

SPSS 计算的被试间的 a 和 a/s 的平方和分别是 10046. 7 和 292. 67; 被试内 P 、 P_a 和 $P_{s/a}$ 的平方和分别是 3272. 27、10. 40 和 37. 33。所以, 区组效应 (a) 的 $\eta_p^2 = SS_a / (SS_a + SS_{s/a}) = 10046. 7 / (10046. 7 + 292. 67) = 0. 97$ (SPSS 输出的 η_p^2 为 0. 97); $\eta_c^2 = SS_a / (SS_a + SS_{s/a} + SS_{Pa} + SS_{Ps/a}) = 10046. 7 / (10046. 7 + 292. 67 + 10. 40 + 37. 33) = 0. 97$ 。时间间隔效应的 $\eta_p^2 = SS_P / (SS_P + SS_{Ps/a}) = 3272. 27 / (3272. 27 + 37. 33) = 0. 9887198$ (SPSS 输出的 η_p^2 为 0. 99), $\eta_c^2 = SS_P / (SS_P + SS_a + SS_{s/a} + SS_{Pa} + SS_{Ps/a}) = 3272. 27 / (3272. 27 + 10046. 7 + 292. 67 + 10. 40 + 37. 33) = 0. 24$ 。正如我们预期的, η_c^2 小于 η_p^2 , 但是区组效应 (a) 是时间间隔效应 (P) 的 4 倍。 η_c^2 和 η_p^2 差异的大小取决于测量因素以及测量因素和操纵因素的交互作用的平方和使分母增加的程度。

从这个例子的计算过程中, 能够看出, 在区组效应 (a) 中, η_p^2 没有考虑到重复测量因素即时间间隔 (P) 以及时间间隔 (P) 与区组效应 (a) 的交互作用 $SS_{Pa} + SS_{Ps/a}$, 所以得到的值小于 η_c^2 。在重复测量因素即时间间隔效应 (P) 中, η_p^2 没有考虑到区组效应 (a)、区组效应 (a) 的被试因素和时间间隔 (P) 与区组效应 (a) 的交互作用 $SS_a + SS_{s/a} + SS_{Pa}$, 所以得到的值小于 η_c^2 。

4 讨论和总结

在报告方差分析的效果大小中, η_c^2 的报告率并不高, 尤其在国, η_c^2 的使用几乎为零, 并且近几年心理学期刊论文中暂时还没有看到关于 η_c^2 的文章的介绍。另外我们还有一些关于 η_c^2 的使用条件需要完善。首先, 还没有文章提出针对 η_c^2 的解释的标准。Cohen 提出的针对 η^2 的标准, 即定义 η^2 (与 Cohen 的 f^2 不同) 0. 02 为小效应, 0. 13 为中等效应, 0. 26 为大效应 [23]。尽管 Cohen 没有考虑重复测量设计, 这一标准对 η_c^2 似乎也是合适的。因此, 在没有其他更好的标准之前, 不管是否包含重复测量设计, 我们一般使用的还是 Cohen 的标准。

其次, η_c^2 的使用也是有局限性的。局限性有两个, 即影响到 $\sigma^2_{Individual-Difference}$ 的设计, 要么是因为控

制实验设置特性的变化, 要么是因为抽样于不同的总体。

比如, 在上个例子中, 第一种情况下的一个例子, 过敏药物的研究中, 一组研究药物在早上进行, 另一组被试自行选择在早上、中午或者晚上, 在第二组中时间是没法记录的, 如果存在一天的时间效应, 那么尽管在两个研究中, σ_D^2 是一样的, 但是效果大小是不一样的。第二种情况下的一个例子, 假如过敏药物的研究中一个研究使用 60 岁以上的被试, 另一个研究年龄范围不受控制。虽然在两个研究中 σ_D^2 是一样的, 但是更大的年龄组中 $\sigma_{s/Cells}^2$ 更小了。正如 Olejnik 和 Algina 所说, 当样本来自不同的总体时, 没有效果大小是可比的 [1]。

还需要指出, 尽管 η^2 、 R^2 高估了解释方差的总体比例, 不过研究者经常报告的还是样本统计量 η^2 、 R^2 , 因为它们的公式简单, 直截了当, 并且随着样本量的增加, 高估几乎可以忽略不计。并且, 就像 ΔR^2 修正了 R^2 的过度估计, ω^2 修正了 η^2 的过度估计 [1, 25]。尽管如此, 却并不提倡使用 ω^2 , 因为 ω_p^2 理论上定义很简单, 实际上很难应用。

对于一些简单的设计, η_c^2 、 η_p^2 和 η^2 没有差别。对于单因素被试间方差分析, 不管是操作因素还是测量因素 (A 或 a), η_c^2 都和 η_p^2 、 η^2 的值是一样的。对于操作因素的被试间设计 (AB 、 ABC 等), η_c^2 和 η_p^2 是相同的, 但是如果包含被试间测量因素, 那么 η_c^2 小于 η_p^2 。同样, 对于单因素 (P) 的被试内方差分析, η_c^2 和 η^2 是相同的, 但是对于其他被试内设计或者混合设计, η_c^2 小于 η_p^2 。

总之, 研究者在报告一个或多个重复测量方差分析的结果时, 建议报告 η_c^2 。容易计算并且, 不像 η_p^2 和 η^2 , 不管是被试间设计还是被试内设计, η_c^2 的值是可比较的。

参考文献

- 仲晓波. (2015). 心理学实验的可重复性. 心理科学, (4), 807-812.
- 聂丹丹, 王浩, 罗蓉. (2016). 可重复性: 心理学研究不可忽视的实践. 中国临床心理学杂志, 24(4), 618-622.

- Olejnik, S., & Algina, J. (2003). Generalized eta and omega squared statistics: Measures of effect size for some common research designs. *Psychological Methods*, 8(4), 434–470.
- Miller, G. (2010). Cognition research. Investigation leaves field in the dark about a colleague's work. *Science*, 329(5994), 890–891.
- Committee, L., Committee, N., & Committee, D. (2012). Flawed science: The fraudulent research practices of social psychologist Diederik Stapel.
- Aarts, A. A. (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), 1–8.
- Franco, A., Malhotra, N., & Simonovits, G. (2015). Underreporting in psychology experiments: Evidence from a study registry. *Social Psychological & Personality Science*, 7(1), 1–5.
- Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, 533(7604), 452–454.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False – Positive Psychology Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, 22(11), 1359–1366.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524.
- Cumming, G., & Fidler, F. (2009). Confidence intervals: Better answers to better questions. *Zeitschrift Für Psychologie*, 217(217), 15–26.
- APA, A. P. A. (1994). Publication manual of the American psychological association. *Fourth edition*, 1994. *Psicothema*, 6(4), 557–560.
- Green, J. L., Hedges, L. V., & Levine, F. J. (2006). Standards for Reporting on Empirical Social Science Research in AERA Publications; American Educational Research Association. *Educational Researcher*, 35(6), 33–40.
- Seastrom, M. (2003). NCES statistical standards. Source: National Center for Education Statistics, Institute of Education Sciences.
- Us, A. P. A. W. D. (2001). Publication manual of the American Psychological Association (5th ed.).
- Us, A. P. A. W. D. (2010). Publication manual of the American Psychological Association (6th ed.).
- Thompson, B., Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement*, 54, 837–847. 1994.
- Murphy and K. R. (1997). Editorial. *Journal of Applied Psychology*, 82(1), 3–5.
- Kendall, P. C. (1997). Editorial Journal of Consulting and Clinical Psychology, 65(1), 3–5.
- Schulz, K. F. (2010). CONSORT 2010 Statement: updated guidelines for reporting parallel group randomized trials. *BMC Medicine*, 8(1):p. :18.
- Jarlais, D. C. D. (2004). Improving the reporting quality of non-randomized evaluations of behavioral and public health interventions; the TREND statement. *American Journal of Public Health*, 94(94), 361–366.
- Snyder, P., & Lawson, S. (1993). Evaluating Results Using Corrected and Uncorrected Effect Size Estimates. *The Journal of Experimental Education*, 61(4), 334–349.
- Kirk, R. E. (1996). Practical significance: A concept whose time has come. *Educational and Psychological Measurement*, 56, 746–759.
- Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *in Behavior Research Methods*.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: L.

A New indicator of the Effect Size of Variance Analysis

Liu Tiechuan Wang Shanshan

(Education Scientific College, Gannan Normal University, Ganzhou 341000)

Abstract: One of the reasons for the low repeatability in the journal papers of psychology is that the results are generally less effective. And always improperly use the index of effect size when reporting. A generally useful effect size statistic—generalized eta squared has been proposed by Olejnik and Algina (2003) to solve the problem of comparability across different research designs, but the domestic literature has not introduced it, and this index has not been used in the recent five years' mainstream psychological journals. We introduce the advantages and disadvantages of different factors based on single factor or multiple factors analysis, including manipulated factors or measurement factors, show that it can easily be computed by standard statistical packages, and recommend that investigators provide it routinely in their research reports when appropriate.

Key words: repeatability; effect size; comparability; generalized eta squared