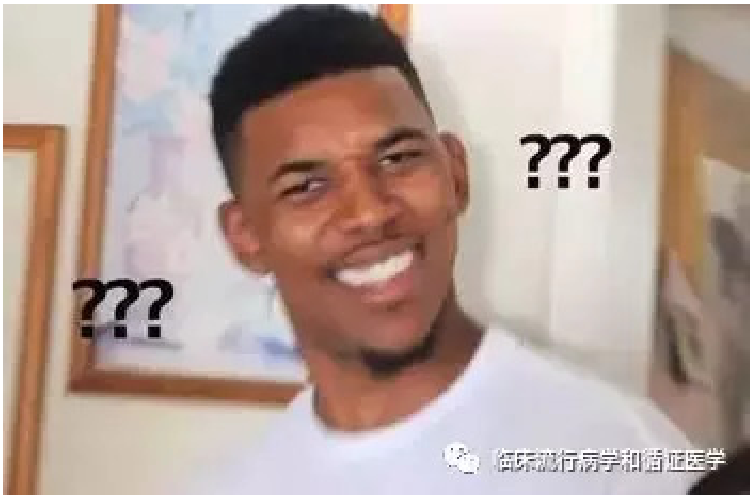
Effect-Size（效应量）——我不是故意让你一脸懵逼

作者：李楠，赵一鸣 来源：临床流行病学和循证医学

本文原地址：https://www.iikx.com/news/statistics/1827.html

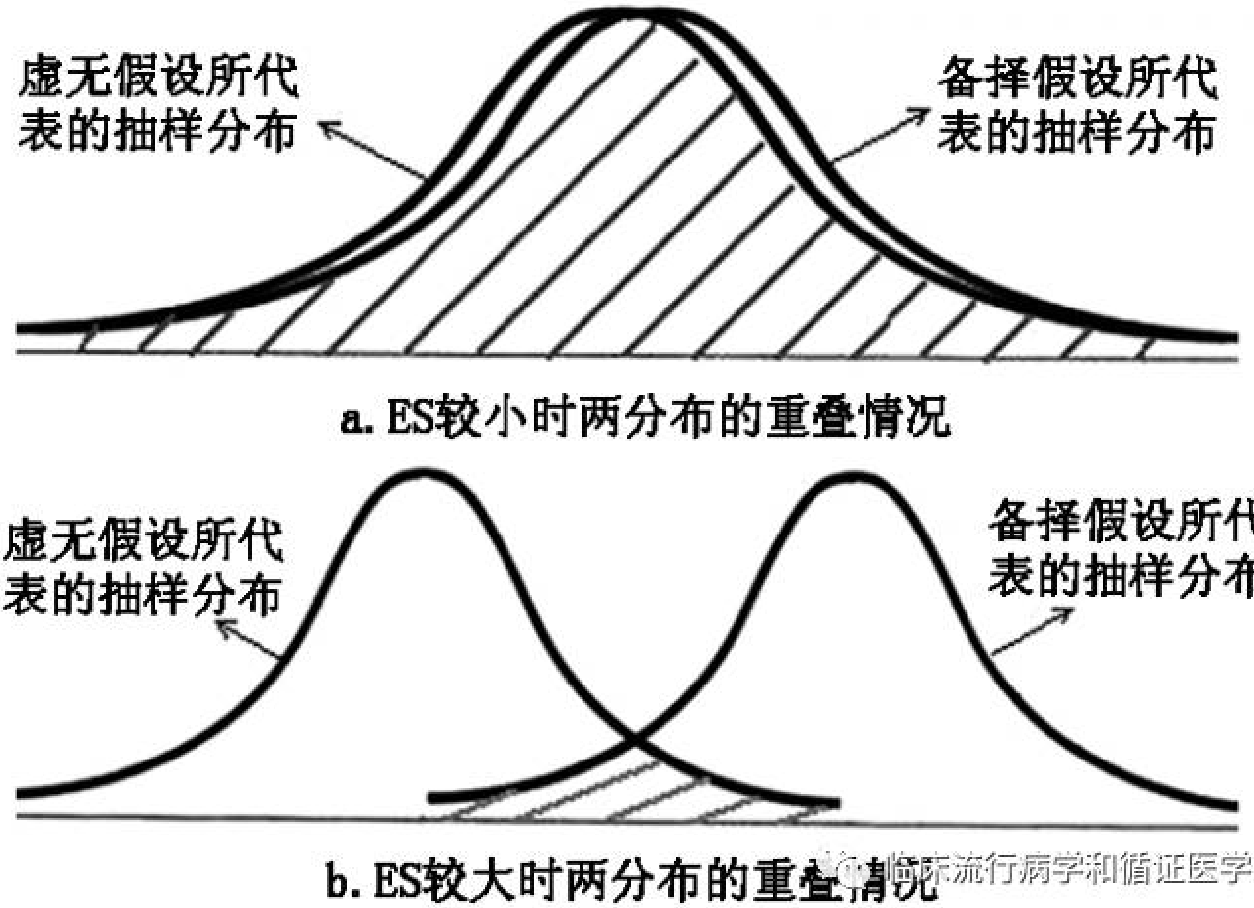
您有没有碰到过这样的问题，文章在投稿后，评审专家让您提供**Effect Size(效应量)**。相信大多数临床医生的表情都是下面这样的。



“ 笑音量?你说笑……笑啥?! ”

无论是我们在进行一个相关性分析时，还是比较组间某一个指标绝对值高低、发生率高低的差异时，我们都习惯了上来就报告p值和统计量。其实这个时候我们已经忽略了一个重要的内容：我们关注了是否存在【效应】(效应是否有统计学意义)，但存在的效应真的【有价值】?真的足够【有临床意义】么?

因此，当我们转换身份成为一个读者的时候，我们除了知道【有没有差别】，还一定想知道【到底差多少】。而后面这个差多少，其实就是效应量。因此我们不难理解，所谓的效应量(也就是效应大小)，其实对应着一个真实的差异或者关联，效应量就是这种真实差异或关联的强度。



比如在这个图中，a、b两个情况下，都有统计学差异。但是到底a的效应更有实际意义，还是b的效应更有实际意义呢?显然是b，也就是b的效应量更大。

也许你觉得好像得到了什么?但是也好像什么也没得到。对了，这感觉就是效应量，效应量就是一个这么抽象的东西。效应量不是一个单一的指标，而是一种真实的差异，这种差异需要通过一些指标(能够反应效应量的指标)来体现。让我们来举个例子：

我们想知道吃不吃维生素对孩子身高的影响。这里面因素是吃/不吃维生素，效应是组间身高的差别。这里面效应量是什么呢?一定是能够反应身高差别的指标啦!但是身高差别都需要什么指标来评价呢?真的只有比较均数么?到底都有哪些指标可以反应组间身高的差别呢?

1、身高均值之差：反应组间集中趋势的差异，最简单粗暴的指标，在实际情况下，身高通常符合正态分布，因此均值之差也是我们衡量的首选效应量。

2、身高变异系数之差：变异系数CV反应的是身高的变异，说白了就是看我们这一组孩子是不是高的高矮的矮。变异系数越小，说明大家的身高差不多。维生素到底是不是会影响到身高的变异情况呢?这个还真没准儿。

3、身高离群值下限之差：反应的是到底是哪一组矮的更厉害。其实补充维生素这类的物质，未必对全人群有效啊，不缺维生素的孩子，多不了可能并不会长得更高。而对于那些却维生素的孩子，补了效果更明显。所以嘛，谁知道到底是不是真正的效应在矮个子里面?

……

虽然这里只是粗暴的打了个比方(各位统计学大拿请随意打脸)。不过，看到这里您一定会明白了，所谓的效应量并不是一个唯一的参数，而且有时候在我们做研究之前也未必完全想清楚了(当然我们希望大家先想清楚)。所谓的效应量，可以理解为真实存在的某种差异，而我们是通过找到合适的效应量评价指标，来说明真实的效应到底有多大。上面的指标只是给大家看看玩儿的，当然对于医学研究，绝大多数的研究中的效应量就是真实的差值(均数差)，或者是RR、OR、HR、相关系数、R ^{2} ……等等。对于一些比较抽象的指标，我们可能无法直观的比较均数差，比如量表的结果。对于这类指标，不同量表可能差值没法直接比较，因此我们就需要得到一些去掉工具影响的大小，这时就需要一些特定的效应量指标来帮助我们了。常见的指标有以下这些。

表 2 效应量一览表

研究目的	研究类型	检验类型	合适的效应量
均值差异比较	两组比较：独立样本或配对样本 t 检验	对比检验	Cohen 的 d 值, Glass 的 Δ 值和 Hedge 的 g 值；点二列相关；I 效应量
	多组比较：单因素或多因素被试间、被试内或混合设计方差分析	对比检验(原假设只有一个等式)	多组比较的差异类效应量；相关效应量 $r_{contrast}$、$r_{alerting}$、$r_{effectsize}$；Contrast 检验的方差比效应量 η^2, ω^2, ε^2 以及它们的偏值
		综合检验(原假设有多个等式)	Omnibus 检验的方差比效应量 η^2, ω^2, ε^2 及它们的偏值；I 效应量
变量之间的相关	两个变量之间的相关研究	对比检验	r、r_{pb}、r_b、$r_{equivalent}$，ϕ 及 Cramer 的 V 系数等基于 χ^2 统计量的相关系数等
	一个变量和多个变量的相关研究(多元回归分析)	对比检验	$f^2=(R_{AB}^2-R_A^2)/(1-R_{AB}^2)$
		综合检验	R^2 或 $f^2=R^2/(1-R^2)$

计算方法可以参考如下的公式：

表 1　标准差异型 ES 指标的计算及使用

估计指标及公式	使用条件
组间设计	
$d = \frac{M_1 - M_2}{\dot{\sigma}}$，$\dot{\sigma}$ 为任意组的样本标准差*	两组别总体方差齐性
$\Delta = \frac{M_1 - M_2}{s_{control}}$，$s_{control}$ 为控制组标准差	理论上能区分出控制组和实验组
$\Delta_1 = \frac{M_1 - M_2}{s_1}$，并 $\Delta_2 = \frac{M_1 - M_2}{s_2}$	理论上无法区分出控制组和实验组，且两组别总体方差不齐性，此时需同时报告两种可能的 Δ 值
$g = \frac{M_1 - M_2}{s_{pooled}}$，$s_{pooled} = \sqrt{\frac{SS_1 + SS_2}{df_1 + df_2}}$	两组别总体方差齐性
$g_{corrected} = \left[1 - \frac{3}{4(n_1 + n_2) - 9}\right]g$	小样本情况下基于 g 值的矫正
组内设计	
$d = \frac{M_D}{\dot{\sigma}}$，$\dot{\sigma}$ 为任意测量条件的样本标准差	非严格意义的组内设计**，且两测量条件总体方差齐性
$\Delta = \frac{M_D}{s_{pretest}}$，$s_{pretest}$ 为前测条件的样本标准差	非严格意义的组内设计，且理论上能区分出前测条件和后测条件
$\Delta_1 = \frac{M_D}{s_1}$，并 $\Delta_2 = \frac{M_D}{s_2}$	非严格意义的组内设计，且理论上无法区分出前测条件和后测条件，两测量条件方差不齐性，此时需同时报告两种可能的 Δ 值
$g = \frac{M_D}{s_{pooled}}$，$s_{pooled} = \sqrt{\frac{SS_1 + SS_2}{df_1 + df_2}}$	非严格意义的组内设计，且两测量条件总体方差齐性
$g_D = \frac{M_D}{s_D}$，$s_D = \sqrt{\frac{SS_D}{df_D}}$***	严格意义上的组内设计
$g_{corrected} = \left[1 - \frac{3}{4N - 9}\right]g$	小样本情况下基于 g 值的矫正

注：* 在 d 值的实际计算中，也常用 $\sqrt{(SS_1 + SS_2)/(n_1 + n_2)}$ 作为两总体联合标准差 σ_{pooled} 的有偏估计量[15,16]，此时 d 和 g 的关系有 $d = g\sqrt{(n_1 + n_2)/(df_1 + df_2)}$。

** 非严格意义的组内设计是指对同一因素的考察在理论上允许以组间设计替换。此时对 ES 的估计类似于组间设计的情况。但对那种关注测量变化量的严格意义的组内设计而言，建议基于差异分数 D 而非原始分数来计算分母标准差。这体现出对研究主题本身的理论思考[14]。

*** g_D 是 Gibbons，Hedeker 和 Davis（1993）将 Hedges 基于组间设计的 g 指标[illegible]$_{pooled}$ $\sqrt{2(1 - r_{12})}$（其中 r_{12} 为两测量条件的积差相关），因此 Gibbons 等人的 g_D 与 Hedges 的 g 指标关系为 $g_D = g/\sqrt{2(1 - r_{12})}$。

表 2　g 值与 t 值、r 值间的转换

设计类型	$g - t$ 转换	$g - r$ 转换
组间设计	$g = t\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$	$g = r_{pb}^2\sqrt{\left(\frac{df_1 + df_2}{1 - r_{pb}^2}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$，$r_{pb}$ 为点二列相关系数（也见表 3）
组内设计	$g = t\sqrt{\frac{2s_D^2}{n(s_1^2 + s_2^2)}}$	当 $s_1^2 = s_2^2$ 时，有 $g = t\sqrt{\frac{2(1 - r_{12})}{n}}$，$r_{12}$ 为两测量条件的积差相关

表 3　关联强度型 ES 指标的计算及使用

估计指标及公式*	使用条件/对应的 NHST	SPSS 操作
非平方尺度		
$\dot{\varphi} = \sqrt{\frac{\chi^2}{n}}$	卡方检验中的四格表分析	运行 Crosstabs 程序，从 Statistics 框中勾选 Phi and Cramér's V
$V = \sqrt{\frac{\chi^2}{\min(R-1, C-1) \times n}}$	卡方检验中的 R×C 列联分析	同上
$r_{pb} = \sqrt{\frac{t^2}{t^2 + df_{within}}} = \sqrt{\frac{F}{F + df_{within}}}$	组间设计的均值差异显著性检验；或组间设计的单因素两水平方差分析；或点二列相关的显著性检验	运行 Correlate 中的 Bivariate 程序，从 Correlation Coefficients 框中勾选 Pearson**
$r_{pearson} = \sqrt{\frac{t^2}{t^2 + df}} = \sqrt{\frac{F}{F + df}}$	组内设计的均值差异显著性检验；或组内设计的单因素两水平方差分析；或积差相关的显著性检验	同上
平方尺度		
$\dot{\eta}^2 = \frac{SS_{effect}}{SS_{total}} = \frac{F_{effect}}{F_{effect} + df_{within}/df_{effect}}$	方差分析	运行 Compare Means 中的 Means 程序，从 Options 框中勾选 Anova table and eta；或同下
$\dot{\eta}^2_{partial} = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$	方差分析中，考察某自变量独立于区组变量（组内设计）和其他自变量时对因变量的边际效应	通过 General Linear Model 运行方差分析，勾选 Options 框中的 Estimates of effect size
$\dot{\omega}^2 = \frac{SS_{effect} - df_{effect} \times MS_{within}}{SS_{total} + MS_{within}}$	平衡设计的方差分析中，自变量数与样本规模之比小于 1∶10 时的矫正指标	手动计算
$R^2 = \frac{SS_{effect}}{SS_{total}}$	多元线性回归	运行 Regression 中的 Linear 程序后系统默认输出
$R^2_{partial} = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$	多元线性回归中，考察某自变量独立于其他自变量对因变量的边际效应	打开 Syntax 窗口，在 Regression 指令中运行/Statistics coeff outs zpp 子命令，对结果中的偏相关系数求平方
$R^2_{adjusted} = 1 - \frac{(n-1)(1-R^2)}{n-v-1}$	多元线性回归中，自变量数与样本规模之比（v∶n）小于 1∶10 时的矫正指标	运行 Regression 中的 Linear 程序后系统默认输出

注：* 对于那些传统心理统计教材中已有的关联指标，如 $\dot{\varphi}$、r_{pb}，我们省略了基于原始数据的计算公式，而尽可能呈现基于 NHST 结果的公式，这些公式同样可进一步用于对二手数据的挖掘整理。

** r_{pb} 的计算在本质上等同于一列变量为连续数据，另一列变量取值为 0,1 中求取 r_{pb} 的途径就是计算两列变量的 Pearson 相关。

参考文献：

1. 卢谢峰, 唐源鸿, 曾凡梅. 效应量:估计、报告和解释[J]. 心理学探新, 2011, 31(3):260-264.

2. 郑昊敏, 温忠麟, 吴艳. 心理学常用效应量的选用与分析[J].心理科学进展, 2011, 19(12):1868-1878.