

智力测验工具的一些标准。现在没有一种测验满足所有这些标准,也没有一种测验可同时满足这些条件。确实智力包括了不同人的不同技能,没有一种测验能完全适合。相反,理想的评定智力的工具应该是对不同类测量的综合。但就是全部综合起来,也不会产生一个确定的 IQ,因为一种工具只对某些

人在某些时候有效。大家所期望的最好的一种评定量表是经过解释后告诉我们的信息能比单一的测量多。

(柳佳 编译自《Testing Intelligence Without IQ tests》by Mr Sternberg)

再分析研究方法浅谈

美国社会心理学家格拉斯(G.V.Glass)大概是最早提出再分析研究方法的学者。他认为,对于数据的分析可以概括为三种不同性质的活动:初步(primary)的分析,第二步(secondary)的分析和再分析(Meta-analysis)。在一项给定研究课题完成以后对所得数据的第一次分析为初步的分析。第二步的分析是在此基础上对研究结果的重复分析。在第二步的分析中往往可以运用到更合适的统计手段,从而能够在原有数据集合的基础上对一些新的问题作出回答。最后,如果研究者想把几项相互独立的研究结果整合起来作出结论,就要用到再分析手段。

再分析研究方法的发展史可以回溯到在社会科学领域普遍采用统计手段的年代。早在1938年,费希尔(Fischer)和皮尔逊(Pearson)提出一种方法,使研究者可以从一系列的研究中得到一个普遍的概率值。1954年,摩斯特勒(Mosteller)和布什(Bush)在他们的《社会心理学手册》一书中曾提出过一种再分析过程。不过,过去很少看到再分析的研究实例,而近年来再分析的研究报告逐渐增加。罗森塔尔(Rosenthal, 1976)的一篇再分析报告,考察了人际间对行为期望的影响。这一报告收集了三百多项独立的研究。格拉斯(1976)运用再分析手段对教学方法、电视教学、社会经济地位与智商的关系作了考察。史密斯(Smith)和格拉斯(1977)对375项有关心理治疗效果的研究作了综合分析。格拉斯由此认为,行为治疗比非行为治疗没有明显的优越性。

可以看出,再分析研究就是收集一系列研究报

告,而这些研究或者是在概念上具有共同的假设,或者在实现自变量或因变量上可以共享同样的操作。再分析研究方法的结论是基于一系列所考察的文献,但它能够给出一个统计上的显著性水平,并由此提出这样一个问题:“显示这些结果的一系列独立的研究,能够产生怎样的概率值?”不同的再分析研究手段在精确程度上是不同的。精确度最低的方法仅仅“数一数”(counts)研究中低于.50 概率水平的数目,然后探讨这个数目是否太大。精确度最高的再分析方法需要用到研究报告的原始数据。在这里介绍的再分析方法是寻求所包括进来的一系列有效研究的概率水平。先将这些水平值转换为 Z 分数(标准常态离差),而后联系到所包括进来的研究数目。介绍这种方法是因为它比“数一数”的方法要精确得多,同时,比用原始数据的方法更简便易行。这一方法由斯托弗(Stouffer 1949)启用,摩斯特勒和布什(1954)对它作了加工修改。

加权的和不加权的斯托弗方法

斯托弗方法组合一系列研究是采用这样的公式:

$$Z_{ma} = \frac{Zs_1 + Zs_2 + \dots + Zs_n}{\sqrt{N_s}}$$

其中, Z_{ma} 是再分析的标准正态离差; $Zs_1 \dots Zs_n$ 是各包括进来的研究的标准离差; N_s 是所包括进来的研究总数。这一方法要求文献考察者采用以下步骤:1. 记录下每个与相关假说有关的研究所报告的概率水平值;2. 在任何一册基本统计书中都有 Z 分数(即标准正态离差)表,由此查出每个概率水平值所对应的 Z 分数;3. 将这些 Z 分数相加并除以研究数目的平方根;4. 把得到的再分析 Z 分

数 Z_{ma} ，通过 Z 分数表，记录下相应的概率水平值。这个 Z_{ma} 表达的是所有归纳进来的研究结果在机遇水平上的概率值。这里需要使用者注意的是， P 值对于假设的大小和方向要在运用这一手段之前选择好。也就是说，在收集的研究报告中，支持假说的双尾 P 值在查相应 Z 分数时应减半；同样，反对假说的 P 值也要减半并要在支持假说研究求和时作减法运算。

斯托弗方法简便易行，优点很多。正如摩斯特勒和布什（1954）指出的那样：如果文献综述者在文献范围内能对不同的研究报告作出加权，这一方法可以构造得更加精确。根据 Z 值所基于样本的大小，或其它合乎需要的特点，如内部和外部的有效性等等，Stouffer方法允许使用者对每一个标准常态离差作出加权。这样，文献综述者可以在进入统计之前，对每一项研究在方法学的精确度上作出判断。如果使用者愿意，那些被看作“严谨稳妥”的研究可以比那些看作“不够严谨稳妥”的研究作出更大的加权。

当样本大小作为加权因素时，加权的斯托弗方法公式：

$$Z_{ma} = \frac{N_1 Z_{s1} + N_2 Z_{s2} + \dots + N_n Z_{sn}}{\sqrt{N_1^2 + N_2^2 + \dots + N_n^2}}$$

这里， $N_1 \dots N_n$ 为所包括进来的各项研究中的被试数。在各项 Z 分数相加以前分别乘以各项研究的被试数，然后除以各项研究被试数平方和的平方根。这一公式给出的加权标准是所包括各项研究的被试数目，因为它是最为客观的加权标准。当然，其它加权标准也有人（如Stickell, 1963）成功地采用过，就不在这里介绍了。运用被试数目作加权标准意味着研究范围越大，它的重要性也就越强。与前面不加权的方法相比，如果少量被试研究得到的 Z 分数比大量被试研究得到的 Z 分数大，那么用这一公式得到的再分析正态离差就相对小些。

描述关系强度 (relationship strength) 的方法

很久以来，对心理学研究的旁观者常常对心理学文献中缺少关于两组受试关系强度的报导感到惋惜。的确，我们评估一项研究重要性的标准通常是两组被试差异的显著性水平，而这一水平值又与研究样本的大小紧密相连。现在问题是不仅要看到两组间“是否”有差异，还应看到这一差异

“有多少”。而它们之间与样本大小无关的关系强度（亦可称做效应大小 Effect-size）应该成为文献综述的组成部分。特别是在已经推断出两组间有相关存在时，评估这一强度就更有意义。

科亨（Cohen, 1977）提出了关系强度的测量方法，它对各种研究设计和文献分析都适用。即：

$$d = \frac{M_a - M_b}{SD}$$

在这里， M_a 和 M_b 分别为所进行考察比较的两组平均数， SD 为所有受试共同的标准差。选择 d 做为指标是因为它简便易行而且“与量表无关”。也就是说，采用不同测验量表的研究结果可以用它互相比较。

这里得到的 d 指标是以两组受试的共同标准差为单位，给出了两组间平均数差异的距差。举例来说，如果 $d = 0.3$ ，就表明两组间平均数相距十分之三个标准差。在科亨的术语中， d 值等于0.2相应地代表两组间关系强度“较小”， d 值等于0.5，两组间关系强度“中等”，而 d 值为0.8则代表强度“较大”。后来，他对这一值又以两组间“重叠分布百分率”（用 u_3 表示）作出改进调整。

然而，读者也许会问，如果所看到的报告中没有给出平均数和标准差，只给出了 t 值，如何得到关系强度值呢？弗雷德曼（Friedman 1968）给出了在这种情况下估计 d 值的方法：

$$d = \frac{2t}{\sqrt{df}}$$

其中 t 为研究报告中的 t 值， df 为有关的自由度。这样同样可以得到两组间的关系强度。

总之，从历史的角度看，再分析研究方法可以初步给出这样一些统计：一、斯托弗 Z 分数；二、加权的斯托弗 Z 分数；三、作为描述关系强度的 d 值等等。近年来，随着计算机的广泛使用，再分析研究方法迅猛发展，出现了许多有关的统计软件包，而所统计的指标也更加广泛和复杂。相信这一方法对我们整理和总结原有研究工作会有积极的促进作用。

（中国科学院心理研究所周林编译，原文载《Journal of Personality and Social Psychology》1979 Vol.37, No.1 作者H.M.Cooper）