

# 元分析

定量地综合某一领域的各项研究成果，一直是心理科学研究发展的重要子目标。虽然早已有研究者尝试对各项在方法、被试、程序、统计方法等诸多方面不同的研究进行分析和综合，但这类分析、综合大多为描述性质，无法给出一个定量的结论。而元分析则在定性分析的基础上引入了定量分析方法，能够在定量层面上综合各项独立研究的成果，从而形成一个综合结论。现在，越来越多的研究者已经开始从传统的文字综述方法转向使用元分析这种对研究进行定量综合的方法了。

## （一）元分析的基本内容

元分析是对已有同类研究结论进行综合评价、分析，整合，以获得普遍性、概括性结论的方法。元分析的提出离不开对有关成功复制的传统观点的重新审视，正是基于对成功的实验复制的新见解，研究者们才提出了元分析的主要指标——效应度。

### 1. 关于成功复制的观点

在心理学研究中，积累研究成果的关键在于对实验的成功复制。一个科学的心理学实验通常是可以被成功复制的、可被验证的。什么情况下才能认为是成功地复制了一个研究？关于实验复制的传统观点有两个主要特征：①将显著性水平作为一项研究的概括统计量；②用二分法来判断实验复制的成功与否。例如，如果两次研究结果都达到或都未达到 $p < 0.05$ ，就认为复制是成功的；若一次是 $p < 0.05$ 而另一次是 $p > 0.05$ ，则认为复制失败。在这种二元决策中，除了 $p \leq 0.05$ 和 $p > 0.05$ 或一些我们认为可以“拒绝零假设”的重要的显著性水平外，其他水平上的 $p$ 的差别都不在考虑范围之内。这种只依据是否达到某一显著性水平来判断实验复制成功与否的方法已受到越来越多研究者的批评。

现在，人们对“成功的实验复制”有了新的认识，该观点的两个主要特征为：①认为一项研究最重要的概括统计量是效应度而非显著性水平；②以连续的方式（如置信区间）来评估一个实验复制成功与否。例如，并不说两个研究是否成功或不成功地复制了对方，而是用从这两个研究中得出的效应度的差异程度来反映

复制成败的程度。这与传统观点中对研究结果的含义的理解是不同的，元分析中的效应度等概念正是建立在这种对研究结果的理解基础之上。

## 2. 元分析的重要指标——效应度

效应度是研究者基于对成功的实验复制新见解，所提出的主要元分析指标。效应度是指实验处理的效果大小，它是元分析中的重要指标，并与传统统计分析方法中的显著性检验有一定联系。在元分析中，对研究结果进行评估需要同时涉及到效应度和统计显著性两个指标。二者之间的大致关系为：显著性检验=效应度×研究规模。从方程中可见，任何显著性检验都可以由一个或多个效应度乘以一个或多个研究规模得到。而且样本数目越多，研究规模越大，结果就越显著。只有当效应度真正为零时，大样本所得的结果才不会比小样本的更显著。但实际研究中很难碰到效应度正好等于零的情况。

在对效应度进行计算时，需要根据实际情况来选择不同的方法。当研究中用以进行显著性检验的统计量不同时（如  $t$ 、 $F$ 、 $\chi^2$  等），所对应的效应度的计算公式也是不同的（具体的公式请参见相关书目）。

在效应度中有两个重要的族： $r$  族和  $d$  族。 $r$  族中包括分别适用于各种不同情况的皮尔逊积差相关（如  $r$ 、 $\phi$ 、 $r_{pb}$ 、 $\rho$  等）；而  $d$  族的三个主要成员则是 Cohen's  $d$ 、Hedges's  $g$  和 Glass's  $\Delta$ （它们的具体计算公式在这里不作介绍）。当原始研究中的效应度估计量是用  $r$  类指数来表示时，应使用  $r$  类效应度来作为检验工具；而当原始研究是在两组间进行比较时，使用  $d$  类效应度估计值更好一些。在元分析中，效应度估计值通常是  $r$  类指数和  $d$  类指数的混合。这两类估计值是可以互相转化的，在元分析中通常需要将所有的效应度估计值转化成一类指数，一般是使用  $r$  族中的  $r$  或  $Zr$ ，或  $d$  族中的 Hedge's  $g$ （或 Cohen's  $d$ ）。

## （二）评价

效应度的提出和应用为评价、分析和综合心理学中的同类研究提供了有效的工具，进而促进了心理学理论和实证研究的丰富和发展。元分析的这种有效性主要体现在以下几个方面：（1）可以对某一研究领域内的同类研究进行定量的综合，

这是元分析最明显的优势，而以往的综合分析（包括最优秀的研究者所做的最好的研究综述）则主要是一些定性的描述，这种定性描述使我们在对研究进行总结时，难免会遗漏掉许多有价值的信息和资料；（2）摆脱了过分强调单一研究结果的趋向，以往人们过于热衷单一研究的结果，他们认为当 $p \leq 0.05$ 时，这个研究就是“好的”、“有价值的”，而当 $p > 0.05$ 时，该研究就什么都不是。而元分析的方法则指出，其实任何一个研究的影响都不应仅仅从 $p$ 水平的基础上来评估，而应该更多的依赖其自身的效应度；（3）降低了人们对某一统计显著性水平的绝对依赖，以往包括现在我们仍在采用的依据某一显著性水平来接受或拒绝零假设的方法是存在问题的，因为有关显著性 $p$ 的统计概率密度函数是连续函数，对于拒绝零假设，两次 $p$ 值为0.06的结果要比一次为0.05的结果更为有力，而10个为0.10的 $p$ 值也比5个为0.05的 $p$ 值更有说服力，从这点意义上来说，依据效应度来进行判断的元分析更为科学。

正是基于上述优点，元分析方法在心理学综合性研究中被广泛应用，不过有一些研究者对元分析仍持悲观态度，他们认为元分析存在下列局限：（1）取样偏差，也就是说，元分析研究多采用已发表的研究论文中的数据作为其分析的原始数据，然而已发表的论文不能代替所有的研究，即元分析在其数据取样上缺乏普遍性；（2）信息的损失，元分析只使用效应度来总结整个领域的研究成果，难免会失去很多有价值的信息；（3）方法和性质的异质性，元分析所纳入的各项独立研究具有不同的自变量、因变量或不同类型的样本单元，虽然有些元分析研究已将这些不同之处作为缓和变量（moderator variables）进行处理，但是各项独立研究间的异质性必定会使其在某种程度上缺乏可比性。

我们认为，尽管批评依然存在，但元分析作为一种可以对研究结果进行定量分析和概括的综合方法，改变了以往仅以单一研究的结果作为依据的研究传统，使得具有较高价值的普遍性结果或结论能够从以往积累的众多研究中被提取出来，这对于实践的积累和理论的发展均具有重要的推动作用。