

<<实用数据再分析法>>

图书基本信息

书名：<<实用数据再分析法>>

13位ISBN编号：9787562442967

10位ISBN编号：7562442967

出版时间：2008-1

出版时间：重庆大学出版社

作者：马克.W.利普西

页数：205

版权说明：本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问：<http://www.tushu007.com>

<<实用数据再分析法>>

内容概要

可以把再分析理解为一种形式的调查研究，不过在这种研究中，调查对象是一些研究报告而不是人。这需要提出一种编码表，需要收集到由多项研究报告构成的一个样本或总体，其中的每一项研究都要被认真地阅读它的编码员“访谈”，并对有关其特征和定量结果的信息进行编码。然后依据对常规统计技术的特定修正技术来分析出现的结果，以便探究、描述在所选择的一系列研究的各项结果中出现的模式。

<<实用数据再分析法>>

作者简介

马克.W.利普西 (Mark W.Lipsey) , 1972年获得约翰霍普金斯大学 (Johns Hopkins University) 心理学博士学位, 现任范德比尔特 (Vanderbilt) 大学公共政策系教授。

他的研究领域: 青少年犯罪的风险和干预问题以及项目评估研究中的方法论质量问题。

其著作有《评估: 方法与技术》(第6版)等。

利普西教授是《评估动态》(New Directions for Evaluation)的前主编, 同时也是《美国评估杂志》(American Journal of Evaluation)等多家著名刊物的编委会成员。

获美国评估协会保罗。

拉扎斯菲尔德 (Paul Lazarsfeld) 奖。

<<实用数据再分析法>>

书籍目录

第1章 导论 再分析的应用情境 效应值——一个关键概念 再分析的优势所在 再分析的缺点 再分析法的近期历史与当代应用 本书概要第2章 问题指定和研究回溯 辨别需要进行再分析的诸项研究结果的形式 研究的合格标准 再论方法论质量 区分、锁定并回溯研究报告 搜索参考文献 回溯研究报告第3章 效应值统计量的选择、计算和编码 效应值统计量及其方差 关于符号的注解 研究结果的类型和可利用的效应值统计量 单变量关系(集中趋势描述) 双变量关系 变量之间的相关 多变量关系 效应值统计量:小结第4章 提出编码框架,对研究报告实施编码 编码计划书的提出 分析单位和编码的等级层次 效应值的编码 研究描述项 关于编码过程本身的编码信息 培训编码员第5章 数据管理 再分析数据文件的生成 在计算机中直接编码 用计算机维护文献目录 再分析数据文件的结构 创建一个用于分析的单个平面文件 合并多重文件用于分析第6章 分析的问题与策略 分析的各个阶段 效应值的修正 分析效应值均值与分布 效应值异质分布的分析 统计上相依的效应值的分析 额外的分析问题第7章 再分析数据的计算技术 均值、置信区间和同质性检验 对效应值异质性分布的分析 加权回归分析 制图技术第8章 再分析结果的解释与使用 解释效应值 关于效应值大小的经验方法 把效应值转换成其他量纲 临床意义和实践意义 解释再分析结果时的注意事项 研究基础在方法论上的适当性 实质性特征与方法论特征的混淆 方差的重要性 研究之间的差与可推广性 抽样偏差 再分析对于实践和政策的含义附录A 以计算机为基础的文献目录服务部门与相关数据库实例附录B 根据合格的研究报告计算效应值的步骤 标准化的均值差效应值 ES_{sm} 的直接计算 与 ES_{sm} 在代数上等价的公式 t 值或 F 比率的精确概率水平 根据一个频次分布计算均值和标准差 基于连续数据的近似值——点二列相关系数估计(X_1-X_2)和 $Spooled$ 二分化的数据 相关系数效应值 ES_r 在定义上的公式 离散数据或分组连续数据的联合频次分布 一个二分测量和一个连续测量 两个二分测量 近似值与概率值 机率比效应值 基于单元频次的计算 基于行比例的计算 基于单元比例的计算 从相关系数和边缘比例导出 2×2 列联表 根据卡方和边缘比例导出 2×2 列联表 从连续性数据中导出机率比附录C 微软Excel效应值计算程序附录D 进行再分析的SPSS宏指令附录E 少年犯挑战项目再分析实例的编码手册和编码表 研究层次的编码手册 样本描述项 研究设计描述项 治疗描述项的性质 效应值层次的编码手册 因变项测量描述项 效应值数据 研究层次的编码表 样本描述项 研究设计描述项 治疗描述项的性质 效应值层次的编码表 因变项测量描述项 效应值数据 推荐阅读 参考文献索引

<<实用数据再分析法>>

章节摘录

第1章 导论 1952年, 汉斯·艾森克 (Hans Eysenck) 论证了精神疗法对病人没有什么有益的影响 (Eysenck, 1952), 从而在临床心理学领域发起了一场激烈的争论。

20世纪70年代中期出现了数以百计有关精神疗法的研究, 所产生的结果既有积极的、无效的, 又有消极的, 对这些令人困惑的结果的述评已经无法解决这场争论。

为了评价艾森克的观点, 吉恩·V·格拉斯 (Gene V. Glass) 对375项有关精神疗法的研究中的治疗组—控制组之间的差异 (treatment—control difference) 在统计上进行了标准化和平均化处理, 称自己的方法为“再分析” (meta—analysis)。

他和同事玛莉·李·史密斯 (Mary Lee Smith) 在现已成为经典的论文中阐述了他们的研究成果, 结论认为精神疗法确实有效 (Smith and Glass, 1977)。

艾森克不相信这种方法, 称之为“超级愚蠢的操练” (an exercise in mega—silliness) (Eysenck, 1978), 并进而试图质疑该方法。

尽管艾森克和其他学者都持批判态度, 但再分析法如今已被广泛接受, 并成为行为科学、社会科学和健康科学领域中用来对经验研究结果进行汇总的一种方法。

大约在格拉斯研发他的再分析法的同时, 罗森塔尔和罗宾 (Rosenthal and Robin, 1978) 在人际关系的预期效果领域、施密特和亨特 (Schmidt and Hunter, 1977) 在职业测试 (employment tests) 的效度推广领域也分别开发着与这种综合性研究类似的一些统计方法。

再分析这个专有名词开始包含由以上这些以及其他学者提出的所有综合性的定量研究方法和技术。

自20世纪70年代的开创性研究以来, 学术界又出现了数以千计的再分析研究, 并且在再分析的方法论方面也有诸多重大进展。

再分析的应用情境 可以把再分析理解为一种形式的调查研究, 不过在这种研究中, 调查对象是一些研究报告而不是人。

这需要提出一种编码表 (调查计划书, survey pro—tocol), 需要收集到由多项研究报告构成的一个样本或总体, 其中的每一项研究都要被认真地阅读它的编码员“访谈”, 并对有关其特征和定量结果的信息进行编码。

然后依据对常规统计技术的特定修正技术来分析出现的结果, 以便探究、描述在所选择的一系列研究的各项结果中出现的模式。

在从各个学科中选择出来的多种学术研究中, 有多种方法可用来汇总、整合和解释学术成果, 再分析不过是其中的一种, 并且它有一种重要的, 但在某种程度上是限定性的应用领域。

第一, 再分析仅适用于经验研究; 不能用它来汇总一些理论性的论文、常规性的研究综述以及政策建议等。

第二, 它仅适用于产生了定量结果的经验研究, 也就是说, 这些研究利用了变量的定量测量, 汇报了诸多用来汇总最终数据的描述统计量或推断统计量。

上述规定就排除了个案研究、民族志研究和“自然主义的”研究。

第三, 再分析是一种对汇总相关的研究结果的诸多统计量进行编码和分析的技术, 这些统计量典型地出现在一些研究报告中。

如果能够得到我们感兴趣的 research 的全部数据集合, 那么通常情况下直接采用常规的步骤来分析它们即可, 这是比较合适的, 并且此方法能够提供丰富的信息, 用不着对一些汇总性统计量进行再分析。

另外, 由于再分析关注的是不同研究结果的聚集和比较, 因而有必要保证那些研究结果具有一种有意义的可比性。

这意味着这些结果必须: (a) 在概念上具有可比性, 即处理的是相同的构项和关系, (b) 以相似的统计形式呈现。

例如, 在有关抑郁症治疗效果的一系列研究中, 如果能够判断各种治疗手段之间可进行有意义的比较, 并且治疗结果采用相同的基本形式——例如都是对治疗组和控制组中抑郁的对比进行测量, 那么在这些情形下, 可以对这些研究进行再分析。

在同一种再分析中一般不包括具有显著差异的主题, 例如把对抑郁症的治疗手段的研究和有关空间可

<<实用数据再分析法>>

视觉方面的性别差异的研究放在一起进行再分析，这就是不恰当的。

在再分析中这一点经常被指称为“苹果和橙子”问题，即试图对那些实际上处理的是不同的构项和关系的研究进行汇总和整合。

与之类似，把来自不同研究设计并以不同的统计量形式出现的诸多研究结果组合起来，即便这些研究处理的是相同的课题，一般来讲这也是不适宜的。

例如，利用治疗组和控制组之间的比较而进行的有关治疗抑郁症的实验研究通常不会与观察研究结合在一起，后者探讨的是抑郁水平与所接收的服务水平之间的关系。

虽然这两类研究在某种形式上都涉及抑郁症与治疗方式之间的关系，但是它们在研究设计、构成这些研究结果的那些定量关系的实质，以及那些结果的含义等方面的差异是如此之大，以至于很难将它们整合在同一项再分析中。

当然，有学者可能针对每类结果分别采取适当的步骤，进而对实验研究结果和相关研究结果分别进行再分析，并且围绕这两种再分析得出一些结论。

这样做是合理的。

在一项再分析中包含的系列研究结果必须来源于实践前提和概念前提方面具有可比性的研究设计，再分析用各种“效应值”（effect sizes）的形式来代表每项研究的结果。

一个效应值就是一个统计量，它对来源于每个相关研究结果的关键定量信息进行编码。

不同类型的研究结果通常要求不同的效应值统计量。

例如，相对于那些对多组对象在因变量的均值上进行的比较研究来说，那些生成双变量关系的研究更能够利用不同的效应值统计量来进行再分析。

与之类似，有些研究结果针对单个对象样本汇报出前一后均值差，这些结果仍然要使用一个不同的效应值统计量，并且还有另外一些更专业的统计量。

在已知各种统计形式具有可比性的情况下，下一步需要界定哪些研究结果出于再分析的目的而在概念上具有可比性，而这种界定在关注者的心里经常是固定的。

在概念上不同的多项结果在一个分析者看来却可能相似。

例如，格拉斯对心理治疗的有效性的再分析（Smith and Glass, 1977）就受到批评，因为他把来自于性质完全不同的领域，如认知行为学、心理动力学、格式塔心理学等的诸多结果混合在一起。

格拉斯声称，他的兴趣在于考察类型广泛的各种心理疗法的全面有效性，并对不同的类型进行比较，因此在再分析中要表述出所有类型的结果。

另外一位分析者的兴趣范围可能稍小一些，例如他可能仅针对恐蛇症的脱敏疗法的研究结果进行再分析。

然而，无论哪种情况，分析者都必须对兴趣的范围有一个界定，并且阐明是把诸项研究包含在再分析之内，还是排除在外。

其他人可能批评这种界定和阐述，但是只要这些界定和阐述是明确的，那么每位评论者都会自行判断这些做法是否有意义。

效应值——一个关键概念 假定一系列定量研究的结果处理的是同一个课题，并且包含的研究设计可相互比较，那么对于想把这些结果编码成为一个数据库，从而能够进行有意义的分析的学者来说，还存在一个重要问题。

除了个别的例外，这种研究的一些关键变量将不会使用相同的操作化（测量步骤）。

例如，假设我们选择了针对抑郁症治疗的有效性的群体对比研究。

有一些研究可能使用贝克抑郁量表（Beck depression inventory）作为输出变量，某些可能使用汉密尔顿抑郁自评量表（Hamilton rating scale for depression），一些可能使用治疗者对抑郁的评估，还有一些研究可能采取有关这种构项的其他特殊但合理的测量。

这些差异很大的测量产生诸多不同的数值，这些值仅仅相对于所使用的特定操作化和量表来说是有意义的，那么以什么方式对它们的定量结果进行编码，从而允许对结果在统计上进行组合和比较呢？

答案与再分析的本质特征有关，实际上，这种特征使再分析成为可能，并且展现了整个再分析过程据以围绕的轴心。

在再分析中，用来对不同形式的定量研究结果进行编码的各种效应值统计量是以“标准化”这个概念

<<实用数据再分析法>>

为基础的。

效应值统计量对诸多研究结果进行统计标准化，从而使得到的数值在涉及的全部变量和测量中保持一致，具有可解释性。

这种情境下的标准化与我们在测验和测量中谈到的标准分具有同样的意义。

例如，我们可能把数学测验的分数转换成百分数，或在一个样本值的标准差基础上转换为标准化的Z值，从而能够与另一类变量，如阅读成绩进行有意义的对比。

约翰尼（Johnny）的数学成绩可能处在第85个百分位上，而在阅读方面则仅仅在60个百分位上。

与这种方式类似，在再分析中，最常用的一些效应值统计量要对所关注的测度值的样本分布的变化进行标准化。

因而，与治疗者的平均评价值以及所有其他关于抑郁症的此类定量测量一样，在贝克抑郁量表中治疗组和控制组之间的均值差也可以用标准差单位来代表。

在标准差单位的量纲中，我们可能对来自不同的测度和操作化的结果进行组合和比较。

利用贝克抑郁量表的一项研究可能表明，治疗组和控制组之间的差是0.3个标准差，而对于采用治疗者自己的评价值的一项研究来说，该差值也许是0.42个标准差。

假定对应的各个样本据以抽取的总体是相同的，我们就可以比较这些数字，在统计分析中用它们来计算均数、方差、相关系数等值，并通常把它们看成是代表同一事件的有意义的指标，就本案例来说，在治疗小组中回应者体验到一定的抑郁量，相对于估计到的总体抑郁变动量而言，在控制组中也有一定的抑郁量，二者之差就是上述的有意义的指标。

因此，再分析的关键在于定义一个效应值统计量，它能以标准化的形式代表一系列研究得到的诸多定量结果，从而允许在各项研究之间进行有意义的数字比较和分析。

这里存在许多种可能性。

把研究结果进行二分处理，找到在统计上显著的结果和不显著的结果，这种二分就是一种基本形式的效应值。

稍有不同的另外一类便是针对每一种统计显著性检验得到的P值（如 $P=0.03$ ， $P=0.50$ ）（Becker，1994）。

然而，这些都不是非常好的效应值统计量值。

较好的值应该既表征关系的大小，又表明其方向，而不仅仅是统计上的显著性。

另外，它们是被明确界定的，从而相对来讲很少与其他问题如样本量等混淆在一起，尽管样本量在显著性检验的结果中是举足轻重的。

为了给有待考察的一系列研究中特定的研究设计、定量结果的形式、变量和操作化等提供恰当的标准化，再分析者应该利用一种效应值统计量。

在某种情境下，可利用多种效应值统计量，但是在实践中，只有少数能够被广泛应用。

经验结果可归为许多类，大多数结果都落入其中的一类，可针对此类提出多种特定的效应值统计量和相关的统计程序并得到广泛认可。

第3章将定义一系列有用的效应值统计量，同时也给出最适用于使用这些统计量的研究条件。

再分析的优势所在 为什么人们应该考虑利用再分析法来总结并分析一批研究成果，而不用常规的研究综述技术？

总的来说，如下四个原因构成了再分析的主要优势。

第一，再分析程序在对研究结果进行汇总的过程方面施行一种有用的准则。

好的再分析本身被实施成为一种结构化的研究技术，因此要求对各步都要进行记录并接受审查。

它涉及制定一种标准的详细说明，由它来界定所讨论的研究结果的总体，进而区分并回溯合格的、有条理的研究策略，对诸多研究特征和结果进行正式的编码，以及对数据进行分析以便支持得到的结论。

通过这种明确并系统的研究总结过程，研究者就能评估作者的假定、程序、证据和结论，而不是无条件地相信结论的正确性。

第二，再分析法表征着一些重要的研究结果，相对于依赖定性总结的常规研究综述过程或者依赖统计显著性的“唱票法”（vote—counting）而言，再分析法的表征方式更为独特和复杂。

<<实用数据再分析法>>

对于一系列研究中的每种相关的统计关系来说，通过对其大小和方向进行编码，各个再分析的效应值便构成一个变量，该变量易受到各项研究中不同质量的结果的影响。

相比之下，那种使用统计显著性来区分那些发现了效应的研究和未发现效应的研究的做法容易引起误解。

统计显著性既反映了所估计的效应的大小，也反映了围绕该估计值的抽样误差，而后者几乎是关于样本量的一个函数。

因而，由于低统计功效的存在，小样本研究可能发现一定数量虽在统计上不显著但有意义的效应或者关系（Lipsey, 1990; Schmidt, 1992, 1996）。

第三，再分析法还能够发现一些隐藏在其他汇总性研究中的效应或关系。

对多项结果进行定性的、叙事性的总结尽管包含丰富的信息，但它本身却不能对各项研究之间的差别和在各项研究结果中的差异进行细致的审查。

然而，在再分析中，具体做法是对各种典型的研究特征进行系统编码，这允许对诸多研究结果和诸如回答者的特征、治疗的性质、研究设计、测量步骤这样的特征之间的关系进行一种解析性的精确考察。

进而言之，通过估计每一项研究中的效应值，汇总各项研究中的估计值（对较大的研究给予较大的权重），再分析会比个体研究产生拥有更大统计功效的综合性的效应估计值。

因此，在诸多研究之间达成共识的一些有意义的效应和关系，以及关系到各项研究之间的差异的差分效应（differential effects）都更容易被再分析法而不是不太系统的、解析性的方法所发现。

第四，针对来自有待汇总的研究结果的大量信息，再分析用有序的方式来处理这些信息。

当研究的项数或者从每一项研究中提取出来的信息量超过某个相当低的临界值的时候，那么记笔记或索引卡片编码等方法就无法有效地记录全部细节。

相比之下，再分析法的系统编码程序，以及为了记录结论信息而构建的电脑化数据库就几乎拥有无限的容量记录来自每一项研究的详细信息，并且能够包含大量的研究。

例如，由本书作者之一所进行的一项再分析就生成了一个数据库，它包含了大约500项研究，每项研究都用150多项信息来记录（Lipsey, 1992）。

然而，我们马上就补充的是，再分析不要求有多项研究，它在某些情况下可有效地应用于少至二项或三项研究结果的分析之中。

<<实用数据再分析法>>

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:<http://www.tushu007.com>