

第一编 心理学研究基础

研究心理学的目的在于理解人们的所思所行。但是,如何来理解人们的所思所行,就要运用科学的心理学研究方法。虽然“研究”并不陌生,但人们对研究的理解却各不相同。研究一词最初来源于英语词汇 research,意思是运用科学方法,以解答问题或验证假设的历程。科学方法的含义是指一套系统化的过程或步骤,心理学研究方法则是指收集与分析心理学资料的方式或手段,在整个心理学研究中有着举足轻重的作用。

本书第一编“心理学研究基础”的目的在于为其他章节的论述打下基础,第一章“心理学与科学”将讨论心理学的性质和科学研究的主要特征以及心理学研究中涉及的伦理问题。第二章“选题与取样”的内容包括心理学研究课题的来源,选题的原则和怎样查阅研究文献,还包括取样的技术。为了增加对事物的了解,科学家们试图发掘变量之间的相互关系模式,这种关系常常呈现为因果关系。第三章“假设与解释”将讨论如何提出假设和检验假设,各种研究设计中的概念和变量的类型,以及研究中的各种范型。第四章“测量”是第三章的延伸,一旦我们确定了术语的定义,接着就要测量术语所涉及的内容。心理学家为收集信息和提供理论解释提供了大量完备的技术,测量就是其中之一。这一章将讨论测量的层次、测量的效度和信度问题以及测量中存在的误差。这些论述为后面章节的分析技巧提供了理论基础。总括而言,本编提供了心理学研究的一些基本概念。在读完第一编后,就应该准备学习心理学研究的具体方法了。

第一章 心理学与科学

心理学是研究心理现象的科学。从研究对象的角度看，它兼具自然科学、社会科学和意识科学的特点，从而决定了心理学研究方法的多样性，多种研究策略都可以应用到心理学的研究中去。不过需要注意的是，真正的心理学研究必须是科学研究。这也是区分心理学与其他貌似心理学的一些非科学或伪科学的重要指标。没有受过科学心理学训练的人常会受影视文艺作品等的影响而对心理学研究产生种种误解，把心理学研究看得很神秘甚至很玄虚。其实，心理学是一门科学，心理学的研究与一般的科学研究并没有本质的不同。要科学认识心理学的研究就必须对心理学的性质、什么是科学研究等问题有一个基本的认识。在本章中，将讨论三个问题：

1. 心理学的性质
2. 心理学的科学研究
3. 心理学研究中涉及的伦理问题

第一节 心理学的性质

一门学科的性质会影响该学科的研究方式与特点，如物理学与教育学的研究就不大一样。因此，要了解心理学（**psychology**）的研究方法，首先需要对心理学的性质有所了解。我们可以从不同角度来认识心理学，以下三点对于科学全面地认识心理学的性质是非常重要的。

一、兼有自然科学和社会科学的性质

19 世纪中叶以后，自然科学的发展特别是生理学的发展为心理学脱离哲学而成为一门独立的学科准备了实验技术和研究方法。1879 年，冯特在德国莱比锡大学建立了世界上第一个心理学实验室，标志着心理学的诞生。从独立的那一天起，心理学就兼具自然科学与人文、社会科学的性质。心理学的这一特点是由心理现象兼具自然性和社会性所决定的。

首先，心理现象的产生有其生物基础。人的神经系统、内分泌系统和遗传基因等自然物质都会影响人的心理。例如，大脑中存在愉快中枢，如果通过插入电极让老鼠压杆以对自己施加刺激，老鼠会疯狂地操作，反应率可高达每分钟压杆 100 次。通过对患有先天肾上腺增生的儿童的研究（Berenbaum & Hines, 1992）

发现，与没有高水平雄性激素的 3~8 岁的女性亲属相比，在胎儿期或新生儿阶段就显现出高水平雄性激素的女孩更爱玩男孩玩具。在 5 个国家进行的 5 项涉及 24 000 对儿童的大规模双生子研究中，同卵双生子与异卵双生子在外向性上的平均相关分别为 0.51、0.18，在神经质上的平均相关分别为 0.46、0.20 (Loehlin, 1992)。这表明遗传对个体人格特点的形成有着重要影响。

其次，心理现象的产生也有环境基础。人不是生活在真空中，人们所处的社会文化环境、人际关系、社会地位、社会舆论、风俗习惯以及时尚潮流等都影响着人们的心理与行为。如，美国的小孩可以对父母直呼其名，但这种行为在中国文化下就绝对不允许。又如，班杜拉等人 (Bandura et al, 1961) 的研究发现，儿童的攻击行为受到成人榜样行为的影响。在这项研究中，观察到成人榜样暴力攻击芭比娃娃的儿童中，男孩每人平均有 38.2 次，女孩平均有 12.7 次模仿了榜样的身体攻击行为，这些特定的身体攻击行为在无攻击行为榜样组和控制组则几乎没有发现。其实，环境因素不仅影响着个体的心理和行为特点，甚至还决定着个体能否表现出正常的心理功能。如，同样是人，被动物抚养长大的人如狼孩、熊孩等就会表现出爬行、吼叫等动物的行为特征，而且智力低下、寿命短暂、语言表达有障碍。

二、心理学是具有不同分析单位的学科

每一项科学研究都是在一定分析单位的基础上进行。所谓分析单位 (unit of analysis) 就是具有所要研究的那些特征的事物。事物从宏观到微观有不同的层次，不同层次的事物又具有不同的特征，这些特征通常是归属于不同的学科来研究，因此不同的学科都有自己的特定分析单位。如，社会学通常的分析单位是群体，生理学通常的分析单位是机体的器官、组织和细胞等。

心理现象的产生与从宏观到微观的多种层次的事物都有关系，各种心理现象的特征也会由于产生这些心理现象的事物所处的不同层次而表现出不同性质的特点，因此心理学的分析单位是多层次的。归纳起来，可以概括为群体、个体和分子三个层次。群体层次的心理学研究是针对两个或两个以上的联合体所具有的心理特征，包括群体决策、群体凝聚力、群体沟通、非正式群体、民族心理、舆论、从众、时尚、流言、合作等问题。个体层次的心理学研究是针对单个个体所具有的整体性的心理特征，包括人格、个性、精神风貌等问题。如卡特尔 (Cattell, Raymond Bernard 1905—1998) 对人格的 16 种基本维度的研究。分子层次的心理学研究是针对某一个别心理现象、行为片段或生物物理与生物化学的特征，包括感觉的元素、短时记忆容量、视觉感受性、愤怒的维度、刺激—反应的形成、神经递质的传递和染色体的变异等问题。

三、既重视内省也重视客观观察

俗话说“人心隔肚皮”，这表明个体的心理活动是非常主观的，常常是只有个体自己才能真正体验到。个人的经验、态度、价值观和计划等心理活动如果个体自己不报告，他人难以准确、全面的了解。因此，通过个人报告并记录自己的意识内容以作为研究资料的内省法（**introspection**）受到科学心理学的创建者冯特 Wundt, Wilhelm 1832—1920）的青睐；现代认知心理学发展出的口语报告法（**verbal report method**），也是利用被试自己报告的心理活动过程和内容来进行研究；在心理测量中通过自陈问卷（**self-report inventories**）让被试以自我评定的方式回答问题以获取资料是使用最多的一种测量方式。总之，心理学研究非常重视通过内省来获取个体的主观资料以进行研究。

个体的心理活动是主观的，但是也有客观反应。与心理活动密切相关的各种行为反应如脑电活动、眼动、皮肤电、反应时、速度、正确率、腺体分泌、神经递质的活动、肢体运动和面部表情等都可以反映个体的心理活动。如，我们都会通过观察表情来了解对方是高兴还是悲伤；我们也可以通过观察个体解决问题的情况来了解他是聪明还是愚笨。总之，通过对各种行为反应的客观观察，不仅可以了解个体自己不能准确报告的心理现象如短时记忆容量、知觉特性、注意广度，还可以了解个体不愿做真实报告的心理活动，如“心理学与生活”栏目中介绍的测谎系统。

因此，心理学研究是既要听其言也要观其行，既重视主观资料也重视客观资料，通过结合这两个方面就能比较客观、全面、准确地认识心理现象。

心理学与生活 1-1 测谎系统的原理及应用

完整的测谎系统由所谓的“测谎仪”和“测谎问卷”两个部分组成。测谎系统的基本原理是：当主试向被试提问，如果被试说谎，其心理上必然会产生紧张、恐惧、焦虑等情绪，这些情绪的产生会引起生理指标的改变，用仪器记录上被试在整个提问期间的生理变化曲线图，就能判断被试是否说谎。

因此，测谎仪也就是记录呼吸率、心率、脑电、血压、皮肤电等生理指标的仪器，其又常被称为“多道生理记录仪”。最初的测谎仪只能测几个生理指标，现在的能测几十种，目前司法心理学家还准备将事件相关电位 P300 波、N400 波甚至脑成像技术等最新的生理测试方法也加进去，以提高测谎的准确性。

测谎问卷技术用于引发被试的情绪，常用的有两种方法：罪识测试法和准绳问题测试法。此外，还有紧张峰测试法、探索测试法等。罪识测试法的原理是：向犯罪嫌疑人呈现与犯罪有关的人、物、事的实物、照片或情景描述时，能引起较

强的生理反应；当呈现无关的实物、照片或情景描述时，激发的生理反应较弱。而对无辜的人来说就不应该存在这种差异。准绳问题测试法的原理是：通过向被试提出包括“相关问题”和“准绳问题”在内的一系列问题，来比较被试在这两类问题上生理变化曲线的差异，以评定被试是否说谎。相关问题就是与调查事件直接相关的问题，即要想鉴别的核心问题。准绳问题是与相关问题在同一范围内的有关社会文化道德的用来触发反应以设定被试紧张水平的基准线的问题。此外，在准绳问题测试法中还可以有“中性问题”、“象征问题”等。

测谎技术在美国应用很早，20世纪80年代以来我国也开始引进和应用这项技术，近几年来，该项技术在我国司法实践的应用越来越多，其不仅被用于了刑侦，也被用于法院对有关证据的鉴别。测谎技术的应用案例很快见诸报端，并引起人们的关注与兴趣，一些不正确的认识与看法也出现了。有研究（潘军、李焰 2001）显示，虽然有关测谎系统有较好的信效度，但其结果也不是绝对准确，而且测“实”比测“谎”难度大，在民事类案件中更是如此（民事类案件中说谎的确定率小于92%，诚实的确定率小于60%）。而且测试结果还会受到被试因素、主试因素（如测谎技能和熟练水平，对结果的不恰当期望等）、事件因素等的影响。因此，测谎技术是司法实践的有益帮助与参考，但也不能绝对化、神圣化，不然反而会导致司法错误。

第二节 心理学的科学研究

科学研究（scientific research）是人类探寻真理的活动，它与一般人了解世界、获取知识有所不同。要了解科学研究，先让我们看看人们在日常生活中有哪些非科学的获取知识的方法，然后再讨论科学研究的特点。

一、一般人的探索方法

我们每天都在与外部世界打交道，而外部世界中总有许多我们未知的东西，因此需要我们不断地进行探索。一般人并不总是使用科学方法来探索世界，他们通常使用的方法归纳起来有常识、传统和权威三种。

（一）常识

常识是个体在日常生活中直接获得的普通知识。它涉及到生活的方方面面，为人们所广泛接受。人们在解决日常生活中的一般问题时所利用的知识，通常都来自常识。例如，我们乘火车旅行、洗脸刷牙、一日三餐、做家务、防火防盗、看病吃药、锻炼身体、判断他人的性别等行为通常都在常识的指导下完成。

常识是人们进行探索活动的工具，人们常常会根据常识来解释和解决自己

遇到的新现象、新问题。如，当一种新的流行病爆发时，人们根据常识就可以知道消毒、隔离以及提高自身抵抗力等方法预防。因此，常识可以成为人们进一步探索世界的起点，帮助人们快速地认识和解决问题，并减少不必要的麻烦。

但是，常识也会阻碍人类的探索。首先，有不少常识是错误的。如，有不少人认为个子矮的人就聪明，就缺乏科学依据。常识还容易让人们不求甚解，仅满足于表面的了解。如，苹果熟了就会自己掉在地上是一个常识，但是在牛顿之前却没有人去深入思考这个常识背后的客观规律。常识还可能将人们的探索活动引入歧途，如“太阳东升西落，而且看上去比地球小”也是一个常识，但是这个常识却让人们长期误以为太阳是围绕地球而运动。

（二）传统

传统是世代相传、具有特点的社会风俗、习惯、道德、艺术和观念等。文化传统告诉我们大量历来如此并且众所周知的知识。我们每个人都都在继承某种文化传统，经常从传统中获取知识。如“门神驱鬼”、“善有善报、恶有恶报”、“艰苦朴素”、“24节气”、“清明扫墓”、“黄道吉日”、“猴年出生的孩子聪明”、“鸡年不宜婚嫁”等等。

传统是人们在长期的实践中总结经验而积累形成的知识，其中很多是有道理的，如24节气就基本符合气候变化的规律，可以据此来指导耕种和生活。因此，继承传统可以节约我们探索世界的时间和精力，为我们的探索活动打下一个良好的基础。

但是，传统也可能妨碍我们的探索活动。传统中的一些错误，甚至是迷信的东西会让我们的探索活动误入歧途。如，看面相知吉凶是一种传统文化，但据此进行面相与行为结果关系的研究就会误入歧途，浪费时间和精力。又如，根据“鸡年是寡妇年，不宜婚嫁”的传统，来研究生肖与婚姻成败的关系也是没有价值的。此外，传统还可能束缚思想，妨碍我们提出创新性的观点和探索新的领域。在传统的领域提出新的观点或者超越传统的领域都常常会被看作是离经叛道，甚至会受到压制，如弗洛伊德刚提出潜意识和性本能概念时，就被认为是大逆不道的胡说，其思想也长期未能得到认可。

（三）权威

权威是一种使人信从的力量和威望。在人们的心目中长者、领袖、专家、政府、书籍以及报刊杂志等都有一定的权威。这些权威可以向我们提供我们所不知道的知识，纠正我们的错误，遇到疑难我们也会主动寻求权威的帮助。

相信权威有助于我们快速地认识世界、解决疑难，是我们探索世界的重要方式。如当生病时，听从医学权威的意见，就可以治愈，而不必自己去探索。又如，当学习中遇到疑难时，听从名师的指点，就可以很快解决问题，甚至达到“听君一席话，胜读十年书”的效果。

但是，“尽信书则不如无书”，如果完全听从、依赖、甚至迷信权威的意见就会严重阻碍探索活动。因为权威的意见并不就一定就正确，特别是在有争议的问题上就更是如此。任何人都会犯错误，权威也是如此。如果权威们超出自己的专业领域去发表意见，就更容易让我们误入歧途。如，电影明星所做的保健品宣传，宗教人士对进化论的评价，文学家对经济形势的预测，等等，都会妨碍我们对有关问题的正确认识。此外，如果完全依赖或听从权威的意见，还会导致懒惰思想，束缚创新性思想的产生和发展。例如，那些完全相信和服从老师的话的学生，通常都是缺乏创造性的。

总之，常识、传统和权威在人们的探索活动中都是双刃剑，一方面，它们为人类的探索提供了基础和工具，可以用它们来指导我们的生活；但另一方面，依靠它们又会妨碍进一步的探索，甚至使探索活动误入歧途。这也是一般人的探索活动不如科学研究可靠、有效的关键原因。那么，为什么在探索活动中依靠科学研究就不会出现上述问题呢？这是由科学研究的特征所决定的。

二、科学研究的特征

前面讨论了一般人的探索方法，这些方法与科学研究的重要区别就在于，它们不具备系统性、重复性、证伪性和开放性这四种科学研究的典型特征。这同时也是心理学的科学研究特征。

（一）系统性

科学研究与一般人的探索活动的一个显著区别在于它是系统地进行研究。任何科学研究都是一项系统工程，心理学研究也是如此，它主要包含 5 个基本步骤：(1) 选择课题和提出假设；(2) 设计研究的方案；(3) 收集资料；(4) 整理和分析资料；(5) 解释结果以及检验假说（参见图 1.1）。从图 1.1 可以看出科学研究的 5 个步骤构成了一个研究系统，可以让科学研究生生不息，不断进行下去。不仅如此，在科学研究的各个步骤也体现出系统性。

第一，在选择课题和提出假设阶段，研究者必须在系统的查阅文献或做实际调查的基础上选择有价值的课题。然后根据已有的科学知识体系对问题的解决提出假设。第二，在设计研究方案阶段，研究者需要根据已有的科学方法摸索出一种可以证实假设的有效方法，并要决定如何去测定假设中涉及的现象，选择什么样的被试作为研究对象，如何控制或排除无关因素的干扰，研究实施的时间安排、物质条件，等等。总之，研究者必须全面考虑研究中的各种问题，设计出一个具有系统性的研究方案才能保证研究的顺利进行。第三，在搜集资料阶段，研究者还必须系统地观察和记录资料。系统观察具有以下特征：(1) 具有明确的研究目的或假设；(2) 有系统、有组织地进行；(3) 通常会借助一定的科学观测工具，将观察与测量结合起来；(4) 要有意识地避免主观和偏见。这与我们日常生

活中的非系统观察是不一样的。我们在日常生活中的观察通常是零散的、资料收集过程相当粗糙，观察到的事实常常缺乏代表性。比如，如果问你们所见到的那个罪犯的长相怎样，很少有人能够清楚地说出来。但是从事这一课题的科研人员就可以拿出种种证据来加以说明。第四，在整理和分析资料阶段，研究者需要对原始资料进行整理和分类，使之系统化和简约化。在资料分析过程中还需要利用统计手段，对数据进行系统的数学分析，以对数据进行有效的描述并对数据的意义进行推断。第五，在解释结果以及检验假说阶段，研究者不能孤立地阐释结果，而应该将研究结果与已知的事实或理论知识联系起来进行系统的解释，以说明研究结果对研究假设的证实情况，并使研究结果成为科学知识系统的新内容。最后，还可以据此发现新的研究课题。

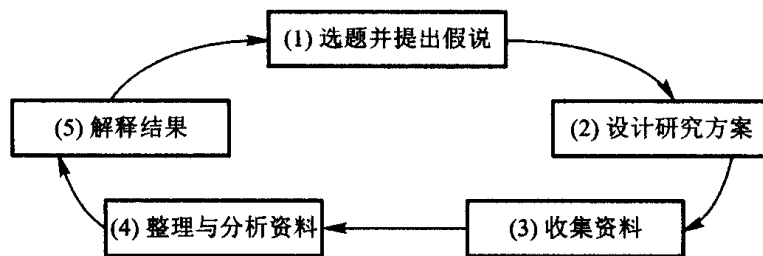


图 1.1 心理学研究的基本步骤

(采自黄希庭 ,1991)

(二) 重 复 性

科学研究不仅是一项系统的活动，而且这个系统活动的过程和结果都是可以重复的，即不同研究者运用同样或类似的程序和方法可以得出一样的结论。如，短时记忆容量是 7 ± 2 这个结论是科学的，就是因为得到了大量研究的重复验证，心理学家们通过使用数字、字母、音节、字词等各种实验材料，向被试短暂呈现后，请被试回忆，结果被试的回忆量都在 5 至 9 这个范围。

要实现科学研究的可重复性，首先就要求研究是具有可操作性的，即要有测量或操纵研究对象的客观方法。显然，如果不能用可感知、可度量的指标和方法来测量和控制研究对象，那么研究过程就无法重复。比如，“鬼怪对人类行为影响的研究”就是无法重复的，因为鬼怪是无法观测和控制的。我们既不能观测到鬼怪的数量和样子，也不能控制鬼怪的出现与消失。短时记忆容量的研究则可以重复，因为可以通过“短暂呈现，立即回忆”的方式控制被试是使用的短时记忆，还可以通过记录被试的回忆数量来测量短时记忆容量的大小。

重复还是证实研究结果的客观性和真实性的重要方法。任何研究都需要人去观测和记录，主观因素和错误就在所难免，因此，即使在同一项研究中通常也需要反复进行操纵和观测才能得出结论。如，测量桌子长度这样的简单工作，都

必须反复测量并取平均数才可靠，仅凭一次的测量就下结论是容易出现差错的。因此，不能重复的研究是不可靠的，不能重复就难以检验其真伪，这样的研究结果可能是观测的误差、偶然、作伪或其他额外因素造成的。比如，有人宣称他观测到了一次鬼怪，但别人观测不到，他自己也没能再次观测到，那么这样的结论就不可靠，这样的研究就不是科学研究。

总之，研究过程可否重复，研究结果可否在相同情况下再次观测到，是判断研究科学性的重要标准。因此，科学研究十分重视方法问题，科研论文都要求对研究方法和过程有详细的交待，以便其他研究者可以重复验证。

(三) 证伪性

所有科学研究都是在一定条件下进行的，因为，科学规律和理论都只能在一定的条件下成立，逾越其适用范围，科学结论就会变成谬误。因此，科学结论不是放之四海而皆准的，它总是在一定条件下被证实，一旦超出了这一条件就会被证伪(falsifiable)例如，韦伯定律表明，在原来刺激强度的基础上，要想增加刺激强度以感觉到刺激有所差别，所增加的强度应是原来刺激的一定比例，其公式是 $\Delta I/I = K$ 。I 是原来刺激的强度， ΔI 是刺激的增量，K 是常数。例如用手提起一个重量，在 300 g 重量的基础上，需增加 10 g，才能刚刚感觉到这一重量比 300 g 重些。因此，提重的韦伯比例是 1/30，这个比例是一个常数。当时认为韦伯定律是一个普遍规律，对任何重量都适用，但是，后来人们发现韦伯定律并不是普遍规律，它只对中等强度的刺激范围有效。

不受任何条件限制，宣称自己的结果是放之四海而皆准的，那肯定是不科学的。例如，弗洛伊德的恋母情结好像是一个普遍真理，一个人孝顺父母是一个恋母情结，一个人打他的父母也是恋母情结，因为他要把压抑发泄出去，这似乎都出自他的童年期经验。

科学是不怕出错、不怕被证伪的，它总是在对已有观点的证伪过程中发展。通过对原有理论的证伪性研究，人们才一步步清楚地认识到有关理论或规律的准确适用范围。

(四) 开放性

学术乃天下之公器，科学研究是对众人开放的。科学研究的开放性(openness)主要表现在四个方面：第一，多视角，即从多种角度来研究客观现象。只要是用科学的方法进行研究，能够提供实证的材料来说明问题，任何角度的研究都是允许的。科学研究从来就不规定看待问题的角度，并且鼓励发散性思维，要求研究者以开放的头脑来看待问题。心理学研究中就更是这样，我们可以从生理的、行为的、认知的、现象学的等多种角度来研究心理现象。第二，公开性。科学研究是一项公开的活动，其方法、过程和结果都是公开的，并强调成果的交流与共享。公布与交流科研成果是科研活动的重要内容和目的。科研工作者也都乐

于将自己的研究成果与方法通过期刊杂志、学术会议和网络等媒介公布，并与大家交流，共同探讨研究的经验与教训。同时，科研工作者也乐于与人共享自己的科学发现，认为科研成果是公共财产，是属于全人类的，创造新知识是人类共同的责任。第三，可争辩性。科学研究承认任何科学知识都只是相对真理，都需要进一步的修正，因此允许并鼓励理性的质疑和争辩。对于任何观点、资料和方法，科研工作者都不会随便接受，科学研究的所有过程和结论都可以受到质疑和反驳。科学通常也就是在质疑与争辩之中发展的，如人本主义的兴起就是源于对精神分析和行为主义的质疑。第四，科学研究无禁区。只要是能够客观观测的事物都能够纳入科学研究的视野，对于科学研究本身来说是没有禁区可言的。这也是科学研究与其他探索世界方式的一个重要不同之处。比如根据宗教的观点，人类的起源问题就是禁区，因为神造人是神圣不可动摇的；根据伦理的观点，克隆也是禁区，因为这会带来伦理问题。

补充讨论 1-1 心理学研究方法的多样性

心理学的种种特点决定了心理学研究的方法是非常丰富的，如奥尔波特（Allport, Gordon Willard 1897—1967）就总结出了 14 大类共 52 种人格心理学的研究方法，如表 1-1 所示。

表 1-1 人格研究方法概观

14 大类	52 种方法
1. 文化模式研究	1. 社会规范分析;2. 成语、格言、文艺作品分析;3. 语言分析;4. 心理描述(形容词核对、量表分析)
2. 生理记录	5. 遗传分析;6. 生物化学相关物;7. 内分泌学研究;8. 体型;9. 面型、动作分析
3. 社会记录	10. 个人档案记录(学校、医院、工职、资历、组织等);11. 工作分析;12. 时间分配;13. 行为频率;14. 社会测量学;15. 拓扑心理学(对人、对阻碍物的反应)
4. 个人记录	16. 日记;17. 自学系统指导;18. 个人信件;19. 主题写作
5. 表情活动	20. 第一印象;21. 外表过细分析(快速摄影分析);22. 外表模式分析;23. 字相学;24. 风格分析
6. 量表	25. 等级量表;26. 记分量表;27. 心理图示
7. 标准化测验	28. 标准化问卷;29. 心理测验(动作测验、迷津测验、语言测验);30. 行为量表(想象、联想、情境测验等)
8. 统计分析	31. 差别心理学;32. 因素分析;33. 内部因素分析
9. 生活情境微型	34. 时间样本;35. 职业微型;36. 欺骗性情境
10. 实验室实验	37. 一元记录;38. 多元记录
11. 预测	39. 外观预报;40. 趋势预报
12. 深层分析	41. 精神科晤谈;42. 自由联想;43. 梦的分析;44. 催眠术;45. 潜意识书写;46. 幻想分析

续表

14 大类	52 种方法
13. 理想型	47. 理解的图式;48. 文艺性格分类
14. 综合法	49. 辨别法;50. 匹配法;51. 全过程分谈;52. 个案分析

(采自陈仲庚、张雨新，¹⁹⁸⁶)

从表 1-1 中可以看出，心理学的研究方法涉及范围非常广泛，包括：从研究个体的外部表现到内心活动的方法，从研究某些个别特征到对个体的整体研究的方法，从收集资料、分析资料的方法到科学方法论。这些方法其实也是一个连续体，在每一项研究中都不可能仅使用一种方法。如，既要用调查或实验的方法收集资料，也要用统计方法、逻辑思维方法来分析资料，整个研究还需要科学方法论的指导。总之，心理学的研究方法可以概括为互为联系互为基础包含三级的系统 如图 1.2 所示。

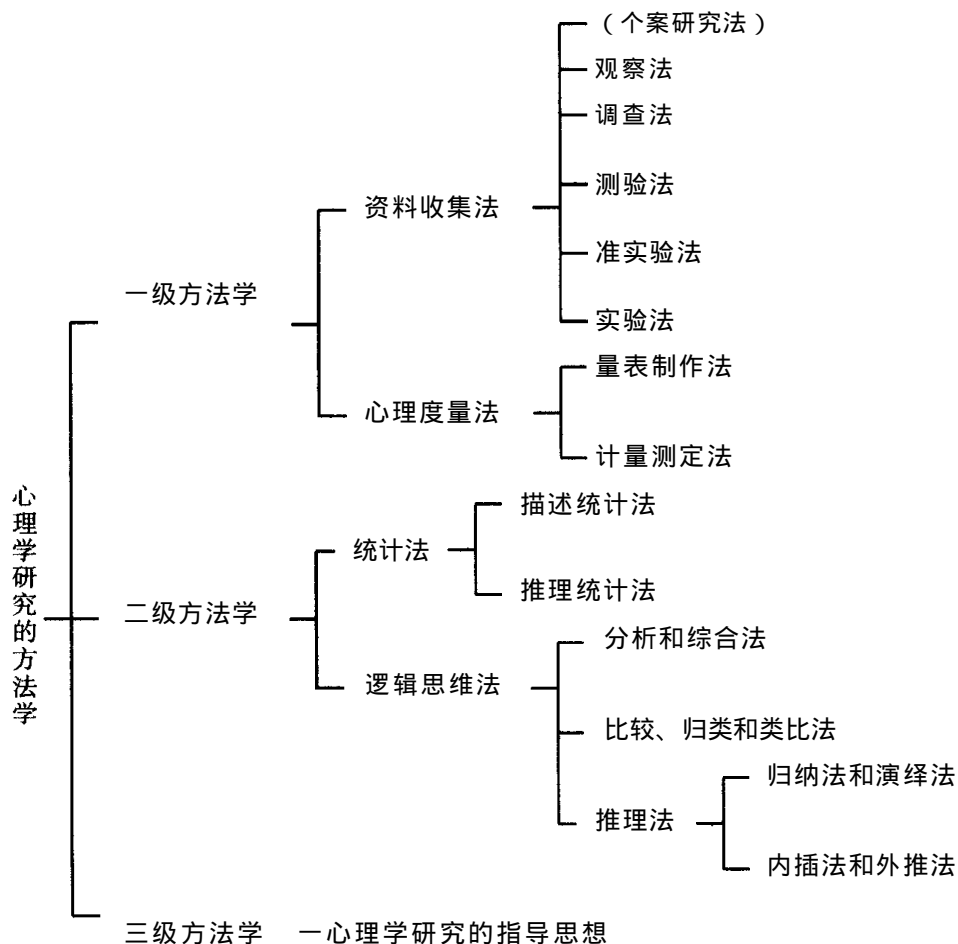


图 1.2 心理学研究方法分类

(采自黄希庭，1987)

第三节 心理学研究的伦理

心理学研究是一项为人类获取新知识，为人类谋福利，促进人类幸福的事业。心理学研究的对象是人和动物，在心理学研究过程中就可能会使被试受到某种身心伤害。比如，津巴多等人（Zimbardo et al,1973）曾经做了一个模拟监狱的研究。研究者在斯坦福大学心理学楼地下室设立了一间模拟监狱，招聘了24名大学生参加研究，其中6人扮演看守，18人扮演囚犯。实验原本进行两周，但是六天半后就不得不中止。因为被试的心理和行为发生了巨大变化，扮演看守的被试日益从侮辱、恐吓以及非人性的处置“囚犯”中得到乐趣；扮演囚犯的被试则变得被动、抑郁、无能为力、极度沮丧和愤怒，其中有一半的“囚犯”要求被释放，而且精神几乎达到崩溃的边缘。可见，这样的研究给被试的心理带来了较大的伤害。另外一些研究则可能给被试带来身体的伤害，如疼痛、辐射等。以动物为被试的研究则可能出现需要使用电击、解剖、在脑中插入电极等操纵，这些操纵也可能对动物造成较大伤害。

为了规范道德行为，帮助研究者解决心理学研究中可能遇到的伦理（ethic）问题，许多国家都制定了有关的道德准则（ethical principles）。一个综合性的广泛适合于心理学家、研究与学习心理学的学生以及涉及心理学研究工作的人的道德准则是美国心理学会2002年发表的《心理学家的道德原则和行为准则》（ethical principles of psychologists and code of conduct）。我国也出台了一些涉及心理学研究与工作的道德准则，如中国心理学会在1992年发表的《心理测验工作者的道德准则》，中国心理卫生协会2000年发表的《心理评估质量控制规定及从业人员道德准则》。

由于心理学研究中涉及的伦理问题颇多，难以一一仔细讨论，下面就对一些基本的道德原则进行讨论。

一、以人为被试的研究

心理学的主要研究对象是人，在研究中必须确保来参加研究的人即被试（subject）的权益不受侵犯，这就涉及到许多伦理问题，其中最常见也是最主要的伦理规范包括知情同意、保护被试以及保障退出自由与保密。

（一）知情同意

研究者应该让被试清楚地知道有关研究的所有必要信息，这些信息包括研究的目的与程序，研究中可能存在的问题与风险等，以便被试决定是否参加研究。被试参加研究应该是出于自愿，不能采用威胁等方法迫使被试参加。如果被试是未成年人还需要得到监护人的同意。

知情同意 (informed consent) 是以人为被试的心理学研究的最基本原则，但这个原则有时候会与科学原则相抵触。如，仅选用志愿者进行的研究可能存在抽样偏差的问题，这与心理学研究所要求的随机取样原则相抵触。又如，当人们知道自己正在被观察与研究时的表现可能会与平常的正常表现有所不同，这就与获得真实表现的科学原则相抵触。此外，如果所有研究都必须事先通知被试，那么自然观察和许多现场研究都无法进行。因此，美国心理学会在《心理学家的道德原则和行为准则》中规定：匿名的问卷调查、自然观察与某些档案研究一般可以不必知情同意。

此外，有的时候由于研究需要，也不能实现完全的知情同意，甚至需要在研究中使用欺骗策略。如在著名的阿希实验 (Asch, 1955) 中，研究者就利用了“假被试”做同谋，这些假被试在比较线条的任务中都将明显不相等的线条说成相等，这导致许多真实被试也做出了同样的回答。通过这个实验阿希获得了关于人们从众行为的许多有效结论。很明显，研究中真被试被“欺骗”了，但是如果不进行这样的研究设计就难以观察到真实的从众行为特点。除了实验研究外，许多调查研究中也涉及欺骗问题，如一些调查中宣称的调查目的与实际目的可能不同，特别是对于一些比较敏感的问题进行调查时，如果真实说明调查意图，就可能造成被试的不真实回答。

对于使用了欺骗策略的研究，应该在研究结束后及时告知被试研究中存在的欺骗以及这样做的目的和价值，并及时消除由于欺骗给被试带来的潜在影响。例如，如果为了研究挫折感而故意让被试做了根本无解的测试题后，应该向被试解释清楚，以免给被试的自尊和学习动机等带来消极影响。如果由于欺骗诱导出了一些违反社会一般道德的行为或反常行为则还应该进一步向被试说明他们的行为是由于研究情境所造成，不是由于他们自身的问题或缺陷导致，并进一步指出他们的行为不是反常行为，大多数人在类似情境中也会有类似的反应，以使被试能够接受有关事实，消除心理负担，并防止助长不良行为。

(二) 保护被试

研究者应该为参加研究者提供保护措施，防止被试在研究中可能受到的身心伤害。在研究设计、研究过程和研究结束以后，研究者都必须特别注意被试的身心健康与安全问题。虽然绝大多数心理学研究对被试不会有任何伤害，但是研究者还是不能对此掉以轻心。研究者应该注意观察被试在研究中的反应，询问有关情况，在研究结束后向被试提供联系方式，以便被试在出现问题时能够获得研究者的建议和帮助。一般来说，多数研究者都会比较注意研究中对被试的身体伤害问题，而对心理伤害，包括对被试情绪的不良影响、自尊的伤害等容易忽视，因此，在研究过程中和研究结束后应该注意关注被试的心理感受，并提供帮助。

必须确保参加研究者的安全，使其身体和精神不受伤害是基本的道德原则。但是，这一原则有时候也很难绝对遵守的。如，对耐痛阈限的研究就是以引起一定的疼痛为基础的；涉及挫折反应的实验研究需要引起被试的挫折感；对恐怖症的研究可能需要被试见到他们所恐惧的东西。有时候，研究对被试产生的伤害可能是研究者在事前都无法完全预料的。如，米尔格拉姆（Milgram, 1963）进行的一项著名的关于权威服从的研究就导致了对被试的伤害。在研究中研究者设计了一个非常吓人的电击装置，告诉被试所进行的是一项关于学习与记忆的研究，并要求被试对做出错误反应的“假被试”进行电击。虽然实际上电击是假的，假被试所做出的各种电击反应也是事先安排好的表演，但是，真被试在实施电击时以为电击是真的。因此，虽然实验设计本身不会给任何人带来身体上的伤害，但却给实施电击的真被试带来了巨大的心理压力，被试因为执行了主试的命令而表现出高度的紧张和焦虑，如使一位原本成熟稳重的商人被试进入了一种近乎精神崩溃的状态。

需要注意的是，如果在研究中确实对被试造成了有害后果，研究者必须要消除这些后果。如，在上述米尔格拉姆（Milgram, 1963）的服从研究完成后，研究者就做了大量缓解被试焦虑、维护心理健康的工作。不保护被试或不消除有害后果会影响心理学研究的名声，这在心理学研究的历史上也是有教训的。如，华生等（Watson & Rayner, 1920）所进行的一项关于条件性情绪反应的著名研究，虽然具有重要的科学价值，但是由于研究过程中对被试的伤害，以及在研究结束后没有采取任何消除有害后果的措施而饱受非议。在这项研究中，研究者使用巨大的声音惊吓，使一名叫艾尔伯特的婴儿对白色的毛茸茸的东西如白鼠、毛皮大衣形成恐惧性条件反射。研究结束后，研究者也没有采取任何消退该条件反射的措施。这样的研究现在肯定是不允许进行的，因为它严重违反了我们的伦理原则。

（三）保障退出自由与保密

在研究之前必须让被试清楚地知道，他们有随时退出研究的自由，并保障被试的这一权利，不能强迫被试完成整个研究，并要避免被试产生退出研究会受到惩罚的误会。例如，如果答应给被试报酬的话即使被试退出研究也应该按承诺给予相应的报酬。又如，当参加实验的是研究者的学生时，研究者需要保证学生有权选择在任何时候放弃，并不会因此受到惩罚性对待。此外，由于儿童常常不可能明确向主试提出退出的要求，那么研究者需要注意儿童在研究中的表现，并提醒他们有退出的权利。

研究者应对所获得的被试资料进行保密，是研究者的基本义务和责任。因为，首先，被试的大量信息是其隐私，应该受到保护。其次，被试信息的泄露可能会给被试的生活带来困扰。如，被试智力测验的分数被公开，可能导致老师对学

生态度的变化，以及其他学生对低智力分数者的嘲笑。因此，在研究开始时研究者就应该明确地告诉被试哪些人可能接触有关的资料，并确保其他人不能随便接触这些资料，并对被试的隐私进行保护。为了确保保密，还应该采取一些必要的措施 包括：①不要求被试署名，不采集与被试身份有关的信息；②不直接将被试与有关资料联系起来，如可以采用代码或化名的方式；③如果在研究报告中需要出现某些被试信息，必须做技术处理；④在研究结束后的规定时间内消除敏感的资料。

当然，为被试保密的原则也不是绝对的，如果研究中确实发现被试存在严重的心理问题、犯罪行为等情况，还是应该根据具体情况向有关部门或人士反映。

二、以动物为被试的研究

虽然心理学研究主要关注人，但不少研究还是会使用动物作为被试。特别是动物可以替代人进行一些不能用人类被试完成的重要研究，以发现重要的科学事实。如，罗森茨韦格等人(Rosenzweig et al,1972)，就使用老鼠作为被试进行研究，发现在丰富环境中生活的老鼠比在贫乏环境中生活的老鼠的大脑皮层更重、更厚，大脑的化学活动水平更高。这样的研究就不能用人来进行，因为我们不能人为地将人放在贫乏的环境中生活，也不能随便解剖人的大脑。

虽然使用动物被试的研究伦理问题不会像使用人类被试那样敏感和复杂，但是动物的权益还是应该受到保护。研究对动物的使用 and 操纵也应该是必要、恰当和道德的，不能虐待动物，同时要适当地饲养和照顾动物。下面就介绍美国心理学会 1992 年发表的《心理学家的道德原则和行为准则》中涉及到研究中动物的使用和照顾的原则，以供借鉴。

1. 使用动物进行研究的心理学家必须人道地对待它们。
2. 心理学家对于动物的获得、照看、使用和处置都必须符合联邦、州、地方的法律与法规以及职业准则。
3. 接受过研究方法的训练并且在照顾实验室动物方面有经验的心理学家要监督涉及动物的所有研究程序，并且要确保动物在舒适、健康和人道对待等方面都获得了恰当的考虑。
4. 心理学家要确保所有在其指导下使用动物的人所接受过的研究方法以及照顾、喂养和处理这类动物方面的教育能够使其胜任相应角色。
5. 在一项研究中起协助作用的人，其责任与行动应与他们各自的能力相适应。
6. 心理学家要尽力减少动物被试的不舒适、被感染、生病和疼痛。
7. 以动物的疼痛、紧张或被施以某种剥夺为条件的研究程序只有在其他替代方案不可行，并且研究目的已被该研究潜在的科学的、教育的或应用价值所证

实为合理时才能使用。

8. 外科手术要在恰当的麻醉条件下进行；在外科手术进行中和结束之后，避免感染和减少痛苦的技术手段要跟上。

9. 当到了需要结束动物生命的时候，要用可接受的方式快速完成，并且尽力减少动物的疼痛。

在心理学的研究中，很容易碰到伦理道德问题。在研究实践中，以上提到的伦理原则有时又难以绝对遵守，那么，如何使我们的研究符合道德规范，如何解决研究中可能碰到的道德困境呢？这是进行一项心理学研究所必须面对的问题，应该说解决这样的问题并没有统一的办法，但是，有两点是必须注意的：（1）警惕研究中的道德问题，合理设计研究。研究者在进行研究之前必须明确意识研究可能涉及的道德问题以及自己所肩负的道德责任，关心和保护参与研究者的利益，确保自己具有操纵有关实验设施的能力，避免出现错误。在进行研究设计时，要考虑自己的研究能力、价值倾向等是否会给被试带来伤害，并充分考虑研究设计的道德可接受性，尽量避免那些需要使用欺骗以及可能引起被试较大不安和痛苦等带来较多伦理问题的设计方案。（2）进行风险一效益评估。如果研究中确实需要使用欺骗或带来一些不可避免的危害，那么就必须进行风险一效益评估。研究者首先应该通过查阅文献，请教有关专家等方式认真考察该研究可能给被试带来的伤害等各种危害究竟有多大，这些危害是否还在伦理尚可接受的范围内。同时，要认真考察研究结果所潜在的科学价值有多大，研究是否有必要进行，是否值得进行这样的冒险。总之，必须要有充分的证据表明研究中的风险是难以避免的，研究带来的效益远大于潜在的危害，该研究才是可以进行的。

补充讨论 1-2 研究报告中的伦理问题

在研究完成后，报告并发表研究成果时 also 需要注意伦理问题。中国心理学会(2002)编写的《心理学论文写作规范》中对于出版的伦理道德进行了比较详细的规定。这些问题主要涉及以下几个方面。

第一，如实反映研究过程和结果。科学研究必须得到重复，因此，研究报告中必须如实报告研究是如何进行的，数据是如何收集的，被试的情况如何，等等。同时，必须完整准确地报告研究的结果，不能任意删除、篡改和捏造数据，不能只报告部分数据。

第二，不能剽窃。不能把他人的思想、观点或工作据为己有，引用他人的观点、资料等必须注明出处并在论文中明确标识。

第三，不能部分或重复发表。一般情况下，不能将一项研究拆分成多篇论文

发表，研究应该作为一个整体进行发表，不能因为研究涉及不同方面就将不同方面的结果拆散发表。更不允许的是同一研究多次发表。因为这会导致学术资源的巨大浪费和学术成果的浮夸。此外，还应该注意的是为了提高命中率而一稿多投的问题，这样做会浪费编辑和审稿人的精力及时间，是不道德的。

第四，著作权与署名问题。一项研究如果有多人参与，著作权与署名问题就会比较突出。有关人员应该就著作权问题进行协商，论文署名的基本原则是根据贡献大小排列，一般性的辅助人员或后勤人员不应该列入作者名单。署名作者在享有研究报告的著作权时，也必须对论文本身及其所引发的各种后果承担责任。

最后，与研究有关的原始数据应该保存 5 年以上，以备查证或重复分析。

►本章提要

1. 心理学是一门兼有自然科学和社会科学的性质、具有不同分析单位以及既重视内省也重视客观观察的学科。

2. 一般人探索世界的常用方法归纳起来有常识、传统和权威三种。

3. 系统性、重复性、证伪性和开放性是科学研究的四种典型特征。

4. 心理学研究主要包含 5 个基本步骤：(1)选择课题和提出假设；(2)设计研究的方案；(3)搜集资料；(4)整理和分析资料；(5)解释结果以及检验假说。

5. 科学研究的开放性主要表现在多视角、公开性、可争辩性和科学研究无禁区四个方面。

6. 以人为被试的研究的伦理规范主要包括知情同意、保护被试以及保障退出自由与保密。

►本章关键术语

心理学(psychology)

开放性(openness)

分析单位(unit of analysis)

伦理(ethic)

内省法(introspection)

道德准则(ethical principles)

科学研究(scientific research)

知情同意(informed consent)

证伪(falsifiable)

►复习与练习

1. 为什么说心理学的性质决定了其研究方法的多样性？
2. 举例说明一般人探索世界的方法的利与弊？
3. 试以科学研究的特征来分析某些心理学的研究。

4. 是否可以在心理学研究中使用欺骗的方法，为什么？

➤ 推荐参考读物

Salkind, N. J. Exploring Research (4th ed.). New Jersey : Prentice Hall, Inc. 2000. 10 ~ 15

Myers, A. & Hansen, C. H. Experimental Psychology(5th ed.). Pacific Grove, CA : Wadsworth Group, 2002

Solso, R. L. & Maclin, M. K. Experimental Psychology : A Case Approach. Boston : Allyn & Bacon, 2002

雷迪·D·高尔, 沃尔特·R·博格, 乔伊斯·P·高尔. 教育研究方法论 (第六版). 许庆豫等译. 南京: 江苏教育出版社, 2002

B·H·坎特威茨, H·L·罗迪格(III), D·G·埃尔姆斯. 实验心理学——掌握心理学的研究. 郭秀艳等译. 上海: 华东师范大学出版社, 2001

➤ 在线资源

想要了解更多关于科学的问题可以访问 <http://xys.freedns.us/pages/science.html> 和 <http://www.oursci.org>.

想要了解更多关于心理学研究中的伦理问题可以访问 <http://www.apa.org/ethics/>.

第二章 选题与取样

选题是科学研究的起点，也是每一项心理学科学研究的起点。科研题目是科研人员对某问题（problem）在理论认识和研究手段方面的概括，是经过充分的理论思考和实践准备之后提出来的。因此，科研题目集中反映了选题者的科学思维、理论认识、实践能力和要达到的预期结果。科研题目是贯穿于整个科研程序中的主题思想，是指导科研工作各项安排的主线。它对于之后的研究工作成败及收效大小，起着决定性的作用。选好科研题目，对于科研人员和科研领导者来说，具有重要的战略意义。要完成一个好选题，必然要涉及查阅文献以及研究对象的性质和数量等。在本章中，我们将讨论以下问题：

1. 心理学研究的课题来源
2. 选题要遵循的原则和标准
3. 心理学文献的查阅
4. 心理学研究的取样

第一节 选题

一、研究课题的来源

心理学研究课题可以有多种来源。主要的来源是：(1)对日常生活的观察，(2)实际的需要，(3)理论，(4)技术发展的推动等。也可以是这些来源的某种结合。

（一）对日常生活的观察

心理学中的许多研究课题都来自学者们对日常生活的观察。例如，俄国生理学家巴甫洛夫（Ivan Petrovich Pavlov, 1849—1936）在研究唾液腺在消化系统中的作用时，观察到实验动物（狗）不但在食物进入口腔时分泌唾液，在食物离口腔还有相当距离的地方，只要引起了动物的注意，看到食物或给食物的人，或者听到给食准备动作的声响，都会产生唾液分泌。这种日常生活中常见的心理性分泌现象，被具有科学素养和好奇心的巴甫洛夫敏锐地抓住了。他设计出一系列实验，进行客观的探索。当时美国著名的生理学家查尔斯·谢灵顿（Charles Sherrington, 1857—1952）曾劝告他应该放弃对这种“心理分泌”的愚蠢研究，但他置之不理。相反，他放弃了对消化功能的研究，用了35年时间来探讨

这种所谓的“心理分泌”，创立了经典的条件作用理论（classical conditioning theory）。从而改变了心理学研究的进程。

又如，卡达西和津巴多（Carducci&Zimbardo,1995）观察到生活中有些人腼腆内向，一直都很害羞，像个性特质；而有些人只是在某些情境下害羞，如在陌生人面前或在没有准备的情况下被推上舞台给大家表演而感到害羞。极端害羞的人沉默寡言，有可能会发展成为对人的害怕和恐惧。性格内向害羞的人采取独居的方式生活，不参加社会活动。性格外向害羞的人，在公共场合表现活跃但内心却害羞，他们喜欢参加社会活动，也有社交技能来有效地完成这些活动，但是，他们总是担心别人是否会真正地喜欢和尊重他们。于是卡达西和津巴多对这一现象进行了调查研究，结果发现，第一种情况可能是天性问题。有研究表明，大概有10%的幼儿“生来害羞”（Kagan,1994），这些孩子向来害羞，在与不熟悉的人或环境接触时就更加害羞了。第二种情况是在儿童期被嘲讽，因一时失误而被大家取笑。这些人小时候在家里是“掌上明珠”，父母宠爱有加，很少与其他人接触。第三种情况是文化上的原因。在我国的文化中女性害羞受到鼓励，因而害羞者较多。目前年轻人都被电子产品所包围着，他们长时间独自一人看电视、打电子游戏、网上冲浪和收发电子邮件，由此产生了社会隔离，减少了与人面对面接触的机会，过量使用网络使不少青年人孤独、隔离和更加害羞。

又如，张增杰（1963）在日常生活中观察到，小学一年级儿童学算术时，相当普遍地发生减式当加式或加式当减式的错误。他怀疑儿童的这种错误很可能是受运算定式和算式强成分掩蔽弱成分的影响。于是他设计了一项研究，即影响儿童感知算式的两个有关因素的实验。第一步是鉴定测验，目的是判定小学一年级儿童感知加减算式错误的性质；第二步为训练实验——结合教学进行为期五周的加减法交替练习。研究结果表明，加强加减法交替练习后，一年级儿童的上述错误明显减少了。

能否从日常观察中发现研究课题的关键是要善于观察、勤于思考。平时，我们要养成对各种心理现象和行为经常问个“为什么”的习惯，这样做往往可以使我们从纷繁复杂的生活大潮中、从变化无穷的人的日常行为中，发现值得研究和探讨的课题。

（二）实际需要

从根本上说，实际的需要是推动一切科学发展的动力，心理科学也不例外。例如，在第一次世界大战后期，由于挑选士兵入伍的需要，促进了美国许多心理测验研究的开展。1917年，美国心理学会指定一个专业委员会，专门研究心理学可以为战争做出哪些服务。该委员会提出一份报告，心理学可为战争做出最实用的贡献是设计一套心理学测试方法，目的是快速挑选合适的军事人员。于是，耶基斯（Robert Yerkes,1876—1956）等召集了约40名心理学家，在两个月的

时间内设计了一套“陆军甲种测验”（言语智力测试题）和“陆军乙种测验”（为应召士兵中百分之四十文盲士兵设计的图形测试题）。到1918年11月，共有170多万官兵参加了测试，并取得了令人满意的效果。在二战时期，不少苏联心理学家参加到保卫祖国的战斗行列里，推动了脑损伤的研究，促进了神经心理学的发展。例如，鲁利亚（Alexander Romanovich Luria, 1902—1977）对脑损伤的研究。他通过长期临床观察，总结了大量的脑损伤病例，以脑的三个基本功能联合区的新范畴来探讨脑在人的各种心理活动过程中的功能组织原则，并相应地把大脑分成三个机能系统。

心理学家常受雇于不同的职业，如研究公众意见和广告宣传，帮助协调陆海空三军的关系，应战斗情景的需要而训练动物，处理战争疲劳、消沉等临床问题等。例如，在1980年进行的一项基于经典条件反射理论的实验中，被试看到某种颜色的钢笔，同时听到悦耳的背景音乐；看到另一种颜色的钢笔，则听到不悦耳的音乐。后来，被试在面对几种钢笔的选择时，往往倾向于选择伴有悦耳音乐的一种。这种原理常出现在电视广告节目中，看似平淡，实则可以引导人们做出连自己也不明究竟的选择。在我国，由于推行计划生育政策，为了提高人口质量的需要，兴起了不少独生子女心理特点的研究。据不完全统计，从1994到2005年，发表在各类中文刊物上关于独生子女心理特点研究的文章就有120多篇。为了临床和其他实际部门的需要，修订和自行编制了智力量表、记忆量表和个性问卷表，等等。例如，1979年至1980年，由龚耀先主持、全国56个单位协作修订的韦氏成人智力量表；1980年至1986年，由林传鼎和张厚粲主持、全国22个单位协作修订的韦氏儿童智力量表；同年，龚耀先和戴晓阳主持、全国63个单位协作修订的韦氏幼儿智力量表等。为最经济有效地选拔飞行员，进行了不少视觉—运动技巧以及其他有关心理品质的心理学研究等，例如，德国航天航空研究院航天心理研究所研制开发了专门用于选拔飞行员的眼（耳）—手—脚协调测试仪（CCT）和多重任务测试仪（TOM）。近年来，随着突发灾难性事件（如911、SARS、伊拉克战争、印度洋海啸等）的增多以及恐怖主义自杀性爆炸的不断发生，应激性障碍和创伤后应激障碍成为研究的重点。例如，李凌江（2003）的研究小组开展了创伤后应激障碍的生物学机制研究，用功能性磁共振等设备观测患者几个重要的脑区，如前额叶、杏仁核、海马的结构和局部血流、脑氧含量的变化；用神经电生理手段观测实验鼠恐惧记忆的脑痕迹，来研究动物和人在受到强烈的精神刺激后脑的结构与功能的改变。初步结果发现：造成创伤性记忆持久处于易化激活状态的原因是大脑功能和结构累积损害的结果，尤其是在杏仁核和海马等脑区。

实际需要推动了心理学研究的发展。在实际需要中产生的课题具有重要的现实意义和应用价值，易于得到社会和政府部门的支持。为了鼓励联系实际需

要的心理学研究，世界各国都实行科学基金制度。如成立于 1950 年的美国国家科学基金，每年资助的科研经费大约 50 亿美金，资金来源于企业、个人所得税及赠款。科研项目分立项、批准、执行三个阶段。其宗旨是：提高科技资金使用效益、培养下一代国家栋梁之材、促进科学事业的发展。其他的基金有英国高等教育基金、法国国家管理教育基金、美国国家自然科学基金等。我国现在也实行这种制度，如国家自然科学基金、国家社会科学基金、教育部博士点专项基金等。这些基金都对心理学研究课题提供资助。例如，2004 年，中国国家自然科学基金会对管理科学与工程面上项目的资助达 96 项，金额 1 330 万元，资助率达 15.82%。获取基金资助的步骤是由研究者提出申请，写明研究问题的意义，预期达到的目标，已有的工作基础，研究的内容步骤，计划经费用途等。基金会组织科学家评审通过并决定拨款金额。专家们评审的一个重要根据就是实际的需要。

（三）理论

理论 (theory) 是一种为某个确定的研究领域提供前后一致的描述并能对全部有关事实作出解释的一套连贯的规范表述。例如弗洛伊德的人格发展理论，皮亚杰的认知发展理论。在心理学中还常用模型这个术语。模型 (model) 比理论这个概念要小，是用来描述或表示所研究的某一心理过程或具有整体功能的行为。例如在认知心理学中选择性注意理论有过滤器模型 (filter model)、能量模型 (capacity model) 和衰减模型 (attenuation model)。心理学中的理论或模型都还需要在科学实验中经过进一步的检验。不少研究课题是为了检验某种理论是否有预测作用而进行的。例如，前面提到的选择性注意理论的发展。选择性注意的早期研究通常涉及对材料的双耳呈现。被试戴着立体声耳机，每只耳朵被输入不同的信息。然后告诉被试只注意一只耳朵并保证执行这种指导语：要求被试注意追随耳。谢里 (Cherry, 1953) 也研究了选择性注意。他让被试大声重复一只耳的听觉信息，忽略呈现给另一只耳的信息。当后来询问被试时，他们通常不能报告未被注意的信息。莫布雷 (Mowbray, 1953) 的研究也发现，给被试同时呈现两个故事，其中一个是视觉的，另一个是听觉的，被试仅能回忆出一个故事的意义。为了解释谢里和莫布雷的发现，英国心理学家布罗德班特 (Broadbent, 1958) 发展出了一个注意模型——过滤器理论 (见图 2.1)。按照这种理论，一个输入刺激在完全加工之前需要通过一个过滤器，过滤器根据刺激的物理特征选择一个输入刺激，拒绝同时呈现的其他所有刺激。布罗德班特的理论虽然有说服力却不可能完全正确。1960 年特瑞斯曼 (Treisman) 报告到被试能够追随信息的语义内容 (即意义)，即使信息进入非追随耳也是如此。于是，特瑞斯曼 (1964) 提出了注意的衰减理论 (见图 2.1)，解决了布罗德班特过滤器理论中存在的一些问题。按照此理论，过滤器并不是按“全或无”的方式来工作的，

不是只允许一个通道（追随耳）的信息通过。而是既允许追随耳的信息通过，也允许非追随耳的信息通过。只是非追随耳的信号受到衰减，强度减弱了，但一些信息仍可得到高级加工。虽然，衰减模型能很好地解释许多与选择性注意相联系的现象，但是它也存在缺点。多伊奇（Deutsch, 1963）提出了一种更简单的观点——后期选择模型（见图 2.1）。他预测所有进入的刺激都被加工。因此，被试能识别几乎所有情况下的信息，即使信息呈现给非追随耳也是如此。

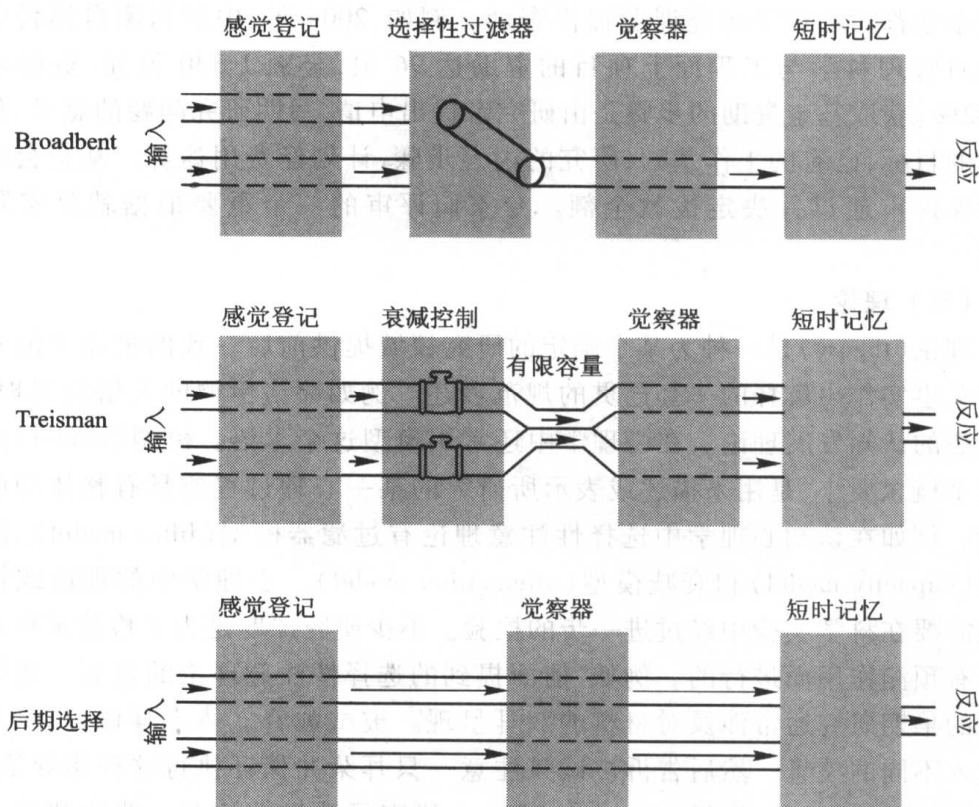


图 2.1 感觉加工中注意和能量的三种模型
(采自 John B. Best 著，黄希庭译《认知心理学》)

任何科学理论的发展都经历了众多的批评、质疑、争论，甚至否定，这些都是科学得以持续发展的动力，在客观上反而促进了已有理论的发展。心理学理论的发展也不例外。例如，1843 年，韦伯（Max Weber, 1864—1920）提出了心理学上第一个定量的差别阈限的定律，即韦伯律。就是说刚能觉察出刺激有差别时，刺激的增量和原来的刺激值的比是一个常数。后人的许多实验表明，韦伯律只是在中等刺激强度范围内才是正确的。1860 年，费希纳（Gustav Theodor Fechner, 1801—1887）提出了对数定律，即感觉强度与刺激强度的对数成正比。之后史蒂文斯（Stanley Smith Stevens, 1906—1973）对费希纳的对数定律又提出了批评，

并于 1957 年提出了刺激强度和感觉量之间的关系的幂定律，即心理量是物理量的幂函数。虽然幂定律弥补了传统对数定律的缺陷，但也难以完全取代它。

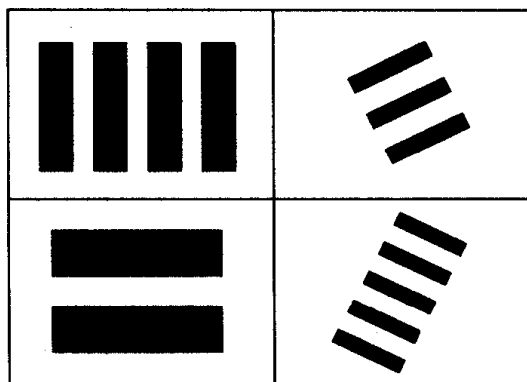
总之，单纯用一两个研究来肯定或否定某一理论，都是不可能的。因为一个理论和某种预测作用之间的关系是非常复杂的。测验成功，一般只是使研究者对此理论和模型可信度提高了一些，只有一个理论多次预测成功，而其他理论都不能圆满解释这个现象时，这个理论就很接近于被接受了。与之类似，如果研究资料不支持某个理论的预期，这个理论也很少被放弃，而可能是被修正以更适合说明研究资料。对理论的验证或预测不成功有时可以表明这个理论应用的限制。例如，很多年前关于问题解决有两种明显相反的理论。桑戴克（Edward Lee Thorndike, 1874—1949）认为问题解决是一个渐进的、尝试错误的过程，而苛勒（Wolfgang Köhler, 1887—1967）则认为问题解决是一种领悟的、全或无的过程。看起来这两种理论是对立的，一个对了另一个就错了，但事实并非如此。经过研究者们多年的研究表明，在某些情况下问题解决是一个渐进的、尝试错误的过程，而在另一些情况下它又是一种领悟的、全或无的过程。总之，理论可以帮助我们理解许多所发现的现象，即使它的某种预期不能得到完全支持时它也有启发我们思考的作用。测试某个理论的预测作用是研究问题的一个重要来源。

（四）技术发展的推动

技术发展是科学进步的持续动力，对心理学研究也是如此。以前由于研究技术手段的限制，对许多问题不可能研究或研究的水平较低。但是，技术的发展使这些问题的研究成为可能。例如，随着光学技术、眼科学以及电子技术和计算机技术的不断发展，对眼睛在人类心理活动中所表现出来特性的研究越来越多。20 世纪 20 年代，因为没有任何仪器，人们只能使用肉眼观察眼动来研究广告心理学，但当现代高科技手段广泛用于眼动记录技术后，眼动仪也发生了革命性的变化。如 EVM3200 型眼动仪，使用了摄像技术、红外线定位技术、电子计算机技术，数据系统记录的容量大、速度快、精度高。它每秒钟可记录 50 次眼动数据，一分钟记录 3 000 个数据。有了技术的推动和研究设备的支持，就可以对眼动进行进一步的分析研究了。

同样，随着脑测量技术的发展，脑认知（功能）成像领域也出现了非常令人振奋的进展，涌现出了一批功能强大的无创性的脑认知成像手段。这些技术被迅速应用到认知神经科学的各个领域，形成一批突破性的成果。例如，脑功能定位技术的进步使研究者在理解脑的视觉基础方面迈出了一大步：他们能追踪视觉信息从眼的视网膜向枕叶初级视皮层的转换。例如，科斯林领导的研究小组（Kosslyn et al., 1999）使用两种方法：正电子发射断层扫描技术（PET 扫描）和透颅磁刺激（Repetitive transcranial magnetic stimulation）。在这项研究中，要求被试记住下面四个象限中呈现的四个图，然后闭上眼睛想象它们的存在。他们

的任务是回答实验者提出的问题。例如关于在四个象限中条形带的相对长度和宽度。设计这一任务是为了使参加实验的人一定会在脑内形成表象。在 PET 研究中，对被试在执行任务时进行 PET 扫描。结果表明，同样的视皮层区被想像的和真实看见的图像所激活。在透颅磁刺激研究中，利用磁刺激脉冲创造一个暂时可逆的脑局部“损伤”，实际上对脑组织没有任何损害，一定脑区能被短暂地失活。结果发现，被试不能完成想象的比较任务，仍可睁眼进行各象限的比较。这一发现较强地支持了初级视皮层主动地参与了视觉表象的形成这一观点。



另外，心理学中许多问题必须在新技术的推动下才能进行研究。例如，不同意识状态（清醒、睡眠、催眠）的 EEG，正负情绪 EEG 的双侧性，不同智力活动的 EEG，不同情绪状态下的面部肌电记录，不同阅读状态下的肌电记录，疲劳对肌电的影响，任务难度对肌电的影响，正负情绪的心率 / 手指脉搏血容体积 / 血压等生理反应的变化，不同情绪调节过程的心跳间期、手指脉搏血容体积、血压等生理反应的变化，说谎的心理生理测量，疲劳的心理生理测量，警觉的心理生理测量，不同情绪状态下的瞳孔反应，任务难度与瞳孔反应，人格特质与瞳孔反应，不同年龄的阅读技巧与眼动分析，Muller-Lyer 错觉的眼动分析，正负情绪的眼动和眨眼分析，正负情绪的 ERPs 左右两半球的不对称研究等。

近年来，在心理学中又出现了一种新的研究方式，即对已出版文献资料进行再分析研究。这是另一种情况的重复研究。关于这个问题我们将在本书第十四章中详细讨论。

二、选题的原则

一旦知道从哪些方面提出问题以后，研究者如何知道他所提出的问题是不是一个好问题呢？弗林克尔和瓦伦（Fraenkel & Wallen, 2004）提出好的研究问题应该具备四个基本特征：（1）问题是切实可行的（即，它在有限的时间、精力或财力下就可以完成）；（2）问题是清楚的（即，大多数人都会同意这个问题中的关键词词的意思）；（3）问题是有意义的（即，它值得研究，因为它将增加有关人类及生存环境的重要知识）；（4）问题是符合道德的（即，它将不会造成对人类生理或心理上的伤害，或者对人类所生存的自然环境或社会环境造成损害）。为了提高心理学研究工作的质量和效益，提高心理学研究的水平，减少失误，下面我们详细讨论其中选题的原则。

（一）课题是可行的

可行性是研究者是否具备研究解决某课题的主客观条件。主观条件主要是研究者的理论水平、研究能力、知识准备、学历水平、时间和精力等；客观条件主要是指外在环境是否适合，有关的研究资料、设备以及必要的资金是否具备。有些研究课题虽然具有重要的理论和应用价值，但对于单个研究者和群体来说，如果主客观条件不满足，这样的课题就必须放弃或者加以修改。例如，一个心理学专业二年级的学生如果选择的题目是“视觉阅读过程的眼动和脑电研究”，就违背了可行性原则。首先，心理学二年级学生在“视觉阅读”、“眼动研究”和“脑电研究”等方面并不具备足够的理论水平和知识准备；其次，他也没有掌握“眼动”和“脑电”的技术，也可能使用不到先进的设备。因此，完成一项研究课题，往往需要三个基本条件和三个要素，即理论条件、物质条件、能力条件和人、财、物三要素。初次从事科研的人员，最好选择一些研究范围小、易完成的，又是本领域急需解决的课题。

（二）课题是清楚的

课题是否清楚对整个研究非常重要，课题清楚的含义就是研究者到底要研究什么问题。例如，“教师觉得为学习障碍儿童设特殊班级如何？”这个课题就不是很清楚。首先，“教师”该如何理解：包含什么样的年龄组（青年、中年还是老年）？是否包含所有学校（公立和私立）？是包含全国所有省份的学校的教师还是仅仅包含某个地区？这些问题都需要明确。其次，“觉得”也很模糊：它是指认识？还是指有什么情绪反应或暗示有行为反应？最后，“特殊班级”和“学习障碍”也需要说明。如，有一个学习障碍儿童的合法定义是：年幼儿童，由于明显的学习和行为失调，不能适应正常的课堂情景。这种失调应该与神经障碍和情绪障碍有关，并且不应由智力落后、文化剥夺或外国语言问题等原因造成。但这个定义本身也包含了一些模糊的词，像“明显的学习失调”，就可以有多种含义。“文化剥夺”也一样。所以将课题定义清楚比我们想象中的要困难。要使研究课题清楚，研究者应该明确界定研究问题中的术语。定义术语（defining terms）的方法有三种。第一种方法使用本质定义（constitutive definition）也就是使用字典来进行定义，研究者只需要用其他的词汇来阐明术语的意义就可以了。第二种方法是通过举例来说明。第三种方法是对重要术语进行操作性定义。有关定义的问题我们将在第三章详细论述。如果研究者不能清晰理解其课题，就不很清楚要寻找什么样的资料，也就无法按计划去收集和分析资料。

（三）课题是有价值的

课题的价值即课题是否是值得研究的，是否具有理论和实践意义。例如，如果有人选择这样的课题：精神分裂症患者的反应时是否不同于正常人？这个课题便没有意义。因为即使他确实找到了他们之间在反应时上有差异，也无法对

这种差异做出确定的解释。如果把课题改为：精神分裂症患者的注意是否不同于正常人这样的课题就有科学意义。因为它与已知的心理过程有着确定的联系。课题的价值主要包括两个方面：一是课题的针对性要强，研究者选取的课题应该是当前心理学改革中具有代表性、被普遍关注和亟待解决的重大问题或热点、难点问题，有利于提高心理学教育质量，能体现出重要的应用价值；二是选择的课题应该符合心理学理论发展的需要，有利于检验、修正、发展、创新心理学理论，体现一定的理论价值。例如，斯佩里 (Roger Sperry, 1913—1994) 和加扎尼加 (Michael Gazzaniga) 对割裂脑的研究就是有价值的。他们证实了大脑不对称性的“左右脑分工理论”，每个人的颅骨内都有两个不同的脑，每个脑都具有复杂的能力。这个结果和一些后续研究对许多领域都有重要而深远的影响。现在有证据表明，大脑的两部分有许多特有的功能。大脑左半球更擅长于言语、写作、数字运算、阅读等等，是主要的言语中枢。大脑右半球则在加工人像、解决空间关系问题、符号推理、艺术活动等方面更有优势。凭借在大脑两半球功能定位方面的成就，斯佩里于 1981 年荣获了诺贝尔医学生理学奖。总之，研究者选题时应该明确课题有无实用价值或科学价值以及价值的大小。

(四) 课题是符合道德的

关于心理研究的伦理道德，我们在第一章已经详细论述。

补充讨论 2-1 心理学研究的主要类型

心理学研究的问题相当广泛，不同的问题或研究目的，各有不同的研究方法。因此，在心理学研究的本质上，也就有各种类型的研究。如理论研究与应用研究、横向研究与纵向研究、定量研究与定性研究等。

就研究的目的来看，心理学研究一般可区分出三种类型：(1) 理论研究；(2) 应用研究；(3) 理论兼应用研究。理论研究 (theoretical research) 是以揭示观察到的现象和事实的基本原理和规律为主要目的，获得新知识而进行的研究。心理学中的理论研究侧重于探究和发展有关心理现象的基本知识，特别是侧重于建立或检验各种理论假设。这类研究力图理解和解释各种心理现象是如何发生、发展和变化的，表现出十分明显的理论倾向，主要探究心理现象之间的因果关系，以增加人们对心理现象的内在规律的认识。例如，心理学家对短时记忆信息提取的研究就是典型的理论研究。此外，心理学家对注意模型、记忆结构、言语的产生、概念形成等问题的研究都是理论研究的代表。

应用研究 (applied research) 是针对某一特定的实际应用目的而进行的研究。通常是为了确定理论研究成果或知识的可能用途，或是为达到某一具体的、预定的实际目的确定新的方法 (原理性) 或途径。心理学应用研究广泛应用于

工业、商业、教育、医疗等领域。例如，闵斯特伯格（Hugo Münsterberg, 1863—1916）最早将实验心理学应用于工业生产的一系列研究。

理论研究与应用研究可以用来研究同一种心理现象。但是两类研究的关注点和使用的研究策略是不同的。比如同样是对外语学习的研究，理论研究会关注外语学习的脑区、脑电、关键期以及学习规律等问题；应用研究则会关注究竟在哪个年级开设外语课最好，如何帮助学生更好记住外语单词等问题。区分理论研究和应用研究的主要标志是具有特定的实际应用目的的研究属于后者。

需要注意的是，虽然可以从定义上把理论研究和应用研究区分开，但在实践中，二者常常是难以严格区分的，两者之间没有明确的界限。因为，理论研究和应用研究是互相联系的。一方面，通过解决实际问题，应用研究可以验证和发展理论，因而有助于心理学理论知识的扩展。另一方面，理论研究的成果总会在实际领域得到应用，有些理论研究成果甚至还可以直接应用。此外，在研究方法上，理论研究和应用研究常常没有显著区别，有些研究往往兼具理论研究和应用研究的特点，因而被称为理论兼应用研究。

根据研究的时间取向，心理学研究可分为横向研究、纵向研究和横向持续研究。

横向研究（cross-sectional research）是指在同一时间里研究不同年龄被试的心理和行为特点以对各年龄段的特征做出分析的研究。它只在一个时间点上收集研究资料，并描述研究对象在这一时间点上的特点，或者探讨这一时间点上不同变量之间的关系。横向研究的优点是简单、经济，样本量大，可对各种类型的研究对象进行描述和比较；其缺点是无法体现过程性和变化性，资料的深度较差，难以确定因果关系，取样程序复杂。

纵向研究（longitudinal study）也称为追踪研究，是指在较长时间内追踪同一群被试以考察其心理和行为特点的变化趋势的研究。纵向研究是跨时间的，需要在一段时间内对同一组被试进行反复观察、测量或实验。例如，普莱尔（Wilhelm Thierry Preyer, 1842—1897）是最早运用纵向研究的心理学家。他以自己的儿子为研究对象，每天三次对儿子进行观察，并辅之简单的实验。他对儿子的心理发展持续研究了三年，并记录了大量的观察事实，即儿童传记。纵向研究的优点是可以描述心理和行为变化的过程，便于探寻变量间的因果关系；其缺点是比较费事，成本大，样本量通常较少且容易亡失，研究所需的时间长，一些无关因素难以控制。

横向研究与纵向研究是用于研究心理现象发展变化的两种基本策略，因此它们常被统称为成长研究（growth study），即关注个体随生长历程，心理和行为模式变化及相互关系的发展性研究方法。需要注意的是：横向研究不仅仅是考察心理现象在年龄上的差异，也可以用于考察不同职业、性别等被试类型在某种

心理现象上的不同。

横向持续研究 (cross - sequential research) 是指在同一时间对多组被试, 从事时间长度适宜 (比横向研究的时间长, 纵向研究的时间短) 来收集资料的一种研究。前已述及, 横向研究和纵向研究各有其优缺点, 为了保持各自的优点而尽量排除缺点, 所以就有了混合横向研究和纵向研究的横向持续研究。

心理现象都有质和量两个方面, 所以我们既可以从质的角度来进行定性研究, 也可以从量的方面来进行定量研究。

定性研究 (qualitification study) 是以解释性的理解、描述为主对心理和行为进行研究的取向。主要是指那些采用非实验的方式, 且不以数量形式表达的研究。定性研究最早被普遍应用于 1910 年到 1940 年期间“芝加哥学派”的研究中, 在这期间芝加哥大学的社会研究人员运用参与观察、深入访谈、分析文件等方法进行有关都市生活的观察研究、少年犯罪的生活史的研究等。定性研究的形式多样, 主要包括参与和非参与观察、无结构和半结构访谈、案例分析、行动研究、历史研究、人种志法等。博格丹 (Bogdan) 和比克林 (Biklen) 将定性研究概括为五个特征: (1) 把自然情境作为研究资料的直接来源。(2) 资料主要是描述性的。(3) 强调研究的过程而不只是关心研究的结果和产品。(4) 倾向于对资料进行归纳分析。(5) 关心被试在自然情境下, 对事物已经发生或已经存在的意见的理解、看法或体验。

定量研究 (quantitative study) 也称量化研究, 用数字计量表达对心理和行为进行研究的取向。量化是用可操作的方法对事件、活动、人物的性质加以数字计量的表达, 如用数字表达感觉的数量、程度或强度上的差别, 量化是心理学研究科学化的重要标志。20 世纪 60 年代, 心理学研究因受自然科学影响, 崇尚实证主义, 更多采用定量研究方法, 包括实验研究和相关研究两类, 都用数量来表现研究结果, 以客观性、系统性和可验证性作为科学研究的标准。定量研究有一套完备的操作技术, 包括取样方法 (如随机取样、非随机取样)、资料收集方法 (如问卷法、实验法) 和以数理统计为基础的资料分析方法 (如描述性统计、推断性统计) 等。定量研究的基本步骤是: 发现问题, 形成假设; 设计研究, 选择方法; 分析资料, 做出结论。这种研究取向主要用于分析心理现象中的各种相关和因果关系, 可以人为地进行控制, 也可以重复进行。

定性研究与定量研究是有着明显区别的两种研究范式, 两者不仅有着不同的哲学基础, 同时对于研究对象、研究方法、研究过程、研究结果的分析、研究者自身的角色等问题均有着不同的看法。定性研究和定量研究各有优缺点。定量研究适合宏观场面对事物进行大规模的调查和预测, 定性研究适合在微观层面对个别事物进行细致、动态的描述和分析。定量研究证实的是总体的平均情况, 定性研究则适合对特殊的情况进行探讨。定量研究对事物进行数量上的计算,

用数字来表达事物和现象，而定性研究则强调使用语言和图像来表达时间的变化过程。定量研究从假设开始，收集数据对假设进行验证；定性研究强调从当事人的角度来了解被试的看法，注意他们的心理状态和意义构建。

第二节 查阅文献

一、查阅文献的意义

初步确定心理学研究课题以后，研究者通常要仔细查阅有关文献（literature）。文献是记录、保存、交流和传播心理学知识的一切材料的总称。查阅心理学文献有重要的意义，在整个研究过程中都必须进行。因为文献记录下了非常丰富的心理学事实、发现、数据、理论、启示和研究方法，能反映出心理学研究最新进展和水平，是心理学研究工作必不可少的信息来源。它不仅可以帮助研究者收集特定问题的各种研究观点和结果，又可以提供对当前研究有用的研究思路和方法。因此，确定课题以后，查阅文献有利于研究者评价和发展初步确定的课题。具体表现在以下三个方面：

第一，有助于寻找知识的空白点。科学研究是为了有所发现，有所前进，那就需要进行前人没有进行过的，或前人进行得还不够全面完整的研究。为此，就必须熟悉文献。否则一个人挖空心思出来的课题和方案，可能是早已研究过的，已有定论的东西。因此，从文献中调查了解某个领域的研究现状，寻找其空白点进行研究，是研究课题的一个通常来源。这里要注意两点：一是要查找文献，弄清研究的问题是否是空白点，二是问题要有意义。即使问题是空白点但其研究毫无价值，也是不值得研究的。

第二，有助于发现矛盾的结果。研究工作的逐渐积累，有时就会发现，研究同一个问题却出现相反的彼此矛盾的结果。例如，通过查阅文献，了解到关于精神分裂症唤醒水平的研究出现两种不同的结果。一种结果是，精神分裂症病人的自主神经系统反应少而弱，唤醒水平较低；另一种结果恰好相反，表明病人处于缓慢的高唤醒水平，自主活动明显强而多。针对这种情况，我国陈仲庚等设置条件，研究结果支持精神分裂症病人唤醒水平更高的学说。但正如前面理论部分所讨论的，并不能因此就否定了唤醒水平过低说，只能说在这种特定的实验条件下唤醒水平过低说是不适合的，需要修正。这种研究问题应当是弄清楚不同研究出现不同结果的差异的来源。

第三，重复已发表的研究。有两种情况，一为简单重复研究，其目的是验证某一已发表的研究结果及其相联系的解释是否真实可靠。一个研究结果，尤其

是与之相联系的解释，绝不是一次研究所能完成的。一般必须经过他人反复验证，能重复出其结果并表明其解释合理方能成立。例如：斯腾伯格（Sternberg, 1970）的短时记忆搜寻的实验，在他发表实验报告后，即有不少研究者对之产生兴趣，完全重复或对实验条件稍加变化予以重复，均得到一致的结果，此实验方式及其主要结果与解释遂被称为“斯腾伯格模式”。这是科学研究中常见的选题方式之一。另一个与之类似的情况是，当研究者开始从事某一领域的研究，对研究内容尚不熟悉时，也可以选择某些经典的研究予以重复，以熟悉有关的工作并同时印证其结果。我国心理学界在建国初期关于条件反射的一些研究就属于这类性质。有时，一些研究既是前人研究的重复，又有创新。例如，在中国儿童身上进行皮亚杰式守恒实验，研究某种守恒作业（如容积守恒）的发展，与欧美儿童比较，观察中国儿童在这种作业的质和量上具有什么特点，由于研究对象的民族不同，该研究便有新的内容了。

总之，文献资料是学习和研究工作的基础，没有文献资料就无法进行科学研究。为了有效地评价和发展心理学研究课题，有效地查阅心理学文献，首先应该清楚文献的来源和查找方法。

二、文献的来源

文献的来源很多。根据文献的保存方式，我们将其分为纸质文献和互联网文献。我们先介绍纸质文献，互联网文献将在查阅方法中详细介绍。

（一）图书

心理学图书是对已发表的心理学科研究成果、研究手段、经验和某一子学科领域的知识系统的论述和概括。主要包括阅览性图书（教科书、专著、文集等）和参考工具书（词典、百科全书、年鉴、手册等）。

教科书（textbook）是为心理学专业的学生和教师编写的专业课本。如《心理学导论》（黄希庭, 1991）、《实验心理学——掌握心理学的研究》（坎特威茨 罗迪格, 埃尔姆斯, 2001）等。这类书中采用的内容经过反复验证，具有较为严格的科学性、系统性和逻辑性，可信度高。如果研究者在开始研究工作时掌握的研究材料很少，可以先从教科书开始查阅。这样做能够迅速而准确地掌握这一问题的基本情况，并且弄清选题的意义、价值和依据。对教科书上讲得很明白的问题，没有必要重复研究；对教科书上有论述，但有许多疑点和争议的问题，可以预测自己对此问题研究的突破点；教科书上很少提到甚至没有出现的问题，说明所要研究问题可能是一个空白点，或者虽然有研究，但未被公认。接下来，可以进一步去查阅专著。

专著（monograph）是对某一心理学问题作系统、深入、全面论述的著作。专著包含的内容广泛而深入，不仅包括了某一专题的详细发展历史和现状，从前和

现在学者的研究成果，而且还包括了从理论到实践的见解、研究方法和结果，特别是还有许多个人研究的第一手资料。例如，斯腾伯格（Sternberg）在 1996 年出版的《成功智力》一书。这本书从全新的视角提出了一套有别于以前的智力理论：成功智力包括分析性智力、创造性智力和实践性智力三个关键方面；分析性智力用来解决问题并判定思想成果的质量；创造性智力帮助人们从一开始就形成好的问题和想法；实践性智力则可将思想及其分析结果以一种行之有效的方法加以实施。成功智力是一个有机的整体，“只有在分析、创造和实践能力三方面协调、平衡时才最为有效。”

手册（handbook）是对本学科某一分支或某一具体领域的研究和进展情况进行全面介绍的工具书，内容包括某一研究问题的历史，一段时期的研究成果、方法、问题和发展方向等。研究者可以从中获得大量有价值的信息。例如，由珀文（Pervin）主编，黄希庭主译（2003）的《人格手册：理论与研究（第二版）》。这本手册全面反映了当代人格研究最新进展，整合了人格领域的经典研究与当代研究，阐述了人格研究的各领域在 20 世纪 90 年代所取得的丰硕成果。此书的出版不仅为我国心理学教学和科研人员提供了一本反映国际水平的人格理论与研究的工具书，也为从事哲学、法学、社会学、经济学、管理学、教育学、文学、美学、医学的教学与研究人员提供了一本可供参考的工具书，同时也可以让读者从中了解到人格心理学研究的现状和发展趋势。其他手册如《实验心理学手册》（Handbook of Experimental Psychology），《儿童心理学手册》（Handbook of Child Psychology），《社会心理学手册》（Handbook of Social Psychology），《临床健康心理学手册》（Handbook of Clinical Health Psychology）等。

年鉴（yearbook）是汇集一年内重要事实文集和统计资料的工具书。年鉴中的统计资料具有重要的价值。如美国加州年鉴公司出版的《心理学年鉴》（Annual Review of Psychology），主要综论心理学领域的研究进展与成果。这些综论资料翔实，参考价值大。

（二）期刊

期刊（periodical）是有统一名称、定期或不定期连续出版的刊物的总称。由于期刊的出版周期短、内容新颖、论述深入、能及时地反映研究动态，所以它是所有文献资料中最重要的。期刊具有以下特点：第一，期刊一般分为月刊、双月刊和季刊，有特殊需要时，还有增刊。第二，期刊所选用的稿件多是本领域专家、学者的最新研究成果；第三，期刊上刊载的论文均是作者科研工作的总结，是第一手资料。这些资料未经别人的加工整理，能很好地保存原作的全貌，若干年后仍然可以作为很有科学价值的参考资料。期刊有多种类型，如学术杂志（journal）、大学学报（acta）、公报（bulletin）、回顾（review）、文摘（abstract）等。

学术杂志：指刊期在七日以上，三个月以下，以一定名称，按期发行的刊物。

例如，中国心理学会和中国科学院心理研究所主办、科学出版社出版的全国性学术刊物（双月刊）的《心理学报》。该刊物主要发表我国心理学家最新、最高水平的心理学科论文。内容涉及心理学各领域（包括认知与实验心理、发展与教育心理、生理与医学心理、企业管理与社会心理、心理测验、心理学史与基本理论、研究方法等）的研究报告、综述、研究简报与评介等。学术杂志种类繁多，既有综合性的杂志，又有专门性的杂志。不同的学术杂志侧重点不同，研究者要了解它们的特点，根据不同的需要进行查阅。下面是心理学领域一些重要的杂志。

- 《人格与社会心理学杂志》（Journal of personality and social psychology）美国
- 《心理学公报》（Psychological bulletin）美国
- 《心理学评论》（Psychological review）美国
- 《实验心理学杂志：学习、记忆和认识》（Journal of experimental psychology: Learning, memory, and cognition）美国
- 《发展心理学》（Developmental psychology）美国
- 《实验心理学杂志：人类知觉与行为》（Journal of experimental psychology: Human perception and performance）美国
- 《实验心理学杂志：总论》（Journal of experimental psychology. General）美国
- 《美国心理学家》（American psychologist）美国
- 《人格评估杂志》（Journal of personality assessment）美国
- 《人格杂志》（Journal of personality）美国
- 《实验社会心理学杂志》（Journal of experimental social psychology）美国
- 《记忆》（Memory）英国
- 《心理学报》（Acta psychologica sinica）中国
- 《心理科学》（Psychological science）中国

注：被 SCI 和 SSCI 收录的心理学及其邻近学科期刊见《心理学大辞典》（林崇德，杨治良，黄希庭，2003）

大学学报：是由学校或学术机关等发行，集合多人作品或报告所出版的学术研究性刊物。全国的多数大学特别是综合性大学和师范院校的学报都有社会科学版或教育科学版，有关心理学方面的文献有较高的学术价值，值得研究人员查阅。

文摘：是文章或图书的内容摘要，是将文献的主要观点、结论和中心内容摘

录出来，按一定方式编排而成的检索工具书。文摘摘录的内容一般是篇名、作者、刊登的文献名、期数、页码及内容提要等。其特点是扼要地摘写文章的论点提要、重要材料，记录科研成果，反映学术动态，积累有关数据，可简明地掌握某一问题研究的历史、现状和发展趋势。文摘的类型有许多种，如期刊文摘、报刊文摘和卡片文摘等。如由美国心理协会出版的《心理学摘要》（Psychological Abstract）和中国人民大学书报资料中心的《文摘卡片》。

（三）会议和学位论文

学术会议是心理学研究者交流科研成果的重要场所。会议论文是他们在各种心理学会议上宣读和发表的论文。这些论文报道了研究者的研究成果，有些论文代表了某一学科或专业的国际或国内的最新研究成果和发展趋势。利用会议论文，可了解有关学科领域的发展现状和水平。会议论文的特征：A. 一般带有会议特征的词，如 Conference（会议）、Congress（大会、综会）；B. 有主办会议的机构特征，如 Society（团体）、Association（协会）；C. 有会议文献类型的特征，如 Paper（论文）、Proceedings（会议录）等。会议文献由于学术性强、内容新，利用率越来越高。

学位论文是作者为获取某种学位而撰写的研究论文。一般有学士论文、硕士论文和博士论文三种。因作者大都比较年轻，思想活跃，撰写论文时参考了一定数量的国内外文献资料，故论文往往带有独创性。

（四）其他

包括各种资料汇编、最新进展、报纸等，这也是文献资料的重要来源。资料汇编常常是某一研究课题的资料集或学术研讨会的论文集。特点是同类资料众多，涉及面广，提供了许多新的检索线索。报纸是以刊登新闻报道和评论为主的连续出版物，一般是每天、每周甚至更长的时间出版。《人民日报》、《光明日报》、《中国青年报》等常刊登心理学方面的内容。由于出版的速度快，研究者能了解到某些研究的新进展、新动态，特别是能了解到社会实践中急需解决的新课题。最新进展一般由著名的专家约请某个专题，各个领域的学者共同撰写而成，论文多是综述性，基本上反映了某一专题进展和成就的全貌。如《发展心理学进展》（advances in developmental psychology）。

三、查阅文献的方法

面对众多的心理学文献，研究者如何在短时间内用较快的速度查阅所需的、尽可能多而全的心理学文献呢？这需要掌握一些常用的方法和技巧。常见查找文献的方法有：参考文献查找法、检索工具法、浏览法、跟踪追溯法等。我们重点介绍参考文献查找法和检索工具法。

（一）参考文献查找法

参考文献查找法又称滚雪球法，研究者根据自己所需查找的有关内容先找出最近发表的一篇文章或出版的一本书，再从已知文章和书籍后所附的参考文献目录查找内容相关的文献，然后又根据这些文献各自所附的参考文献目录，掌握更多有价值文献的方法。严格说来，各类杂志上发表的心理类文章和有关书籍的后面均附有丰富的参考文献，可供读者查找。通过参考文献来查找的优点是所查文献的针对性强、直接而集中，效率也高。特别是研究者找到的首篇文献是所研究专题的文献综述时，他既可以对该专题获得全面概括的初步了解，也可以快速获得该专题充分的参考文献。其不足是参考文献不够全面，而且会受到首篇参考文献作者水平和所能涉及到的资料范围的影响。

参考文献查找法的关键是如何找到最近发表的首篇文章或最近出版的书。一般有以下途径：一是请专家推荐。因为专家对本领域的进展情况比较了解。二是经常浏览比较权威杂志的目录。三是利用检索工具法找到首篇文献。

（二）检索工具法

检索工具法是利用已有的检索工具来查找文献资料的方法。主要有使用传统纸笔方法的手工检索和借助计算机和互联网进行的电子检索。

1. 手工检索

手工检索又称传统检索，是用人工查找信息的方式。其主要类型有各种书本式的目录、题录、文摘和各种参考工具书等。检索人员可与之直接“对话”，具有方便、灵活、判断准确，可随时根据需求修改检索策略，查准率高的特点。但由于全凭人的手工操作，检索速度受到限制，也不便于实现多元概念的检索。手工检索工具有目录卡片、目录索引和文摘三种主要方式。

目录卡片是简要记录文献资料要点的卡片。如题目、作者、出处、年代等。图书目录卡的检索又分为分类目录、书名目录和著者目录三个目录分类体系。在对各类图书馆或资料室所收藏的文献图书进行检索时，经常采用目录卡片检索法。例如，如果知道《心理学研究方法》这本书的书名，但不知道作者是谁，也不知道属于哪一类，应该去查“书名目录”。如果想查“弗洛伊德”的书，但不知道书名，应该去查“著者目录”。

目录索引是汇集了一定时间内各类文献的题目与出处的一览表，其特点是分类明确统一、按时间顺序编排，便于检索。目录索引主要有综合目录索引、报刊目录索引、专业目录索引和专题目录索引等。目前可供心理学研究者使用的目录索引主要有：

一、心理学和邻近学科期刊每年最末一期刊登的全年目录索引，如《心理科学》每年第六期刊登的全年总目录；

二、有关资料部门编撰的综合性或专业性目录，如中国人民大学书报资料中

心的《报刊资料索引》系列刊，该索引是按月度或年度编排的大型检索工具书，将本年度内复印报刊资料专题系列各刊每期选印文献的目录与没有选印的题录集中按专题和学科体系分类编排。每个条目包括篇名、作者、原载报刊等项目；

三、各大图书馆、资料室围绕某一专题或某一普遍关注的问题专门编制的目录索引。

关于文摘，我们在前面已经有所讨论。

2. 计算机检索

计算机检索就是利用计算机对存储的文献进行检索。研究人员将大量的文献资料按照一定的格式输入到计算机中，经过计算机的加工处理，以一定的结构存储在计算机的内部或外部存储介质上，如磁盘、光盘、磁带等，成为电子文献。查询者按照自己对文献的需求，编写成检索提问式，按一定的要求输入计算机，由计算机对检索提问进行处理，并与已存储在计算机外部介质上的电子文献资料进行检索运算，最后计算机检索系统将检索结果按要求显示或打印输出，这就是计算机查询，也称为电子文献检索。目前，计算机检索已被广泛地应用在文献检索工作中了。

利用计算机检索文献信息既可以利用单机检索，也可以利用计算机网络检索。单机检索是利用已有的静态数据库进行检索（主要是以光盘为载体的数据库）。而利用计算机网络检索除了可以检索专业的动态数据库以外，也可以检索其他网络资源。联机检索更具有优越性。

计算机检索有很多优势。一是速度非常快，通常只需几秒钟查询的结果就会呈现；二是费用相当便宜，有些甚至是免费的；三是可以获得打印的检索结果，甚至摘要和全文；四是可以同时检索多个关键词。

下面列出一些著名的心理学文献检索数据库：

名 称	链 接
ProQuest	http://www.ilproquest.com/
EBSCOhost	http://eadmin.epnet.com/
APA	http://www.apa.org/
OCLC	http://www.oclc.com/
Elsevier Science	http://www.sciencedirect.com/
ERIC	http://www.eric.ed.gov/
PNAS	http://www.pnas.org/
中国高等教育文献保障系统管理中心	http://www.calis.edu.cn/
中国期刊网	http://www.cnki.com.cn/
超星数字图书馆	http://www.ssreader.com/
万方数据库	http://www.wanfangdata.com.cn/

除了用专门的心理学数据库来查阅文献外，研究者也可以用专门的搜索引擎 (Search Engines)。方法很简单，研究者只要找到一台联网的计算机，通过网页浏览器，在地址栏键入相应搜索引擎的地址，然后在呈现的对话框中输入想查询的内容，就可以找到几乎与所有问题有关系的信息。搜索引擎的优点是非常迅速和方便，每天 24 小时均可查询，而且可以得到多方面的资料。缺点是搜索到信息不可靠，而且杂乱无章，需要很多时间来进行检索。下面是世界著名搜索引擎列表。

http://allsearchengines.com/	http://www.lycos.com/
http://www.yahoo.com/	http://www.webcrawler.com/
http://www.topclick.com/	http://www.northernlight.com/
http://www.hotbot.com/	http://www.excite.com/
http://www.google.com	http://www.baidu.com

第三节 取样

一、取样与选题的关系

研究者选择被试总是在选好了研究课题以后，开始进行研究设计之前。涉及选择被试和研究课题的有关问题是：是使用人类被试还是非人类被试？被试应具备哪些机体特征？用哪一种取样方法能获得被试所能代表的总体？这些问题的解答是以课题的性质及研究成果的概括程度而定。

第一，依研究课题的性质选择被试。研究对象选用人类还是非人类被试？应该根据课题的性质而定。许多心理学研究选用人类被试，是因为它关心的是人类的心理和行为。只要处理不违背教育性原则和道德原则，研究者可以使用人类被试。否则，就应当考虑选用非人类被试。例如，“限制活动对视知觉的影响”，这样的实验就不能选用人类被试，而只能选用动物被试。

第二，依研究成果的概括程度选择被试。在科学研究中，研究者通常是假设所得的结果，推广其适应范围，从而得出一种概括性的陈述。这就涉及研究的总体以及用于研究的该总体的数目问题。在从事一项研究时必须考虑研究成果的概括程度来选择被试。心理学研究的总体可能是很大的群体，也可能是一个小群体，或仅具有某些特性的成员。例如，所有的生物，世界上所有的人，某个国家所有的人，某个民族所有的人，所有的大学生，某个年龄组的所有成员，某个学校的学生……都可以作为心理学研究的总体。因此，必须考虑研究结果的适用范围，对总体加以界定。

由于种种条件的限制，研究者不可能也没有必要对总体都加以研究，而仅仅从其组成的个体中进行取样，使样本能代表总体。而样本的代表性在很大的程度上取决于总体分类的精确性、样本的足够性和总体的异质性。

二、取样的类型

取样的方法很多，根据是否按随机原则进行操作，取样可分为两类：非随机取样（**non-random sampling**）和随机取样（**random sampling**）。

（一）随机取样

随机取样，是按随机和独立的原则进行的取样过程。随机取样的一个基本原则是如果总体中的每一个体被抽取为样本的概率相同，那么从这个总体中抽取的样本，就具有对该总体的代表性。具有代表性的样本通常称为等概率取样方法样本。

随机取样中最常用的方法有以下几种：

1. 简单随机取样

最简便的取样方法是简单随机取样（**simple random sampling**）。简单随机取样又称纯随机取样，即对研究总体单位不进行任何分组排列，只按随机原则直接从总体中抽取一定的样本，以使总体的每一个样本都有被同等抽取的可能性。常用的具体抽取方式有抽签法和随机表取样法。

（1）抽签法，也称抽彩法。把总体中的每一个个体都编上号码并做成签，充分混合后从中抽取一部分，抽取的签对应的个体就组成一个样本。例如，假设要从 320 个人的总体中选取 50 个人作为样本。首先，制作写有 320 个人姓名或序号的标签，进行充分混合，然后选取 50 张标签，以此作为随机样本。抽签法要使标签充分混合，但这并不容易，而且当总体的量无限大或者足够大时，制作标签很不现实，所以这种方法在总体规模很大的研究中用的并不是很多。

（2）随机表取样法。随机表是由一些任意的数字毫无规律地排列而成的数字表。随机表取样法就是运用严格制作的随机数字表（表 2-1 为随机表的一部分）。随机进入包含总体数目的随机数字区，选取所需要的样本数目。随机表取样法的一个关键是“任意”进入随机表的数字区，从而确定一个起点。例如，假设要从 1 000 个人的总体中抽取 100 人作为样本。首先，将总体中的所有个体编码，本例中编码为 1~1 000。接下来要确定所选的随机数需要几位数字。如总体中个体的最大编号小于 10，只取一位数，如最大编号在 10~99 之间，取两位数，依次类推。本例中总人数为 1 000，所以需要有四位数才能保证所有的人有选中的机会。最后，按随机原则，决定以某行、某列的数字位起点，由左至右，或者由上而下，找出号码。如果查到的号码小于总体的最大编号，就选择进入样本，如果大于总体的最大编号或重复的数字就忽略，继续往下查找，直到抽

满所需的样本量为止。

表 2-1 随机数字表

	00 ~ 04	05 ~ 09	10 ~ 14	15 ~ 19	20 ~ 24	25 ~ 29	30 ~ 34	35 ~ 39	40 ~ 44	45 ~ 49
00	88758	66605	33843	43623	62774	25517	09560	41880	85126	60755
01	35661	42832	16240	77410	20686	26656	59698	86241	13152	49187
02	26335	03771	64005	88133	40721	06787	95962	60841	91788	86386
03	60826	74718	56527	29508	91975	13695	25215	72237	06337	73439
04	95044	99896	13763	31764	93970	60987	14692	71039	34165	21297
05	83746	47694	06143	42741	38338	97694	69300	99864	19641	15083
06	27998	42562	65402	10056	81668	48744	08400	83124	19896	18805
07	82686	32323	74625	14510	85927	28017	80588	14756	54937	76379
08	18386	13862	10988	04197	18770	72757	71418	81133	69503	44037
09	21717	13141	22707	68165	58440	19187	08421	23872	03036	34208
10	18446	83052	31842	08634	11887	86070	08464	20565	74390	36541
11	66027	75177	47398	66423	70160	16232	67343	36205	50036	59411
12	51420	96779	54309	87456	78967	79638	68869	49062	02196	55109
13	27045	62626	73159	91149	96509	44204	92237	29969	49315	11804
14	13094	17725	14103	00067	68843	63565	93578	24756	10814	15185
15	92382	62518	11752	53163	63852	44840	02592	88572	03107	90169

简单随机取样在理论上最符合随机原则，在研究者对研究总体中各类别比例不了解的情况下，最好用简单随机取样，这样能选出有代表性的样本。但这种方法在实践中也有一些局限，当总体很大时，抽签法要将每个个体都编上号码，这样工作量太大，几乎不大可能。另外，简单随机取样忽略了总体已有的信息，当样本量较小时，就有可能发生取样的偏向，降低样本的代表性。例如，在 100 名学生中有 80 名男生、20 名女生，如果现在要随机抽取 10 名学生作为样本，就有可能全部抽到男生。

2. 分层随机取样

分层随机取样 (stratified random sampling)，也称比率取样、分类取样或分组取样，是按照总体已有的某些特征，将总体分成几个不同的部分（每个部分叫一个层或一个子总体），然后在每一个层或子总体中进行简单随机取样。总体中通常存在性别、年龄、民族等层次。例如，假设从一所大学学生总数 10 000 人中取样 200 人为被试，学校学生分布状况为新生 30%、二年级学生 30%、三年级学生 20%、四年级学生 20%。分层应该是以年级来进行，按 3 : 3 : 2 : 2 来选取各年级的学生。又如，某中学共有 2 700 名学生，男生 1 620 人，女生 1 080 人，欲

抽取 120 人作为研究样本。如果用简单随机取样，有可能抽出的男生过多，也可能女生过多。如果想调查全校学生的平均身高，男女差异就很大，男生过多或女生过多，都会对调查结果产生较大的影响，因此最好采取分层取样，按男女比例 3:2 来选取。在男生中用简单随机取样抽出 $120 \times 0.6 = 72$ 人，在女生中用简单随机取样抽出 $120 \times 0.4 = 48$ 人，此 120 人就构成研究的样本。

分层随机取样具有以下优点：

(1) 样本的代表性强。由于分层是按总体的某些特征来进行的，层与层之间的变异大，而各层内的变异小，这样根据层的规模抽取的样本代表性更强。

(2) 取样更灵活。由于各层的比例可能存在不一致，分层就可以按不同的比例在各层抽取，取样更加灵活。

(3) 参数估计更加准确，因为取样的误差往往发生在各层内，而很少发生在层与层之间，分层的目的是减少层内的变异，误差也随之减少，这样参数估计就更准确。

分层随机取样是把异质性较强的总体分成几个同质性较强的子总体，再抽取不同的子总体的样本分别代表该子总体，所有的样本代表总体。分层时可以考虑以下几个标准：(1) 以调查所要分析和研究的主要变量或相关的变量作为分层的标准。如要研究男女大学生在心理承受能力上的差异，可以根据男女来分层。(2) 以保证各层内部同质性强、各层之间异质性强、突出总体内在结构的变量作为分层变量。例如，某项研究要考察中学生在时间管理倾向上的差异，根据以往的研究结果，重点中学、一般中学学生时间管理倾向有显著差异，因而进行分层取样时就按重点中学、一般中学分为两层。(3) 以那些有明显分层区分的变量作为分层变量。

3. 系统取样

系统取样 (systematic sampling) 又称等距取样或机械取样，是以某种系统规则来选择样本的方法。系统规则如“每隔 10 个”、“每第三列”等等。具体步骤是先将总体中的个体按顺序排列起来，从总体中随机取一起点，然后从该起点开始每隔若干个抽取一个，直到抽满先定的样本量为止。例如，如果总体名册中包含 10 000 个个体，研究需要样本容量为 1 000 个个体的样本时，研究者可以选择每第十个个体作为样本。系统取样的关键是要确定好取样距离 (sampling interval) 和取样比率 (sampling ratio)。取样距离是两个被选择个体间的标准距离，计算方法是取样距离 = 总体大小 / 样本大小。在上述例子中取样距离为 $10\,000 / 1\,000 = 10$ 。取样比率是选择样本个体与所有总体个体数量的比率，计算公式是取样比率 = 样本大小 / 总体大小。上述例子中的取样比率为 $1\,000 / 10\,000 = 1/10$ 。

系统取样法比较简单，易于实施，在实际生活中应用较多。缺点是样本对总

体代表性不够，名单中个体的排列方式可能使系统取样产生问题。这种排列方式问题通常被称为周期性问题。如果个体名单是以与取样间隔一致的循环方式排列的，系统取样可能产生一个有重大偏差的样本。例如，一个有关二战的经典研究中，研究人员从名册中每隔 10 个士兵抽出一个来进行研究。然而士兵的名册是依下列的组织方式来编排的：首先是中士，接着是下士，其后才是二等兵；用一班一班的方式进行编排，每个班 10 个人。因此名册中每隔 10 个便是一位中士。系统取样就会得到一个完全是中士的样本，同样，这种方式也可能会取得一个完全不含中士的样本。因此，要进行系统取样，应该先考查个体排列的基本特征。如果个体有特定的排列顺序，必须确定该顺序是否会使样本产生偏差。

4. 聚类取样

许多研究都需要从不易获得取样名单的总体中取样，例如：一个省、一个城市、一个国家的人口、全国的大学生等等。在这些总体中取样，取样设计更加复杂，通常需要采用聚类取样 (cluster sampling)。聚类取样又称整群取样，是将总体按照某种标准（如班级、地区）划分为若干子群体，每个子群体作为一个取样单位，用随机的方法从总体中抽取子群体，将抽中的子群中的所有单位合起来作为总体的样本。例如，要调查某学校学生的时间管理倾向和控制源的关系，假定该学校有 20 000 人，400 个班级，准备抽取 300 人作为调查对象，这样可以将班级作为取样单位（即取样的子群体），随机抽取 6 个班级，然后对 6 个班级的所有学生都进行调查。

聚类取样的特点是调查单位比较集中，调查工作的组织和进行比较方便，问题是调查单位在总体中分布不均匀，准确性相对随机取样要差。因此，在类别相对较多、每一类中个体相对较少的情况下采取此种方法效果最好。

聚类取样和分层取样既有相似之处又有不同。相似之处是这两种方法的第一步都是根据某种标准将总体划分为一些子群体。不同之处是分层取样要求各子群体之间的差异较大，而子群体内部差异较小；聚类取样要求各子群体之间的差异较小，而子群体内部的差异性很大。换句话说，分层取样是用代表不同子群体的子样本来代表总体中的群体分布；聚类取样是用子群体代表总体，再通过子群体内部样本的分布来反映总体样本的分布。

5. 多段取样

多段取样 (multi-stage sampling) 也称多级取样或分段取样，先将总体按某种标准（如学校、地区）分成若干层（组），称为第一层，然后从这一层中随机抽取几个层作为第二层，依次类推，直至从最后一层中用随机取样的方式抽取一定数量的样本作为研究所需样本的取样方法。

例如，研究者要调查中国高校教师的教师效能感，可采用三阶段取样的方法：先以省份为取样单位，抽出若干省；然后以学校为取样单位，从抽中的每个省

中抽出若干学校，最后以教师个人为单位，从抽中的学校中随机抽取最终样本。

多段取样适用于总体范围大、对象层次多的研究，能有效地节省调查所需的人力及费用。但由于每级取样时都会产生误差，故这种取样方法的误差较大，这是它的主要不足。在同等条件下减少多段取样误差的方法是：相对增加开头阶段的样本数而适当减少最后阶段的样本数。所以，当研究者的人力和经费允许时，应尽量扩大开头阶段的取样规模。

聚类取样和多段取样常结合起来使用，统称为多段聚类取样。

（二）非随机取样

心理学研究常会遇到无法选择概率样本的情形，如总体的规模太大，研究的人力物力有限，研究者只是进行探索性的研究等，这时进行随机取样并不适当。在这些情况下，就应该采用非随机取样。非随机取样是事件或元素被抽取的概率未知，总体中的每个个体进入样本的概率未知的一种取样分类。常用的非随机取样有以下四种：方便取样、立意取样、定额取样和滚雪球取样。

1. 方便取样

方便取样（convenience sampling）也称偶遇取样，研究者选择在方便的时间和地点将所遇到的人作为研究样本的方法。方便取样采用就近的原则，如在街道、商场、车站、广场等人群密集的地方对过往的行人进行公众观的调查，电视台和报社的记者就某些重大事件对公众进行的街头采访等。方便取样的另一种情况是研究者采用招募的方式来选择被试。通常，那些自愿参加研究和容易招聘的被试成为了研究的样本。

方便取样根据研究者的意愿和便利来抽取样本，受到研究者的欢迎。但根据这种方法取得的样本缺乏对总体的代表性，偶然性很大。只有容易见到、方便的对象能够被研究者抽取。因此，方便取样得到的资料，通常很少有实际上的价值，它可以作为问卷的前测，但不应该代表总体。

2. 立意取样

立意取样（purposive sampling）也称主观取样、判断取样，研究者根据自己的经验，如对总体构成要素和研究目标的认识，主观判断选取可以代表总体的个体作为样本。立意取样是基于研究者对总体的了解和经验，从总体中抽取有代表性和典型的单位作为样本，受主观因素影响较大，样本的代表性取决于研究者对总体的了解程度和判断能力，主要用于试测和可行性研究。例如，在过去的五年中，一个中西部学区的教师协会的领导们在大多数问题上代表了本区四分之三教师的观点。因此，这一年，学区主管决定只访谈这个协会的领导，而不再从本地区的所有教师中取样本（弗林克尔，瓦伦，2004）。

立意取样主要优点在于可以充分发挥研究人员的主观能动作用，有目的选择样本。特别是当研究者对研究总体的情况比较熟悉、研究人员的分析判断能

力较强、研究方法与技术十分熟练、研究的经验比较丰富时，采用这种方法往往十分方便。例如，在问卷设计阶段，为检验问卷设计的条目是否得当，常有意选择一些观点差异悬殊的人作为调查对象。又如，研究者特意找那些偏离总体平均水平的被试作为调查对象，研究是什么原因使他们发生偏离，目的在于发现问题，提出假设，而不是对总体作出概括。但是由于它仍然属于一种非概率取样，所以，其所得样本的代表性往往难以判断。

3. 定额取样

定额取样 (quota sampling) 也称配额取样，是按特定的标准（性别、年龄、职业、受教育程度、教育背景等）将总体中的个体分成若干类或层，然后在各层中取样。例如，假设某高校有 20 000 名学生，其中男生占 60% 女生占 40% 文科和理科学生各占 50%，一年级、二年级、三年级、四年级学生分别占 40%、30%、20%、10%。现要用定额取样的方法依上述三个标准（男女、文理和年级）抽取一个规模为 1 000 人的样本。依据总体的构成和样本规模，可得到下面的定额表(表 2-2)：

表 2-2 定额取样人数分配表

		男生(60%)								女生(40%)							
		文科(30%)				理科(30%)				文科(20%)				理科(20%)			
年级		一	二	三	四	一	二	三	四	一	二	三	四	一	二	三	四
人数		120	90	60	30	120	90	60	30	80	60	40	20	80	60	40	20

4. 滚雪球取样

滚雪球取样 (snowball sampling) 是先从总体中合适的调查对象开始进行调查，再通过他们得到更多的调查对象，如同滚雪球一样。例如，想要研究下岗职工的心理健康状况，可以就近的找几个厂的下岗工人作为调查对象，然后通过他们提供的情况再找其他的调查对象，依次类推，使样本像滚雪球一样，由小变大。

滚雪球取样有时产生的取样误差会很大，如总体不大的情况下，取样几次后就会接近饱和状况，即后面调查的被试都是前面已经调查过的人，而且有很多被试仍然无法调查到，或者某些被试由于某些原因而故意的漏掉某些符合要求的被试。滚雪球取样适合于探索性研究和特定的总体成员难以找到的情况。

补充讨论 2-2 取样的有效性

当研究者想将某个特定的研究结果推广到研究中所使用的特定总体时，他必须保证抽取的样本能够代表它所属的总体。样本能很好的代表总体，取样的有效性就能够得到保证。因此，取样的有效性取决于样本的代表性。当研究者

抽取的样本特征大体接近试图研究的总体特征时，样本就具有代表性。但样本不需要在每个方面都具有代表性，代表性只需局限在与研究的实质性需要相关的特征。取样的有效与以下因素有关：

第一，取样的方法对取样有效性有决定性的影响。通常，随机取样比非随机取样有效。随机取样具有两项特点：其一，随机取样较其他取样方法更具代表性。它能更有效地降低取样误差；其二，随机取样遵循的概率理论使研究者能够估计样本的精确度和代表性。但是用随机取样的方法抽取的样本也无法绝对完美的代表总体。

第二，样本容量的大小也影响取样的有效性。样本容量是样本中所包含个体或单位的数目，常用 n 表示。样本容量的确定是取样设计的重要内容之一，样本容量直接影响到样本本身的代表性，也影响到调查的费用、人力和时间，样本太大会浪费人力、物力和财力，增加不必要的工作量，甚至使研究半途而废，样本太小则会影响研究的效果。从理论上而言，总体越大，样本也应该越大，样本平均数才能越接近总体平均数。实际上，选取较大的样本，并不一定能显著的提高研究的准确性，只有在总体相对小的情况下，样本的大小才比较明显的影响参数估计的准确性。

第三，研究总体的同质性程度。总体的同质性是指总体内各个个体的差异程度，一般用方差或标准差来表示。对同一总体来说，抽取相同量的样本，标准差越大，同质程度就越小，取样的有效性会降低。反之，标准差越小，同质程度就越大，取样的有效性会提高。

第四，研究的时间和经费。取样的有效性也应该考虑样本规模是否与准备用于研究的时间、人力、费用等相适用。例如，访谈研究比问卷研究需要更多的时间、人力和费用，因此只能选取小样本，降低有效性。在某些情况下，研究时间和经费的限制会使研究者无法按理论上的要求来确定规模，因此，研究者只能减小样本规模，损失精确度，或者增加时间和经费，保证精确度。

► 本章提要

1. 心理学研究课题可以有多种来源。主要的来源是：(1)对日常生活的观察；(2)实际的需要；(3)理论；(4)研究文献的提示；(5)技术发展的推动等；也可以是这些来源的某种结合。

2. 好的研究问题应该具备四个基本特征：(1)问题是切实可行的；(2)问题是清楚的；(3)问题是有意義的；(4)问题是符合道德的。

3. 文献是记录、保存、交流和传播知识的一切材料的总称。心理学文献就其保存方式可分为纸质文献和互联网文献。纸质文献的常见类型有图书、期刊、

会议和学位论文等。

4. 查阅文献常用参考文献查找法、检索工具法。检索工具法分为手工检索和计算机检索。

5. 取样和选题的关系密切，要依据课题的性质和研究结果的概括程度来选择被试。取样的方法很多，根据是否按随机原则进行操作，取样可分为两类：非随机取样和随机取样。

6. 简单随机取样又称纯随机取样，即对研究总体单位不进行任何分组排列，只按随机原则直接从总体中抽取一定的样本，以使总体的每一个样本都有被同等抽取的可能性。分层随机取样，也称比率取样、分类取样和分组取样，是按照总体已有的某些特征，将总体分成几个不同的部分（每个部分叫一个层或一个子总体），然后在每一个层或子总体中进行简单随机取样。

7. 系统取样，又称等距取样或机械取样，是以某种系统规则来选择样本的方法。聚类取样，又称整群取样，是将总体按照某种标准（如班级、地区）划分为若干子群体，每个子群体作为一个取样单位，用随机的方法从总体中抽取子群体，将抽中的子群中的所有单位合起来作为总体的样本。

8. 多段取样，也称多级取样、分段取样，先将总体按某种标准（如学校、地区）分成若干层（组），称为第一层，然后从这一层中随机抽取出几个层作为第二层，依次类推，直至从最后一层中用随机取样的方式抽取出一定数量的样本作为研究所需样本的取样方法。

9. 方便取样，也称偶遇取样，研究者选择在方便的时间和地点将所遇到的人作为研究样本的方法。立意取样，也称主观取样、判断取样，研究者根据自己的经验，如对总体构成要素和研究目标的认识，主观判断选取可以代表总体的个体作为样本。

10. 定额取样，也称配额取样，是按特定的标准（性别、年龄、职业、受教育程度、教育背景等）将总体中的个体分成若干类或层，然后在各层中取样。滚雪球取样，是先从总体中合适的调查对象开始进行调查，再通过他们得到更多的调查对象，如同滚雪球一样。

►本章关键术语

文献(literature)

教科书(textbook)

专著(monograph)

手册(handbook)

年鉴(yearbook)

期刊(periodical)

随机取样(random sampling)

简单随机取样(simple random sampling)

分层随机取样(stratified random sampling)

系统取样(systematic sampling)

取样距离(sampling interval)

取样比率(sampling ratio)

杂志(journal)

学报(acta)

公报(bulletin)

回顾(review)

文摘(abstract)

非随机取样(non - random sampling) 滚雪球取样(snowball sampling)

聚类取样(cluster sampling)

多段取样(multi - stage sampling)

方便取样(convenience sampling)

立意取样(purposive sampling)

定额取样(quota sampling)

►复习与练习

- 下面列出了一系列的选题，用本章所介绍的选题原则对它们进行评价。
 - 父母教养态度与儿童在 2 ~ 4 岁期间的问题行为
 - 同性教师与大学生的学习兴趣
 - 大学生的压力与心理健康
 - 青少年吸烟行为的预防与干预
 - 不同注意条件下大数与小数的加工差异
- 根据选题的原则，请从日常观察、实际需要、理论等方面至少提出 5 个心理学的研究问题。
- 从《心理学报》或《心理科学》中选择 3 篇研究报告，分析并评价这些研究所采用的抽样方法。
- 假设某研究者想调查全市中学生学习心理健康状况，该市共有 400 所中学，共 320 000 名学生，这些中学分布在 9 个区县中，其中重点中学有 40 所，一般中学有 320 所，较差的中学有 40 所。现在要求从全市的中学生中抽取 1 600 名中学生进行调查，以了解他们的学习心理健康状况。请设计一种取样方案。

►推荐参考读物

- 弗林克尔，瓦伦. 教育研究的设计与评估. 蔡永红等译. 北京：华夏出版社，2004
- 张力为. 体育科学研究方法. 北京：高等教育出版社，2002
- 艾尔·巴比. 社会研究方法. 邱泽奇译. 北京：华夏出版社，2000
- Seymour, S. Applied Sampling. Handbook of survey Research edited by P. H. Rossi, J. D. Wright, & A. B. Anderson. New York : Academic Press 1983. 145 ~ 194.

►在线资源

想了解更多取样的知识，可以访问 <http://www.tardis.ed.ac.uk/~kate/qm-cweb/scont.htm>

想获取更多的心理学网络资源，可以访问 <http://www.psychwww.com/>

第三章 假设与解释

在日常生活中人们经常需要对某种心理或行为的原因作出说明，即进行解释。例如，为什么有些人很聪明，为什么有些人会紧张，等。这也是心理学研究的目标。只不过心理学研究要解释的是系统收集而来的纷繁复杂的观察资料。

观察资料的收集不是随便进行的，而通常是在研究者的假设指引下进行。假设是科研者最重要的思想方法，其主要作用是提出新实验或新观测。在心理学研究中绝大多数观测都是以验证假设为目的而进行的。假设的另一个重要作用是使我们看清事件的重要意义，如果没有假设，那么这一事件就不说明问题，例如在进行现场调查时，一个用行为论武装头脑的人就比没有任何假设的人能看出更多东西。

从假设的提出到最后观测结果的解释都不是随便进行的，这些活动受心理学家研究心理现象的视角即范型的指引。人们在认识世界、看待问题时总是从某个角度或立场出发。由于心理现象的复杂性使得心理学家从不同的视角来研究心理现象。

了解心理学家研究心理现象的视角，如何提出假设以及如何进行解释是深入认识心理学研究方法的重要途径。在本章中，我们将讨论以下问题：

1. 心理学研究的范型
2. 理论与假设
3. 解释

第一节 范型、假设与解释

范型(**paradigm**)是指看待问题的基本角度或框架。它为研究者提供观察事物的基本方式和信念，决定或影响着研究者所选择的研究内容、方法、策略和所提出的理论。这个概念最早由科学哲学家库恩(Kuhn, 1962)提出。比较成熟的自然科学如物理、化学、生物学等都已经形成了比较统一的研究范型。而心理学迄今还没有形成一个统一的研究范型。在心理学中主要的研究范型有生物学范型、心理动力学范型、行为论范型、人本论范型和认知论范型。

一、生物学范型与研究

生物学范型(**biological paradigm**)引导心理学家在大脑、神经系统、内分泌系

统及基因中寻找心理和行为的原因。认为心理现象即使是最复杂的心理现象（如自我、羞怯）都能够最终依据脑细胞或其生化过程加以解释。采用该范型的研究者共同关心的是心理和行为受生物因素决定的方面。

例如，罗森茨韦格等（Rosenzweig et al, 1972）做的“大脑会因经历的不同而改变”的著名研究就是探讨脑与心理现象的关系。在这项研究中，研究者共进行了 16 次实验，每次实验都使用了 12 组老鼠，每组有 3 只老鼠，都是从同一胎所生的老鼠中选择出的雄鼠。在每次实验中每组的 3 只老鼠都被随机分配到 3 种不同的实验条件中，即每次都有 12 只老鼠被安排到每一种实验条件中。这三种不同的实验条件分别是：①标准的实验室笼子，即有几只老鼠生活在足够大的空间里，笼子里总是有适量的水和食物；②贫乏环境，即一个比较小的笼子，老鼠被放置在单独隔离的空间里，笼子里总有适量的水和食物；

丰富环境，即 6 - 8 只老鼠生活在一个有各种可供玩耍的物品的大笼子里，每天从 25 种新玩具中选择一个放在笼子里。研究者让老鼠在不同环境里生活了 4 到 10 周时间，之后对它们的大脑进行解剖。为了避免实验者的偏见，解剖时不让解剖人员知道老鼠是在哪种环境下成长。解剖后对老鼠大脑的各个部分进行测量、称重和分析。结果发现，在丰富环境中生活的老鼠的大脑皮质比在贫乏环境中生活的老鼠的要显著的更重和更厚，如图 3.1 所示；丰富环境下老鼠大脑中的乙酰胆碱更具活性；丰富环境下的老鼠的神经元更大；在丰富环境中长大的老鼠的大脑中的神经突触比在贫乏环境中长大的老鼠的要大 50%。

又如，在探讨内分泌系统与心理现象的关系的一项研究（Terkel & Rosenblatt, 1968）中，研究者给 4 组（每组 8 只）未成熟的雌性老鼠分别注射了分娩未超过 48 小时的老鼠的血浆（母性血浆）、发情前期老鼠的血浆、发情末期老鼠的血浆以及盐溶液，并让这些雌性老鼠接触 5 只 5 至 10 天大小的幼崽，观察这些雌性老鼠出现护崽等母性行为的潜伏期。结果如表 3 - 1。从表中可以看出，在母性血浆和来自幼崽的刺激联合影响下，平均两天多的时间，那些注射了分娩未超过 48 小时的老鼠的血浆的雌性老鼠就出现了护崽等母性行为，比注射发情前期或末期的老鼠的血浆、盐溶液的老鼠出现母性行为的时间要短得多。这一结果表明分娩后出现的母性行为有体液基础，刚做母亲的老鼠的血浆中携带某种物质，能够促使未成熟的雌性老鼠对幼崽做出母性行为。

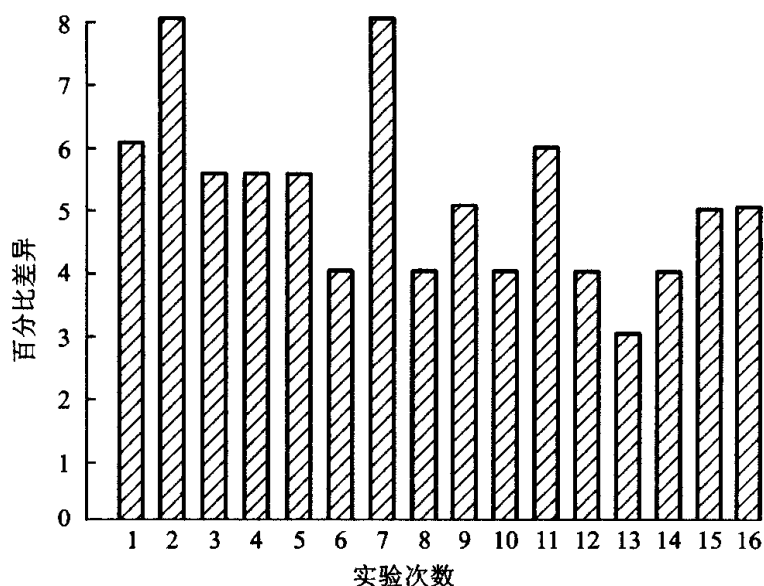


图 3.1 丰富环境与贫乏环境相比老鼠大脑皮质增加的比率
(采自 Rosenzweig et al,1972)

表 3 -1 护崽行为发作的平均潜伏天数

组 别	N	Mean	SE
母性血浆	8	2.25	0.97
发情前期血浆	8	4.62	1.21
发情末期血浆	8	7.00	2.96
盐溶液	8	4.00	1.41

(采自 Terkel & Rosenblatt,1968)

在探讨遗传基因与心理现象关系的一项研究 (Bouchard et al,1990) 中 ,研究者对 56 对来自美国等 8 个国家的分开抚养的同卵双胞胎 (MZA) 以及那些共同成长的同卵双胞胎 (MZT) 进行了将近 50 个小时的包括生理与心理特征的一系列测试。结果如表 3 -2 所示 其中 $R(MZA)$ 表示分开抚养的同卵双胞胎在这些特征上的相关系数 , $R(MZT)$ 表示共同成长的同卵双胞胎在这些特征上的相关系数 , $R(MZA) / R(MZT)$ 表示这两种相关系数的比值 , 其反映了分开抚养和共同成长的同卵双胞胎在这些特征上的相似性。从表中可以看出 , 分开抚养的同卵双胞胎在一些主要的心理特征上都比较相似 , 分开抚养和共同成长的同卵双胞胎在这些特征上的相关系数非常近似。这种结果暗示 , 许多个体差异都与遗传因素有着密切关系。

表 3-2 分开抚养的同卵双胞胎 (MZA) 与一起抚养的同卵双胞胎 (MZT)
在某些特征上的相关系数的比较

特 征	R (MZA)	R (MZT)	相似性 R (MZA) / R (MZT)
生理	—	—	—
脑电波	0.80	0.81	0.987
血压	0.64	0.70	0.914
心率	0.49	0.54	0.907
智力	—	—	—
韦氏成人智力量表	0.69	0.88	0.784
瑞文智力测验	0.78	0.76	1.030
人格	—	—	—
多维人格问卷 (MPQ)	0.50	0.49	1.020
加利福尼亚人格问卷	0.48	0.49	0.979
兴趣	—	—	—
史特朗-康久尔兴趣问卷	0.39	0.48	0.813
明尼苏达职业兴趣量表	0.40	0.49	0.816
社会态度	—	—	—
宗教信仰	0.49	0.51	0.961
无宗教信仰社会态度	0.34	0.28	1.210

(采 自 Bouchard et al,1990)

二、心理动力学范型与研究

心理动力学范型 (psychodynamic paradigm) 引导心理学家在与生俱有的本能和生理驱力中寻找心理和行为的原因，例如剥夺状态、生理唤起、冲突等都给心理和行为提供力量。而有机体的需要一旦得到满足，其驱力就降低，行为反应也就停止了。

例如，该范型的创始人弗洛伊德 (Sigmund Freud, 1856—1939) 对小汉斯 (Hans) 的个案研究 (Freud, 1909/1955) 就反映了这些特点。首先看一看案例中介绍的小汉斯的主要症状：

在三岁的时候，小汉斯对自己的阴茎特别感兴趣，把它称之为“小摇铃”。他喜欢摸它，也经常注意别人的阴茎，还问他母亲是否也有一个“小摇铃”。他看到马的阴茎很大，然后对母亲说：“我想，你（母亲）这么大，你的小摇铃一定和马的一样大。”他发现动物有阴茎，而桌子、椅子没有。他对许多事情都感到好奇。汉斯三岁半时，他的妹妹出生了，他说：“她的小摇铃太小了。等她长大后，

反正就会变大。”后来，他母亲告诉他不要摸自己的阴茎并威胁说要把它割掉。当汉斯四岁半时，在一次洗完澡，他母亲在他阴茎周围扑痱子粉时，他问为什么她不摸他的阴茎。他母亲解释说这里脏，而汉斯却笑着说它很好玩。大约在六个月后的一天，汉斯同他的保姆一起外出散步。当一架马车驶过时，他大哭起来，要赶紧回家找母亲。此后不久，他对马十分恐惧，怕马会咬他。在此之前，他常做噩梦，他母亲常把他抱过来同自己一起睡。他开始怕马会闯进他的房间，并很快发展成为典型的对马的恐惧症。他对马嘴边的黑圈和马的眼罩感到恐惧；对拖着笨重货车的马以及马跌倒了并跌坏了腿等都感到恐惧。

弗洛伊德在考察了这一个案的详细资料后，就对小汉斯的马恐惧症作了解释。简言之，小汉斯对他的阴茎很感兴趣，他对母亲有强烈的性依恋，想母亲摸他阴茎并同她一起睡觉。会咬人的马象征他父亲，马嘴的黑圈和眼罩象征他父亲的胡子和眼镜，马会咬他，象征被阉割。从这些要素及其他要素中，弗洛伊德认为小汉斯的恋母情结十分强烈，同时把父亲看做他的情敌。他十分惧怕因这些欲望会受到父亲阉割惩罚，因此他内心希望父亲也像那匹马一样跌倒而死。他对马的恐惧，在潜意识中是想避开父亲、即避开马，以减轻其对被阉割的焦虑。总之，小汉斯的恐惧症最主要的起因是恋母情结的冲突。

此外，投射测验也反映了该范型的研究特点。投射测验 (projective test) 是提供较模糊的刺激情境，让被试在不受限制的情况下自由地作出反应，使其将隐藏在潜意识中的欲望、动机投射出来。其中，罗夏 (Rorschach, 1942) 编制的罗夏墨迹测验 (Rorschach inkblot test) 就是其中的经典之一。罗夏 (Hermann Rorschach, 1884—1922) 认为给被试呈现模棱两可的知觉刺激，其注意力会被转移，心理防御会减弱，这时对刺激的解释就能反映其潜意识倾向，即潜意识被投射出来了。罗夏选择了随机墨迹图作为自己的投射测验的刺激图。这种图形的制作方法非常简单，就是将墨汁滴在一张纸上，将纸对折，墨迹就会在纸的两半形成对称的图形。不过，并不是这样的图形都能使用。经过大量测试，罗夏最后确定了 10 张图形 其中 5 张图是黑白色，2 张是黑白色加红色，3 张是彩色。图 3.2 是测验图片示例。测验的实施非常简单，即每次向被试呈现一张图片，并问：“这可能是什么？”被试可以自由观看该图片，如改变方向、拿远拿近。主试则记录下被试对每个图形的所有反应。罗夏给出了几条评分规则：①总共反应了多少次？反应时间？拒绝解释的次数？解释是仅根据形状，还是包括了颜色？是否包含了运动？图形被知觉为一个整体还是几个部分？哪些部分被分离出来？被试对此有何解释？被试看到了什么？罗夏对从正常人到精神病患者的不同群体进行了测试，并发现了一些差异。例如，被试通常对 10 张图片总共有 15 至 30 个反应，抑郁症的被试一般给出的答案较少，快乐的被试给出的答案较多，精神分裂症患者的答案数目则有很大的个体差异。回答看到了动物是最常

见的，抑郁症患者回答动物的比率最高，而艺术家则最低。对于在测验中发生概率很小的独创性回答在精神分裂症的被试中最为常见，在中等智力水平的正常被试中则最少见。

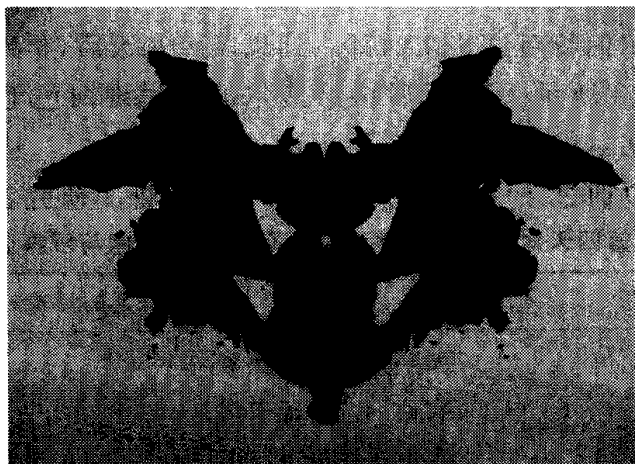


图 3.2 罗夏墨迹图示例

三、行为论范型与研究

行为论范型 (behavioral paradigm) 引导心理学家在特定的环境中去探寻特定行为模式是如何产生的，他们认为人类的所有行为都源于学习。

该范型的创始人华生 (John Broadus Watson, 1878—1958) 进行的一项著名的关于恐惧习得的研究 (Watson & Rayner, 1920) 就反映了这些特点。该研究中，研究者使用的被试是一名叫艾尔伯特的 9 个月大的孤儿。为了了解艾尔伯特是否惧怕某种刺激，研究者给他呈现白鼠、猴子、狗、有头发和没有头发面具以及白色羊绒棉。研究者发现艾尔伯特对许多动物和物体都感兴趣，愿意接近它们，并不时触摸它们，从来都没有表现出丝毫的恐惧。因此，研究者把这些物品作为中性刺激。然后，研究者在艾尔伯特身后用锤子敲一根 1.2 米长的铁棍，以确定艾尔伯特是否对巨大的声音产生恐惧反应。结果发现，这种声音的突然出现，使艾尔伯特受到惊吓哭泣。在艾尔伯特 11 个月时，研究者开始真正进行建立恐惧条件反射的实验。实验开始时，先向艾尔伯特出现白鼠，当他对白鼠感兴趣并正试图伸手触摸它时，研究者突然敲响铁棍，突然的巨响使艾尔伯特十分害怕。这一过程重复了 3 次，一周后，继续重复同样的过程。当白鼠和令人恐惧的声音配对呈现 7 次以后，单独向艾尔伯特呈现白鼠，而不呈现声音，发现艾尔伯特对白鼠产生的极度恐惧使他大哭起来，转身背对老鼠，并向远离它的方向移动，他爬得如此之快，以至于研究者不得不冲过去抓住他，以免他从桌子边掉下

去。一周后，再次对艾尔伯特进行测试，发现他仍然对白鼠感到恐惧。之后，研究者又进一步研究了这种恐惧的泛化问题。研究者给艾尔伯特呈现一种与白鼠类似的动物即白兔，发现艾尔伯特尽可能远离白兔，先低声抽泣，然后大哭起来。同一天，研究者依次给艾尔伯特呈现狗、白色毛皮大衣、一袋棉花和华生头上的灰白头发，发现艾尔伯特对这些东西都感到恐惧。几天后，再次对艾尔伯特进行测试 结果如表 3-3 所示。31 天后再给艾尔伯特呈现白色毛皮大衣、白鼠、白兔和白狗等，发现艾尔伯特对这些东西仍然感到十分恐惧。从以上结果可以看出，恐惧反应是可以通过条件发射方式习得的，这验证了研究者的假设。

表 3-3 第四天对艾尔伯特测试时呈现刺激的顺序与观察到的反应

呈现的刺激	观察到的反应
1. 积木	像平常一样玩积木
2. 白鼠	害怕、后退（没有哭）
3 白鼠 + 噪音	害怕并哭泣
4 白鼠	害怕并哭泣
5. 白鼠	害怕、哭泣并离开
6. 兔子	害怕，但不如前面表现得强烈
7. 积木	像平常一样玩积木
8. 兔子	害怕，但不如前面表现得强烈
9. 兔子	害怕，但不如前面表现得强烈
10. 兔子	有点害怕，但还想触摸它
11. 狗	害怕 回避
12. 狗 + 噪音	害怕并离开
13. 积木	像平常一样玩积木

四、认知论范型

认知论范型 (cognitive paradigm) 引导心理学家用信息加工的观点来研究感觉、知觉、注意、表象、学习、记忆、思维和语言等心理过程。认为认知活动是个体行为的原因与结果，用信息加工的模型来研究心理过程才能充分了解心理与行为现象。

例如，阿特金森和希夫林 (Atkinson & Shiffrin, 1968) 提出的记忆系统结构模型 (如图 3.3 所示) 就反映了该范型理论的特点。该理论强调记忆结构的区别，认为记忆有感觉登记、短时贮存和长时贮存 3 种贮存。外部信息最先输入感觉登记器 (sensory register, SR)。它具有通道特异性，按感觉道分别存储视觉的、听觉的和其他各类感觉刺激。感觉登记器的容量很大，所有的新刺激都可以短暂

地贮存于此，但会很快消退。那些被提取到的信息会进入短时贮存（short-term storage, STS）。短时贮存的信息会以听觉、言语或语言认知的形式重新进行组织。短时贮存的容量极其有限，材料在其中的贮存时间要比感觉登记长，在没有复述条件下，可以保持 15 至 30 秒。借助复述，短时贮存的信息可以转移到长时贮存（long-term storage, LTS）中。长时贮存的容量非常大，贮存时间也相当长，是一个真正的信息库，其中的信息有听觉的、口语的、言语的以及视觉的等编码形式。长时贮存的信息也可能会因消退、干扰或强度减少而不能被提取出来。

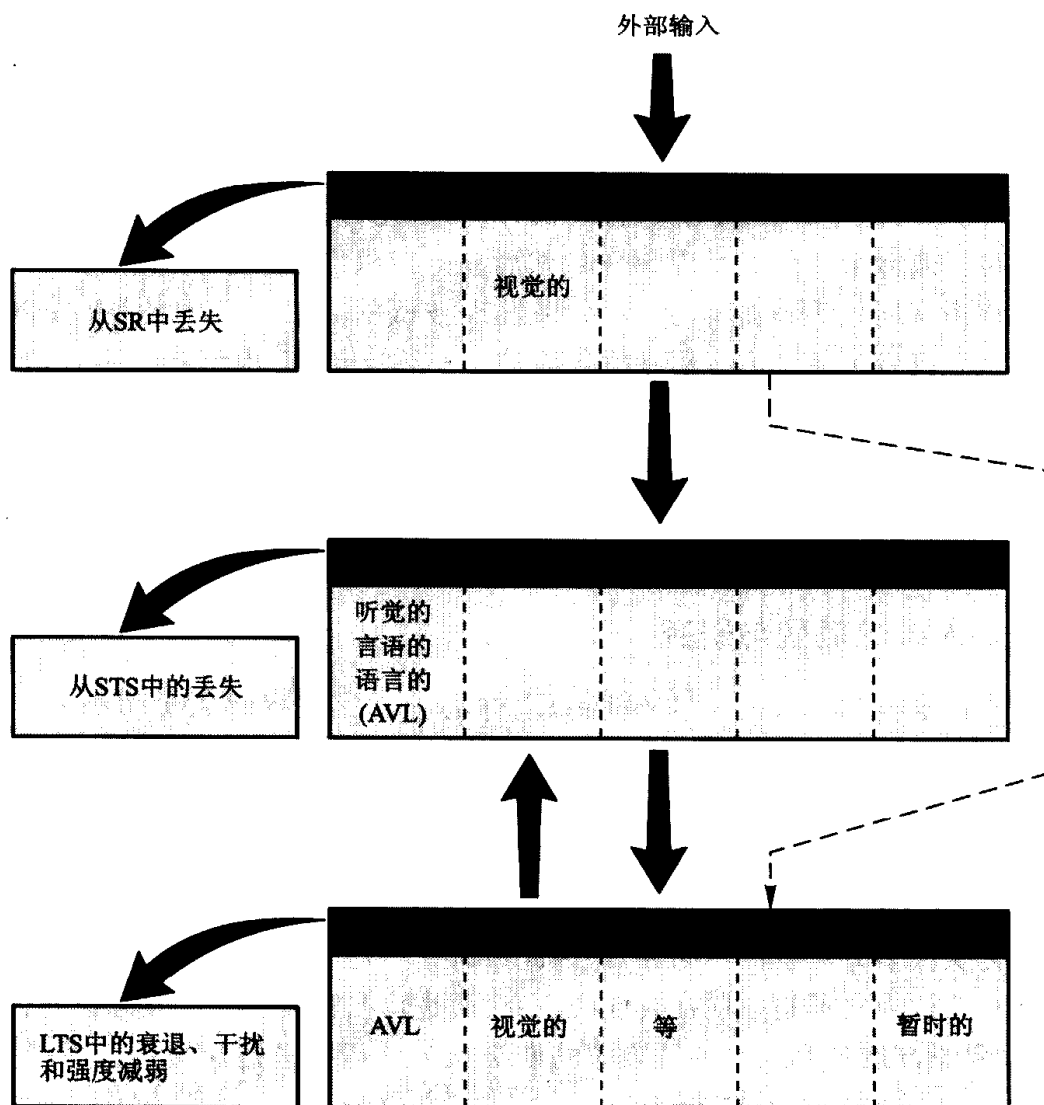


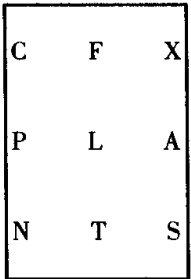
图 3.3 记忆系统结构模型

（采自 Atkinson & Shiffrin, 1968）

这个理论模型不是凭空构想出来的，而是建立在有关记忆的大量研究的基础之上。其中关于感觉登记部分，就有斯珀林（Sperling, 1960）的经典研究所支

撑。斯珀林的这项研究也很好的反映了认知论范型的研究特点。

在这项研究 (Sperling, 1960) 中，研究者先在一个实验中用速示器给被试呈现分为上、中、下三行，每行 3 个字母，共 9 个字母的卡片，如图 3.4 所示，卡片的呈现时间为 50ms。然后要求被试立即报告所记住的字母，结果发现被试通常仅能报告出 4 - 5 个，即 50% 左右的字母。之后，他改变实验方式，为了防止被试在报告先前的字母时忘了后面的字母，



C	F	X
P	L	A
N	T	S

图 3.4 实验用卡片示例

他创造出了一种新的报告方法，即部分报告法 (partial report)。这种方法不要求被试报告全部记住的内容，而是仅将记住的东西的一部分报告出来。具体是这样，斯珀林给卡片的 3 行字母的每行都配以不同的声音，即给最上面一行配以高音，给中间一行配中音，给下面一行配低音，并事先告诉被试每种声音所代表的那行字母。刺激卡片仍然呈现 50 ms，然后立即随机出现一个声音信号，要求被试报告该声音所代表的那行字母，结果发现被试的回忆率基本能达到 100%。虽然被试只报告了某个声音信号所代表的那一行字母，但由于这个声音信号是在卡片呈现完后随机出现的，因此可以推断被试其实是记住了卡片上的所有字母的。之前的实验中仅能报告出 50% 的字母是由于被试在报告时迅速遗忘了部分内容。斯珀林据此提出，存在一种感觉记忆，它的容量相当大，但信息保留的时间相当短暂。

五、人本论范型与研究

人本论范型 (humanistic paradigm) 引导心理学家去探寻人性积极的一面。他们把人视为自己命运的行动者，每个人都是自由的行动者，可以自由选择和控制自己的人生目标。

马斯洛 (Abraham Maslow, 1908—1970) 对自我实现的研究 (《动机与人格》马斯洛著，许金声等译，1987) 就能体现这些特点。马斯洛在自序中就首先承认这项研究从很多方面看都不同寻常，因此该研究不是按照常规的研究方式进行。马斯洛甚至认为它不是一项社会性的研究，而是旨在解决各种个人道德的、审美的、以及科学问题的私人性的冒险。马斯洛认为该研究只是让自己信服并且获得教益，而不是对其他人进行论证。

在研究之初马斯洛对什么是自我实现的人并没有给出什么明确的界定，也不知道应该选择哪些人作为研究对象。他对自我实现的定义、对自我实现者的挑选以及对自我实现特征的研究，三者几乎是同步进行、相互促进和不断完善的。他首先 (《动机与人格》马斯洛著，许金声等译，1987) “以个人或文化的非技术的信仰作为开始，对自我实现症候群的各种扩大用法和定义进行比较，然后再

仔细地给它下定义”，“根据修正过的通俗定义，第一批研究对象小组选出来了，其中一组质量高，一组质量低”，然后再“以临床风格对这些人进行尽可能仔细的研究，在经验研究的基础上，最初修正过的通俗定义又按照现在手中的材料进一步修改。这样就得出了第一个临床定义。按照这一新定义，对最初的研究对象进行重新筛选，一些人保留，一些人淘汰，一些新的成员补充进来。然后，又继续对第二种水平的研究对象小组进行临床研究，如果可能，进行试验和统计研究。这又导致了对第一个临床定义的修改、订正和补充。然后，又根据这一个新的定义进行再筛选。”经过这样不断的重复，马斯洛认为其关于自我实现的概念已经从最初模糊、不科学的状态变得越来越精确，在特性上也越来越便于操作，因而也越来越科学。他也选择出了符合要求的被试，同时，他对自我实现者的特征也越来越明确了。可以把马斯洛确定的自我实现者的挑选标准归纳为以下几条：①没有神经病、精神变态性格、精神病或这方面的冒险倾向；②满足或征服了生理需要、安全需要、归属与爱的需要和自尊的需要等需要；③“似乎在竭尽所能使自己趋于完美”，充分开发和利用自己的天赋和潜能。最后他选出了 49 位他认识的人以及一些历史人物如爱因斯坦、斯宾诺莎、林肯等。这些人物又可以分为三类：①非常理想的实现者；②很有希望的实现者；③有某些不足的部分的实现者。根据对这些人的研究，马斯洛归纳出了自我实现者的 15 种特征，见表 3-4。

表 3-4 自我实现者的 15 种特征

自我实现者的特征
1. 能够准确地知觉现实
2. 悦纳自己、他人和周围世界
3. 能自然地表达自己的情绪和思想
4. 超越以自我为中心，而以问题为中心
5. 具有超然独立的性格
6. 对于自然条件和文化环境的自主性
7. 对平凡的事物不觉厌烦，对日常生活永感新鲜
8. 具有高峰经验
9. 爱人类并具有帮助人类的真诚愿望
10. 有至深的知交，有亲密、温暖的家人
11. 有民主的性格，能尊重他人的人格
12. 道德标准明确，能区分手段与目的；绝不为目的而不择手段
13. 具有哲理的、善意的幽默感
14. 具有旺盛的创造力，不墨守成规
15. 对现有文化更具批判精神

人本论范型注重对个人积极的内心生活和经验的描述，因此自我价值感（self - worth）的研究也能反映该范型的特点。自我价值感是个人在社会生活中，认知和评价作为客体的自我（me）对社会主体（包括群体和他人）以及对作为主体的自我（I）的正向的自我情感体验（黄希庭、杨雄,1997）。通过使用开放式问卷与封闭式问卷调查以及使用探索性因素分析、验证性因素分析等方法进行的系列研究（黄希庭、杨雄,1997 黄希庭、余华,2002 黄希庭、凤四海、王卫红,2003），我们建构出了一个多维度多层次的自我价值感模型。这个模型包括总体自我价值感、一般自我价值感和特殊自我价值感三个层次，一般自我价值感层次又分为社会取向和个人取向两个维度，而特殊自我价值感层次又分为人际价值感、心理价值感、道德价值感、生理价值感和家庭价值感五个维度。通过这个过程还编制出了适合中国人特点的自我价值感量表，该量表为 5 点量表，共有 56 道题 要求被试就每一题在“完全不符合”、“大部分不符合”、“部分符合部分不符合”、“大部分符合”和“完全符合”五个选项中选择一个。使用这个量表可以了解人们对自己积极体验的程度。表 3 - 5 是该量表题项的示例。

表 3 - 5 自我价值感量表的题项示例

题 号	题 项
1	我喜欢主动与他人交往。
9	我很受周围人的欢迎。
15	在学校里，我常常当干部。
21	无论什么事，我是按自己的标准来衡量好坏的。
31	我有能力支配自己的生活。
52	我的父母都器重我。

库恩(Kuhn,1962)指出科学研究中的范型不会是一成不变的。这 5 种心理学主要的研究范型也是如此。在心理学的发展过程中这 5 种研究范型早已不是泾渭分明，而是相互渗透、相互影响了。特别是近 10 年来，关于研究范型整合的呼声和研究实践是越来越多。例如，心理动力学范型和人本论范型开始都不使用甚至反对使用实证的方法进行研究，后来也开始应用实证的方法。行为论早期反对研究意识，后来也研究心理过程，并提出在刺激一反应之间还应该有机体变量。新精神分析将弗洛伊德理论中所强调的生物决定论加以修正，增添了社会文化因素（人际关系）。认知论研究也开始关注内隐学习（ implicit learning）、内隐记忆（ implicit memory）等潜意识问题；认知论研究还与神经生理研究结合形成认知神经科学。

补充讨论 3-1 理性认识的客观性问题

科学研究是人类利用自己的理性来认识客观世界的一种有效方式。科学研究崇尚客观性原则，认为科学研究应该客观的进行，并可以发现客观规律。但是通过前面的讨论你会发现，人们在进行科学研究时却无法避免主观性的问题。因为研究者采用的研究范型影响了他们所选择的研究课题、提出的假设、使用的方法、解释结果的方式以及建构理论的角度。例如同样是对攻击问题进行研究，生物学范型的研究会从脑、遗传基因等角度来进行解释，心理动力学范型下的研究会从性本能角度来考察，行为论范型下的研究会考虑行为的强化，认知论范型下的研究则关心攻击者的认知特点。那么，人类对外部世界的理性认识能否客观呢？

人是用主观的头脑去认识客观的世界，因此任何一项科学研究都不能保证绝对的客观。但是这也绝对不意味着科学研究就都是主观的。科学研究通过一系列的措施来确保客观性，这些措施也成为科学研究的重要特征，如重复性、证伪性、开放性等，这些在第一章中已经讨论过了。

除此之外，保证科学研究客观性的重要手段就是对研究范型进行整合。每一种研究范型都是人们认识心理现象的一个角度，只用一个角度去认识心理现象，主观性就很强，如果多个角度综述起来，就能比较全面、客观的认识心理现象。正如盲人摸象一样，每个盲人都认识到了大象的一部分，如果用这一部分来概括整个大象就是主观的，但是把所有盲人所认识的大象整合起来，就能得到一个比较客观的认识。

对于心理学研究中的五种研究范型也是这样。这些范型在人类心理的研究重点、研究方法和研究水平上各有侧重。因此，五种研究范型都有其合理的因素，都看到了其他范型所忽略的观点，同时也都忽略了其他范型所揭示的一些重要方面，因此也有其不足之处。仅使用一种研究范型不能够客观认识心理现象，综合各种范型来研究心理现象则可以取长补短。

总之，导致某种心理现象出现的原因很多，每种研究范型都可以帮助我们认识其中一部分原因，只要综合使用多种角度和多种方法来研究心理现象就能提高研究的客观性和科学性。

第二节 理论与假设

前已述及，范型为研究者提供了看待问题的视角。不过，范型并不能直接解

释现象，要解释和预测有关现象就需要应用理论。理论（theory）是对某个研究领域提供完整、一致的描述并能对有关的全部事实进行系统解释的一套连贯的规范性陈述。理论可以将一些表面看起来没有什么关系的现象联系起来，并解释和预测现象。

一、演绎式理论与归纳式理论

建构理论所使用的逻辑方法会影响研究者如何提出理论。根据研究者是采用演绎逻辑还是归纳逻辑来建构理论可以将理论划分为演绎式理论和归纳式理论。归纳式理论是采用归纳的逻辑，根据一系列特定的观察资料而概括出的理论。这种理论的获得方式是从个别现象到一般规律。例如，上述马斯洛对自我实现的研究就是通过归纳他所找出的自我实现者的典型特征来建构出自我实现的理论。归纳式理论是进行质的研究的研究者通常采用的理论建构方式，有关内容可以参见第九章。当然，并非仅有质的研究才是建构归纳式理论的唯一方法，定量的研究也可以用来发展归纳式理论。例如，上述中过滤器模型和感觉记忆理论都是研究者通过归纳定量的资料而提出的。

运用演绎的逻辑，采用假设检验方式获得的理论称为演绎式理论。它是传统的科学研究中理论建构的典型模式。它与归纳式理论的显著区别在于，在研究开始阶段研究者就已经对现象做出了尝试性的理论解释，即提出了假设，然后才去进行观察、收集和分析资料以检验假设。如果假设得到验证，就说明之前提出的尝试性的理论解释是合理的，也就建立起了演绎式理论。例如，上述华生（Watson & Rayner, 1920）关于恐惧习得的研究就是获得演绎式理论的典型方式。在这项研究中，华生首先提出了“情绪反应是个体对环境中某种特定刺激的条件反射”的理论假说。根据这一假说，他进一步假设，如果有一种刺激（如巨大的声响）能自动导致个体产生某种特定情绪反应（如恐惧），而这种情绪体验（如恐惧）总是伴随着其他事物（如一只白鼠）出现，那么该事物就会与该种情绪建立联系。如上所述，他的实验确实证明通过条件反射个体会害怕原本不害怕的事物，也就是说害怕这种情绪反应可以通过条件反射习得。

在心理学研究中演绎式理论与归纳式理论都是建构理论的有效方式，它们没有高低之分。虽然我们平常可能会更多见到建构演绎式理论的研究，但归纳式理论同样有价值。其实，两者常常也是密不可分的，归纳式理论可以成为演绎式理论的源泉，演绎式理论则可以检验或发展归纳式理论。在心理学研究中，心理动力学范型和人本论范型中的大量理论都是归纳式理论，而且其中不少产生了巨大的影响力，如弗洛伊德关于潜意识、人格结构的理论，马斯洛的自我实现理论。

二、理论与假设的元素：概念与变量

(一) 概念与变量

我们在提出假设、建构理论以及对心理现象进行解释的过程中总是要用某些概念（例如依恋、潜意识、自立、自尊、好奇、动机、态度）来代表所要研究的现象，并用这些概念构成的体系来解释与概括所观察到的结果。因此，概念（concept）是理论与假设的基本元素，是进行研究的基础。

心理学研究中的概念大致可分为两类：一类是具有操作性定义（operational definitions）的概念，如速度、准确率、反应时。另一类是心理学家“发明”出来的尚无操作性定义的概念，如弗洛伊德理论中的本我（id）、自我（ego）和超我（superego）阿德勒（Adler, Alfred 1870—1937）提出的自卑情结（inferiority complex）和优越情结（superiority complex），马斯洛提出的高峰体验（peak experience）和约拿情结（Jonah complex）。

那些具有操作性定义，可以用量化的方法加以测量并赋予两个或两个以上值的概念就是变量（variable）。所谓操作性定义是一种规定，是根据一组产生被定义事物的操作所下的定义。也就是用一些可感知、度量、操纵的事物、方法来对概念进行具体的界定。如，可以把“智力”下一个操作性定义为“韦氏智力测验测得的分数”，“记忆”的操作性定义为“能够准确回忆出识记过的材料的内容的数量”。这样智力、记忆等概念就成为具有多值特点的变量，研究者也就可以对其进行测量或操纵。又如，上述罗森茨韦格等（Rosenzweig et al, 1972）的研究中“贫乏环境”的操作性定义是“一个比较小的笼子，老鼠被放置在单独隔离的空间里，笼子里总有适量的水和食物”；丰富环境的操作性定义是“6—8只老鼠生活在一个有各种可供玩耍的物品的的大笼子里，每天从25种新玩具中选择一个放在笼子里”。可见，通过操作性定义能使许多不能直接观察的心理现象成为可观测的变量（Myers & Hansen, 2002）。

定量的科学研究通常都以变量为研究对象，质的研究则可以针对尚无操作性定义的概念进行探究。关于质的研究，本书将在第九章中详细讨论。

(二) 心理学研究中的主要变量

1. 刺激变量、机体变量和反应变量

心理现象纷繁复杂，心理学研究中涉及的变量也是多种多样，但它们都属于三类最基本的变量，即刺激变量（stimulus variable）、机体变量（organism variable）和反应变量（response variable）。刺激变量简称S变量，是指已知的对有机体的反应发生影响的刺激条件，包括研究者可以操纵和控制的环境特征。有机体接受的刺激具有多种形式（黄希庭，1991）：自然性刺激和社会性刺激，前者如声、光、电、温度、气味等刺激，后者如言语、表情和动作所代表的社会意义如蔑

视、关怀、期待 ;②具体性刺激和抽象性刺激，前者如具体的人、事、物及其变化，后者如文字、符号、讯号 ;③外部刺激和内部刺激，前者来自机体外部环境，后者来自有机体内部的变化，如内分泌激素的变化、通过药物使有机体发生的变化以及头脑中浮现的思想、欲望等。

机体变量也称为被试变量 (subject variable)、属性变量 (property variable)，一般是指被试本身对行为反应有影响的特征。这些特征可以分为稳定特征与暂时特征，前者如年龄、性别、智力等，后者如动机、疲劳、情绪波动等。还可以分为生理上的特征和心理上的特征，前者如健康状况、身高、体重，后者如态度、动机。不同的机体变量的操纵方式是不一样的，一般来说性别等稳定特征是被试本身所固有的，实验者难以改变其特性，只能选择，这样的机体变量也称为属性变量 (attribute variable) 因此，属性变量一般不能成为因变量。而习惯强度、内驱力、动机、疲劳等暂时特征是可以控制和改变的。此外，有些机体变量可以定量控制，如饥饿程度可以用禁食的时间来度量，而有些如善良程度则难以定量分析。

反应变量 (response variable) 指由刺激引起的有机体行为的任何变化。有机体对刺激的反应是多种多样的，人对刺激的行为反应可以分为言语行为反应和动机行为反应。有机体的各种反应通常可以通过以下几个方面来度量：①速度，如完成一项作业所需要的时间或单位时间内完成的作业数量，简单反应的时间或潜伏期 ;②准确性，如走迷津进入盲路的次数，计算正确的题目数，击中靶心的次数 ;③难度或量表分数、评定分数，有些任务可以制定一个难度量表 (难易的等级或水平)，看被试能达到什么水平，如记忆广度、比纳 - 西蒙智力量表；④次数或机率，是指单位时间内被试做出反应的次数，如在心理物理实验中根据机率来确定阈值 ;⑤幅度和强度，如膝跳反射的幅度、光线的强弱。

2. 自变量、因变量和控制变量

从实验研究的角度看，又可以把变量分为自变量、因变量和控制变量。自变量 (independent variable) 指在实验中被实验者加以操纵的变量，如声音的响度、光的明亮、室内的温度、喂给老鼠的食丸数量。因此，刺激变量和机体变量都能够成为自变量，只要研究者认为能够影响行为，并能够对其进行操纵。在实验中，实验者相信操纵自变量就会引起行为反应。当自变量水平的变化导致了行为变化时，就可以认为自变量是有效的，或行为是在自变量的操纵之下。如果自变量操纵行为失败，就可能是实验者的研究假设有问题，也可能是未能有效操纵自变量，还可能是因变量的测量存在问题以及其他多种原因造成的。

因变量 (dependent variable) 是在实验中被试对刺激做出的反应，也就是实验者观测和记录下的随着自变量变化而变化的被试行为反应值，如被试反应的时间、速度、准确率与次数等。因此，在实验中因变量一般就是反应变量。因变

量依赖于实验者对自变量的操纵，例如增加奖金会导致生产效率上升，而减少奖金会导致生产效率下降。这里生产效率这种行为反应就依赖于实验者对奖金的操纵，因此生产效率是因变量。

控制变量(control variable)指在实验中实验者需要通过保持恒定等方法使其中立化的潜在变量。控制变量实际上就是潜在的自变量，如果实验中对它们控制不充分，它们就会随着自变量一起发生系统性的变化，可能减弱或增大自变量对因变量的效应。在任何一个心理学实验中，需要控制的变量都比实际控制的变量要多得多。例如，上述奖金与生产效率关系的研究中就需要控制富裕水平、智商、工作技能、工作态度、工作群体等很多因素。控制这些与实验无关的变量的方法很多，其中保持恒定是最直接和常见的方法。

三、假设检验

如上所述，演绎式理论建构的典型方式就是假设检验。要了解如何去建构演绎式理论就需要知道假设及其检验的过程。

(一) 什么是假设

假设(hypothesis)是指科学研究中任何一种对某些问题做出尝试性解释的陈述。这种尝试性解释至少有可能获得为检验所必需的证据来加以检验，即要么能予以证实，要么能予以否定。因此假设首先必须是一个能用科学研究的事实来加以检验的陈述，即是说，我们可通过科学研究来证明其对或错的那种陈述，而不是意见、价值判断或规范的陈述。例如，关于“每个大学生每天都应该吃早餐”的陈述，就是一个意见性的陈述。这是一个关于应该是什么的陈述，而不是一个可通过调研的事实以检验其对或错的陈述。而“98%的大学生每天都吃早餐”的陈述，则是一个能够用事实来加以检验的陈述。

显然假设不是随心所欲设想的或价值的陈述(虽然，研究人员的价值观会影响其对假设的选择)，而仅仅是研究人员认为是事实，但仍然是尝试性的和未经证明的一种陈述。为了使这个陈述得到证明，就必须对它加以检验；而要使它受到检验，就必须尽可能准确地加以陈述。因此假设中所涉及的概念必须要给出操作性定义。例如，我们假设“智力与主观幸福感之间存在着正相关”。要回答这个问题时，我们可以用韦克斯勒智力量表(Wechsler Intelligence Scale)上的得分和总体主观幸福感量表(General Well - Being Schedule)上的得分来界定“智力”和“主观幸福感”，然后检验这个假设是对的或是错的。

(二) 假设的类型

可以从不同的角度对假设进行分类。这种分类也反映了假设的一些基本特性。

从统计中假设检验的角度可以将假设划分为备择假设(alternative hypothe-

sis) 与虚无假设 (null hypothesis) 两类。备择假设亦可称为研究假设 (research hypothesis), 是陈述变量间所期望的关系, 通常是研究者期望证实的假设。备择假设不能直接检验, 其真实性可以通过对虚无假设的检验来推论。虚无假设亦称无差假设、零假设、原假设, 是陈述变量间没有关系或差异, 是根据样本统计量决定予以拒绝或接受的假设, 常以 H_0 表示。如“饲养在贫乏环境中的动物与饲养在丰富环境中的动物相比, 它们在大脑发育和化学物质等方面将表现出明显的不同”和“工作满意度会影响工作绩效”都是备择假设, 而相应的虚无假设则是“饲养在贫乏环境中的动物与饲养在丰富环境中的动物相比, 它们在大脑发育和化学物质等方面将表现为没有差异”和“工作满意度与工作绩效无关”。虚无假设不表达研究者对研究结果的预期, 是传统假设检验的出发点, 它通过统计技术判断所要研究的变量关系是否是偶然的。设定虚无假设是为了运用否定假设逻辑和反证的方法来对研究假设进行论证。科学研究中通常都使用这种论证方法, 因为科学研究是要发现事物之间的关系, 而这种发现不会非常容易, 因此进行假设检验就必须十分小心, 最谨慎的做法就是去检验虚无假设。

从假设是否定向的角度可以将假设划分为定向假设 (directional hypothesis) 与非定向假设 (nondirectional hypothesis) (Salkind, 2000)。定向假设是对变量关系的方向有所示意的假设; 非定向假设是对变量关系的方向没有示意的假设。例如, 上述的“饲养在贫乏环境中的动物与饲养在丰富环境中的动物相比, 它们在大脑发育和化学物质等方面将表现出明显的不同”就没有示意关系方向的假设, 因此是一个非定向假设。又如“贫穷家庭的学生的自立水平优于富裕家庭的学生”是一个定向假设, 而“贫穷家庭的学生的自立水平与富裕家庭的学生有差异”则是一个非定向假设。定向假设提供的信息更准确、具体, 一旦证实价值也更大, 当然同时风险性也就更大。采用定向假设还是非定向假设要根据具体情况来定。如果有证据表明可以期望一个有方向的结果, 那么就应采用定向假设; 如果感觉可能存在某种关系但又没有证据表明结果的方向性, 或者正在进行的是一个探索性的研究, 则应该采用非定向假设。

从假设是否表述因果关系的角度可以划分为实验假设 (experimental hypothesis) 和非实验假设 (nonexperimental hypothesis) (Myers & Hansen, 2002)。实验假设是对变量间因果关系的猜测性陈述。例如, 上述的“饲养在贫乏环境中的动物与饲养在丰富环境中的动物相比, 它们在大脑发育和化学物质等方面将表现出明显的不同”就是一个实验假设, 该假设涉及的是经历与大脑变化之间的因果关系。又如“生理唤醒状态决定情绪状态”也是一个实验假设, 它猜测的是生理唤醒状态与情绪状态间的因果关系。非实验假设是对变量间相关关系的猜测性陈述。如, “数学成绩与语文成绩相关”就是一个非实验假设, 该假设没有指明数学成绩与语文成绩之间有任何因果关系。实验假设通常出现在实验研究

中，而非实验假设则通常出现在调查研究等相关研究中。因此，假设与研究类型、研究设计是相对应的，不同类型的假设决定了应该采用的研究类型和设计方式。

（三）假设检验的过程

已经讨论了假设的基本问题，那么该如何进行假设检验呢？针对不同的问题可能会有不同的方式，但其基本过程是一致的，如下图所示。

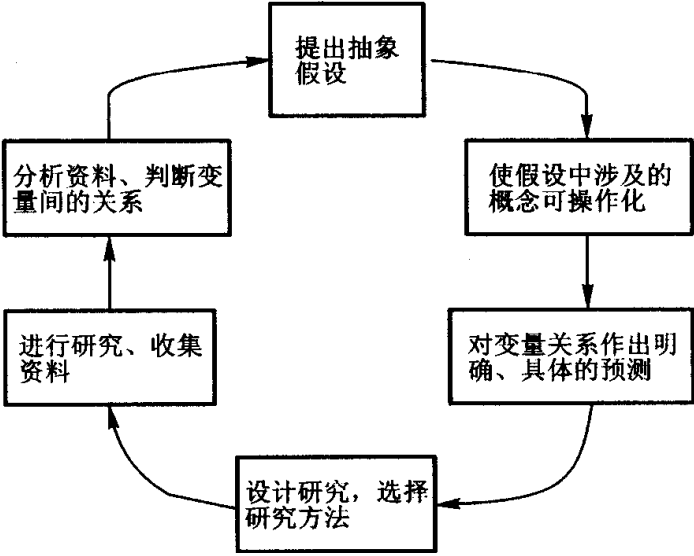


图 3.5 假设检验的过程

如图 3.5 所示，进行假设检验首先要从前人的研究、有关理论或日常观察等途径提出一个假设。这样的假设一般都是比较抽象，难以具体操作。因此应该通过操作性定义的方式使假设中涉及的概念可操纵化。经过这个步骤，假设的抽象语言描述就变成了实际可以操作的程序。这个时候还需要把变量间的明确关系确定下来，然后根据对变量关系的明确预测去设计研究。之后，根据设计去进行研究，并观测变量的变化，收集有关的资料。最后，使用统计手段等方法分析资料，然后判断收集到的资料能否验证假设。如果不能验证，则重新提出假设进行研究。

例如，上述的罗森茨韦格等（ Rosenzweig et al,1972 ）的研究就是这样。研究者首先有了一个“ 大脑会因经历的不同而改变 ”的假设。之后将“ 不同经历 ”和“ 大脑改变 ”这两个关键变量都做了操作化处理，得出了具体的预测即“ 饲养在贫乏环境中的动物与饲养在丰富环境中的动物相比，它们在大脑发育和化学物质等方面将表现出明显的不同 ”。之后，如上所述，研究者选择了恰当的研究方法、收集资料的方法，并进行了有效的分析，最后指出研究结果支持了研究的

假设。

通过这个过程也可以看出，心理学研究所检验的假设常常不是最初用抽象语言表达的理论假设，而是根据这种理论假设推导出的一个具体的可以操作的预测，因此，就算这个预测得到验证也不能说明你的假设得到“完全证明”。例如，上述研究就曾遭到批评，认为可能并不是丰富的生活环境使大脑发生了变化，而是因为触摸（换玩具时需要把它们移开，因此就会每天触摸老鼠两次）或贫乏环境中老鼠因被隔离而产生的压力导致的。虽然在研究者后来的研究和其他的有关研究表明这些批评不成立（Hock, 2004），但也表明要检验一个假设并不容易，研究中任何疏漏都可能导致失败。此外，同一个假设常常可以推导出多个甚至是无穷个具体的预测，一个预测得到验证，只能说明该假设得到了事实的支持，但还不能说该假设就绝对正确了。同时，即使某个具体的预测被推翻了，也不要急于完全否定该假设，也许该假设只是不太完备需要修改，但并不一定就完全错了。

第三节 解释

解释(explanation)是科学研究的重要目标之一。心理学家解释心理现象的最典型的做法就是指出造成一些现象的原因。例如心理学家在解释为什么“有些人聪明，有些人不那么聪明”的现象时就是通过指出造成人们智商差异的原因如遗传、教育等来进行解释的。但是人们进行推理和解释时常常会犯错误（见补充讨论 3-2）。

补充讨论 3-2 推理与解释中的错误

无论是在科学研究还是日常生活中，人们对现象进行解释时往往会犯以下一些错误。

1. 以偏概全。就是把少量类似的情况当成普遍情况，特别是当我们去寻找普遍规律时。如，当你连续碰到三个重庆人都很豪爽时，你可能就会得出重庆人豪爽的结论。科学研究中使用足够大的调查样本的方法来防止过度概化的错误。此外，进行重复验证(replication)也可以防止这种错误。重复验证就是进行重复观察或重复研究，看每次观察或研究得出的结果是否一致，如果是其他研究人员独立进行的重复观察或重复研究就更好。

2. 赌徒谬误。就是认为一个事件发生的可能性依赖于先前的事件。如，你抛掷硬币时，如果连续 8 次都是正面向上，那么你就可能相信第 9 次是背面向上的可能性要比正面向上的可能性大，因为你会认为“总该有一次是背面向上”。

但是实际上每次抛掷硬币时，背面向上的可能性都是 50%。除了赌徒外，日常生活中人们也会有这样的推理方式，例如在预定的春游日之前连续出现好天气的话，人们就会担心春游时很有可能会下雨。

3. 虚假的二元对立。二元对立就是用非此即彼的思维方式来认识事物，而排除其他可能的存在。其实，很多事情既不是处于完全对立的两极也不是只有两种可能。如，对一个人的判断往往习惯上把他或她视为好人或坏人，实际上每一个人都是相当复杂的，难以用简单的好或坏来概括。

4. 滥用例外。是指当发现观察到的事实与我们的解释相违背时，人们常用“例外”来解释，认为此种现象是特殊情况，甚至将之忽略或神秘化。如，许多迷信的人在求神拜佛并未灵验时就会用例外来解释，认为是某些神秘的原因导致不灵验，并继续迷信。但是根据逻辑原则，只要发现一个反例就能推翻整个观点。

5. 偏见。主要指人们对现象解释的先入之见。这种思维方式不是依事物本身特点而是以想当然的先入之见来进行推理。例如人们往往认为长得凶恶的人可能是罪犯，而长得漂亮的人则不大可能是罪犯。又如，有的女性因为被某个男人骗了以后就认为所有男人都不是好人。

一、解释的逻辑

当我问你，你为什么读大学，你回答说是因为你想读大学，那么我可以追问你为什么想读大学，你可能会说因为你想得到一个好工作，那么我可以继续追问你为什么想要得到一个好工作……。我打破砂锅问到底的这种思维方式反映了解释现象时所使用的一种逻辑方式，即决定论逻辑。这种逻辑不相信个体能够完全自己选择和控制自己的行为，认为原因背后还有原因。采用决定论逻辑，就默认个体的心理与行为主要是由其不能控制的因素决定的，就需要不断的探究现象背后复杂的因果链条。

心理学的研究大多是根据决定论逻辑来提出假设和解释结果的，如生物学范型就认为心理现象是由脑、神经、遗传、内分泌等生物学因素决定的；心理动力学范型认为人们的大多数行为产生于本能和生理驱力；行为论范型则认为所有行为都源于学习，刺激—反应是人类行为的主要模式。

使用决定论逻辑可以帮助我们提出研究假设、进行研究设计并找出导致心理现象的种种原因，但是这种逻辑可能会忽视人的主观能动性，从而会导致对研究结果的错误解释。最明显的例证是实验者效应和安慰剂效应。实验者效应（experimenter effect）是指实验者有意无意地将实验目的及对结果的期望等传递给了被试，从而造成被试反应朝着有利于证实实验假设方向的改变。安慰剂效

应(placebo effects)指仅仅通过给被试安慰剂就能产生正面、有效的结果的现象。实验者效应和安慰剂效应表明,被试不是被动接受操纵,而是有主观能动性的,他们的自由意志就可以成为研究效应出现的原因。不注意这个问题就会导致对研究结果的错误解释。

绝大多数的心理学研究中都强调寻找心理现象背后的普遍规律。例如,我们寻找记忆的规律、寻找注意的规律、寻找童年期心理发展的规律等等。这种力求找出能够对某一类现象进行解释的一个或少数几个因素的做法称为通则式(nomothetic)的解释逻辑。这种解释逻辑不针对个案,而是寻求可以解释一类现象的最主要原因。利用这种逻辑有助于我们迅速且经济地把握一类现象。例如,发现短时记忆的容量是 7 ± 2 个组块这一规律,我们就可以以此来解释涉及短时记忆容量的心理现象。通则式解释通常与概率有关,它指出某些因素的存在会使某些现象出现的概率很高或很低。例如,“高智商的学生的数学成绩显著高于低智商的学生”,就是说智商这个因素会导致数学成绩好这种现象出现的概率非常高。

但是,根据通则式逻辑来进行解释,往往会忽视个体差异,使解释变得不完备。例如,虽然从概率上说孩子聪明,取得好成绩的可能性就大,但聪明并不能为学习成绩好提供完备的解释,因为我们总能观察到笨鸟先飞和聪明反被聪明误的情况。此外,受通则式逻辑的影响,一些研究者喜欢用平均数来处理研究结果,这就容易导致错误的结论。例如,一个班的语文成绩的平均分是 80 分,但是除了极个别学生的成绩是 80 分左右以外,绝大多数学生的成绩与 80 分相差很远,那么用 80 分来说明学生的成绩就有问题。

总之,决定论逻辑和通则式逻辑是进行心理学研究的两种主要思维逻辑,但是这两种思维逻辑并不完美。在解释研究结果时必须注意解释逻辑中潜在的问题,才能防止草率和做出错误的结论。

二、关系的确定

对现象进行解释的过程,就是确定变量间关系,特别是确定因果关系的过程。

变量之间有两类基本的关系。一是相关关系(correlation),即变量在发展变化的方向与大小方面存在关联。二是因果关系(causality)即一个变量是另一个变量的原因,另一个变量是结果。

相关关系包括三种情况,一是变量变动的方向相同,即变量 A 变大时变量 B 同时变大,变量 A 变小时变量 B 同时变小。这种情况被称为正相关。二是变量变动的方向相反,即变量 A 变大时变量 B 同时变小,变量 A 变小时变量 B 同时变大。这种情况被称为负相关。三是,变量之间没有关系,即零相关。相关关系

在说明变量间关系时给出的信息非常有限，它仅能说明变量之间有关系，但不能说明变量间关系的特点，不知道究竟哪个是因，哪个是果。如，工作满意度与工作绩效之间有相关关系，但是难以确定究竟是工作满意度导致工作绩效变化，还是工作绩效导致工作满意度变化。同时，表面的相关关系还可能由第三变量所决定的。如，夏天冰棍的销售量与淹死人数之间的相关就是由于天气因素造成的。此外，相关关系还可以是人为造成的。如，两个测验分数之间的高相关可能只是由于它们包含了一些相同的题目造成的。

确定变量之间有关系相对容易，但远不能让人满足，科学研究的重要任务是要确定变量之间有没有因果关系，因为只有揭示因果关系才能认识事物的规律和本质。

变量间的因果关系不会自己显现，需要研究者根据收集到的资料进行逻辑推断。显然，这是一个主观的过程，因此一不小心就会犯逻辑错误，做出错误的判断，特别是当研究者非常希望揭示因果关系的时候。

如何根据收集到的研究资料做出正确的因果判断呢？这涉及到判断因果关系的标准问题，不少学者探讨过这个问题，其中拉扎斯菲尔德（Lazarsfeld, 1959）提出的判断因果关系的三条标准为许多研究者所遵从，可以成为我们进行因果判断的依据。这三条标准是：①变量发生在时间序列上是前因后果。这是做因果推断最基本的条件。因为显然作为原因的变量应该比作为结果的变量先出现，而不是相反。如中枪导致死亡，必然是先中枪后死亡。当然达到这一标准还不够，它只是判明因果关系的必要条件而非充分条件。比如，我们早晨起床后先刷牙后吃饭，不能据此就认为刷牙是吃饭的原因。两个变量之间有实证的相关。即要能够实际观测到两个变量之间有稳定的共同变化关系。也就是说，一个变量的每一个变化都可以引起另一变量相应的、可以预见的变化。但是，如上所述，变量之间有相关也并非一定是因果关系，因此，存在相关仍是判断因果关系的必要条件而非充分条件。两个变量间的实证相关不因为其他变量的存在而存在。即可以排除其他解释变量间实证相关的因素，也就是可以排除全部额外变量的干扰。如果变量之间的明确先后关系和实证相关不会受到其他任何因素的干扰时，我们就可以认为这两个变量是有因果关系的。

可见，进行有百分之百把握的因果判断的条件是比较苛刻的，一不小心就会做出错误的判断。特别是心理学研究中存在大量的非实验研究，根据这样的研究结果是不能得出因果判断的。即使是实验研究也可能存在控制变量干扰实验结果等诸多问题，因此要慎下因果关系的结论。虽然科学研究总是希望获得因果关系，但是也不是说相关关系的研究就没有价值，其实由于心理现象远比自然现象复杂，心理学中的大量研究得出的都是相关关系的结论，正是这些研究为更深入的研究打下了基础，同时也为我们预测变量变化提供了指导。

补充讨论 3-3 充分原因与必要原因

科学研究中最完美的因果关系是一个变量成为另外一个变量的充分必要条件,即有 A 必有 B 有 B 必有 A。如,“三内角相等的三角形是等边三角形”就是这样。但是在实际的心理学研究中这样的因果关系是很难找到的。

对于事物的原因可以分为必要原因和充分原因两类(巴比,2000)。必要原因是指对产生效应必不可少的条件。如,要考上研究生就必须通过英语考试,但仅通过英语考试还不能考上研究生;要去做心理咨询就必须掌握心理学的知识,但是仅有心理学知识而不会应用也做不好心理咨询。充分原因是指某一条件出现就一定会产生预期的效应。如,剧烈运动导致心跳加快,缺考导致没有成绩。但是充分原因并不一定是唯一原因,如心跳加快还可能是害怕导致的,没有成绩也可能是考试作弊导致的。

心理学研究中必要原因和充分原因都可以当作判断因果关系是否存在的依据,但是这样的判断不能绝对化。许多心理现象都是多因一果的,因此不能因为找到了某个原因就认为寻找因果关系的工作已经完成了。你必须首先要判断你找到的是充分原因还是必要原因,然后再考虑是否还存在其他原因,特别是是否还存在其他更重要的原因。

►本章提要

1. 范型是指看待问题的基本角度或框架。它为研究者提供观察事物的基本方式和信念,决定或影响着研究者所选择的研究内容、方法、策略和所提出的理论。在心理学中主要的研究范型有生物学范型、心理动力学范型、行为论范型、人本论范型和认知论范型。

2. 归纳式理论是进行质的研究的研究者通常采用的理论建构方式。演绎式理论是运用演绎的逻辑,采用假设检验方式获得的理论。

3. 心理学研究中的概念大致可分为两类:一类是具有操作性定义(operational definitions)的概念,另一类是心理学家“发明”出来的尚无操作性定义的概念。操作性定义是一种规定,是根据一组产生被定义事物的操作所下的定义。那些具有操作性定义,可以用量化的方法加以测量并赋予两个或两个以上值的概念就是变量(variable)。

4. 刺激变量简称 S 变量,是指已知的对有机体的反应发生影响的刺激条件,包括研究者可以操纵和控制的环境特征。机体变量也称为被试变量、属性变量,一般是指被试本身对行为反应有影响的特征。反应变量指由刺激引起的有

机体行为的任何变化。有机体对刺激的反应是多种多样的，人对刺激的行为反应可以分为言语行为反应和动机行为反应。

5. 自变量指在实验中被实验者加以操纵的变量。因变量是在实验中被试对刺激做出的反应，也就是实验者观测和记录下的随着自变量变化而变化的被试行为反应值。控制变量指在实验中实验者需要通过保持恒定等方法使其中立化的潜在变量。

6. 假设是指科学研究中任何一种对某些问题做出尝试性解释的陈述。备择假设亦可称为研究假设，是陈述变量间所期望的关系，通常是研究者期望证实的假设；虚无假设亦称无差假设、零假设、原假设，是陈述变量间没有关系或差异，是根据样本统计量决定予以拒绝或接受的假设，常以 H_0 表示。定向假设是对变量关系的方向有所示意的假设；非定向假设是对变量关系的方向没有示意的假设。实验假设是对变量间因果关系的猜测性陈述；非实验假设是对变量间相关关系的猜测性陈述。

7. 决定论逻辑默认个体的心理与行为主要是由其不能控制的因素决定的，需要不断的探究现象背后复杂的因果链条。通则式的解释逻辑力求找出能够对某一类现象进行解释的一个或少数几个因素。

8. 变量之间有两类基本的关系。一是相关关系，即变量在发展变化的方向与大小方面存在关联。二是因果关系，即一个变量是另一个变量的原因，另一个变量是结果。

9. 判断因果关系的三条标准是：①变量发生在时间序列上是前因后果。

两个变量之间有实证的相关。即要能够实际观测到两个变量之间有稳定的共同变化关系。③两个变量间的实证相关不因为其他变量的存在而存在。

►本章关键术语

范型(paradigm)	生物学范型(biological paradigm)
心理动力学范型(psychodynamic paradigm)	投射测验(projective test)
罗夏墨迹测验(Rorschach inkblot test)	行为论范型(behavioral paradigm)
认知论范型(cognitive paradigm)	部分报告法(partial report)
人本论范型(humanistic paradigm)	自我价值感(self - worth)
理论(theory)	概念(concept)
操作性定义(operational definitions)	变量(variable)
刺激变量(stimulus variable)	机体变量(organism variable)
反应变量(response variable)	自变量(independent variable)
因变量(dependent variable)	控制变量(control variable)
假设(hypothesis)	备择假设(alternative hypothesis)

虚无假设 (null hypothesis)

定向假设(directional hypothesis)

非定向假设 (nondirectional hypothesis) 非实验假设 (nonexperimental hypothesis)

解释 (explanation)

安慰剂效应(placebo effects)

实验者效应 (experimenter effect)

通则式(nomothetic)

相关关系 (correlation)

因果关系 (causality)

►复习与练习

1. 什么是范型？在心理学研究中主要有哪些范型？
2. 对比演绎式理论与归纳式理论。
3. 对比概念与变量。
4. 心理学研究中的主要变量是哪些？
5. 什么是假设？假设检验的基本过程是怎样的？
6. 心理学研究的解释逻辑主要是什么？应用这些逻辑进行解释时应注意哪些问题？
7. 如何判断因果关系？

►推荐参考读物

Myers A. & Hansen C. H. Experimental psychology (5th ed.). Pacific Grove, CA: Wadsworth Group, 2002

雷迪·D·高尔, 沃尔特·R·博格, 乔伊斯·P·高尔. 教育研究方法导论 (第六版). 许庆豫等译. 南京: 江苏教育出版社, 2002

巴比. 社会研究方法. 邱泽奇译. 北京: 华夏出版社, 2000

►在线资源

有关心理学研究范型的进一步讨论可以参见: [http://www. pep. com. cn/xlxjz/index. htm](http://www.pep.com.cn/xlxjz/index.htm).

[http://202. 120. 92. 199/dadapaper/xinli0401guo. htm](http://202.120.92.199/dadapaper/xinli0401guo.htm) 和 [http://info. psych-cn. com/](http://info.psych-cn.com/)

第四章 测量

测量是心理学研究中不可或缺的部分。我们可以在心理学研究的开初提出完美无缺的假设并设计出研究方案，但是如果不能对概念或变量进行有效的测量，那么这样的科学研究肯定是要失败的。通常情况下，测量是将需要测量的概念或变量用数字加以表示的过程。对智力的测量是大家所熟知的：我们用某个智力量表之类的测验以一个数字来表示某个人的智力水平。在本章中，我们将讨论以下问题：

1. 测量及其层次
2. 测量的效度和信度
3. 测量的误差

第一节 测量及其层次

一、何谓测量

心理学研究中涉及到许多概念或变量，有些容易被测量而有些则难以被测量。如果某种性质能被直接观察到，并且大家都熟悉又无争议，这种概念或变量就比较容易定量测量，例如一个人的体重和年龄就是如此。而另一些概念，特别是像人格、态度等心理变量或概念，要加以测量就困难得多。虽然已经发展出一些量表来测量人格、态度等心理变量，但像自信、后悔、羞怯这类概念不能被直接观察到（虽然它们的影响是我们能直接观察到的），如果要研究这些问题我们就得想办法来对它们进行测量。

测量（**measurement**）是一个为确定某个特定分析单位的特定属性值的过程。在第一章里我们曾讨论过分析单位的概念。心理学中最常用的分析单位可能是分子，但也有不少是用其他单位，如个人、群体、组织等。我们在测量概念或变量时，首先要确定分析单位，然后要详细说明所要测量的单位的特征。例如，如果我们确定的分析单位是一个分子（如自信），那就要详细地说明所要测量这个分子的种种特征，其中有些是可见的有些是不可见的。可见的特征如面部表情、遇到特定情境时的行为反应。不可见的或不能观察到的特征如内心的认知、情绪体验等。分析单位的这种潜在的可测量的特征，有各种不同的提法，如维度、属性、因素等。

测量并不仅限于作数值上的或定量的阐明，也可以作定性的阐明。我们对一些概念的定性不是用数字而是用定性的标示或用名称来表示其相应范畴的。任何用数字测量的概念，一般我们称之为定量概念或定量变量。

有些概念是定性变量而不是定量变量。在心理学研究中，我们可以想出许多定性变量，如家庭类型（核心家庭、非核心家庭）、父母教养方式（民主型、独裁型、溺爱型、忽视型）或团体类别（正式团体、非正式团体）。在观察研究中，我们经常使用定性变量。心理学研究中所观察到的许多资料是属于定性类别的，给每个类别的资料一个名称，使之区别于其他类别。例如：克雷奇默尔（Kretschmer, 1921）依据体型把人分为瘦长型、矮胖型、强壮型、发育异常型；威特金（Witkin, 1949）把人们的认知风格分为场依存型和场独立型；奥尔波特等人（Allport, Vernon & Lindzey, 1960）依据价值观把人们分为理论型、经济型、政治型、社会型、审美型、宗教型；马斯洛（Maslow, 1954）把人们的需要分为生理需要、安全需要、归属和爱的需要、自尊需要、自我实现需要等等。

定性概念的类别也可以用数字而不是用名称来标示，但这些数字并不具有数字系统的属性——它们是不能用加减乘除进行运算的。例如，一个宾馆的房间号码是不能使用算术运算的。我们不能将 231 号房间和 101 号房间像 231 斤和 101 斤一样的方法进行相加。对定性变量唯一可作数字运算的是计算每一类别的出现率和百分率（如，男性或女性被试的百分率）。

二、测量的层次

美国心理学家史蒂文斯（Stevens, 1961）提出的测量层次的分类，包括定名测量、定序测量、定距测量和比率测量，如今已被学者们所广泛采用。

（一）定名测量

定名测量也称定类测量（nominal measurement），是把事物的属性归于不同类别的测量。前述所有定性的测量都是定名测量，不管类别是由名称（男性、女性）还是由数字（231 号房间、101 号房间）划分的。定名测量本质上是一个分类系统。定名测量的变量（如性别）至少有两个类别（否则就不是变量），而且这两个类别在性质上是截然不同的，是包罗无遗且相互排斥的。所谓包罗无遗，是指欲测量的每样东西都必须归属于其中一个适当的类别。所谓相互排斥，是指这每样东西都仅仅符合于一个类别。因此，每样东西都必须有一个类别（包罗无遗），但只明显符合一个类别（相互排斥）。现实生活中，性别是定名变量，它有两个类别——女性和男性。这一变量的类别是包罗无遗且相互排斥的，因为每个人都符合于一个类别，且仅属于一个类别。因此，最好的定名变量或类别变量是那些类别明显不同又没有模棱两可难以分类的变量。

(二) 定序测量

定序测量(ordinal measurement)是对事物的等级和顺序的测量，即对一系列东西进行排序。除前述的定名测量外，还有三种定量的测量层次：定序测量，定距测量和比率测量。这些测量的后一个层次都是建立在前一个层次的基础上的，即后一个测量层次都具有前一个层次的属性再加上它自己的附加属性。定序测量像定名测量一样，由相互排斥且包罗无遗的类别组成。但是，这些类别不在同一层次上，也不具有等等的值；它们具有不同的值，因而可以依次排列等级，例如最年老，次年老，最年轻等。

在定序测量中，常常会遇到平分的情况，或者说具有相同等级的类别。例如 如果我们将参加甲 A 男子排球赛的所有排球队，单纯按胜败的场数记分，从得分多的到得分少的按等级排列。假设 10 个队的输赢场数记录如下：

队 名	获胜场数	失败场数
河南	4	6
北京	6	4
辽宁	8	2
上海	10	0
广东	0	10
江苏	5	5
八一	8	2
湖北	4	6
浙江	9	1
山东	4	6

根据上述记录，我们可以把这些队按等级排列为以下次序：上海队，浙江队，辽宁队，八一队，北京队，江苏队，湖北队，山东队，河南队，广东队。在进行排序时，我们遇到了平分问题：辽宁队和八一队是平分；河南队、山东队和湖北队也是平分。对于这些平分，我们该怎么办呢？通常的做法是将同分的队同等排列。

队 名	次序排列
上海	1
浙江	2
辽宁	3.5
八一	3.5
北京	5
江苏	6
湖北	8
山东	8
河南	8
广东	10

辽宁队和八一队获胜场数相等，我们不能把它们中的一个队排为 3，另一队排为 4。如果这样做，就掩盖了同分的真相，使人误以为一个队的获胜场数高于另一个队。但也不宜将这两个队都排为 3 或都排为 4，因为定序测量系统每一个等级只有一位。等级从 1 到 10 总数是 55。如果排两个 3 或两个 4，各个等级相加的总数便成了 54 或 56。为了保持测量系统的连贯性，应采用以下等级排序法：同分（3 和 4）相加除以同分队数，即 $(3 + 4)/2 = 3.5$ 。同样，如果是三个队（河南队、山东队、湖北队）同分（它们的等级是 7、8、9），也是把它们的排列等级相加除以同分的队数： $(7 + 8 + 9)/3 = 8$ ，这三个同分队的每一队都排为等级 8。由此可见，同分的队为偶数时，排列等级是有小数点的；而同分队为奇数时，则等级排列是整数的。

在定序测量中，我们可以说等级 A 的人在所测量的属性值上比等级 A - 1 的人高，等级 A - 1 的人在属性值上又比等级 A - 2 的人高（在定名测量中我们却不能这样说）。不仅如此，我们还知道，如果 $A > A - 1$ ，而 $A - 1 > A - 2$ ，则 $A > A - 2$ 。但我们仍然不知道 A 的值比 A - 1 大多少，因而我们也不能确定 A 与 A - 1 之差与 A - 1 与 A - 2 之差是否相同。能得到这一信息的测量为定距测量。

（三）定距测量

定距测量（interval measurement）是对事物的数量差别和间隔距离的测量。以年龄为例，不仅只是把人们分为不同的年龄类别或把人们从最年老到最年轻进行排序，而且还能够按每人所活的年数来加以测量。通过定距测量，我们可以测定从一个等级到下一个等级在年龄上有多少单位之差。这一信息，无论是定名测量或定序测量都是无法得到的。我们可以看到，在最年老者 and 次年老者之间（如 99 岁和 98 岁）仅有一个单位之差，但在次年老者和第三年老者（98 岁与 66 岁）之间则有 32 个单位之差。所谓一个单位之差，不论它发生在量表的何处，它都是相同的。即是说，一岁之差就是一岁之差，不管它发生在 78 岁和 77 岁之间还是在 13 岁和 12 岁之间。

定序测量和定距测量之间的区别就在于：在定距测量中，我们不仅可以说哪个测量等级较高，而且还可以说这个等级高多少单位。无论它发生在量表的何处，单位的值总是相同的，因此，定距测量的资料便具有真正数量的意义，就可以对它们进行加、减运算了。

（四）比率测量

比率测量（ratio measurement）是对事物的比例或比率关系的测量。比率测量除可以进行加、减运算外，还可以进行乘、除运算。要进行乘、除运算就要求有一个绝对的、固定的零。前面我们曾用“年龄”这一变量来说明定距量表的概念。但在方法上，“年龄”已超过定距量表的要求，因为年龄有一个绝对的零（零

是非任意的，但负的年龄是没有的）。是否有一个绝对的（而非任意的）零是比率测量与定距测量之间的唯一区别。体重也有一个非任意的零却不存在负值，因而是一个比率变量。一个人不会比零（出生）更年轻，其体重也不会少于零。如果有这样一个固定的、非任意的零为基点，而单位又是始终不变的，那么就可以进行乘或除了。例如，30岁是10岁的三倍，而15岁是30岁的1/2。一个测量是否是比率测量关键在于是否有一个绝对的零，即使这样的量表并无负值（像对一个家庭的人数，一个人的年龄、体重、身高的测量都是比率测量）。

表4-1列出了四种测量层次的数据的特性及适合的运算方法。

表 4-1 四种测量尺度及观测数据特性

	定 名	定 序	定 距	比 率
基本特征	反映性质、类别	等级、位次、大小； 反映相对数量	有相等单位的 连续数量	有绝对零点和相等单 位的连续数量
适用运算	归类、计数	比较、排名	+ / -	+ / - / × / ÷ / 开方等
统计方法	众数、次数、百 分比、品质相 关、 χ^2 检验	中位数、等级相 关、秩次、非参数 检验法	\bar{x} 、 σ^2 、 r 、 t 、 F	\bar{x} 、 σ^2 、 r 、 t 、 F 、几何均 值、差异系数
举例	性别、种族、职 业、教养方式	喜好程度、排名； 优劣评价	智力测验和人 格测验结果	长度、重量、时间；正 确回答次数

（五）离散测量与连续测量

对测量的另一种分类是将测量分为离散和连续测量。离散测量（discrete measurement）所得到的数据是不包含小数的，而连续测量（continuous measurement）所得到的数据是包含小数的。实验组或控制组的人数、反应的次数是离散变量，而一个人的年龄是连续变量（如21.5岁）。任何有小数的测量都是连续测量。一次计数的结果一般是离散测量。离散的意思是从一个整数到下一个整数，而连续的意思是在两个整数之间可能有很多的值。离散测量意为被打断或打破的测量，而连续测量则意为连续的测量不被打断。

前已述及，通过观察法所搜集到的大多数数据都会是定名的或定量的。计数资料一般是离散的。用观察法搜集到的定性资料几乎总是分类的，因而是离散的。定量的资料可以是连续的，也可以是离散的。虽然定名测量似乎可以被视为在测量某个潜在的连续体，但这种测量通常是离散的。定距测量和比率测量所获得的资料既可以是离散的，也可以是连续的。

第二节 测量的效度

一、何谓效度

对于任何测量手段，我们必须提出一些基本的问题：它测量的是什么？它所测得的数据与人们所认同的特征是否有关？记分之差在何种程度上反映了意欲测量特征的真正差异？这就涉及到测量的效度问题。

测量的效度 (**validity**) 是指测量得到的记分是否反映了欲测的特征及其程度。这个定义有两层含义：一是指测量手段实际上测量了所研究的概念，而不是其他的概念；二是指该概念被准确测量的程度。显然，如果测量的不是要测量的概念，那么就谈不上对该概念进行准确测量的程度了。

例如，我们想测量智力，小陈的智力水平假定是 100 个单位。我们编制出的智力量表有一定的误差，用它测小陈的智力是 75 个单位，而实际上他是 100。这个量表测量的是应该测量的概念（智力），尽管有误差，如果我们能加以改善，消除它的误差，这个量表还能成为有效的智力量表。但是我们编制出的不是智力量表而是测其他概念（如自信心）的量表，用它来测量小陈的智力，那就是无效的，即使知道碰巧得到的分数是 100，也仍然是无效的。因此，一个测量工具是否测量了欲测的概念是非常重要的。

二、测量效度

翻阅研究文献你就可以发现有许多效度鉴定程序。下面介绍最主要的三种效度鉴定程序。

（一）内容效度

内容效度 (**content validity**) 是指测量在多大程度上包含了欲测的内容范围。例如，卡米尼斯和泽勒 (**Camines&Zeller, 1979**) 指出，数学能力测试不能只包含加法，还应该包括减法、乘法、除法等等。又如，假设研究者要测量一般的“偏见”，那么测量的内容能否反映种族偏见、宗教偏见、性别偏见以及许多其他的偏见等等。由于内容效度主要靠推理判断而非定量的方法来确定，所以也叫逻辑效度。

在内容效度的证据中，一个关键的要素就是题目取样的适当性。因为研究者使用的测量方法不可能包含所要测量的心理行为领域的全部可能的材料或情境，而只能选择一个有代表性的样本，通过观察被试对少数项目的反应来推测他在总体中的表现。如果所选择的测题过难或过易，或偏重于某部分内容，就会使测验对总体估计的准确程度下降。例如，教师要求学生在学期结束时掌握 1 000 个英语单词，为了检验学生的学习情况，教师编制了一个包含 100 个单词的词汇

测验。显然，只有当这 100 个单词能代表所要求掌握的 1 000 个单词时，测验才会有比较高的内容效度。如果在该测验中，只有简单的题目，或者非常难的题目，该测验就会失去代表性，因而也不能提供有效推论的信息。要想获得较高的内容效度，就应该明确定义想要测量的内容范围，而测验项目应是所界定的内容范围的一个有代表性的样本。

估计内容效度的方法有：

第一，专家逻辑分析。由于内容效度的确定没有可用的数量化标准，通常是由专家根据测验题目和双向细目表（two - way specification table）做系统的比较判断，分析测验题目是否恰当地代表了所规定的内容，以此作为效度的证据。

逻辑分析的过程是先确定测验内容的总体范围，描述有关知识和技能及所用材料的来源。然后编制双向细目表，确定内容和行为各自所占的比例以及各题的测验目标。双向细目表是按两种分类标准编制成的表，可以从纵横两向度表示两个变量间的关系，也是设计和编制测验的一种结构或框架。以编制学科成就测验为例，双向指教学目标和教学内容，其中教学目标分为知识、理解、应用、分析、综合、评价六个层次，教学内容因学科和教材而异。最后再评定测题与内容的符合程度。如果题目具有较好的代表性，则说明测验有较高的内容效度。采用专家逻辑分析方法的问题是不同的专家对同一测验的内容效度的判断可能不一致。为了提高逻辑分析的客观性，应该尽量对测验的编制过程和测验目标进行详细说明，并编制评定量表（或专家调查表），以便专家对测验做出客观的评定。

第二，统计分析。克伦巴赫（Lee J Cronbach, 1916—2001）提出内容效度可以进行数量估计。具体方法是从同一教学内容总体中抽取两套独立平行的测验，分别对同一组被试进行测验，两种测验的相关系数可用来估计内容效度。若相关系数大，则内容效度高；若相关系数小，则两个测验中至少有一个测验的内容效度低。

第三，重测法。这是类似重测信度的一种评定方法。具体程序是研究者对一组被试进行前测。然后，让被试学习新的知识或参与某种训练课程，课程结束后将测验再实施一次。如果后测的成绩显著地优于前测的成绩，说明测验到的内容是新近学的知识和受到的训练，表明了该测验有较高的内容效度。

内容效度的使用范围广泛，在成就测验中尤为重要，因为成就测验主要是考察被试知识和技能的掌握程度，内容效度也同样适合某些用于选拔和分类的职业测验。

（二）准则效度

准则效度（criterion validity）也称效标关联效度或实证效度，是以测验分数与作为外在标准的效标行为之间的关联程度来表示的一种效度。例如，通过高

考选拔了一部分学生进入大学学习，若有证据表明进入大学的高分学生在大学的学业成绩上确实优于低分学生，则可以认为高考有较高的准则效度。在准则效度中 效标(criterion)是衡量测验结果有效性的参照标准，它是体现测验目的、独立于测验之外的一种行为变量。常用的效标有：学业成就、行为的等级评定、临床诊断、实际工作表现、先前有效的测验、团体对照等。

准则效度有两种形式的证据——同时效度(concurrent validity)和预测效度(predictive validity)。当效标数据与测量数据在同一时间获得并进行比较时，获得的效度证据就是同时效度。例如，研究者用自尊问卷对大学一年级学生进行调查，并将他们的得分与几乎在同一时间获得的教师对学生的自尊评价进行比较，就构成了同时效度的证据。同时效度反映测验能在什么程度上取代效标。如果效标数据后于测量数据获得时，获得的效度证据就是预测效度。例如，研究者在开学时对一组大学生实施学科能力倾向测验。到期末时，再比较他们在测验上的得分与某学科的期末考试成绩，就构成了预测效度的证据。预测效度反映测验能在什么程度上预测效标。

估计准则效度的主要方法有：

第一，相关法。相关法就是用相关系数来描述同一组被试在某个测量工具上的得分与他们在效标测量上的得分之间的关系，这种数量指标也被称为效度系数(validity coefficient)。相关系数的获得有两种情况：如果测验分数与效标测量分数都是连续变量，用积差相关公式求相关系数；如果测验分数是连续变量，而效标测量分数是二分变量时，用二列相关公式求相关系数。

第二，区分法。具体程序是让被试接受一个测验，然后让他们学习一段时间，再根据学习成绩（效标测量）的好坏将其分成两组。回过头来分析这两组被试原来接受的测验的分数差异。区分法的关键是用先进行的测验得分来区分由效标测量所定义的团体。例如，某学校通过入学考试录取了一批学生，经过一段时间的学习后，根据学习成绩的高低将其分为合格和不合格两组被试，然后通过检查他们的入学测验成绩，运用统计方法来检验两组被试在入学测验上的平均分数是否有显著差异。若差异不显著，则说明测验肯定是无效的；若有显著差异，也不保证测验一定有效。

第三，命中率。如果研究者用测验的分数来进行决策，那么测验是否有效的指标就是正确决策的比例，即命中率。假设被试在测验成绩上有两种可能：通过测验（或被接受）和没有通过测验（或被拒绝），在效标测量上也有两种可能：成功和失败。在实际生活中就有可能出现四种情况：通过测验并在效标测量上成功；通过测验但在效标测量上失败；没有通过测验但在效标测量上成功；没有通过测验在效标测量上也失败。如果将正确的预测称为命中，不正确的预测称为失误，测验成绩和效标测量就会出现如下关系（表 4 - 2）。

表 4-2 测验和效标的关系

测 验	效标预测失败(-)	效标预测成功(+)
通过(+)	(A) 失误	(B) 命中
不通过(-)	(C) 命中	(D) 失误

判断决策正确性就有两个指标：

1. 总命中率(P_{CT})。是指正确决定数目(命中)对总决定数目(N)的比例：

$$P_{CT} = \frac{\text{命中}}{\text{命中} + \text{失误}} = \frac{B + C}{A + B + C + D} = \frac{\text{命中}}{N}$$

P_{CT} 值越大说明测验越有效。

2. 正命中率(P_{CP})。是所有被选择的人成功的比例：

$$P_{CP} = \frac{\text{成功人数}}{\text{选择人数}} = \frac{A}{A + B}$$

P_{CP} 值越大说明测验越有效。

(三) 结构效度

结构效度(construct validity)是测验对某种理论构想或特质所能体现测量的程度。通常,心理学研究中的智力、焦虑、自我和动机等都是假设性质的概念或特质,无法直接度量,这些构想只能在理论上通过具有操作性定义的测验来加以测量。例如,比纳—西蒙智力量表的制订,比纳(Alfred Binet,1857—1911)首先详细研究了智力的结构,他认为智力行为是一种连锁性的过程,包括判断、推理、解决问题等。他根据这一理论编制的智力测验,确实测量出被试的判断、推理和解决问题的能力,可以认为他所编制的测验具有较好的结构效度。

确定测验的结构效度,有三个步骤:(1)提出理论假设,建立理论框架,以解释被试在测验中的表现;(2)依据理论框架,推导出各种与测验成绩有关的假设;(3)以逻辑和实证的方法来验证假设。

估计结构效度的方法有：

第一,测验内方法。就是通过分析测验的内部构造来获取效度证据。例如,通过分析测验的内容、被试对题目所作的反应、测验题目的同质性以及分测验之间的关系来判断测验的结构效度。

第二,测验间方法。就是通过计算测验与标准化测验的相关来获取效度证据。如果两个测验测量相同的心理特质,那么对同质的被试进行测验时,二者的得分应该有较高的相关。因此,如果某一个新测验与同类的大家公认有效的已有测验之间,在测验结果上有很高的相关,说明这两个测验测的是相同特质,即新测验也有较高的结构效度。例如,在评价新编的智力测验时,通常要与斯坦

福一比纳智力测验或韦氏智力测验进行比较。

第三，因素分析。其原理是：通过对一组测验项目进行因素分析，找出影响测验的共同因素。每个测验在共同因素上的负荷量（即测验与各因素的相关）就是测验的因素效度，测验分数总变异中来自有关因素的比例即是该测验结构效度的指标。关于因素分析的具体内容，我们将在第十三章中详细讨论。

第四，多特质—多方法矩阵。坎贝尔和菲斯克（Campbell&Fiske,1959）提出了一种分析测验结构效度的方法——多特质多方法矩阵（multitrait - multimethod matrix），此程序要求采用两种以上的方法测量两种以上的特质。特质与方法间有多种搭配，其两两间相关系数组成多特质 - 多方法矩阵。在该矩阵上，如果以不同方法测相同特质所得分数之间的相关系数 $r_{tt'}$ ，高于以相同方法测量不同特质所得分数的相关系数 $r_{tt''}$ ，且高于以不同方法测量不同特质所得分数的相关系数 $r_{t't''}$ ，则该测验有较高效度，其中 $r_{tt'}$ 为聚合效度， $r_{tt''}$ 为区分效度。聚合效度（convergent validity）是一个测验与测量相同特质的其他高效度测验相关的程度。而区分效度（differential validity）是一个测验与其他不同特质的测验相区分的程度（二者测验分数有低相关）。

三、研究效度

前述内容效度、准则效度和结构效度都是指效度鉴定过程的最终结果。与前者不同，后面提到的两种效度是在实验过程中发展出来的，是用于阐明实验中的误差来源（内部效度）和实验结果的适用范围的（外部效度），后来这两个术语也被应用于其他的研究类型之中。

（一）内部效度

内部效度（internal validity）是研究中研究变量（处理）被精确估计的程度，也就是研究中自变量和因变量关系的确定程度。如果在测量中，排除对结果产生影响的无关因素，因变量所产生的变化都是由研究者所操纵的自变量所引起，说明这个研究具有好的内部效度。反之，研究的内部效度较低。

影响内部效度的因素：

1. 历史

历史（history）是研究过程中发生的对研究结果可能产生影响的一个或多个特定因素和事件，这些事件可能会影响被试的反应。例如，长时记忆实验中被试的反复默念，或一些活动的安排与实验内容相似或完全相逆，从而影响实验结果，使遗忘的检测结果不真实（孟庆茂，常建华,1999）。历史对内部效度的影响有两种情况。一是在两次研究之间，一些特殊的事件可能发生，而研究者又不能对其加以控制，这些特定事件就有可能和自变量一起对被试的行为产生促进或干扰作用。例如，在一项持续考查美国和伊拉克人民对恐怖主义的态度实验

过程中，“9·11”事件发生了，这个事件就会对实验的结果产生不利的影响；二是在实验组和控制组实验中，某些特殊事件只对其中的一个组产生影响而对另一个组没有影响。例如，在有关发现学习和接受学习学习效果的比较研究中，如果在考试之前，一个令人悲伤的消息传到了发现学习的那个组，使大家沉浸在悲伤的气氛中，因此这个悲伤的消息有可能对发现学习组产生不利影响，而对接受学习组没有影响。

一般说来，前后研究之间的时间间隔越长，特定事件影响内部效度的可能性越大。

2. 成熟

成熟 (maturation) 是被试在研究过程中随时间推移所发生的生理和心理变化。在长期的实验研究中，被试的年龄增长会带来一系列的身心变化，如社会经验增加、学习的结果等等；在短期的实验中，被试则可能出现疲劳、饥渴、情绪变化等。这些都可能和实验中真正的自变量混淆起来对实验结果产生影响，降低实验研究的内部效度。例如，在一项“抓握练习对6~8岁儿童动作发展影响”的实验研究中，研究者发现，通过一年的实验后，练习与儿童动作灵活性的提高有显著的相关。然而，6~8岁儿童成长十分迅速，他们在动作灵活性上的提高可能只不过是成长的结果。

3. 统计回归

统计回归 (statistical regression) 也称向平均数回归，是指在重复测量时，初测时获高、低极端分数者的成绩会出现向平均数运动的趋势，即高分者随着时间的推移成绩下降，低分者随着时间的推移成绩升高。例如，某国家队教练从那些在各种竞赛的最后预赛上跑得最快的人中挑选队员。然而，她发现，在下次赛跑中，他们的平均成绩下降了，她认为这是由于跑道条件的差异所造成的，实际上是统计回归产生了影响。又如，教师对20个学生进行数学测验，被试的分数范围从52到100，于是他对分数最低的四位同学进行特殊的数学辅导，半个月后，对分数最低的四位同学和分数最高的四位同学进行与上次同样的测验，结果发现低分者平均分提高了，而高分者平均分降低了，统计回归产生了影响。

实验分组常常涉及某些具有极端特性的个体作为被试（如高自尊组和低自尊组）时，统计回归就会对实验的内部效度产生影响，研究者对统计回归的效应和自变量本身的效应难以区分。要避免统计回归的影响，研究者应该在匹配被试时考虑两组被试是否同质，如果不同质，就应该以几组被试各自的相对水平进行匹配。此外，研究者也可以利用统计控制进行协方差分析，在某种程度上也能降低统计回归对内部效度的影响。

4. 被试的亡失

被试亡失(subjects mortality)是指在研究进程中,一些难以预知的因素会导致样本中有较多的被试中途退出或死亡(动物实验)。研究者无法对研究结果进行正确的统计分析,从而无法做出正确的解释,尤其可能发生在周期较长的实验研究中。例如,坎加斯和布拉德韦(Kangas&Bradway,1971)在不同时间对同一组被试进行智力测验。其测验的具体时间是: 1931 年,被试的平均年龄只有 4 岁多时首次进行的斯坦福一比纳测验,以及后来的 1941 年、1956 年和 1969 年进行的重测。最初的样本是由旧金山海湾地区的儿童组成,他们是斯坦福一比纳量表修订时所需要的全国标准化群体中的一部分。研究中运用了两种形式的测验。在 1941 年,用同一个量表对 138 名被试进行了再测;在 1956 年,111 名被试受到了韦克斯勒成人智力量表和斯坦福一比纳测验的再测。在 1969 年,只剩下 48 名被试同意被再测。

在一些干预研究中,被试由于一种或多种原因(疾病、受到歧视、对研究漠不关心、其他活动需要等)而退出实验;又如,在资料和数据收集时,被试缺席或者没有完成测验,忘记寄回问卷等等,这样会严重影响统计分析的结论。被试亡失不仅仅影响到研究结果的推广,而且可能导致错误的研究结果。

5. 被试特征

被试特征(subject characteristics)也称选择(selection)偏向 是在研究中没有用随机取样的方法选择被试或进行分组,使所选择的几组被试的特征在实施实验处理之前就存在差异,不相等。例如,在各种对照组和实验组研究中,各组中被比较的被试可能在诸如年龄、性别、能力和社会经济背景等方面存在差异,如果对此不加控制的话,这些变量就会“解释掉”任何发现的组间差异(弗林克尔,瓦伦,2004)。研究中可能影响研究结果的一些被试特征的例子有:

- 年龄 ● 性别 ● 民族 ● 力量 ● 成熟 ● 动机 ● 灵活性
- 文化程度 ● 智力 ● 态度 ● 阅读能力 ● 社会经济地位
- 信仰 ● 心理健康 ● 强壮 ● 风俗习惯 ● 职业

6. 研究工具的使用

研究工具(research instrument)是指研究过程中实验仪器的使用和实验人员的变化。主要包括仪器使用不当,出现耗损、失灵、测验材料出现问题或主试身心发生变化等。例如,在进行“简单反应时”的实验中,经过多次的按键反应之后,反应键的灵敏度降低,被试需要更大的力气才能将键按下。又如,在高考阅卷过程中,当评分者一份接着一份地大量批阅试卷时,就会产生疲劳,对测验的评分也在不同的时间里变得不同,特别是对主观题的评分(刚开始很严格,后来变得宽松)。要降低研究工具使用对内部效度的影响,研究者应该将所有程序都标准化,每次研究之前对研究工具进行校准,实验人员也要得到充分的休息,保持良好的生理心理状态。

7. 测验

测验 (test) 对内部效度的影响主要是指先前的处理对被试在随后的处理中的反应产生影响。特别是在干预研究中，研究者在干预前要对被试进行测验，前测所产生的实际效果使被试在后面的测验中更为熟练，产生练习效应。例如，假设有研究者想知道使用新教材的效果。在使用新教材之前，研究者对学生进行了前测，然后，又在一个月的研究结束时，又对他们进行了后测，发现后测成绩有显著的提高。但后测分数的提高有可能是被试在前测上的“练习”本身所造成的。同样，前测后测的研究也可能导致疲劳效应。

实施前测本身不一定产生可能的测验效果，相反，产生测验效果的是，前测和实验处理之间出现了交互作用。前测的使用使被试对后测更加“警觉”和“有意识”，因而使他们对随后所发生的处理更加敏感和做出更积极的反应。

8. 主试和被试的交互作用

主试和被试交互作用是在研究的过程中，主试和被试的人为因素引起实验结果出现偏差。例如，被试会猜测实验的意图和研究目的，尽量表现自己，配合主试。主试有意无意地给予被试暗示，让被试的反应符合实验目的，主试对被试的偏见也会影响被试如何反应和实验者的客观判断。

心理学实验中，要有效消除主试和被试间不恰当的交互作用，保证实验内部效度的通用手段是采用单盲法 (single-blind method) 甚至双盲法 (double-blind method)，也就是说，主试知道实验目的，而被试不知道实验目的，或者主试和被试都不清楚实验的目的，可以最大程度上避免主试的暗示和被试的顺从。

9. 无关变量的交互作用

上述各种因素有可能产生交互作用，共同对内部效度产生不利影响。例如，在临床心理、发展心理研究中经常出现选择与成熟之间的交互作用，如果参加研究的实验组和对照组被试年龄不同，那么除了选择本身会影响内部效度以外，由于两个组的年龄不同，被试的成熟程度就会出现差异，选择和成熟就会产生交互作用。

(二) 外部效度

外部效度 (external validity) 就是将研究结果推广到研究外情景的程度，即实验结果的普遍代表性和适用性。詹金斯 (Jenkins, 1979) 的记忆实验四面体模型提供了评价这种普遍性的维度：(1) 用其他的被试人群也能得到同样的结果吗？(2) 用其他的实验材料也能得到同样的结果吗？(3) 用其他不同类型的测验会出现同样的结果吗？(4) 用不同的实验处理和不同的自变量操作方式也能出现同样的结果吗？如果能够将实验当中得到的结果推广运用到实验以外的情景，该研究有较好的外部效度。反之，该研究的外部效度不好。

影响外部效度的因素有：

1. 实验条件的人为性。心理学实验是在严格控制条件下进行的，这可能导致在实验室得到的某些实验结果难以用来解释日常生活中的行为现象。例如，在特定的实验情景中，被试在实验处理前就了解到了实验的安排，在实验过程中受到了周围环境的暗示，就可能导致霍桑效应（Hawthorne effect）和安慰剂效应（placebo effect）的出现，被试表现出与平时不同的行为，从而降低了实验结果的普遍意义。

2. 实验处理的多重性。同一被试多次接受实验处理时，前一实验处理通常会对后一实验处理产生积极或消极的影响，发生重复处理的干扰，使被试产生练习或疲劳效应。例如，在被试内设计（within - subjects design）中，相同的被试要接受多种实验处理，这些处理相互会产生影响。有多重处理的实验结果难以推广到没有重复实验处理的情境。例如，坎贝尔和斯坦利（Campell&Stanley, 1966）证实，艾宾浩斯（Ebbinghaus）的遗忘曲线（forgetting curve）受到多重处理的干扰，因为他的发现只对已经学习了多列无意义音节的被试来说是正确的，而不能运用到只学习了一列无意义音节的被试。

3. 被试取样的代表性。被试的代表性是影响实验结果推广的一个重要因素。只有参与实验的被试具有代表性，从样本推论总体才可靠。我们从第二章了解到只有从总体中随机取样，才能保证样本的代表性，但在实际中很难做到，特别是当总体非常大甚至无限的情况下。例如，“大学生”是一个无限总体，包括中国的、外国的大学生，过去、现在、将来的大学生，就无法对其进行随机取样。

同样，由于取样的方法不当或取样错误，研究者选择参与实验的样本具有某种心理特征，这种特征与实验处理发生作用共同对实验结果产生消极或积极的影响，导致该样本的实验结果难以推论到总体。例如，如果选择一些智力水平较高的学生，进行发现学习和接受学习的比较，实验结果发现前一种学习方法优于后一种学习方法，但这一结果不能推论到实验外的其他智力水平的学生中去。

4. 测量工具的局限性。实验者对自变量和因变量的操作性定义往往是以所使用的特定测量工具的测量结果来加以考虑的。例如，把人格作为一个实验变量，实验者常以某种人格量表所测得的分数来评定其人格特征。但人格的测量工具有各种不同的形式（艾森克人格问卷，卡特尔 16 种人格因素问卷等），所测量出的分数并不代表同一种人格特征。如果在实验时采用的是某一种人格量表，那么所得出的实验结果便不能推论到采用其他人格量表的情况中去。

补充讨论 4 - 1 一种特殊的内部效度：统计结论效度

统计结论效度（statistical - conclusion validity）是测量中有关决定实验处理效应的数据分析程序的有效性和准确性的指标。统计结论效度的基本问题是研

研究误差的变异来源与如何适当地运用统计显著性检验。例如，采用小样本的研究数据时，由于样本成分与测量数据都波动比较大，稳定性差，如果依赖统计显著性水平做出推论是不可靠的。在这种情况下，应该运用功效分析（power analysis），考察一定的样本范围、变异程度和 α 水平上能够检验出多大的效应。

研究的统计结论效度主要取决于两个方面的条件：一是数据的质量，数据分析程序的效度是以数据的质量作为基础的，数据质量差的研究是谈不上统计结论效度的，包括数据的量表特征（称名量表数据、顺序量表数据、等距量表数据和比率量表数据）、数据的分布特征（ t 分布、 F 分布、卡方分布和正态分布等）以及数据测量的信效度。二是统计检验方法，数据分析中所采用的各种统计方法，都有其明确的统计检验条件的要求，一项研究中统计检验条件不明确或者被违反，就会显著降低统计结论效度。

影响统计结论效度的因素主要有几个方面：一是统计功效低。统计功效也称统计检验力（power of Statistic），是当虚无假设不正确时拒绝虚无假设的能力。它反映某个统计检验正确辨认真实差异的能力，其值等于 $1 - \beta$ ， β 越小，检验力越大。二是违反统计方法的使用条件。心理学研究的大多数统计检验都要满足一定的条件，然后才有可能对数据分析的结果作出有意义的解释。因此，对实验的数据进行统计检验前，应该了解统计检验的前提条件。例如，对于顺序量表所测得的数据应运用非参数统计方法进行分析， t 检验（ t -test）必须在总体正态分布（normal distribution）的条件下进行，若是独立样本，还要求两个总体方差齐性（homogeneity of variance）。三是测验信度低。低信度的测验将夸大估计标准误（Standard error of estimate），它是预测误差大小的指标。四是实验处理实施的可靠性。由于在同一实验中不同的研究者之间存在差异或者同一研究者在不同的时间采用不同的方式来实施实验处理，这也会降低统计结论效度。另外，研究背景中无关的随机因素和被试的随机异质性也会影响到统计结论效度。

第三节 测量的信度

一、何谓信度

信度（reliability）是测量的可靠性程度，表现为测量数据和结果的稳定性以及测验内容的一致性。一项科学的心理学研究，不论其过程由谁操作，或进行多少次同样的操作，其结果应该稳定可靠，即重复测量的结果要保持稳定、一致，否则便不可信。例如，假如你用直尺多次测量身高，星期一你的身高是 180 cm，星期二你仍然是 180 cm，那么你用的直尺就是可信的，因为它两天都测量你的身

高是 180 cm。假如它测量你的身高星期一是 180 cm 而星期二是 160 cm 或者星期一是 166 cm 而星期二是 173 cm，这个直尺就是不可信的。又如，假设在高考阅卷中，一位教师对某作文的评分是 35 分，另一位教师的评分是 36 分 相差不多，这种评分具有一致性，是可靠的。但是如果一位教师对某作文的评分是 35 分，而另一位的评分是 60 分，二者的差异太大，难以令人信服。测量的数据和结果是否可靠与变量的控制和误差密切相关，变量控制严格，测量误差小，测量结果的可靠性就高，变量控制随意，误差大，测量结果的可靠性就低。可靠性高的测量就可以重复验证，心理学研究的科学性才可以得到保证。

测量的信度同时要考虑两个问题：（1）测量结果一致性的程度。即不同时间和条件下测量结果的一致性有多大？（2）测量结果不一致的原因。即什么因素影响了测量的信度？

二、两种信度评定方法

前已述及，一种测量方法的可信度就是测量结果的一致性和稳定性。如何来检验这种一致性和稳定性呢？通常我们采取两种形式：并行法和重复法。

（一）并行法

并行法（parallel - form method）又称对等法，通常的做法是根据测量的目标，同时做两次在内容、难度和题型上尽可能类似的测验，然后分别用来测量两组同质的学生，两组学生的得分之间的相关就是所求的测量的信度。此种信度旨在测定分数的对等性或稳定性。

并行法中常用复本信度来作为测量信度的指标。复本信度（alternate - form reliability）是对同一组被试几乎同时实施同一性质的两个测验（复本），所得结果一致性程度。即在最短的时间间隔内，用同一测验的两个复本对同一组被试测试，以测验分数的相关系数表示的信度。复本是同一个测验的几种不同形式。具体地说，能以相同程度测量同一心理特质的测验就是复本，也叫平行测验。例如，斯坦福一比奈智力测验（1937 年版）就有 L 和 M 两种测验，同时进行这两种测验所得的信度系数是 0.91。复本信度系数的计算用皮尔逊积差相关。

以复本信度系数作为测验信度指标的基本前提是所用的两个复本必须是真正等值的。即复本除了具体项目不同外，在作为测量目标的特质、测验长度、项目类型、难度、区分度等方面一般应是相等或相似的。复本信度的缺点在于：容易出现练习效应，测验性质有可能发生改变，增加了测题的编制和筛选难度。

（二）重复法

重复法（repeat method）就是对原有的测量过程进行复制。即采用相同条件、相同方法进行重复测量（两次以上）的方法。运用重复法，测量可靠性的问题就可简单归纳为：如果重复测量，其结果与前面的测量相同吗？例如，在 1980

年 美国的杰佛里·赛克(Sacks) 与他的同事们进行了两次问卷调查，内容都是关于被试的生活习惯与行为的，总共有 207 名被试参加了这次调查，第二次调查与第一次调查在时间上相隔了 3 个月。但是，问卷的内容与问题完全一致。调查的结果显示，只有 15% 的被试在两次调查中提供的信息与情况完全一致。赛克报道说：“大约 20% 的被试在第二次调查时报了不同的身高，每三个被试中就有一个把父母的年龄加以改变，每五个吸烟者与饮酒者中就有一个记不清他们以前的消费习惯。甚至有一个被试已故的母亲，在第二次调查中神奇地复生，还有一个被试在第一次调查时为 55 岁，但是在第二次调查时却只有 50 岁零 3 个月。”很显然，重复调查的结果不一致，可靠性存在问题。大多数经典的测量是能够重复验证的。例如，社会心理学的一项经典研究“米尔格莱姆服从研究”自 1963 年发表以来，米尔格莱姆(Stanley Milgram) 本人用耶鲁大学以外的其他类似被试重复了他本人的实验过程，被试都是义务参加实验的大学生自愿者和一些女性被试，每次得出的结果都是相同的(哈克，2004)。重复法可以对研究过程、研究工具、研究结果的信度水平做出直观的判定。因此，它是判定测量信度的基本方法。

我们常用重测信度(test - retest reliability) 来作为重复法的信度指标。重测信度是同一量表在同一样本中先后测量两次或多次，其测量结果的一致程度，是反映测验分数随时间的变化是否发生变化的信度指标。使用这种方法时，两次测量所使用的工具是完全一样的。例如，艾森克人格问卷简式量表中国版(钱铭怡等 2000) 的修订，相隔 3 周的重测信度分别为：P 量表 0.67、E 量表 0.88、N 量表 0.80、L 量表 0.78(所有 $P < 0.01$)。Wechsler 儿童智力量表(WISC - R) 的各分量表和全量表的重测信度在 0.90 ~ 0.95 之间。同样我们用皮尔逊积差相关来计算重测信度。

例 4 - 1 用同一测验在 30 天内对 10 名学生测验两次，第一次测验的成绩用 X 表示，第二次测验的成绩用 Y 表示，具体分数如下表 4 - 3) 求重测信度系数。

表 4 - 3 某 10 名学生的两次测验成绩表

	1	2	3	4	5	6	7	8	9	10
X	61	70	75	74	69	90	65	82	72	65
Y	65	64	90	71	75	92	75	88	80	72

$$\sum X = 732 \quad \sum Y = 772 \quad \sum X^2 = 54412 \quad \sum Y^2 = 60504 \quad \sum XY = 57263, N = 10$$

代入皮尔逊积差相关公式可得：

$$r_{xy} = \frac{10 \times 57263 - 732 \times 772}{\sqrt{10 \times 54412 - 732 \times 732} \times \sqrt{10 \times 60504 - 772 \times 772}} = 0.87$$

三、信度和效度的关系

信度和效度存在以下关系：

1. 信度为效度的必要而非充分条件

如图 4.1 所示，根据经典测量理论， $S_x^2 = S_T^2 + S_E^2$ ，当随机误差的变异数（ S_E^2 ）减小时，真分数的变异数（ S_T^2 ）增加，测验信度（ S_T^2/S_x^2 ）随之提高。信度的提高只给有效变异数（ S_T^2 ）的增加提供了可能。至于是否能提高效率，还要看系统误差变异数（ S_V^2 ）的大小。可见，信度高，效度不一定高。但一个测验要有高的效度，真分数的变异数（ S_T^2 ）必须占较大的比重，即测验的信度必须高。例如，用智力测验来测量智力是有效的，也是可信的；但用智力测验来测量成就动机，虽然多次测量的结果是一致的，即有高的信度，效度却很低。

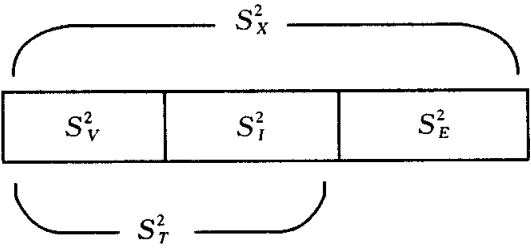


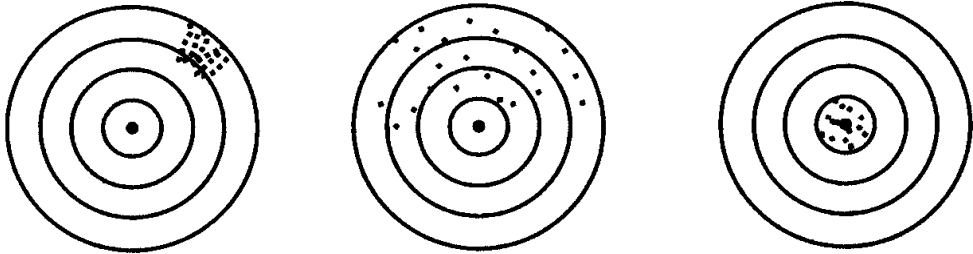
图 4.1 信度和效度的关系
(采自郑日昌,1987,113 页)

2. 效度受到信度的制约

根据信度和效度的定义（ $r_{xx} = \frac{S_T^2}{S_x^2}, r_{xy} = \frac{S_{VT}}{S_x^2}$ ），以及 $S_x^2 = S_T^2 + S_E^2$ 可以得到：

$$r_{xy}^2 = r_{xx} - \frac{S_T^2}{S_x^2}, \text{ 可见 } r_{xy}^2 \leq r_{xx}, \text{ 一个测验的效度总是受到其信度的制约，总小于信度。}$$

信度和效度的关系就像打靶的成绩一样，有以下三种关系：信度高，但缺乏效度；没有信度也没有效度；信度和效度都高。如图 4.2。无论是缺乏信度还是效度，都将使测量的质量受到影响。



信度高,但缺乏效度 没有信度,也没有效度 信度和效度都高

图 4.2 信度和效度的比喻

补充讨论 4-2 测验的内部一致性

如果测验没有复本，也不可能进行重复测量时，如何考察测验的信度呢？这时我们也可以采取另一种形式来估计测验的可信度，即考察测验内部各题目所测内容的一致性（内部一致性）。如果一个测验可靠，这个测验所包含的项目就应该是前后一致的。

内部一致性信度（internal consistency reliability）是反映测验内部所有题目间一致性程度的信度指标。题目的一致性有两层含义：一是所有测验题目反映的是同一特质；二是各个测题之间有较高的相关。分半信度系数、克伦巴赫 α 系数、库德-理查森公式、因素分析等都可作为估计测验内部一致性程度的指标与方法。

分半信度是反映测验内部两半题目之间一致性程度的信度类别。如对一组被试实施某一测验，按题号将测验分成由奇号题组成的“半个测验”，与由偶号题组成的“半个测验”，这两个“半个测验”上得分的相关系数，经校正后就得到分半信度系数。求测验的分半信度，首先要将测验项目分成对等的两半。对等有两个要求：一是测验的两部分在难度、区分度及测验目标上基本是相同的；二是被测者以同等的态度来对待两部分测验，即在完成两部分测验过程中，练习、疲劳、情绪等因素对被试产生了同等的影响。因此，将一个测验分成两部分时，常用的是奇偶分半法，即奇数题为“半个测验”，偶数题为“半个测验”。特别是测验项目是按由易到难排列时，这种分法可以将测验分为大致相等的两半。

克伦巴赫 α 系数（Cronbach's α coefficient）是由美国心理学家克伦巴赫提出的一种计算测验题项内部一致性的指标，等于所有可能的分半信度系数的平均值，通常情况下取值在 0、1 之间。克伦巴赫 α 系数是反映测验项目的一致性程度和内部结构的良好性，对于由客观性题目和主观性题目组成的测验，有些题目是多重计分的情况下，常使用的一种信度估计方法，内部一致性也可以用库德-理查森公式（Kuder - Richardson formulas）来估计。

第四节 误差

一、何谓误差

误差（error）就是对正确的偏离。在心理学中常常用到这个术语。在对该术语应用时，研究者主要关注的是误差的来源；对该术语的修饰词起着确定和指

出误差源的作用。例如，绝对误差（ absolute error ）是指一量度的观察值与真值之间的差别的绝对值（即不考虑正负）。例如，对某一标准时距的高估 2 秒与低估 2 秒所产生的绝对误差是一样的。又如，抽样误差（ sampling error ）是指某总体内的一个统计数的真值与从该总体内的一个样本中得出的那个统计数的估计值之间的差异。测量误差（ measurement error ）是一种因测量手段的偏差所造成的常误。

二、误差类型

从前述缺乏效度或信度的讨论中可以了解到，测量误差只是在心理学研究中可能出现的多种误差之一。在心理学研究过程中可能出现各种误差，表 4 - 4 列出的是不同类型的误差及其在各研究阶段可能出现的情况。甚至在研究开始之前，研究者也可能由于选择了一个不相关的或不重要的研究题目而在一定意义上就造成了误差。在表 4 - 4 的第一阶段中，如果提出的概念和假设模棱两可，难以下操作性定义，测量得到的结果就会导致缺乏内容效度，从而使整个研究价值降低。由于第一阶段工作导致缺乏了内容效度，因而也必然导致缺乏信度，因为后者是依前者而定的。但是，即使概念明确，定义充分，假设恰当，如果在第二阶段中建立起的测量工具（如问卷）中的项目表达不清，被试不理解，其结果也是缺乏信效度。如果不同的被试对同一题项的意义有不同的理解，也会导致无效的和不可信的资料。

表 4 - 4 心理学研究过程中的误差类型

研究阶段	误差类型
1. 提出概念和假设 (包括操作性定义的界定)	缺乏内容效度
2. 确立研究手段(例如问卷)	缺乏信度(所问问题的用词错误或含混不清)
3. 取样	缺乏外部效度(取样误差)
4. 资料搜集	因没有控制下列因素而造成的误差： (a) 环境 (b) 被试的个人特征(疲劳,等等) (c) 被试和研究者之间的关系 (d) 研究手段的缺点(录音声音有缺点,设备失灵,等等) (e) 主持问卷者误解了指导语
5. 编码	编号或记录错误
6. 资料分析	误用统计或误解资料

在第三阶段中，取样不充分会导致结果缺乏代表性，造成研究结果仅适用于特定的样本，而缺乏外部效度。

在第四阶段中，资料搜集时有许多可能的误差源，包括无法控制的外部环境因素、被试的内部因素（如期盼、倦怠、注意力分散）、研究者和被试间的关系，或者设备失灵（如，录音机出毛病）等。

在第五阶段中，把问卷中的答案进行编码或将其转输到计算机上，可能会产生笔误或输入错误（在第四阶段中，被试回答问卷时也会产生这类误差）。因被试的字迹难辨会导致编码错误，还有一种可能是将一个人的回答错编成另一人的等等。

最后是资料分析阶段。这里的误差可能源于对资料的错误解释。造成这种误差，是因为从资料中得出了不正确的结论，或者明明资料并不说明什么而偏说它们说明了什么。这种误差之所以发生，是因为研究者对资料的分析有倾向性，或未受过适当的训练。这一阶段另一个最常见的错误是选择了错误的统计方法来分析资料，因而导致结论错误。

在心理学研究中可能导致的各种误差，其严重性（它们对研究结果所造成的损害程度）和所能克服或纠正的程度都各不相同。误差可能是偶然性的，也可能是系统性的。由偶然因素造成的误差被称为随机误差（random error）。这种误差原则上是偶然性的、无任何特定型式的。这类偶然误差，只要样本数目足够大，往往能互相抵消，机误的总和接近于零。我们当然希望什么误差都没有，但偶然误差似尚可容忍，因为它们对我们整个的结论没有或几乎没有影响。

系统误差（systematic error）则不是偶然性的，而是定型式的，因素间也不能互相抵消掉。不过，既然这类误差是定型式的，它们就较易被发现、去除或在无法去除时加以纠正。例如，在邮寄问卷研究时，回收的问卷中被试在回答教育水平一栏上只有 60% 的回答率，这种无回答现象显然不是由于偶然疏忽所致，而是由于被试教育水平较低所致。一旦找到了造成此种误差的原因，那么我们就能够加以纠正。

► 本章提要

1. 测量是一个为确定某个特定分析单位的特定属性值的过程。包括定名测量、定序测量、定距测量和比率测量。

2. 定名测量是把事物的属性归于不同类别的测量。定序测量是对事物的等级和顺序的测量，是对一系列事物进行排序。定距测量是对事物的数量差别和间隔距离的测量。比率测量是对事物的比例或比率关系的测量。

3. 测量的效度是指测量得到的记分是否反映了欲测的特征及其程度。包括内容效度、准则效度和结果效度。

4. 内容效度是指测量在多大程度上包含了欲测的内容范围。准则效度是以测验分数与作为外在标准的效标行为之间的关联程度来表示的一种效度。结构效度是测验对某种理论构想或特质所能体现测量的程度。

5. 内部效度是研究变量（处理）被精确估计的程度，也就是研究中自变量和因变量关系的确定程度。影响内部效度的因素有历史、成熟、统计回归、被试亡失、被试特征、研究工具的使用、测验、主试与被试的交互作用等。

6. 外部效度就是将测量结果推广到研究外情景的程度，即测量结果的普遍代表性和适用性。影响外部效度的因素有实验条件的人为性、实验处理的多重性、被试取样的代表性和测量工具的局限性。

7. 信度是测量的可靠性程度，表现为测量数据和结果的稳定性和测验内容的一致性。检验信度的方法有并行法和重复法。

8. 误差就是对正确的偏离。测量误差是一种因测量手段的偏差所造成的常误。误差可能是偶然性的，也可能是系统性的。

►本章关键术语

测量(measurement)	结构效度(construct validity)
定名测量(nominal measurement)	聚合效度(convergent validity)
定序测量(ordinal measurement)	区分效度(differential validity)
定距测量(interval measurement)	内部效度(internal validity)
比率测量(ratio measurement)	外部效度(external validity)
离散测量(discrete measurement)	信度(reliability)
连续测量(continuous measurement)	并行法(parallel - form method)
效度(validity)	复本信度(alternate - form reliability)
内容效度(content validity)	重复法(repeat method)
准则效度(criterion validity)	重测信度(test - retest reliability)
效标(criterion)	误差(error)
同时效度(concurrent validity)	随机误差(random error)
预测效度(predictive validity)	系统误差(systematic error)
效度系数(validity coefficient)	

►复习与练习

1. 如何提高测量的信度和效度？
2. 对下列变量进行测量属于哪种层次（定名测量、定序测量、定距测量、比率测量）？
英语成绩 智力 态度 人格 性别 宗教信仰 体重 学习速度 自尊
3. 如何有效的降低测量误差？

►► 推荐参考读物

1. 周文钦. 研究方法—实证性研究取向. 台北: 心理出版社, 2002
2. 维尔斯曼. 教育研究方法导论. 袁振国主译. 北京: 教育科学出版社, 1997
3. Salkind, N. J. Exploring Research(4th ed). New Jersey :Prentice Hall, Upper Saddle River, 2000
4. 贝利. 现代社会研究方法. 许真译. 上海: 上海人民出版社, 1986
5. 凌文轻、方俐洛. 心理与行为测量. 北京: 机械工业出版社, 2003

►► 在线资源

想进一步了解测量的知识, 可以访问 <http://trochim.human.cornell.edu>

想查阅大量的信度和效度的研究报告, 可以访问 <http://www.ncme.org>

第二编 心理学研究设计

研究设计是整个研究工作的计划和安排,它是科学研究中的关键环节。研究设计是否科学、合理,不仅直接影响到科学研究的进程,而且还影响到研究结论的可靠性和科学性。本编“心理学研究设计”,将着重介绍和讨论心理学研究中的各种研究设计模式。心理学的大部分研究结果是在实验条件下获得的,第五章“真实验设计”着重介绍和讨论了心理学研究中各种真实验设计模式及其评价。人生活在复杂的社会环境中,实验室“真空”状态的研究,不能有效的揭示出人类的心理和行为规律,因此心理学的研究逐渐从实验室里人为的实验环境转向现场研究。这种研究的变化,对实验设计提出了新的要求,从而促进了准实验设计理论与方法的发展与应用。第六章“准实验设计”将对准实验设计的基本模式进行介绍。但是心理学研究的主要对象是人,实验研究和准实验研究仍不能很好的揭示在复杂多变的社会环境中的心理和行为活动以及人的心理的本质,第七章和第八章“非实验设计(Ⅰ)(Ⅱ)”和第九章“质的研究”就如何针对人的研究设计进行了探讨。

第五章 真实验设计

真实验研究设计是相对准实验研究设计和非实验研究设计而言，是实验类研究中条件控制最为严格的一种，有时也简称实验研究设计。真实验设计的基本逻辑是，根据随机化的原则把被试分配到不同的实验条件中去，所形成的这些组具有同质性或是等组（equivalent groups）；也就是这些组在相同的条件下完成相同的任务，他们的成绩在统计上应该是相等的。如果这些组的成绩有所差异，则可以推论这些差异是由于不同的实验条件所造成的。真实验研究设计的特点主要体现在，能够随机的选取并分配被试，能够在有效控制无关因素干扰的基础上操纵自变量的变化，能够精确的测量因变量的变化。因此只要能够严格控制无效变异来源的实验设计，都可以成为真实验设计。在一个好的实验设计中，自变量是唯一正在被操纵的变量，而各组中的所有其他条件都应当保持恒定。也就是，除自变量外，如果实验组和控制组的处理非常相似，那么因变量之间的差异一定是由自变量引起的。例如，在视错觉的实验研究中，被试眼睛的颜色、身高、运动技能以及关于足球的知识等变量可能不会影响被试的知觉，研究者通过随机化的方法就可以平衡掉这些个体差异。但是，被试的视力和呈现刺激的亮度等变量很可能会影响到实验的结果，研究者要对这些变量进行控制。

可以从不同的角度把真实验设计划分为不同的类型。例如从控制无关变异方法的角度，可分为完全随机、随机区组和拉丁方设计。从自变量数目的角度，可分为单因素和多因素设计。从被试是否接受所有实验处理的角度，可分为被试内、被试间和混合设计。本章重点介绍单因素完全随机设计、多因素完全随机设计、单因素随机区组设计和多因素随机区组设计。

在讨论各种实验设计之前，我们把本章中所采用的符号及其含义做一个简单说明：

(1) X ：表示一种处理，即研究者操纵的实验变量（自变量）；在比较不同的实验处理时，以不同的下标来表示不同的实验处理，如 X_0 、 X_1 、 X_2 、 X_3 。

(2) O ：表示处理前或后的一种观测或量度。

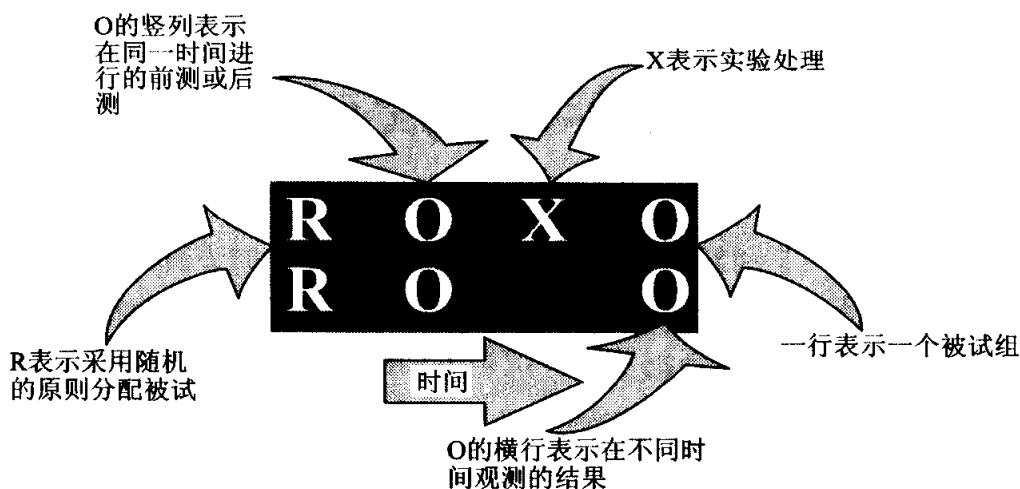
(3) 自左向右：表示时间的次序或先后。

(4) 同一横行的 X 或 O 表示 X 或 O 作用于同一组被试。

(5) R ：表示对被试进行随机化的选择和分配。

(6) M ：表示把被试进行配对。

下面以实验组控制组的前测后测设计为例说明这些符号的含义，



在本章中，我们将讨论以下问题：

1. 各种实验设计的基本模式、统计方法及其优缺点。
2. 心理学研究中为什么要尽可能采用多因素实验设计。

第一节 单因素完全随机设计

单因素设计只有一个自变量（因素），而随机化设计则是指采用随机化的方法分配被试到各个实验处理中。单因素完全随机设计是指研究者在实验中只操纵一个自变量，并采用随机化的原则把被试分配到自变量的不同水平上的一种实验设计。根据自变量水平的多少，单因素完全随机设计可分为两等组模型和多等组模型；根据有无实验前测，可分为后测模型和前测后测模型；根据是否进行配对分组，可分为随机等组模型和随机配对等组模型。

一、实验组控制组后测设计

（一）实验组控制组后测设计模式

实验组控制组后测设计只有一个自变量，并且自变量只有两个水平，其设计的基本模式如下：

$$\begin{array}{ccc}
 R_1 & X & O_1 \\
 R_2 & - & O_2
 \end{array}$$

该设计的基本模式表示：首先采用随机分配的方法将被试分为同质的两组，两个组在理论上完全相同，然后随机选择其中的一组作为实验组接受实验处理，另一组作为控制组不接受实验处理。在实验处理后，两组接受相同的后测，并对所获得的观测结果的差异进行比较，以推论实验处理的效果。例如，为了探寻观

看暴力电视是否导致攻击行为的增多 (Eron, Huesmann, Lefkowitz & Walder, 1972), 实验中, 实验组儿童观看暴力动画片, 在同样长的时间里, 控制组儿童则观看非暴力的动画片。结果发现, 观看暴力动画片的儿童与同伴们交往时变得具有更多的攻击性, 而观看非暴力动画片的儿童的攻击表现则没有变化。因此, 经常观看暴力电视的儿童具有更多的攻击行为的倾向。

对于该设计实验结果的统计分析, 可采用独立样本的 t 检验的方法进行数据统计分析。

(二) 实验组控制组多组后测设计模式

如果在一个实验中, 实验因素具有三个或三个以上的处理水平时, 上述设计的结构模式可变为:

R_1	X_1	O_1
R_2	X_2	O_2
R_3	X_3	O_3
...
R_n	X_n	O_n
R_{n+1}	-	O_{n+1}

这种模式和实验组控制组后测模式的区别仅在于增加了自变量的水平, 即由两个水平变为多个水平。这种设计也称为随机多组后测设计。在这种设计模式中, 随机选取并分派被试组成等组, 其中可以有一个组是不接受实验处理的控制组, 其他各组分别接受不同的实验处理; 也可以所有的组都接受不同的实验处理, 对各组可能出现的结果差异进行比较。

例如在上例暴力电视是否导致攻击行为的增多的研究中, 如果我们想进一步考察电视对儿童的行为的影响, 也就是通过设置助人、中性和暴力三种电视内容考察对儿童行为的影响。这种设计就是实验组控制组后测模式。同样, 实验中, 随机形成的三个实验组, 在同样长的时间里, 其中一个组的儿童观看暴力动画片, 另一组儿童则观看中性动画片, 第三组的儿童观看助人动画片。然后, 考察观看不同类型动画片的儿童与同伴们交往时行为的变化。

对于单因素完全随机多等组后测设计的数据分析, 可采用单因素方差分析的统计方法。如 F 检验达到了统计显著性水平, 表明在所有处理条件中至少有两个处理条件的差异达到了显著性水平。随后还需要进一步分析这些处理中哪些处理间具有显著的差异, 进行有关单因素方差分析的事后多重比较 (post hoc test)。如果 F 检验没有达到统计显著性水平, 则可以得出所有处理条件间的结果不具有可靠的差异, 也就不需要进行事后多重比较。

需要注意的是, 对单因素完全随机多等组后测设计, 不能直接进行处理间的

两两 *t* 检验比较。这样做的后果会使统计检验的 I 型错误的概率增大。心理学研究中，尤其是很多初学者很容易犯这样的错误，应引起大家的注意。

（三）实验组控制组后测设计的评价

实验组控制组后测设计的优点：（1）采用随机选取和分配被试的方法形成两个或两个以上的等组，从而可以控制选择、被试消亡以及选择和成熟的交互作用等因素对实验结果的影响。（2）由于安排和使用实验组和控制组，并以两组后测成绩的比较为依据，因此可以控制历史、成熟、仪器的使用等因素对实验结果的干扰。（3）对实验组和控制组不进行前测，可以避免因前测经验而产生对后测的练习、熟悉和疲劳效应，从而控制前测对后测的反作用效应。此外不进行前测还可以简化实验程序，节省人力、物力和时间。总之，单因素完全随机等组后测设计，可以对影响实验设计内部效度的干扰因素进行有效的控制。

单因素完全随机等组后测设计的局限是，只有一个自变量，所以不能分析多个自变量及其交互作用对因变量的影响。这是所有单因素实验设计共同具有的研究局限。

实例学习 5-1 实验组控制组后测设计实例分析

为了验证社会心理学中印象形成的“首因效应”，即对某个人的最初信息（首因信息）比后面的信息（近因信息）在印象形成中更重要。阿什（Asch, 1952）设计了一个实验来进行验证。随机选取并分配被试到两组中，然后呈现一系列描述个人的形容词。其中，一组首先呈现肯定的形容词，然后呈现否定的形容词；另一组则相反，先呈现否定的形容词，然后呈现肯定的形容词。呈现给两组被试的形容词完全相同，只是顺序相反。随后，要求被试写出他们对这个人的大致印象。阿什研究的实验设计如下表所示：

阿什研究的实验设计

自变量		因变量
形容词的呈现顺序(先肯定后否定)	⇒	对某人的印象
形容词的呈现顺序(先否定后肯定)	⇒	对某人的印象

研究结果表明，呈现先肯定后否定形容词的被试组对人产生的是积极肯定的印象，而呈现先否定后肯定形容词的被试组产生的是消极否定的印象。因此验证了印象形成的“首因效应”。

在该实验设计中，自变量是肯定和否定形容词的呈现顺序（先肯定后否定，先否定后肯定），因变量是被试对个体的描述。其他变量则保持一致，例如随机

选取并分配形成两个等组，两组所呈现的形容词完全一致等。由于两组处理的唯一差别就是形容词的呈现顺序，因此可以得出印象形成的差别可归因于形容词的呈现顺序，也即印象形成中的“首因效应”。

二、实验组控制组前测后测设计

（一）实验组控制组前测后测设计模式

实验组控制组前测后测设计是在单因素完全随机等组后测设计基础上的扩展，即对两个等组施加了前测，其基本模式为：

$$\begin{array}{cccc} R_1 & O_1 & X & O_2 \\ R_2 & O_3 & - & O_4 \end{array}$$

例如，沃坦阿贝等（Watanabe, Here & Lomax）采用实验组控制组前测后测设计探讨教学训练对培养学生根据报纸标题预测报道内容的能力。实验中，随机抽取 46 名中学生随机分配为实验组和控制组。在接受训练前，要求学生阅读 20 个文章标题并预测其内容。然后对实验组进行为期 3 周的阅读教学训练，而控制组则进行常规阅读训练。3 周的训练结束后，两个组同时接受后测测验，即根据 20 个标题来预测所报道的内容。统计结果表明，经过阅读训练的实验组显著高于控制组，表明阅读训练可以提高学生根据标题预测其内容的能力。

（二）实验组控制组前测后测设计的数据分析

实验组控制组前测后测设计可采用两种统计方法进行数据分析。（1）独立样本的 t 检验。这里首先要计算出两组前测的平均数，并进行比较是否相同。如果两个平均数没有达到显著性差异，则对两组后测分数的平均数进行比较。通过对后测得到的两组平均数差异进行独立样本的 t 检验。如果两个组前测的平均数存在显著差异，则不能直接比较两组后测平均数的差异，这时需要对两组后测的增值平均数进行检验，即先计算出两组前测后测差异的平均数，再进行独立样本的 t 检验。（2）单因素的协方差分析。在单因素协方差分析中，把两个组的前测成绩作为协变量，通过统计控制前测对后测的影响效果后，来比较两组后测的差异。

（三）实验组控制组前测后测设计的评价

实验组控制组前测后测设计的优点主要体现在：（1）由于采用随机选取并随机分配被试的方法形成等组，从而可以控制选择、被试亡失以及选择与成熟的交互作用等因素对实验结果的干扰。（2）由于安排了实验组和控制组，对于实验结果的分析是以实验组和控制组后测成绩的比较为依据，因此在前测到后测的阶段内，所发生的一切可能影响实验结果的因素对实验组和控制组的影响是基本相同的。

从而可以控制历史、成熟、仪器的使用和统计回归等对实验结果的影响。

实验组控制组前测后测设计的局限性主要表现在，由于实验组和控制组都参加前测，使得被试的前测经验可能影响后测的敏感性，可能产生熟悉效应或疲劳效应，从而影响后测的可靠性。此外，进行两次测验，从人力、物力和时间上也不是很经济。因此，使用该种设计的所得实验结果不能推论到没有使用前测的样本群体中。

三、所罗门四组设计

(一) 所罗门四组设计的基本模式

所罗门四组设计 (Solomon four-group design)，也称重迭实验设计，是由所罗门 (R. L. Solomon) 于 1949 年提出的一种具有两个实验组和两个控制组的随机设计，其基本的设计模式为：

<i>R</i>	<i>O</i> ₁	<i>X</i>	<i>O</i> ₂
<i>R</i>	<i>O</i> ₃		<i>O</i> ₄
<i>R</i>		<i>X</i>	<i>O</i> ₅
<i>R</i>			<i>O</i> ₆

上述模式显示，将被试随机分成四个组，其中随机组 1 和 2 接受前测，随机组 1 和 3 接受相同的实验处理，而随机组 2 和 4 不接受任何实验处理，作为控制组，并且 4 个随机组都接受实验后测。

所罗门四组设计的主要特点是把“有无前测”作为一个变量纳入实验设计中。在实验设计中，将此变量所造成的变异量从总变异量中排除出去，以此来检验实验处理所产生的效果是否显著。与实验组控制组前测后测相比，它增加了两个后测组；同样，与实验组控制组后测相比，它增加了两个前测组。因此，事实上这种设计是实验组控制组前测后测设计和实验组控制组后测设计合并的结果。

所罗门四组设计可能产生的实验处理和前测影响的实验结果可用图 5.1 和图 5.2 来表示。图中共有 6 个点，分别代表 6 个观测结果的平均值，为了能够清楚地看到前后测变化的趋势，我们只把具有前、后测的两个被试组的成绩连接起来。而没有连接的两个点表示另外两个只有后测的被试组的结果。图 5.1 表明，由于采用随机选取和分配被试，因此两个组的前测成绩非常接近。对于后测成绩，两个实验组的成绩明显高于控制组的成绩。同时在两个实验组，具有前测的实验组的后测成绩并没有与只有后测的实验组的成绩具有显著的差异。在两个控制组中也表现出同样的趋势。因此，该结果表明存在显著的实验处理效应，而没有前测的影响。

图 5.2 表明，具有前、后测的实验组成绩显著高于具有前、后测的控制组，同样只有后测的实验组成绩也显著高于只有后测的控制组，表明具有显著的实验

处理效应。但是，具有前测的实验组和控制组的后测成绩分别显著高于只具有后测的实验组和控制组，表明前测对后测结果的影响。

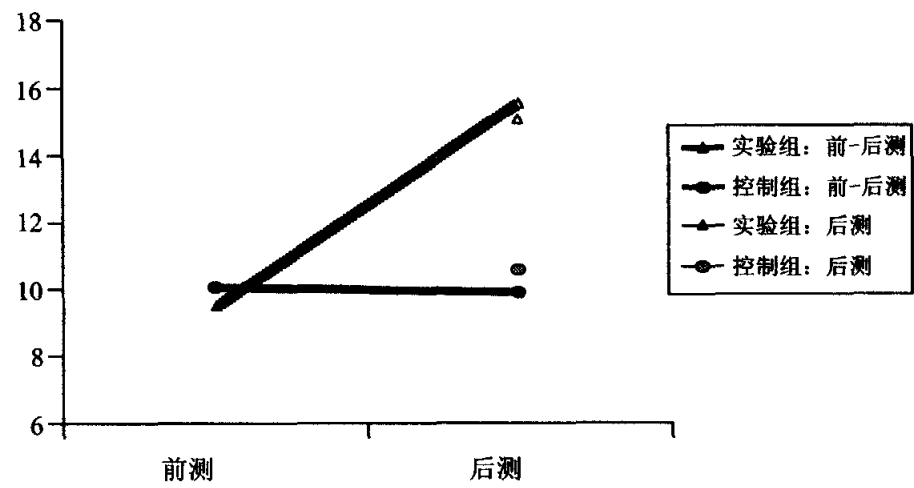


图 5.1 有处理效应无前测影响

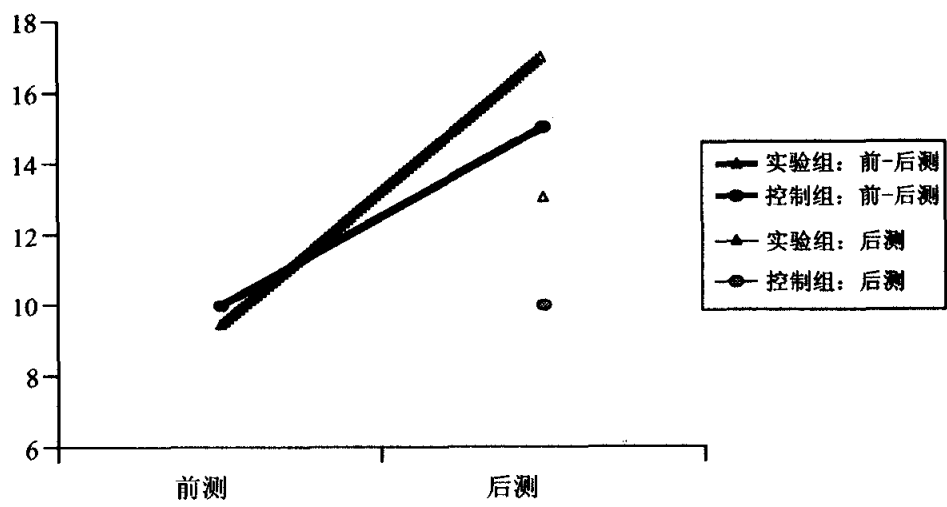


图 5.2 处理效应与前测效应均存

(二) 所罗门四组设计的数据分析

所罗门四组设计的数据分析根据不同的情况，可以采用不同的数据分析方法（张力为, 2002）。

第一，如果对于前测的影响，或者前测和实验处理的交互作用允许忽略不计，则可以使用单因素方差分析（ANOVA）对四个组的后测平均数 O_2, O_4, O_5 和 O_6 ）进行比较和检验。

第二，如果不能够确信是否可以忽略前测效应，则可以把前测成绩作为协变

量，采用协方差分析(ANCOVA)来比较 O_2 和 O_4 。采用 F 或 t 检验来比较 O_5 和 O_6 。如果单因素的协方差分析和 t 检验均达到了统计显著性水平，则可以得出实验处理的效应。否则，应当考虑前测以及前测与实验处理的交互作用的影响。

第三 采用 2×2 方差分析的方法，来分析实验处理的主效应和前测效应，以及实验处理和前测的交互作用。其分析的基本模式如表所示。

表 5-1 2×2 方差分析基本模式表

	无实验处理	有实验处理
有前测	O_4	O_2
无前测	O_6	O_5

其中，纵列的平均数可以估计实验处理的主效应，横行的平均数可以估计前测的主要效应，根据各个交叉格中平均数（ O_2, O_4, O_5 和 O_6 ）可以估计前测与实验处理的交互作用是否显著。

（三）所罗门四组设计的评价

所罗门四组设计除了具有前两种实验设计的优点外，还具有以下的优点，第一，所罗门四组设计实际上是进行了四个实验，从而可以通过检验 $O_2 > O_1, O_2 > O_4, O_5 > O_6$ 以及 $O_5 > O_3$ 来验证实验处理 X 的效果。如果比较的结果确实发生上述情况，那么就有充分的理由推断，结果的差异是由实验处理造成的。此外，所罗门四组设计还能够考察测验、历史和成熟等因素对因变量的影响。

所罗门四组设计是心理和行为科学研究中一种理想的研究设计。此种设计在内部效度和外在效度方面均无缺点而言。但是，在研究过程中很难同时找到四组同质的被试。这也是所罗门四组设计应用的局限所在。因此，在研究的初级阶段一般不宜采用这种研究设计，除非就实验假设作决定性检验的时候才考虑加以使用。

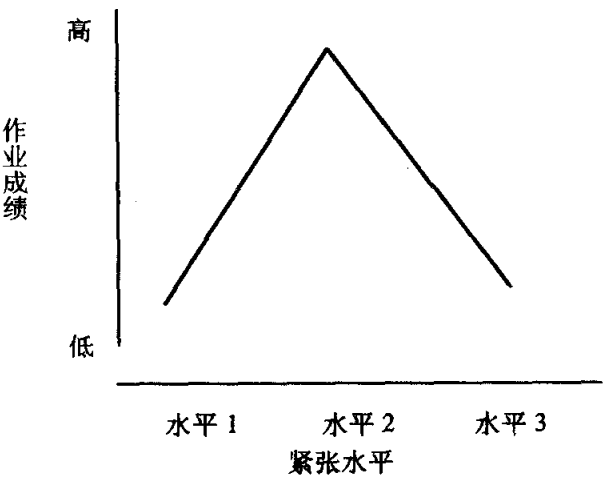
补充讨论 5-1 自变量表该设置几个水平？

在最简单的实验设计中，通常只有一个自变量，并且这个自变量也只有两个水平。但是，这种两个水平的自变量能够提供足够的证据来反映自变量和因变量之间的关系吗？

心理学研究中，自变量与因变量之间的关系并不都是直线关系，例如，著名的耶基斯-多德森(Yerkes & Dodson)的倒“U”曲线，表明紧张和作业成绩之间的曲线关系。下图是一个假想的自变量与因变量之间的关系。如果自变量只设置两个水平，例如水平 1 和水平 2，那么就会得出随着自变量的增加，因变量也

会增加的结论；同样，如果只选择了水平 2 和水平 3，就会得出随着自变量的增加因变量而减少的结论。如果选择了水平 1 和水平 3，可能的结果就是自变量的变化并不影响因变量。

那么，根据什么来确定自变量的水平呢？一般可借鉴前人的研究，如果预期自变量和因变量之间是线性关系，那么选择 3 - 5 个水平就可以了；如果两者之间的关系比较复杂，则至少选择 5 个水平。同时，自变量的变化范围不宜过小，可以通过相关的研究来确定合适的大小。至于自变量各个水平的分布，则可以根据实际情况而定。如果自变量和因变量的关系近似线形关系，则各水平平均分布即可；如果自变量与因变量的关系符合对数关系，则各水平间的间距应按对数变化。



第二节 多因素完全随机设计

我们的心理和行为通常都不是由单一原因引起的，而是由多个原因的复合效应而引起的。例如，对道路交通标志的反应不仅取决于标志本身的亮度，而且还与你对标志意义的理解（不同国家的交通标志是不同的），你所处于的意识状态（疲劳或饮酒），视觉敏度，驾驶速度等因素有关。因此，前述只有一个自变量的实验设计便不能满足心理学研究的需要。于是，研究者采用了复杂的多因素实验设计。本节和第四节将分别介绍两种重要的多因素实验设计：多因素完全随机设计和多因素随机区组设计。

在介绍有关的多因素设计之前，我们先对有关的概念进行简单了解。

1 因素与因素设计

所谓因素（factor）就是指实验中的自变量。实验中只有一个自变量的称为

单因素实验，有一个以上自变量的就称为多因素实验。研究者通过对因素的操纵来考察其对因变量的影响。其中，因素的类别称为水平（level），因素的水平可以是定量的如“年龄”、“声音的强度”等，也可以是定性的如“性别”、“人格类型”等。

因素多于一个的实验设计称为因素设计（factorial design），并根据各个因素的所有可能组合来设计组别，并同时考察所有因素对因变量的影响。

2. 主效应与交互作用

实验设计中，每个因素对因变量的单独效应称为主效应（main effect）。在多因素实验中，估价一个因素的主效应时，应该忽略实验中其他因素的不同水平的差异。

在一个多因素的实验中，研究者常常需要估价因素的不同水平之间的复杂的变化关系。当一个因素的水平在另一个因素的不同水平上变化趋势不一致时，称两个因素之间存在交互作用（interaction effect）（舒华,1994）。

一、多因素完全随机实验设计的基本思路

所谓多因素完全随机实验设计，是指研究者在同一个实验里同时操纵两个或两个以上自变量，并把被试完全随机地分配到各个处理的组合中，以观察自变量以及自变量之间交互作用效果的实验设计。包含两个以上自变量的因素设计一般称为多因素设计（multiple factorial design）或析因设计。由于多因素的实验设计可以对两个或两个以上自变量之间的交互作用进行分析，因而可以获得比单因素实验更多的信息。

在因素设计中，通常采用英文大写字母表示因素（自变量），用与大写字母相对应的小写字母及下标代表因素（自变量）的水平，各个因素的不同水平之间的相互结合和相互作用常用乘号（ \times ）表示。例如，在两个因素且每个因素各有两个水平的因素设计中，以 A 和 B 表示两个因素， a_1 、 a_2 和 b_1 、 b_2 分别表示因素 A 和 B 的两个水平，共组成 a_1b_1 、 a_1b_2 、 a_2b_1 和 a_2b_2 四种实验条件。此外，描述因素设计的一般形式为：（第一个自变量水平的数目） \times （第二个自变量水平的数目） \times （第三个自变量水平的数目） $\times \cdots \times$ （第 n 个自变量水平的数目）。仍以上面的两个因素并且每个因素各有两个水平的设计为例，可以将该实验称为双因素实验设计，也可称为 $A \times B$ 因素设计，还可称为 2×2 因素设计。如果实验有 A 、 B 、 C 三个因素，每个因素分别有 3、4、5 个水平，则该实验可称为三因素实验设计，也可称为 $A \times B \times C$ 因素设计，还可称为 $3 \times 4 \times 5$ 因素设计。其中最简单的因素设计是 2×2 因素设计，即有两个自变量，每个自变量都有两个水平。

在多因素实验设计中，实验处理的个数就是各个因素水平数的乘积。例如在一个二因素的实验设计中， A 因素有 p 个水平， B 因素有 q 个水平，一共就有

$p \times q$ 个实验处理。从理论上讲，一个多因素实验中因素的数目及其每个因素的水平数可以是任意多，但是这样会造成过多的人力物力投入，而且会造成对实验结果的解释尤其是交互作用解释的难度增大。因此研究者一般将采用因素的数目限于 2~3 个。

因此，本节以两个自变量的因素设计为例，介绍多因素完全随机实验设计，至于更多因素的多因素实验设计，其基本思路与两因素的设计基本相同。

二、 2×2 因素设计示例

(一) 2×2 因素设计的基本模式

2×2 因素设计是多因素设计中最基本，也是最简单的一种设计类型。该设计包含两个自变量，每个自变量各有两个水平，共组成 4 种实验条件。其基本模式如下图所示：

		Xb		
		Xb_1	Xb_2	
Xa	Xa_1	$O_1 : O_{11}, O_{12} \cdots O_{1n}$	$O_2 : O_{21}, O_{22} \cdots O_{2n}$	Oa_1
	Xa_2	$O_3 : O_{31}, O_{32} \cdots O_{3n}$	$O_4 : O_{41}, O_{42} \cdots O_{4n}$	Oa_2
		Ob_1	Ob_2	

图 5.3 2×2 因素设计示意图

上图中 Xa_1 、 Xa_2 和 Xb_1 、 Xb_2 分别代表 A 、 B 两个因素的两个水平。 O_1 、 O_2 、 O_3 和 O_4 分别表示各组内被试成绩的平均值， Oa_1 和 Oa_2 分别表示在 Xa_1 和 Xa_2 水平上的被试成绩的平均值，同样 Ob_1 和 Ob_2 分别表示在 Xb_1 和 Xb_2 水平上的被试成绩的平均值。其中， Oa_1 、 Oa_2 和 Ob_1 、 Ob_2 被称作边缘值 (marginal mean)。

通过 Oa_1 与 Oa_2 的差异比较，可以确定 A 因素的主效应；同样，通过 Ob_1 和 Ob_2 的比较可以确定 B 因素的主效应。此外，通过 O_1 、 O_2 、 O_3 和 O_4 之间的比较，可以确定 A 和 B 因素的交互作用，即在 Xa_1 条件下比较 O_1 和 O_2 之间的差异， Xa_2 条件下比较 O_3 和 O_4 之间的差异；以及 Xb_1 条件下 O_1 和 O_3 之间差异， Xb_2 条件下 O_2 和 O_4 之间的差异。

(二) 2×2 因素设计的数据分析

2×2 因素设计的数据分析，主要采用两因素的方差分析，分别考察 A 因素和 B 因素的主效应，以及 A 和 B 两因素的交互作用。此外，还要根据 A 和 B 两因素的交互作用是否达到统计显著性水平，来决定是否进行简单效应分析。

(三) 2×2 因素设计的图示分析

假如有研究希望考察两种治疗恐怖症方法（认知疗法和行为疗法）的效果

是否与每周治疗的时间长短有关。在这个实验中，治疗方法（因素 A）分为两个水平：认知疗法和行为疗法；每周的治疗时间（因素 B）分两个水平：1 小时和 4 小时。经过一段时间的治疗后，治疗的效果以临床心理学家的评估为依据。下面我们分别讨论可能出现的结果（如图 5.4 所示）。

(1) “虚无”结果

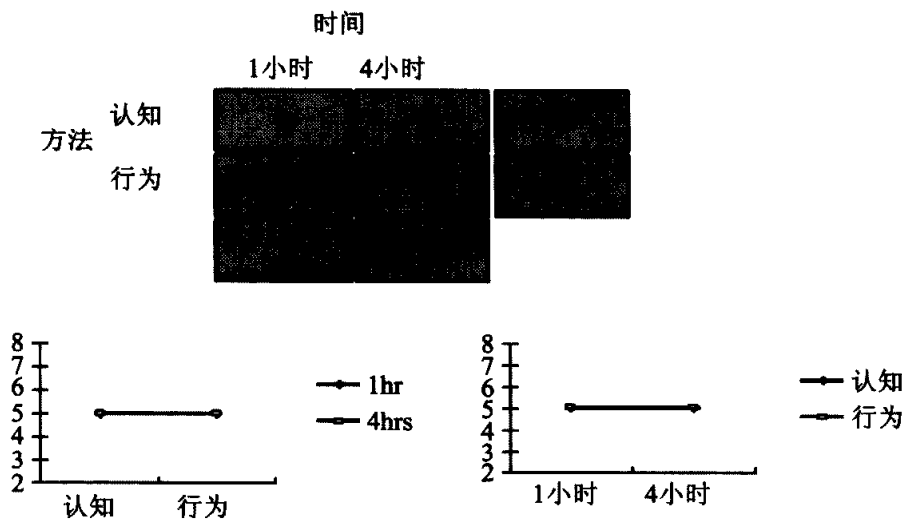


图 5.4 “虚无”结果

所谓“虚无”结果 (null outcome)是指所有的实验处理都对因变量没有效应。从上图中可以看到，所有四个组的平均成绩都是 5，所有行和列的平均值也都是 5，并且图中的线相互重叠。因此既不存在显著的主效应，也不存在显著的交互作用。对于恐怖症的治疗效果，上述假想结果表明认知治疗和行为治疗的效果是一样的，每周治疗 1 小时和 4 小时的治疗效果也是一样的，并且两者之间不存在交互作用。

(2) 因素主效应

在没有交互作用的情况下，因素主效应可以表现为是只有一个因素的主效应显著和两个因素的主效应都显著。首先，我们来看只有一个因素的主效应显著的情况。

图 5.5 中可以看到，认知和行为治疗在所有治疗时间的水平上的平均值都是 6，而治疗时间在所有治疗方法上的平均值表现出差异，每周 1 小时的平均值是 5 每周 4 小时的平均值为 7。如果这种差异达到显著性水平，表明存在显著的治疗时间的主效应，说明不管采取认知治疗还是行为治疗，每周 4 小时治疗效果显著优于每周 1 小时的治疗效果。

同理 图 5.6 清楚的显示了治疗方法因素的主效应，说明不管每周治疗时间

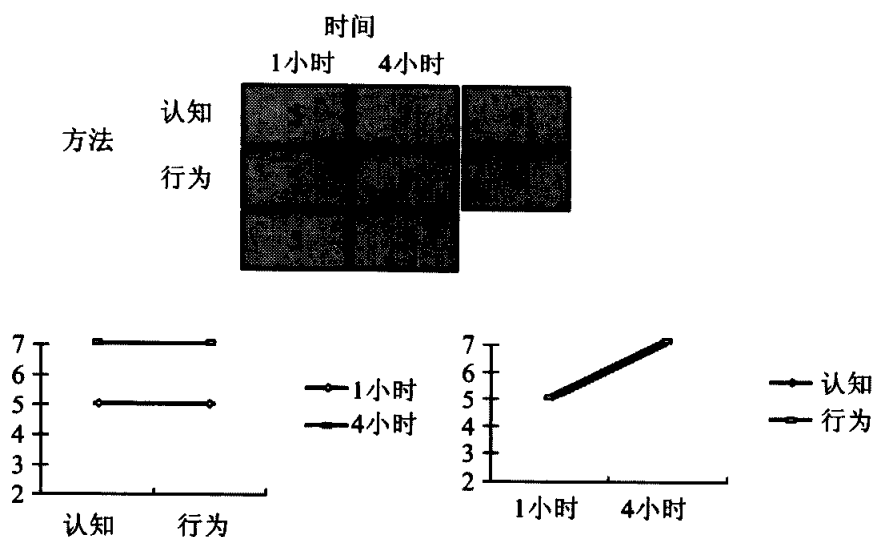


图 5.5 治疗时间因素的主效应

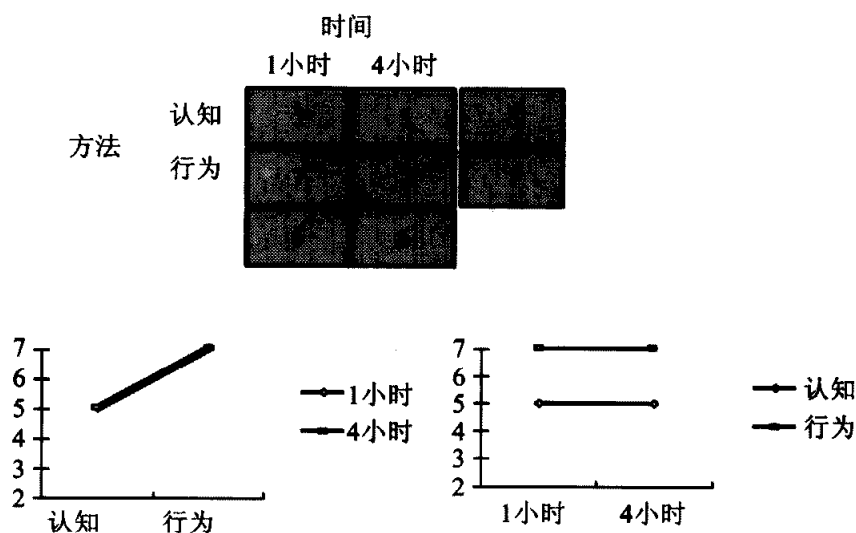


图 5.6 治疗方法因素的主效应

的长短，行为治疗的效果显著优于精神分析治疗的效果。

图 5.7 显示，在治疗方法因素中，行为治疗的效果显著优于认知治疗。在治疗时间因素中，每周 4 小时的治疗效果显著优于每周 1 小时的治疗效果。并且，治疗方法的效果并没有随着治疗时间的变化而变化 ($7 - 5 = 2, 9 - 7 = 2$)，同样，治疗时间的效果也没有随着治疗方法的变化而变化 ($7 - 5 = 2, 9 - 7 = 2$)。这种情况说明在治疗方法和治疗时间之间没有交互作用发生。途中的两条直线平行或接近平行，也说明这两个因素之间没有交互作用。

(3) 两因素间交互作用

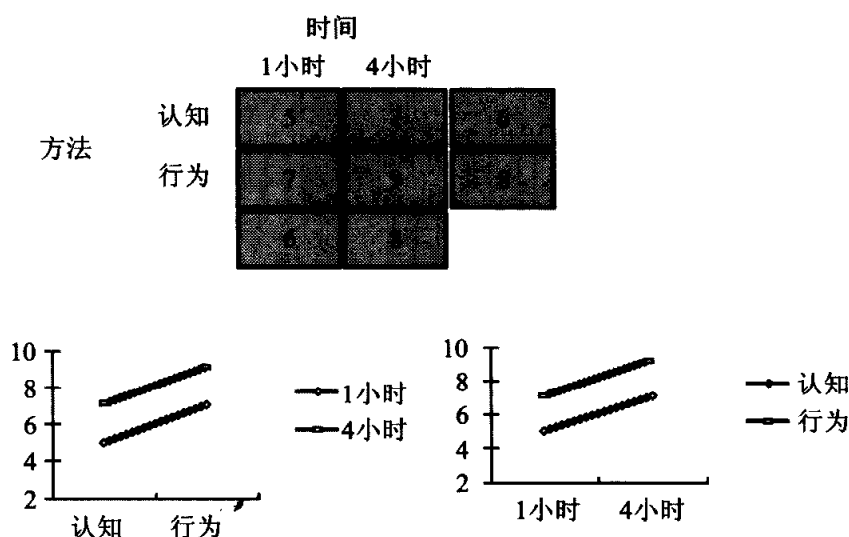


图 5.7 治疗时间和治疗方法的主效应

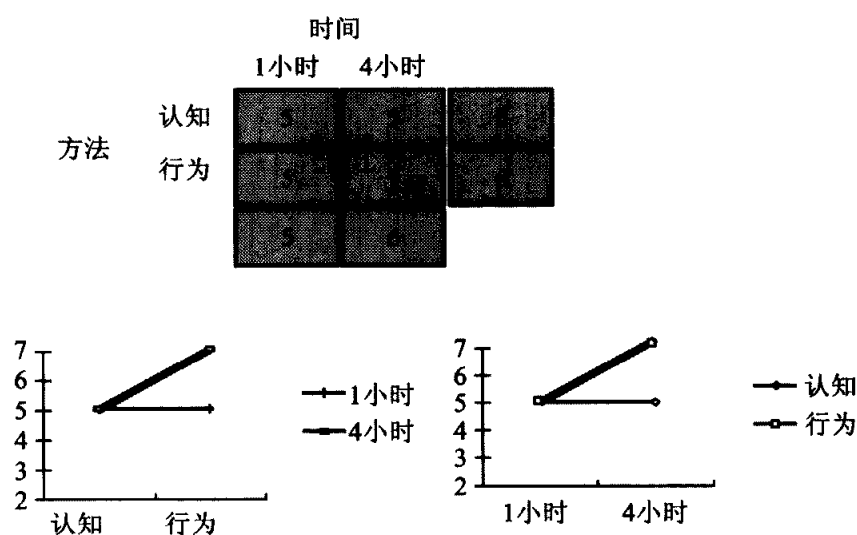


图 5.8 治疗时间和方法的有序交互作用

因素间的交互作用表明因素某个水平上的效果会随着另一个因素水平的不同而不同。因素之间的交互作用可以分为有序交互作用 (ordinal interaction) 和无序交互作用 (disordinal interaction)。有序交互作用指的是，虽然因素水平间的效果会因其他因素的不同而不同，但各个水平之间的相对位置保持不变。而无序交互作用则不存在这种关系。图 5.8 所示的是一种有序交互作用的情景，认知治疗的效果在治疗时间上没有存在差异，但是行为治疗每周 4 小时的效果优于每周 1 小时的效果；同样，每周 1 小时的效果在认知和行为治疗上不存在显著的差异，但在每周 4 小时的行为治疗效果明显优于认知治疗效果。图 5.9 所示

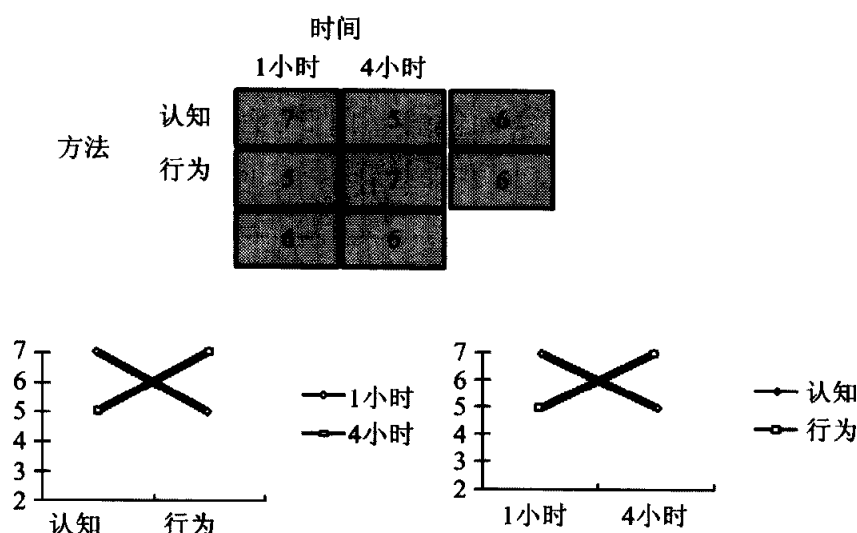


图 5.9 治疗时间和方法的无序交互作用

的是一种无序交互作用的情景，认知治疗每周 1 小时的效果明显优于每周 4 小时的效果，但是行为治疗的效果却是每周 4 小时的明显优于每周 1 小时；同样，每周 1 小时的认知治疗效果明显优于行为治疗效果，而每周 4 小时行为治疗效果明显优于认知治疗。

另外从图形上看，有序交互作用可以用两条虽不平行、但是不交叉的直线（折线）来表示，如图 5.8 所示。无序交互作用是两条不平行且交叉的直线（折线）来表示，如图 5.9 所示。

三、多因素完全随机设计的评价

多因素完全随机实验设计是单因素实验设计的扩展。单因素设计中每种被试的分配方法，以及无关变量的控制程序，同样都可以应用在多因素设计中。因此，多因素设计不仅具有单因素实验设计的优点，而且还具有单因素实验设计不具备的其他优势。这主要表现在：第一，可以同时获取两个或多个自变量对因变量的影响，因而具有节省人力、物力和时间的优点。例如可以在原有单因素设计的基础上，再增加一个或几个变量，这样在不增加被试的情况下，就可以获得两个或更多的自变量效果的信息及其间复杂关系的信息；第二，可以探讨不同自变量间的交互作用。在复杂环境中，某一心理和行为现象产生的原因是多方面的，并且这些原因相互交织以复杂的形式表现出来。因此，多因素设计可以使研究者分析各个自变量及其交互作用引起因变量变化的信息，从而更加准确的分析影响因变量的各种自变量及其相互关系的作用。

多因素设计的局限性在于：这种设计在各个实验处理的组合、被试分配以及

统计分析上，都是比较复杂的。对于三个以上因素的实验设计，实验结果的统计分析是比较困难的。特别是多个因素间的交互作用如果达到统计显著性水平，对交互作用的解释就变得相当复杂和困难。因此，研究者通常除了采用单一因素设计外，最经常采用的是随机二因素设计，进行二向方差分析。

实例学习 5-2 多因素完全随机化设计实例分析

为了研究专业知识对记忆的重要作用，Chi(1978)进行了如下的实验研究。在研究中，要求儿童和成人分别对所呈现的数字和棋子进行回忆。棋子任务要求被试看一些在正常下棋中常出现的棋子；数字任务是要求被试进行标准数字测验。其中，儿童是10岁的专业象棋棋手，成人是象棋初学者。因此，这是一个 2×2 因素设计。年龄因素包括成人和儿童两个水平；回忆任务包括棋子和数字回忆两个水平。

研究的实验设计如下所示：

		年龄	
		成人	儿童
回忆任务	棋子	成人/棋子 回忆成绩	儿童/棋子 回忆成绩
	数字	成人/数字 回忆成绩	儿童/数字 回忆成绩

实验结果见图 5.10。统计分析结果表明，象棋知识有助于棋子信息的记忆，但对数字的回忆没有显著的影响。初学象棋的成人虽然数字的回忆成绩好于儿童，但回忆棋子的数量却少于儿童。图 5.10 清楚的显示了这种关系。

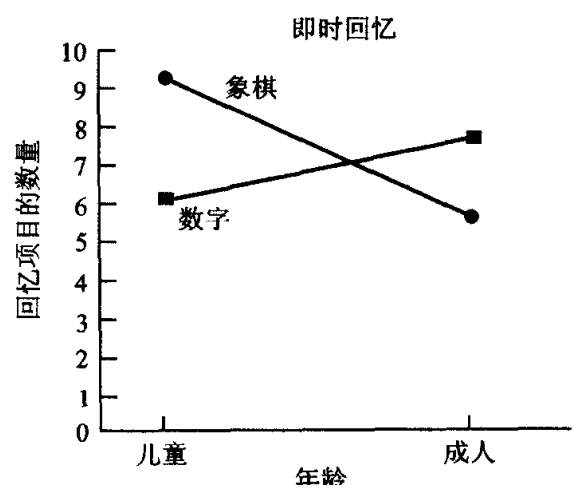


图 5.10 儿童和成人对数字和棋子的回忆成绩

第三节 单因素随机区组设计

在进行心理和行为研究时，接受实验处理的被试可能是一个人、一只大白鼠或一只鸽子等。在实验前，没有经过选择的被试在反应时间、记忆力、学习能力、问题解决的能力等各个方面都是存在差异的。因此对于所观测到的因变量的变化，除了所施加的实验条件的影响外，可能存在的个体差异也可能存在显著的影响。

这部分我们要讨论的实验设计称为随机区组实验设计（**randomized block design**）。“区组”的概念最早来自于农业实验。早期对不同品种的农作物进行试验时，即使在同一种植区内，土质、水分等土壤因素也有所差异。因此，为了保证各种品种的作物种植受土壤因素的影响相同，就把种植区按照土壤因素划分成一些小的“区域”，每块区育种的土壤因素基本相同。然后再把每一“区域”分成小区，每个小区种植一个品种，来比较在同一“区域”下农作物品种的差异，类似这样的“区域”，就称为一个区组。以后，区组的概念就应用到农业以外的其他研究领域。

一、随机区组设计的基本原理

随机区组设计的基本原理，就是把实验单位划分为若干区组。一般来讲，区组是根据实验的要求来划分的，对于那些可能影响因变量，但又不是研究者所关注的变量都可以作为区组变量来考虑，以控制其对因变量的影响。个体差异（智力、情绪、人格等）和环境变量（时间、地点、仪器等）都可能是需要考虑的区组变量。实际上，一个随机区组设计类似于一个双因素的实验设计。但是两者之间又有明显的区别。例如，在一个两因素的实验设计中，可以把被试完全随机的分配给两个因素的各个水平的组合，而在随机区组实验中，区组因素的各个水平通常表示受试者的个体差异水平，而另一个因素的各个水平则表示各种处理的差异。因此，这时就不可能把被试完全随机的分配到这两个因素的组合中。被试只能在区组因素的某个水平内随机的分配到各个处理的水平上。

随机区组设计的目的在于使区组内的被试差异（或环境因素）尽量减小，而对区组之间的差异则根据设计要求设定。每种实验处理出现在每个区组中，这时，区组之间的差异并不影响各实验处理平均数之间的差异，区组之间的差异可以从误差项中剔除。实际上，研究者是将区组视为无关变量而加以控制，以便更精确地考察实验处理效应。如果区组之间没有差异，则随机区组设计无助于分辨实验处理之间的差异，这种情况下，无需采用随机化区组设计。

二、随机区组设计的被试分配

随机化区组设计的原则是同一区组内的被试尽量同质（朱滢,2000）。每一区组内被试的人数分配有 3 种情况：

第一，一名被试作为一个区组。每名被试（区组）均接受全部实验处理，在接受实验处理的顺序上需采用随机方法。

第二，每个区组内被试的人数是实验处理数目的整倍数。例如，在不同声音刺激对学生引体向上成绩影响的例子中，实验处理分为 A、B、C 和 D 共 4 种，每个区组内被试的人数可根据需要和条件确定为 4 的整倍数，即 4、8、12 等等。如果选取的被试人数为 12 人，则可将这 12 名被试分为 3 个小组（区组）每组内有 4 人，区组内不同的被试接受不同的实验处理。再如，选取 24 名被试，也分为 3 个小组，每个区组内有 8 名被试，每 2 人接受同样的实验处理。

第三，区组内的基本单元不是一名或几名被试，而是以一个团体为单元。例如，以某学校同一年级的不同的班作为不同的区组，每个班都接受所有实验处理。或者，以不同学校为实验对象（表示不同的区组），同一学校的几个班（班级个数等于实验处理个数）成为一个区组，每个班随机接受一种实验处理。

总之，每一区组应该接受全部实验处理，每一种实验处理在不同的区组中重复的次数也应完全相同。

三、单因素随机区组设计的基本模型

单因素随机区组设计的研究中，只有一个自变量，自变量有两个或多个水平，研究中还有一个无关变量，也有两个或多个水平，并且自变量的水平和无关变量的水平之间没有交互作用。如果无关变量是被试自变量时，首先将被试在这个无关变量上进行匹配，然后随机分配到不同的实验条件中。这样，区组内的被试在此无关变量上更加同质，被试接受不同的实验处理时，可看作不受无关变量的影响，主要受处理的影响，而区组之间的变异反映了无关变量的影响。可以表示为如下基本模式：

$$\begin{array}{ccccc} & & R & X & O_1 \\ & & & & \\ O & M & & & \\ & & R & & O_2 \end{array}$$

其中 O 表示对全体被试进行前测，并根据前测的成绩对全体被试进行匹配 M ，并将匹配的被试随机分配，形成不同的区组。 O_1 和 O_2 表示实验处理后的后测成绩。

此外，环境因素也是可以考虑的区组变量，例如每天的时间、每年的季节、地

点、仪器等方面的因素也可以形成区组，以减少误差的影响，事件是一个特别有效的区组变量，因为它常常还会带来一些附加的变量，如身体的生理周期、疲劳等等。

单因素随机区组设计与多组随机后测设计十分相似，不同的是单因素随机区组设计要求将被试分为不同的区组，每一区组接受所有的实验处理，而每名被试只随机接受一种水平的实验处理。然后，观察实验处理后每个区组对不同实验处理的反应，并将这种反应作为后测成绩。单因素随机区组设计的基本模式见下表(表 5-2)。

表 5-2 单因素随机区组设计基本模式表

区 组	实验处理					区组 平均
	X_1	X_2	X_3	...	X_n	
1	O_{11}	O_{12}	O_{13}	...	O_{1n}	O_1
2	O_{21}	O_{22}	O_{23}	...	O_{2n}	O_2
3	O_{31}	O_{32}	O_{33}	...	O_{3n}	O_3
...
m	O_{m1}	O_{m2}	O_{m3}	...	O_{mn}	O_m
实验处理平均	$O_{.1}$	$O_{.2}$	$O_{.3}$...	$O_{.n}$	$O_{..}$

上述模式表示 $X_1, X_2, X_3, \dots, X_n$ 表示自变量的 n 个水平，即 n 个实验处理；区组的个数为 m ； $O_{i1}, O_{i2}, O_{i3}, \dots, O_{in}$ 表示第 i ($i = 1, 2, \dots, m$) 个区组在 n 个实验处理中的观测值； $O_1, O_2, O_3, \dots, O_m$ 表示 m 个区组的平均值； $O_{.1}, O_{.2}, O_{.3}, \dots, O_{.n}$ 表示 n 实验处理中观测的平均值。

例如，在研究一种新的英语单词记忆方法的有效性时，由于学生原有的英语水平会影响到单词的记忆效果。因此，在随机选取的 60 名学生中，研究者通过前测，根据学生已有的英语水平、英语单词量等基本条件把 60 名学生分为 3 类：优、中和差。这样被试分成 3 个区组，每个区组各 20 人。然后在每个区组中随机选取 10 个进行新记忆方法的训练，而剩下的 10 个进行传统方法的学习。最后，根据训练后的测验成绩的统计分析，来推论新的单词记忆方法的有效性。

四、单因素随机区组设计的数据分析

单因素随机区组设计的数据分析可通过多因素方差分析进行处理，也就是实验处理和区组分别作为因素来考虑。实验处理效应是研究者关注的重点，区组因素则作为无关变量加以控制。虽然随机区组设计中假设区组因素和实验因素之间不存在交互作用，但是研究者可以尝试分析两者之间的交互作用。如果

实验因素和区组因素的交互作用达到统计显著性水平，研究者应该进一步考虑修改自己的研究假设，把区组因素作为一个影响因素，而不是无关变量来考虑。

第四节 多因素随机区组设计

一、随机化区组设计的基本模式

多因素随机区组设计与单因素随机区组设计在理论上基本相同。只是单因素随机区组设计中的每个区组接受一个因素各个水平的实验处理，而多因素随机区组设计中的每个区组则接受两个以上因素的各个水平组合的实验处理，见下表(表 5-3)

表 5-3 2×3 随机区组设计基本模式

区 组	实验处理的组合					
	Xa_1b_1	Xa_1b_2	Xa_1b_3	Xa_2b_1	Xa_2b_2	Xa_2b_3
1	O_{11}	O_{12}	O_{13}	O_{14}	O_{15}	O_{16}
2	O_{21}	O_{22}	O_{23}	O_{24}	O_{25}	O_{26}
...
n	On_1	On_2	On_3	On_4	On_5	On_6

在该设计中，因素 A 有两个水平，即 a_1 和 a_2 因素 B 有 3 个水平，即 b_1 、 b_2 和 b_3 构成了 2×3 共 6 种实验处理。 O_{i_1} 、 $O_{i_2} \cdots O_{i_6}$ 表示第 i ($i = 1, 2, \cdots n$) 个区组在 6 种实验处理中的观测值， On_1 、 $On_2 \cdots On_6$ 表示第 n 个区组在 6 种实验条件下的观测值。

例如，有研究者想要研究在红、黄、绿三种灯光下的刹车反应时有无差异，同时又把灯光的强弱也作为一个影响因素。因此在该实验中，包括两个自变量，即灯光的颜色，分红、黄、绿三个水平和灯光的强度，分强、弱两个水平，共构成六种实验条件。现在随机选取 24 名受试者，考虑到个体差异对反应时的影响，可把每个人的反应作为一个区组，分别接受六种条件的实验处理，其中每个被试先接受哪一种实验处理可采用随机的方法或拉丁方的方法进行排列。最后通过对不同条件下反应时的差异来探讨灯光的颜色和强度以及两者的交互作用对反应时的影响。

二、多因素随机区组设计的数据分析

多因素随机区组设计的数据也可通过多因素方差分析进行处理，即将实验

处理 *A* 作为 *A* 因素 将实验处理 *B* 作为 *B* 因素 将区组作为 *C* 因素。实验处理 *A* 和实验处理 *B* 的主效应及其交互作用是研究者关注的中心，区组因素则作为无关变量加以控制。与单因素随机区组设计一样，多因素随机区组设计假设区组因素和实验处理之间不存在交互作用。但是研究者可以尝试分析区组因素和实验处理之间的交互作用，如果达到统计的显著性水平，就可以进一步修改原有的理论假设，把区组因素作为一个实验因素加以考虑，以提高实验研究的外部效度。

三、随机区组设计的评价

随机区组设计与完全随机设计相比，其主要优点是，第一，可以控制个体差异（或环境因素）对实验结果的影响，区组设计是将这种个体差异（或环境因素）作为无关变量，根据个体差异（或环境因素）情况将被试（或环境因素）分成几个不同的区组，并在统计分析中将这种个体差异（或环境因素）的影响从组内误差中分离出来，从而更加有效地反映实验处理的作用；第二，随机区组实验设计可用于具有任何处理水平数量的实验中，且区组的数量也不受限制，因而有较大的灵活性。

随机区组设计的局限是：（1）如果实验设计含有多个处理水平，可能会在寻找同质被试以形成同质组时遇到困难；（2）随机区组设计比完全随机设计有更多的限定，例如，随机区组设计的前提是，自变量与无关变量之间没有交互作用。如果存在自变量与无关变量之间的交互作用，则不应使用随机区组设计（舒华，1994）。

►本章提要

1. 实验研究是心理学研究中探寻因果关系最有效的方法。
2. 根据随机化原则选取并分配被试，有效地控制无关因素的干扰，操纵自变量的变化，并精确地测量因变量的实验设计称为真实验设计。
3. 单因素完全随机设计是只有一个自变量（因素）并采用随机化的方法分配被试到各个实验处理的实验设计。
4. 根据自变量水平的多少，单因素完全随机化设计可分为两等组模型和多等组模型；根据有无实验前测，可分为后测模型和前测后测模型；根据是否进行配对分组，可分为随机等组模型和随机配对等组模型。
5. 多因素完全随机实验设计，是指研究者在同一个实验里同时操纵两个或两个以上自变量，并把被试完全随机分配到各个处理的组合中，以观察自变量以及自变量之间交互作用效果的实验设计。包含两个以上自变量的多因素设计一般称为多因素设计或析因设计。由于多因素的实验设计可以对两个或两个以上

自变量之间的交互作用进行分析，因而可以获得比单因素实验更多的信息。

6. 随机区组设计的目的在于使区组内的被试差异（或环境因素）尽量减小，而使区组之间的差异尽量大。每种实验处理出现在每个区组中，这样区组之间的差异并不影响各实验处理之间的差异，区组之间的差异可以从误差项中剔除。

7. 单因素随机区组设计要求把被试分为不同的区组，每个区组接受一个因素各个水平的实验处理。并且假设区组变量和实验变量之间不存在交互作用。

8. 多因素随机区组设计要求把被试分为不同的区组，每个区组则接受两个以上因素的各个水平组合的实验处理。并且假设区组变量和各个实验变量之间不存在交互作用。

►本章关键术语

实验设计(experimental design)

真实验设计(true experimental design)

实验组控制组后测设计(experimental/control group posttest design)

实验组控制组前测后测设计(experimental/control group pre-posttest design)

所罗门四组设计(Solomon four group design)

因素设计(factorial design)

主效应(main effect)

交互作用(interaction effect)

多因素完全随机实验设计(completely randomized multiple factorial design)

完全随机化区组设计(randomized blocked factorial design)

►复习与练习

1. 什么是实验设计，什么是好的实验设计？
2. 单因素完全随机设计包括哪些设计类型，其各自的优点和局限性是什么？
3. 为什么在心理学研究中要尽可能的采用多因素实验设计？
4. 完全随机化设计和随机区组设计的区别是什么？
5. 试用多因素随机区组设计来设计一个实验。

►推荐参考读物

黄希庭. 心理学实验指导. 北京：人民教育出版社，1987

舒华. 心理与教育研究中的多因素实验设计. 北京：北京师范大学出版社
1994

张力为. 体育科学研究方法. 北京：高等教育出版社 2002

朱滢. 实验心理学. 北京：北京大学出版社 2000

Martin, D. W. Doing psychology experiments. Monterey, CA: Brooks/Cole, 1991

Solo, L. & MacLin Experimental Psychology : A Case Approach. Boston: Person Education Company, 2002

➤ 在线资源

<http://www.socialresearchmethods.net/kb/>

第六章 准实验设计

在心理学研究中，有时无法运用随机化原则和方法来分配被试。例如想知道一种新的思维训练方法是否可以提高学生的数学成绩，由于实验前班级的学生是固定的，不可能采用随机化的方法形成实验组和控制组，这时准实验设计便是一种可行的办法。准实验设计（**quasi-experimental design**）是介于真实验设计和非实验设计之间的一种实验设计。准实验设计的概念是由心理学家坎贝尔和斯坦利（Campbell&Stanley,1966）提出。这种研究设计适应了心理学理论和研究发展的实际需要，因而在心理与行为的研究中得到了广泛的应用，成为心理与行为研究的主要实验设计方案。准实验研究虽然不需要采用随机化的方法来分配被试，但是能够严格的操纵自变量和控制无关变量，适合更广泛的研究目的。在心理与行为的研究中，许多情况下都难以满足真实验设计的条件，坎贝尔和库克（Campbell&Cook）于1979年发表的《准实验研究：现场背景的设计和分析》（*Quasi-Experimentation: Design and Analysis Issues for Field Settings*）这一专著，为准实验设计提供了充实的理论基础，并系统、全面的提出了准实验设计原则和统计分析方法，使准实验设计的理论与方法日趋完善。

70年代以来，心理学研究方法论的重点出现了引人注目的变化。心理学的研究重点逐渐从实验室里的人为环境转向现场研究。也就是，心理学家日益重视心理学的研究对现实生活的意义，重视心理学研究的“生态效度”。这种研究方法论的变化，对实验设计提出了新的、更高的要求，从而促进了准实验设计理论与方法的发展与应用。准实验研究作为一种有效的研究手段，不仅适用于现场背景的研究，也可以在某些模拟的实验室中进行。因此，准实验设计正在成为心理学研究中最有前途和最有应用价值的实验设计。

在本章中，我们将讨论以下问题：

1. 实验设计与准实验设计的区别。
2. 各种准实验设计的模式、统计分析方法及其评价。

第一节 单组准实验设计

在心理与行为研究中，有时研究条件或问题的性质无法满足真实验设计所要求的控制组设置，这时就可以采用单组准实验设计的办法进行研究。本节主要介绍时间序列设计和相等时间样本设计这两种类型的单组准实验设计。

一、时间序列设计

(一) 时间序列设计的基本模式

时间序列设计(time-series design)是指对一组被试进行周期性的一系列观察，并在测量的这一时间系列呈现实验处理 (X)，然后观测呈现实验处理后的一系列观测结果，并将这些结果与实验处理前的一系列观测结果进行对比分析，从而推断实验处理是否产生效果。其基本实验模式如下：

$$O_1 \quad O_2 \quad O_3 \quad O_4 \quad X \quad O_5 \quad O_6 \quad O_7 \quad O_8$$

从该实验模式中可以看出，时间序列设计是只有一个实验组的单组设计。要求研究者分别在实验处理前后进行一系列连续的观测，最后通过对比前测和后测的差异来推断实验处理的效果。根据研究的内容和目的不同来确定实验处理前、后的观测次数，并且前、后的观测次数也可以不同。但总的来说，观测的次数越多越好。有学者 (Orwin,1997) 推荐在实验处理前的观测次数不少于 50 次。时间序列设计可能产生的部分结果如图 6.1 所示。

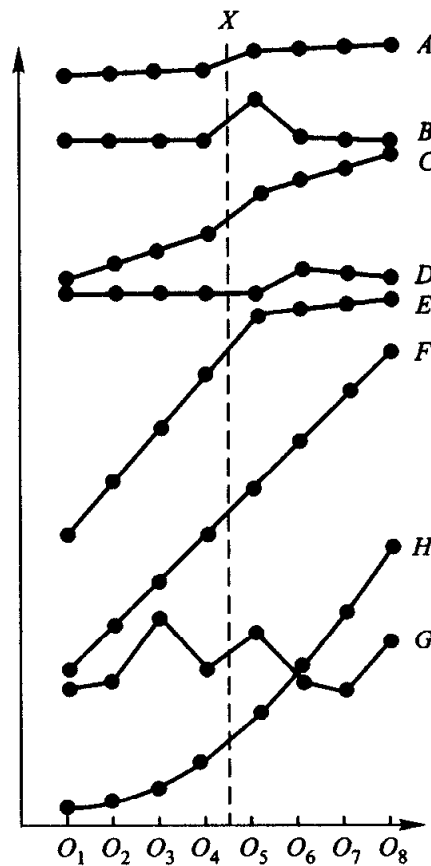


图 6.1 时间序列设计中可能出现的部分结果
(采自朱滢,2000)

其中 $O_1 \sim O_8$ 为观测成绩，竖虚线为实验处理的引入时间。从上图中可以看出 A 线和 B 线在实验处理 X 前后出现跳跃，可以推测被试在实验处理前后的观测成绩有所提高，其中 A 线的结果说明实验处理具有稳定的正效应，而 B 线的结果说明实验处理的作用是暂时的；从 $C、D、E$ 线上可以推测，实验处理前后的观测成绩可能存在差异；从 $F、G、H$ 线上可以推测，实验处理前后的观测成绩可能不存在显著的差异。虽然在从 $F、G、H$ 线看到，实验处理之前的 O_4 和实验处理之后的 O_5 之间的变化幅度很大，时间序列设计的结果分析，不能只看与实验处理邻近的前后两次观测成绩的差异，而是要从实验处理前后的整个发展变化趋势来评估实验处理的效果。

例如，坎贝尔 (Campbell) 于 1969 年采用时间序列设计研究了康涅狄格州的交通死亡人数与实施严惩制度的关系。从下图 (图 6.2) 中我们可以看到，在 1955—1956 年实施严惩制度后，比 1955 年以前的交通死亡人数下降了。而且继后的几年里 (1957, 1958, 1959)，交通死亡的人数仍在继续下降。

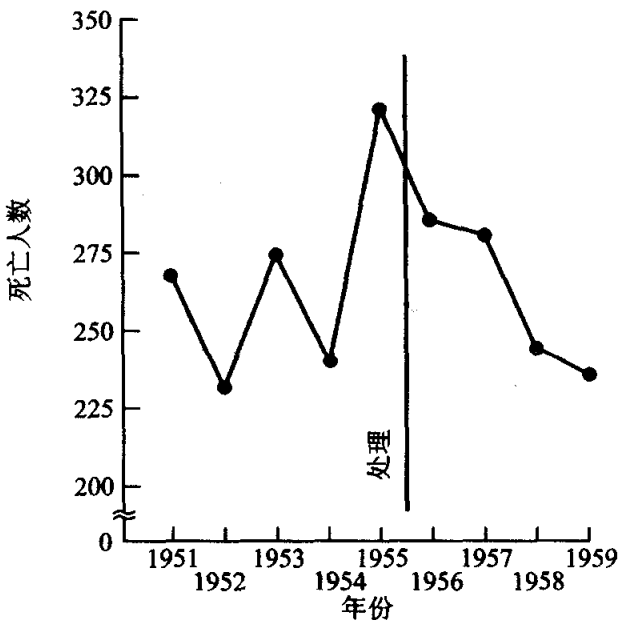


图 6.2 1951—1959 年康涅狄格州的交通死亡人数

(二) 时间序列设计的数据分析

对于时间序列设计结果的数据分析，需要对实验处理前后的一系列观测成绩进行检验和比较，来检验实验处理的效果。应当注意的是，不能仅仅使用引入实验处理的前一次和后一次的观测成绩 (如图 6.1 中的 O_4 和 O_5) 进行比较。

如果假设实验处理的效果是暂时性的 (如图 6.1 中的 B 线所示)，可根据 $O_1、O_2、O_3、O_4$ 的观测结果，采用直线回归或曲线回归的方法，把在没有引入实验处理

时的 O_5 可能的结果估计出来，然后通过相关样本 t 检验的方法检验估计值和实际观测值之间的差异显著性。如果假设实验处理的效果是稳定的或持久的（如图 6.1 中的 A 线所示），需要考察实验处理前的回归直线与实验处理后的回归直线的差异，也就是需要确定 $O_1 \sim O_4$ 各点的回归直线和 $O_5 \sim O_8$ 各点的回归直线的差异显著性，这种差异显著性检验主要涉及回归直线的截距和斜率的问题。在实际问题中，回归直线的截距和斜率的差异显著性检验可采用相关的统计方法。

此外，有研究者建议采用“自回归整合移动平均模型”（autoregressive integrated moving average models, ARIMA）及有关的技术进行时间序列分析，并做出模型，这种方法可以较准确的评估实验处理的效果。

（三）时间序列设计的评价

时间序列设计的优点主要体现在：第一，可以较好的控制无关变量对实验处理效果的影响。因为时间序列设计是在一定时间内经过一系列顺序的观测，能够提供对成熟的控制。由于观测的间隔时间基本相同，因此在每个间隔内成熟的发展基本上是相同的。此外，对于历史的影响，虽不能排除，但也能够减少到最低程度。第二，可以较好的控制测验因素的影响。由于每个被试的成绩都是经过反复测验而得到的一系列结果，这样就能够降低由于只做一次测验而出现的有偏样本成绩的概率。因此，时间序列设计通过实验处理前后的一系列观测，能够在较长的时间跨度内评价实验处理对观测变量的影响，并进一步检验自变量和因变量之间的因果关系。

时间序列设计的缺点是，由于这种设计是在没有设置控制组的情况下所进行的研究，因此不能排除与实验处理同时发生的偶发事件的影响，也不能排除那些与自变量同时出现的附加变量的影响。此外，该设计不能有效的控制测验与实验处理交互作用的影响。最后，由于实验过程中对被试的反复观测，容易使被试产生疲劳、厌烦而导致增加或降低被试对实验处理的敏感性，从而影响实验处理的效果。

二、相等时间样本设计

（一）相等时间样本设计的基本模式

相等时间样本设计（same-time-sample design）是指对一组被试选取两个相等的时间样本，在其中的一个时间样本中出现实验变量（ X_1 ），在另一个时间样本中不出现实验变量（ X_0 ）的实验设计。其基本模式如下：

$$X_1 O_1 \quad X_0 O_2 \quad X_1 O_3 \quad X_0 O_4$$

其中 O_1 、 O_3 表示被试接受实验处理 X_1 后的观测结果， O_2 、 O_4 表示被试接受常规安排后的测验结果。通过对两种实验条件下观测结果的分析比较，考察 X_1

和 X_0 产生的效应差异。

(二) 相等时间样本设计的数据分析

相等时间样本设计的数据分析可以进行三个方面的统计分析。第一，可以对两种实验条件下的观测结果进行比较。即用 O_1 和 O_3 的平均数与 O_2 和 O_4 的平均数进行比较，以比较实验处理 X_1 和控制条件 X_0 的不同效果。第二，可以对两种实验条件下的顺序效应进行分析。这种分析方法如下：

	第一次实施	第二次实施
X_1	O_1	O_3
X_0	O_2	O_4

从上述模式中可以看出，通过 O_1 和 O_2 的平均数与 O_3 和 O_4 的平均数的比较，来确定顺序效应的大小。第三，可以对实验条件和顺序效应的交互作用进行分析，进一步检验不同的实验条件在不同时间序列上产生的不同效应。

(三) 相等时间样本设计的评价

相等时间样本在控制影响内部效度方面的因素十分有效，其原因是影响内部效度的历史、成熟等因素得到了较好的控制。因为，在不同实验背景中，出现众多一致的无关事件并与实验处理共同影响实验结果的可能性很小。

但是它在控制外在效度的影响因素方面并不理想。这主要体现在：

第一，实验安排的反作用效果会影响设计的外在效度。采用相等时间样本设计时，由于不同的实验处理实施于同一组被试，很容易使被试知道自己在接受实验。因此该实验的研究结果不能推论到无这种反作用效果的群体。

第二，选择偏差和实验变量的交互作用也会影响该设计的外在效度。由于该种设计只选用一个被试组，研究者选到的被试很可能对某种性质的实验敏感或不敏感。因此，所得到的实验结果只能适用于与参加实验的被试具有相同性质的群体。

第三，重复实验处理也会影响该设计的外在效度。由于该种设计需要对同一组被试多次重复的施加实验处理的影响，可能会使被试产生练习或疲劳效应。另外，在实验中，实验处理的循环出现，可能会使实验处理的效果持续到没有实验处理的观测成绩中，导致对实验处理效果的低估。因此，该设计的实验结果不能推论到没有重复实验处理的情境中。

第二节 多组准实验设计

在心理和行为研究中，要使实验结果尽量少的受到无关因素的干扰，使准实验设计的结果具有较高的内部效度，应在条件允许的情况下尽量采用具有实验

组和控制组的多组准实验设计。多组准实验设计主要包括不等组实验组控制组前测后测设计、不等组实验组控制组前测后测时间序列设计和平衡设计。

一、不等组实验组、控制组前测后测设计

（一）不等组实验组、控制组前测后测设计的基本模式

不等组实验组、控制组前测后测设计，是指研究者不能按随机化原则用等组法来分配实验组和控制组时的情况下，经常采用的一种准实验设计类型。这种设计主要用来评价某种处理的有效性。由于不能采用随机化的原则来形成实验组和控制组，因此在实验前两组就存在某些差异，因此称为不等组实验组、控制组前测后测设计。有时候研究者不能随机确定哪个为实验组，哪个为控制组的时候，也采用这种准实验设计类型。其基本的模式如下：

$$\begin{array}{ccc} O_1 & X & O_2 \\ & \cdots & \\ & O_3 & O_4 \end{array}$$

从该模式中可以看出，该实验模式包括一个实验组和一个控制组，其中接受实验处理的为实验组，没有接受实验处理的为控制组，并且对两组均进行前测和后测。需要注意的是，准实验设计类型的实验组和控制组不是随机分组，因此两组被试在实验前是不对等的，用虚线表示。在该设计中，研究者通过前测可取得两组是否等价或具有某种差异的指标，以提供固定整组在控制机体变量和因变量方面的最初数据，作为两个整组之间进行比较的基础。

（二）不等组实验组、控制组前测后测设计的数据分析

由于实验组和控制组在前测成绩上就有可能存在差异，因此不能采用简单的比较两组后测的成绩来评估实验处理的效果。一般情况下，不等组实验组、控制组前测后测设计结果有两种数据分析方法可供选择。通常采用的方法是参数检验中的 t 检验或非参数检验中的曼 - 惠特尼 U 检验或中位数检验，通过对两组前测后测差异分数的平均数进行检验，即对 $O_2 - O_1$ 和 $O_4 - O_3$ 的结果进行比较，来推测实验处理的效应。另一种方法就是采用协方差分析的方法，将前测结果作为协变量，后测分数作为因变量，通过控制前测成绩对后测成绩的影响之后，来估计实验处理的效果。

（三）不等组实验组、控制组前测后测设计评价

不等组实验组、控制组前测后测设计评价的优点是：第一，由于设计中增加了控制组，从而可以控制历史、成熟、测验以及仪器等因素的干扰。第二，由于两组都有前测成绩，因而为研究者提供对选择偏差的控制方法。

不等组实验组、控制组前测后测设计评价的缺点是：由于没有使用随机化的方法来选择或分配被试，因此实验组和控制组是不对等的，因而选者、成熟以及

选择与实验处理的交互作用可能会降低实验的内部效度。又由于实验组控制组都是用前测，因而该实验的结果不能直接推论到无前测的情境中。

二、不等组实验组、控制组前测后测时间序列设计

（一）不等组实验组、控制组前测后测时间序列设计的基本模式

不等组实验组、控制组前测后测时间序列设计，是在单组时间序列设计和不等组实验组、控制组前测后测设计基础上，结合形成的一种多组准实验设计类型。这种设计既具有时间序列设计的特点，即在实验中对被试组进行一系列的前测和后测，又具有不等组实验组、控制组前测后测设计的特点，即在实验中采用了未经随机分配的不相等组的实验组和控制组，所以该设计又称为多组时间序列设计。其基本模式如下：

$$\begin{array}{ccccccc} O_1 & O_2 & O_3 & X & O_4 & O_5 & O_6 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ O_7 & O_8 & O_9 & & O_{10} & O_{11} & O_{12} \end{array}$$

在该模式中 X 表示对实验组施加的实验处理。 $O_1、O_2、O_3$ 表示实验组在实验处理前的一系列观测成绩； $O_4、O_5、O_6$ 表示实验组在实验处理后的一系列观测成绩。 $O_7、O_8、O_9$ 表示在对实验组做一系列前测的同时，对控制组所做同样观测的观测成绩； $O_{10}、O_{11}、O_{12}$ 表示在对实验组做一系列后测同时，对控制组所做同样观测的观测成绩。虚线表示两组为固定整组，即不是随机分配被试所形成的实验组和控制组。

例如，在坎贝尔关于康涅狄格州的交通死亡人数与实施严惩制度的关系的研究中，也可以采用不等组实验组控制前测后测时间序列设计。也就是，除了对康涅狄格州在实施严惩制度前后的交通死亡人数进行观测外，选择另一些条件相似的州的交通死亡人数进行比较，结果见下图（图 6.3）。从这个图中可以看出，尽管 1951 年至 1959 年 4 个州的交通死亡率都稍有下降，但是康涅狄格州的下降最显著，特别是 1955 年对驾驶员实施严惩制度之后。因此，研究结果表明 1955 年的严惩制度对交通死亡率的确有显著作用。

（二）不等组实验组、控制组前测后测时间序列设计的数据分析

不等组实验组、控制组前测后测时间序列设计的数据分析，有两种方法可供选择。第一种方法，分别计算实验组和控制组的前测成绩的平均数，以及实验组和控制组的后测成绩的平均数；然后计算出实验组前测成绩和后测成绩平均数的差异，以及控制组前测成绩和后测成绩平均数的差异。采用独立样本的 t 检验或 z 检验对实验组差值和控制组的差值进行比较。

第二种方法，进行回归直线的截距和斜率的差异显著性比较。对实验组的前测成绩、后测成绩以及控制组的观测成绩分别进行回归分析，得出三条回归直

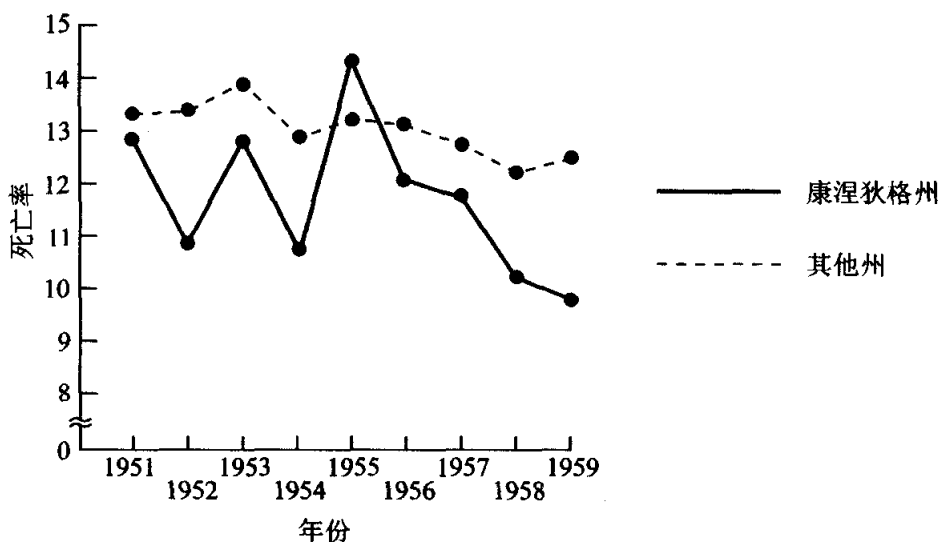


图 6.3 康涅狄格州在实施严惩制度前后与其他州的交通死亡人数的比较

线。通过检验实验组前测成绩的回归直线和控制组的回归直线的差异来比较实验组和控制组在选择上的偏差。通过实验组后测成绩的回归直线与前测成绩的回归直线以及控制组的回归直线来确定实验处理的效应。

(三) 不等组实验组、控制组前测后测时间序列设计的评价

不等组实验组、控制组前测后测时间序列设计结合了时间序列设计和不等组实验组、控制组前测后测设计的特点，主要优点体现在既能对一组的一系列的观测成绩的变化趋势进行了解，也能对两组的前测和后测的系列观测成绩的趋势进行比较，以估计实验处理的效果。同时，该设计还能对历史、成熟、选择与成熟的交互作用等影响进行有效的控制。该设计的局限性主要体现在，系列的前测和后测所引起的反作用效果对实验处理效果的影响。

三、平衡对抗设计

(一) 平衡对抗设计的基本模式

平衡对抗设计 (balanced design) 又叫轮换设计或拉丁方设计 (Latin square design)。这三个名称是从其作用、方法和模式三个不同的角度来说明这种设计的意义。

所谓平衡对抗设计，是指在实验中，由于前一个实验处理往往会影响到后一个实验处理的效果，而该实验设计的作用就在于提供对实验处理顺序的控制，使实验条件均衡，抵消由于实验处理的先后顺序的影响而产生的顺序误差，因而也可称之为抵消法设计。

所谓轮换设计 是指在实验中 由于学习的首因效应 先实验的内容 被试容

易记住；又因为学习的近因效应，对于刚刚学过的内容，被试回忆的效果一般也较好。因此，在实验方法上，有必要使不同实验条件出现的先后顺序轮换，使情境条件以及先后顺序对各个实验组的机会均等，打破顺序界限。

所谓拉丁方设计，是指平衡对抗设计的结构模式，犹如拉丁字母构成的方阵。例如四组被试接受 *A*、*B*、*C*、*D* 四种处理，其实验模式为：

组 1	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
组 2	<i>D</i>	<i>A</i>	<i>B</i>	<i>C</i>
组 3	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>
组 4	<i>B</i>	<i>C</i>	<i>D</i>	<i>A</i>

上述模式中可以看出，每种处理即表中的字母在每一行和每一列都出现了一次而且仅出现了一次。像这样的—个方阵列就称为一个拉丁方（Latin square）。要构成一个拉丁方，必须使行数等于列数，并且两者都要等于实验处理的种数。

在只有两个实验处理的情况下，通常采用的平衡对抗设计是以 *ABBA* 的顺序来安排实验处理的顺序。或者把单组被试分为两半，一半按照 *ABBA* 的顺序实施处理，另一半按照 *BAAB* 的顺序实施处理。

（二）平衡对抗设计的数据分析

平衡对抗设计的数据分析可以采用方差分析的方法，分别检验实验处理的主效应、不同组别的主效应和实验处理顺序主效应。在上述的拉丁方设计的方差分析中，可以分别对 4 列的总和进行比较，4 行的总和进行比较，对 *A*、*B*、*C*、*D* 四种处理的效果进行比较。通常情况下，拉丁方设计的统计分析，主要涉及 3 种主效应，研究者最关心的是实验处理的主效应，而一般不考虑因素之间的交互作用。

（三）平衡对抗设计的评价

平衡对抗设计由于采用不同实验条件出现的先后顺序轮换，使用情境条件以及先后顺序对各个实验组的机会均等的方法，可以基本上控制历史、成熟和测验因素的影响。

平衡设计的缺点主要表现为，由于不能采用随机分配被试的方法，因此人为的选择偏差就可能产生组间的差异，并且这种差异还可能与历史、成熟等因素产生交互作用，造成对实验处理效果的影响。

►本章提要

1. 准实验设计是介于真实验设计和非实验设计之间的一种实验设计。准实验研究不需要运用随机化原则和方法来选择和分配被试，需要严格的操纵自变量和控制无关变量，因而适合更广泛的研究目的。

2. 单组准实验设计包括时间序列设计和相等时间样本设计。

3. 时间序列设计是指对一组被试进行周期性的一系列观察，并在测量的这一时间系列呈现实验处理，然后观测呈现实验处理后的一系列观测结果，并将这些结果与实验处理前的一系列观测结果进行对比分析，而且推断实验处理是否产生效果。

4. 相等时间样本设计是指对一组被试选取两个相等的时间样本，在其中的一个时间样本中出现实验变量，在另一个时间样本中不出现实验变量实验设计。

5. 多组准实验设计主要包括不等组实验组控制组前测后测设计、不等组实验组控制组前测后测时间序列设计和平衡设计。

6. 不等组实验组、控制组前测后测设计是，研究者不能按随机化原则用等组法来分配实验组和控制组时的情况下，或不能随机确定实验组和控制组时，采用一种准实验设计类型。

7. 不等组实验组、控制组前测后测时间序列设计，采用了未经随机分配的不相等组的实验组和控制组，并在实验中对被试组进行一系列的前测和后测。

8 平衡对抗设计又叫轮换设计或拉丁方设计，该设计使不同实验条件出现的先后顺序轮换，使情境条件以及先后顺序对各个实验组的机会均等，抵消由于实验处理的先后顺序的影响而产生的顺序误差。

► 本章关键术语

准实验设计 (quasi-experimental design)

单组准实验设计 (single-group quasi experimental design)

多组准实验设计 (multi-groups quasi experimental design)

时间序列设计 (time-series design)

相等时间样本设计 (same-time-sample design)

不等组实验组控制组前测后测设计 (nonequivalent experimental/control group pre-posttest design)

不等组实验组控制组前测后测时间序列设计 (nonequivalent experimental/control group pre-posttest time-series design)

平衡设计(balanced design)

► 复习与练习

1. 实验设计和准实验设计的区别。
2. 各种准实验设计的基本模式、统计方法和评价。

► 推荐参考读物

黄希庭. 心理学实验指导. 北京：人民教育出版社,1987

张力为 .体育科学研究方法 . 北京 :高等教育出版社 2002

Campbell D. & Stanley J. . Experimental and quasi-experimental designs for research. Chicage :Rand McNally 1963

Cook T. D. & Campbell D. T. . Quasi-Experimentation: Design and Analysis Issues for Field Settings. Chicage:Rand McNally 1979

➤在 线 资 源

<http://www.socialresearchmethods.net/kb/>

第七章 非实验设计(I)

非实验设计(**non-experimental design**)，又称前实验设计(**pre-experimental design**)。非实验设计通常是一种描述性或相关性研究策略，用于识别和检验自然存在着的变量及其相互关系。非实验设计由于不能满足真实验设计所要求的随机选择和分配被试的基本原则，又不能主动操纵自变量和有效地控制无关变量，所以非实验设计很难有效地推断出自变量和因变量之间的因果关系。但非实验设计可以使研究者对各种变量之间存在的相关关系做出因果关系的假设，并在后续的实验研究中检验这种假设，因此非实验设计也是真实验设计的组成部分或重要元素。

在心理学研究中，由于一些特殊的原因，研究者无法随机选择和分配被试，或者无法有效地操纵自变量，这种情景下可以考虑选择非实验设计的研究策略。例如，为了研究灯光的照明强度对产量的影响，研究者分别对两家工厂进行实验，其中一家采用高强度的灯光，另一家采用原来的照明强度。由于两家工厂的工人已经确定，不能采取随机分配被试的方式来形成实验组和控制组，在这种情况下就只能采用非实验的研究设计。

同时，非实验设计还为研究者提供了多种研究方法，如自然观察法、访谈法、调查法、作业分析法、测验法、个案法等等，研究者可根据研究的性质和目的，采用其中一种或几种方法结合起来使用。此外，非实验研究基本上是在自然环境或现场情境中进行，研究的结果具有较好的推广价值，因此非实验设计可能会有较高的外部效度。

在本章中，我们将讨论以下问题：

1. 单组后测设计基本模式、统计方法及其优缺点。
2. 单组前测后测设计基本模式、统计方法及其优缺点。
3. 固定比较设计基本模式、统计方法及其优缺点。
4. 事后回溯设计基本模式、统计方法及其优缺点。

第一节 单组后测设计

一、单组后测设计的基本模式

单组后测设计(the one-group posttest design)是指，在实验设计中只有一组

被试，只实施一种实验处理，并且在实验处理之前不进行前测，然后通过后测得到该组的后测成绩。这种实验设计的基本模式为：

$$X \qquad O$$

在上述模式中， X 是研究者施加的实验处理或者某种因素， O 是在施加实验处理后所测量到的结果，以评定该组接受处理后的效果。

例如，一个研究者想要考察一种新的教学方法是否能够提高学生的学习成绩，通过一学期的教学后，测量学生的学习成绩。由于这种设计不提供任何比较，所以研究者不能将新教学方法后的成绩与使用前的成绩进行比较，也不能与其他使用传统教学方法的学生学习成绩进行比较。而且，由于没有前测，研究者也无法知道学生在采用新教学方法前的成绩。因此，也就无法知道学生的学习成绩是否真的是由实验处理（新的教学方法）所导致的，也就是这种单组后测设计并不能给研究者提供任何有关实验处理和观测结果之间关系的任何结果。因此，在心理学研究中这种设计很少使用。

二、单组后测设计的评价

单组后测设计缺乏对照组，而且不符合控制影响内部效度的大多数因素原则，也就是除了实施的实验处理外，不能排除历史、成熟、选择等无关因素对后测成绩的影响，因此很难对后测结果进行解释，也就无法推断实验结果 O 和实验处理 X 之间的因果关系。但是这种研究程序操作简单，并且可以为后续的真实实验设计提供研究方向。

以上述新的教学方法是否能够提高学生的学习成绩的研究为例，来分析除了实施新的教学方法外，还有哪些可能的因素会对学生的成绩产生影响。首先，由于实验组不是随机选取，可能存在选择偏差，因此很难确定实验组的学生是否与其他同年级的学生在新方法教学前处于同等的水平。其次，在整个学期的学习中，实验中的学生由于采用了新的教学方法，与没有使用新方法的学生相比，可能会有更高的学习兴趣，从而具有更好的学习成绩。最后，在整个一学期的学习过程中，学生心理、生理状态的变化，也可能影响学习成绩。这些因素都会影响到新的教学方法的效果。

虽然单组后测设计的结果不能进行有关因果关系的推论，但可以在这种设计的基础之上，为进一步的实验研究提供参考。

第二节 单组前测后测设计

一、单组前测后测设计的基本模式

单组前测后测设计 (the one-group pretest-posttest design) 是对单组后测设计的一种改进。但同单组后测设计一样，也没有设置相应的控制组进行比较。只是在实施实验处理前对观测组实施了一次前测。通过前测的结果，获得有关该组的某些信息，并作为与实验处理的结果进行比较的基线，以此评估实验处理的效应。

其基本的实验模式为：

$$O_1 \quad X \quad O_2$$

从该模式中可以看出，在被试接受实验处理 X 之前，先对有关的变量进行一次前测 O_1 ，然后施加实验处理 X 的影响，在实验处理之后，再对有关的变量进行后测 O_2 。

例如，在上例中，研究者在进行新的教学方法前，对学生的学习成绩进行了前测，在实施新的教学方法后，对学生的学习成绩进行后测。这样，单组后测设计就变成了单组前测后测设计。虽然这种设计优于单组后测设计，研究者至少知道了学生的学习成绩是否发生了变化，但是还不能推论出学习成绩的变化是由新的教学方法所导致的。因为在这个过程中学生的学习兴趣、学习内容的难易等等都会影响到最后学习成绩。所以研究者仍无法知道前后学习成绩的差异是来自于新的教学方法，还是其他的因素。

二、单组前测后测设计的数据分析方法

单组前测后测设计的数据分析方法，一是采用参数检验中的相关样本 t 检验或 z 检验进行数据分析。检验实验组前测成绩的平均数和后测成绩的平均数有无统计显著性差异，并根据实验组人数的多少来选择 t 检验或 z 检验。二是采用非参数检验中的符号秩次检验。三是采用非参数检验中的符号检验。后两种方法检验功效相对较低，对于满足统计检验假设的依赖程度也较低。

三、单组前测后测设计的评价

单组前测后测设计的优点是，第一，因为增加了前测，可以在实验处理之前提供有关被试的基线信息和其他有关的信息。第二，通过前测和后测，研究者可以得到同一被试在实验处理前后两次观测条件下行为变化的直接数据，能够明显地观测到实验处理的效果。并且可在一定程度上归因于实验处理的效果。第三，因为在该设计中，被试既做实验组，又做控制组，因而便于估计被试个体态度

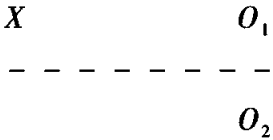
对实验效果的影响。因此当研究者能够操纵实验条件，了解被试在实验处理前后发生的变化，而被试又能够定时提供观测条件时，单组前测后测设计是一种非常实用的非实验设计。

虽然单组前测后测设计由于增加了前测，为后测成绩的变化提供了基线并在一定程度上控制了选择变量，但由于没有设置控制组，所以仍不是一种合理的研究设计。单组前测后测设计的局限性主要体现在，由于没有控制组的比较，不容易控制历史、成熟以及统计回归等问题，使研究者难以判断前测后测的差异是由于实验处理引起的，还是实验处理以外的成熟、历史等因素引起的，从而影响实验的内部效度。再有，实验过程中前测也有可能对后测（处理效果）产生影响。例如，前、后两次观测的时间间隔如果很短，被试可能由于前测而产生练习或疲劳效应等，影响后测的效果；如果前后测的时间间隔较长，则可能会出现对前测的保持与遗忘的个体差异等问题，致使后测结果受到影响，导致实验的内部效度的降低。

第三节 固定组比较设计

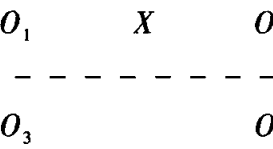
一、固定组比较设计的基本模式

固定组比较设计(the static-group comparison design)也称静态组设计或整组比较设计，该种设计采用实验组和控制组两组被试，但是这两组被试的形成并不是研究者根据随机化的原则和方式进行分配，而是在实验前就已经存在。固定组比较设计，与上述单组后测设计以及单组前测后测设计不同，其主要特点是使用了不接受实验处理（ X ）的控制组，以便于与接受实验处理的控制组进行比较。该设计的基本模式如下：



在该模式中，虚线表示两组均为没有经过随机化分配所形成的固定组。虚线上面的为实验组， O_1 表示研究通过操纵实验处理（ X ）后的后测成绩。虚线下面的为控制组，控制组没有接受任何实验处理， O_2 为控制组的观测成绩。

在固定组比较设计中，也可以对实验前的两个固定组进行前测，形成固定组前测后测设计(the static-group pretest-posttest design) 其设计模式如下：



其中 O_1 和 O_3 表示两个固定组的前测成绩， O_2 和 O_4 表示两个组的后测成绩。虽然这种设计增加了前测，可以通过计算后测和前测成绩的差来对被试变量进行一定程度的控制，但是仍不能完全排除被试变量的影响。因为这种前、后测的差异有可能会依赖于被试的前测成绩，也就是，前测分数高的组可能提高的更多（或在某些条件下提高的更少）。例如，对于学习成绩好的学生，成绩从 90 分提高到 95 分，比学习成绩差的学生，成绩从 60 分提高到 80 分的程度要小的多。因此，被试变量还是有可能造成某种程度的影响。

二、固定组比较设计的数据分析

对于固定组比较设计的数据分析，可以采用四种分析方法，一是独立样本 t 检验，比较实验组和控制组的均数是否存在显著差异。二是采用非参数检验中的曼 - 惠特尼 U 检验法，三是非参数检验中的中位数检验法，四是 χ^2 检验。

三、固定组比较设计的评价

由于在固定组比较设计中使用了实验组和控制组的比较，因此该种设计在一定程度上可以提高实验的内部效度。因此，该种设计基本上能对历史等因素进行控制，因此如果出现与实验处理同时发生其他无关因素的影响，则对两组后测成绩的影响程度在理论上应该是一致的。同时，由于使用了控制组，可以对成熟等因素进行有效的控制。此外，如果没有前测，还可以有效的控制测验效应或仪器等因素的影响。

固定组比较设计的局限性主要表现在，由于不能随机选择和分配被试，导致实验前的实验组和控制组可能不是等组，因而不能对两个固定组在实验前的差异进行估价，从而影响实验的内部效度。再有，固定组设计中的选择与成熟的交互作用，以及选择与实验处理的交互作用也会影响到该种实验设计的内部效度。

第四节 事后回溯设计

事后回溯设计 (backdating design) 是指所研究的对象是已经发生的事件，而且在研究过程中，研究者不需要选择与分配被试，也不需要设计实验处理或操纵自变量，而是通过观察存在的条件或变量，将这种已经自然发生的变量与某种可能的结果联系起来进行分析，从中发现某种可能存在的变量间的关系。例如，如果要研究某地区含酒精饮料的销量和当地犯罪率之间的关系，由于不能够随意操纵该地区含酒精饮料的销售量，所以只能采用事后回溯式设计，根据业已存在的当年含酒精饮料的销售量和犯罪率的数据来分析两者之间的关系。

事后回溯设计的基本模式如下：

③ O

在该模式中，③是研究者无法操纵或改变的自变量，O是研究者观测的结果。研究者通过O与③来推论变量间的关系。事后回溯设计与前面提到的实验设计是一个相关的研究过程。前面的研究设计通过对被试实施不同的实验处理来观测结果在不同条件下的变化，并对组与组之间的差异进行比较。而事后回溯设计则是从已经发生并存在差异的被试开始，追溯产生差异的原因。事后回溯设计主要包括两种类型，即相关研究设计和准则组设计。

一、相关研究设计

（一）相关研究设计的基本模式

相关研究设计（correlational study design）是从一组被试收集两组数据，其中一组数据作为观测到的结果，另一组数据则是被追溯的原因，通过研究两组数据之间的关系，来阐明两者之间的关系是属于正相关、负相关或是无相关。相关研究设计的基本模式如下：

O_1 O_2

使用相关研究的目的，是为了说明在实验变量之间可能存在的某种关系，并为进一步提出变量间可能的因果关系奠定基础。两个变量之间的关系可根据相关系数（correlation coefficient）的高低来确定。相关系数用小写字母 r 表示。相关系数是两个变量之间相互关联程度的估计，用 0 与 1 之间的数字表示。0 表示毫不相关，1 表示完全相关。相关系数又有正负之分，用 + 或 - 的符号表示。正相关表示一个变量的数量增加时，另一个变量的数量也随之增加；当一个变量的数量减少时，另一个变量的数量也减少。负相关表示当一个变量的数量增加时，而另一个变量却反而减少。但是，研究者应当特别注意的是，两个变量间的相关关系并不意味着两者间必然存在因果关系。即使两个变量间出现了较程度的相关关系，也不能代表两个变量之间的因果关系。例如我们发现看电视的时间数与学生的学业成绩有相关，至于是否因看电视的时间太多，而影响了学业成绩的问题，则须进一步的研究才能加以确定。对于存在一定关系的两个变量来说，两者间的关系可能有以下四种情况：

1. O_1 是 O_2 的因， O_2 是 O_1 的果，或者一部分的因或果。
2. O_2 是 O_1 的因， O_1 是 O_2 的果，或者一部分的因或果。
3. O_1 和 O_2 都受到未被测量的第三个变量的影响，是第三个变量的果或部分果。
4. O_1 和 O_2 互为因果。

例如，一位研究者想考察工人的生产效率和压力的相关，研究者可以通过观

察人们在生活中所经受的压力以及人们在工作上的表现。其中，压力可以操作定义为人们在压力量表上的得分，而工作效率可以操作定义为每日生产产品的数量。然后，研究者可以选取大量的被试，测量出一个变量的值，并计算出他们之间的相关关系。一个很大的负相关关系意味着压力增大时，生产效率下降。因此，当知道人们生活压力的分数，研究者就可以对人们的生产效率建立合理的预期。但是，高的相关只表明这两列数据以一种系统的方式建立了联系，并不意味着一个事件可以导致另一个事件的发生。相关可能反应一部分因果关系，也可能根本就没有因果关系的存在。例如，压力和生产效率之间高的负相关可以意味着（1）压力从根本上是人们工作做得差的原因；（2）不高的生产效率使人们经受了更大的压力；（3）具有某种人格特质的人，他们更可能经受压力，同时也在工作中表现不佳。这种情况是第三个变量影响到其他两个的变化。还有最后一种情境，假设工作环境的变化，导致工人更大的压力，同时也导致了生产效率的下降，这时相关关系依然存在，但两者之间根本就不存在因果关系。

（二）相关研究设计的数据分析

相关研究设计的数据分析，可根据所取得的变量性质的不同而采用不同的分析方法：一是当所研究的两组数据都是连续型数据时，可采用皮尔逊积差相关的方法进行统计；二是，当所研究的两个变量中一个为连续型变量，另一个为二分变量时，可采用点二列相关进行统计；第三，当所研究的两个变量都是等级变量时，可采用斯皮尔曼等级相关进行统计。由上述方法计算出的相关系数 r 的绝对大小会受到样本容量的影响，因此相关系数 r 是否达到显著性水平，可以从 r 临界值表中直接查出来。

（三）相关研究设计的评价

相关研究设计作为非实验设计中的一种典型方法，其优点主要体现在可以提供各个变量或现象之间相关的程度和方向（正相关或负相关）。此外，还可以在相关研究的基础上，进一步设计严格的实验方案来探讨变量间的因果关系。其缺点是，研究者不能操纵的相关的实验过程，而只是对固定组被试的特征及反应效果进行事后的分析，因而相关研究的结果，只能得出相关关系而不能得出因果关系的推论。虽然相关研究不能确定因果关系，对探究纵向发展过程不能提供太多的资料，但在研究横向关系上却是一种很重要的方法。相关研究可以弥补因实际情况的限制而无法使用实验法之不足。此外，在心理测验的编制上，相关研究也是不可或缺的。例如智力测验、性向测验、兴趣测验、人格和态度测验等的编制在项目准备、量表编制和建立常模等方面都需要使用相关研究。

二、准则组设计

(一) 准则组设计的基本模式

准则组设计(criterion-group design)和相关研究设计类似,都是对业已发生的事情进行研究的一种非实验设计。准则组设计要求研究者通过确认某些被试(准则组)具有一种状态的特征,而另一些被试(非准则组)不具有这种状态的特征,在此基础上去追溯可能存在的原因。相关研究设计和准则组设计是相辅相成的。相关研究设计主要在于探讨两个或多个变量之间是否存在一定的关系,而准则组设计则强调存在一定关系的两个变量之间,表现出什么样的不同状态。在非实验设计中,这两种设计是经常交替使用的。

准则组设计的基本设计模式如下:



在上述模式中, O_1 表示对具有某种特征的准则组的观测结果, O_2 表示不具有这种特征的非准则组的观测结果, \textcircled{C} 表示研究者所要追溯的准则组和非准则组差异的原因。

例如,富兰克林(Franklin Janoff-Bulman, and Roberts, 1990)为了研究离婚家庭儿童对态度的影响。他们采用准则组设计探讨了离婚家庭的大学生和正常家庭的大学生对世界、他人、自我价值以及人际关系的态度。结果发现,两组学生在对世界、他人、自我价值的态度上并没有显著的差异。两组的差异主要表现在对待人际关系的态度上,尤其是对爱和婚姻的态度。来自离婚家庭的学生表现为更能接受离婚和对自己将来的婚姻充满消极的态度。采用这种研究设计,研究者证实重大的生活事件似乎并没有对儿童的整个态度系统产生广泛的、消极的影响。相反,离婚似乎只是影响了他们对于自己将来婚姻的态度。但是这种研究设计并不能形成因果关系的推论,因为除了研究的变量(是否来自离婚家庭)外,其他的变量也可能存在显著的影响。在上例中,导致对婚姻消极态度的原因还可能与父母本身对婚姻的态度有关。如果父母具有对婚姻的消极态度,这种消极态度就有可能导致父母的离婚并影响儿童对婚姻的态度。

(二) 准则组设计的数据分析

准则组设计采用的数据分析方法,根据满足统计分析的条件,可采用参数检验中的独立样本 t 检验或方差分析,也可以采用非参数检验中的曼-惠特尼 U -检验法和中位数检验法。

(三) 准则组设计的评价

准则组设计的优点主要表现为,通过准则组和非准则组的对比,可以确定与

准则组有关的特征以及研究准则的伴随情况。其缺点是，准则组所具有的经验特征不能由研究者随机分配被试和操纵自变量而形成。也不能有效的控制无关变量的影响，而是在事件发生之后推测这种现象的可能原因或可能条件。因此，对于准则组和非准则组差异的原因的解释具有较大的不确定性。

总之，相关研究设计和准则组设计是相辅相成的。除了上述介绍的两种事后回溯设计的优点和局限性外，事后回溯设计比实验设计具有较少的人为干预，因而更接近自然条件，因此在无法进行严格实验干预的条件下，事后回溯设计是一种较好的选择。而且，事后回溯设计也比实验设计经济，节省人力和物力。事后回溯设计的最大局限性体现在研究者不能随机分配被试和操作自变量。但是，尽管事后回溯设计的性质决定了不能操纵自变量和控制无关变量，但是从统计的角度能够对无关变量的影响进行一定程度的控制，从而在某种程度上提高事后回溯设计的内部效度。

► 本章提要

1. 非实验设计通常是一种描述性或相关性的研究策略，用于识别和检验自然存在着的变量及其相互关系。非实验设计由于不能满足真实验设计所要求的随机选择和分配被试的基本原则，又不能主动操纵自变量和有效的控制无关变量，所以非实验设计很难有效的推断出自变量和因变量之间的因果关系。但非实验设计可以使研究者对各种变量之间存在的相关关系做出因果关系的假设，并在后续的实验研究中检验这种假设，因此非实验设计也是真实验设计的组成部分或重要元素，也称作前实验设计。

2. 非实验设计一般包括单组后测设计、单组前测后测设计、固定比较设计和事后回溯设计等几种类型。

3 单组后测设计是指，在实验设计中只有一组被试，只实施一种实验处理，然后通过后测得到该组的后测成绩。

4. 单组前测后测设计是对单组后测设计的一种改进，在实施实验处理前后分别对被试组进行观测。

5. 固定组比较设计也称静态组设计或整组比较设计，该种设计采用实验组和控制组两组被试，但是这两组被试的形成并不是研究者根据随机化的原则和方式，而是在实验前就已经存在的。

6. 事后回溯设计是指所研究的对象是已经发生的事件，而且在研究过程中，研究者不需要选择与分配被试，也不需要设计实验处理或操纵自变量。而是通过观察存在的条件或变量，将这种已经自然发生的自变量与某种可能的结果联系起来进行分析，从中发现某种可能存在的变量间的关系。

7. 事后回溯设计主要包括两种类型，即相关研究设计和准则组设计。相关

研究设计是从一组被试收集两组数据，来阐明两者之间的关系是属于正相关、负相关或是零相关。准则组设计要求研究者通过确认某些被试（准则组）具有一种状态的特征，而另一些被试（非准则组）不具有这种状态的特征，在此基础上去追溯可能存在的原因。

►本章关键术语

非实验设计(non-experimental design)

单组后测设计 (the one-group posttest design)

单组前测后测设计 (the one-group pretest-posttest design)

固定比较组设计 (the static-group comparison design)

事后回溯设计 (backdating design)

相关研究设计 (correlational study design)

准则组设计(criterion-group design)

►复习与练习

1. 非实验设计为什么在心理与行为研究中得到重视？
2. 各种非实验设计的基本模式，统计方法及其评价。

►推荐参考读物

黄希庭，心理学实验指导，北京：人民教育出版社，1987

张力为，体育科学研究方法，北京：高等教育出版社，2002

周谦，心理科学方法学，北京：中国科学技术出版社，1994

►在线资源

<http://www.socialresearchmethods.net/kb/>

第八章 非实验设计(Ⅱ)

心理学的研究对象主要是既有生物属性又有社会属性的人，因此，我们不但要研究人的生物属性，更要研究人的社会属性。这显然与以物为研究对象的物理、化学、生物学等自然科学是完全不同的。作为心理学研究对象的人，不能像自然科学那样进行精确的控制与操纵，既不能把人像老鼠一样关在笼子里观察其行为的变化，也不能随意切除大脑的一部分来观察其行为的变化。因此，就需要采用一些有别于自然科学的特殊研究手段和方法。访谈研究、问卷研究、个案研究及跨文化研究就是属于这一类型的研究手段。事实证明，访谈、问卷、个案研究等方法所提供的有关心理活动过程的信息，比传统心理实验中采用最多的反应时、正确率等指标所提供的信息更为丰富、完整和深化。但是这并不表示实验研究已经过时，只是由于心理学研究对象的特殊性，通过这些方法，来弥补实验研究的缺陷，以便更全面和深入地了解人们的态度、情感、思想观念和主观感受，从而对他们的各种心理活动和行为进行全面深入的分析、研究。在本章中，我们将讨论以下问题：

1. 访谈研究设计的含义及其过程
2. 问卷研究设计的含义及其过程
3. 个案研究设计的类型
4. 跨文化研究设计的基本过程

第一节 访谈研究设计

访谈研究 (**interview survey**) 是研究者通过与研究对象的交谈来收集心理特征与行为数据资料的一种研究方法。它是心理学研究中运用最广泛的研究方法之一。访谈研究在心理学研究中的特殊意义和作用是由心理学研究对象的特殊性所决定的。目前，访谈研究越来越多的运用在心理学的各个研究领域。

一、访谈研究的含义

访谈研究是根据特定的科学目的，根据设计和编制原则来实施资料收集的过程。因此，访谈不同于日常生活中的“聊天”或者一般的交谈。访谈研究是研究者与研究对象之间的一种社会交往过程，在这个过程中访谈者和访谈对象之间形成了一种社会互动关系。因此访谈研究的结果是访谈双方之间的一种社会

过程和社会交往的产物。访谈双方的心理特征、态度、动机、知觉和行为等相互作用和影响，以及访谈所处的情景、信息传递的性质等都会影响到访谈的效果。图 8.1 表示访谈过程中谈话人之间的交互作用关系（王重鸣，1990）。

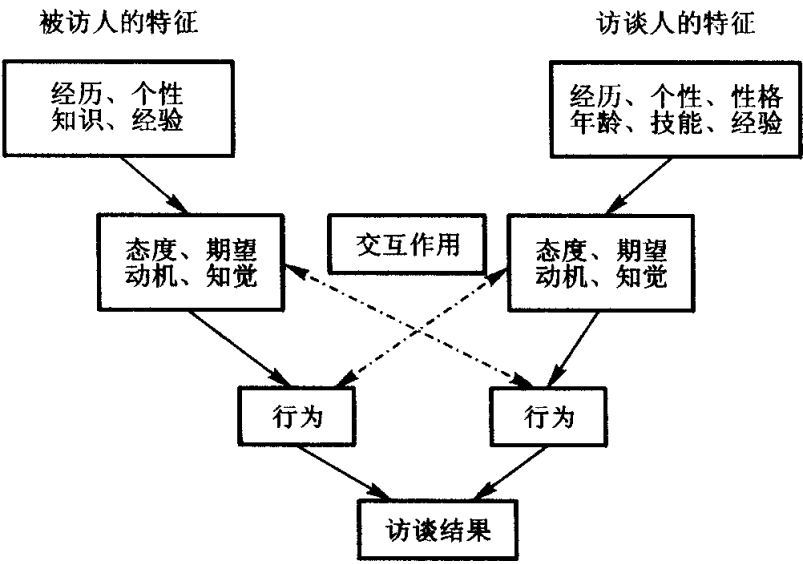


图 8.1 访谈过程中谈话双方的相互作用

二、访谈研究的类型

访谈研究根据研究目的、性质或对象的不同而划分为不同的种类。

（一）结构访谈和非结构访谈

根据访谈内容和过程有无标准化的程序，可分为结构访谈和非结构访谈。

结构访谈指按照统一的设计要求，具有固定结构的问卷而进行的访谈。结构访谈对选择访谈对象的标准和方法、所提出的问题、提问的方式和顺序、被访谈者回答的方式，访谈记录的方式等进行了标准化，有时甚至对于访谈的时间、地点、周围环境等外部条件，也要求基本保持一致。因此，结构化访谈又称为标准化访谈（standardized interview）。结构访谈的优点在于结果便于统计分析，对于不同访谈对象的回答易于进行对比分析。但是，这种访谈方法缺乏弹性，使访谈者难于根据当时的具体情况，灵活地采用适当的方式进行访谈，难于对问题进行深入探讨。

非结构访谈又称非标准化访谈，该方法对访谈对象的条件、所要询问的问题等只有一个粗略的基本要求，访谈者可以根据访谈时的实际情况灵活地作必要的调整。非结构访谈有利于发挥访谈者和被访谈者的主动性、创造性，有利于适应各种被访谈者的具体和特殊情况，有利于拓宽和加深对问题的研究，也有利于处理原来访谈设计方案中没有考虑到的新情况、新问题。但是，非结构访谈的结

果难以进行定量分析，对不同被访谈者的回答难以进行对比分析，此外，该方法对访谈者的要求比较高。

根据结构访谈与非结构访谈的上述特点，在心理学的实际研究中，研究者们常常是将它们结合起来加以应用，以便互相补充、取长补短，提高研究的科学性水平。一般来说，综合采用两种方法时，往往先进行非结构访谈，然后再进行结构访谈，否则非结构访谈将受到结构访谈的很大影响，不利于了解被访谈者原有的心理活动与特点。此外，也有按非结构访谈→结构访谈→非结构访谈的顺序进行的。

（二）直接访谈与间接访谈

根据访谈时访谈者与访谈对象的接触方式，可分为直接访谈和间接访谈。

直接访谈是指访谈者与被访谈者进行面对面的交谈，因此又称面对面的访谈。直接访谈的突出特点是，访谈者与被访谈者直接发生相互影响、相互作用。访谈者不但能广泛、深入地探讨有关问题，了解被访谈者的思想、态度、情感和其他各种情况，而且还能亲自观察被访谈者的有关特征和他们在访谈过程中的许多非言语信息，从而加深对谈话内容的理解，利于判断访谈结果的真实可靠性。但是，由于直接访谈首先是人与人之间的直接交往过程，访谈者与访谈对象相互直接作用可能会对访谈结果产生影响。

间接访谈法就是访谈者通过一定的中介物与被访谈者进行非面对面的交谈。目前，间接访谈的主要方式就是电话访谈。电话访谈适用于访谈内容较少、较简单的调查研究，其优点是收集数据资料省时省力，对于某些不适宜于面对面对谈的问题可以通过电话访谈来进行。但是由于需要借助一定的中介物，所以适用范围有限。因此，访谈者难以深入探讨有关问题，更不能直接观察被访谈者的有关特征和各种非言语信息，这不利于对访谈结果的分析与解释。

（三）个别访谈和集体访谈

根据访谈对象的人数，可分为个别访谈和集体访谈。

个别访谈通常只有一个访谈者与一个访谈对象。个别访谈的特点主要表现在，访谈对象能够有较多的机会与访谈者进行交流，因此能够对问题进行深入、全面的了解。

集体访谈是由访谈者同时对多个访谈对象的访谈。集体访谈具有很多个别访谈所不具备的优势。集体访谈是多个访谈对象同时就某个问题发表自己的看法，因此可以为访谈对象提供一个相互交流的机会。由于人们在集体环境中的表现往往与个人单独的表现不同。因此，在集体访谈中访谈者可以了解访谈对象在集体互动中的表现。同时，也正是由于处在集体环境中，访谈对象可能会对某些问题隐瞒自己的看法。

在实际研究中，个别访谈和集体访谈可以结合起来进行，以提高研究结果的

效率，并且在不同的条件下获得的研究结果可以相互补充、相互验证。

三、访谈研究的设计

访谈研究的设计主要包括研究问题的确定、访谈程序的制定、访谈对象的选取，访谈人员的选取与训练等许多方面。

（一）访谈目的的确定

首先就是要明确访谈研究的目的，并将其进一步具体化——即确定访谈研究的各种具体变量。一般来说，研究目的指明研究所要达到的总目的，因而也就对研究的范围、对象等作出了一些限定。但是，研究目的往往是比较笼统、概括的。对同一研究问题，我们可以从不同角度去研究。因此，进行访谈设计时，首先需要将一个比较笼统的大的研究目的和问题具体化成一个限定的研究目的和问题，并提出自己对研究问题的各种具体假设。然后，再根据这一具体研究问题，详细列出研究所涉及的所有变量的类别与名称，进一步明了回答研究问题，检验研究假设需要收集哪些方面的信息。要做好这一步工作，最好列出一个研究变量简表，这一方面有助于我们思考，另一方面可以防止遗漏。

明确访谈研究的目的，详细列出研究涉及的所有变量，对访谈设计以后的工作是十分重要的。为此，需要认真查阅与访谈内容有关的文献，特别是研读有关研究报告，从中吸取有重要参考价值的东西。有时还需要深入实际作些初步的了解和调查。在此基础上，再选择访谈研究的侧重点，详细列出研究变量。这一步工作如果未做好，将直接影响到以后的设计工作和整个研究的质量。

（二）访谈问题形式的设计

详细地说明研究的具体目的和研究变量之后，研究者就需要考虑访谈问题的具体形式。在访谈设计中，访谈问题的形式主要有两种：一是封闭式问题（又称固定选择式问题、限定性问题），二是开放式问题（又称非限定性问题）。封闭式问题要求访谈对象在事先确定的几个选择答案中选择一个自己认为最适合的答案。比如，“你认为自己在待人方面是非常热情，还是一般，还是很冷淡？”和“你目前与单位领导的关系是充满矛盾还是很协调？”这两个问题就是封闭式问题。开放式问题则是访谈对象根据自己的想法，用自己的语言来作出回答。这类问题的例子如“请谈谈你在待人方面的情况”、“请谈谈你目前与单位领导的关系如何？”

选用何种访谈问题的形式，要根据访谈研究的具体目的、访谈对象的具体情况、访谈者的有关知识经验而定。一般来说，以下两种情况是比较普遍的（董奇，1992）：一是在进行一项访谈研究时，如果研究者对访谈对象有关情况不了解，常常需要在研究初期采用开放式问题，以取得有关基本情况和资料，进行定性分析；研究后期则在此基础上再设计出若干封闭式问题，去收集有关数据资

料，以便进行定量分析。二是大多数访谈问卷在开头部分安排了一些封闭式问题，以取得访谈对象的有关基本情况，如性别、婚姻状况、教育水平、职业、职务等，随后再安排一些开放式或封闭式的问题或两种形式并用。

开放式问题的优点是有助于访谈对象充分表达自己的思想、情感，有利于访谈者了解额外信息，对不明确的答案进行追问等等。其主要缺点是计分困难，带更多的主观性。封闭式问题易于计分，能更好地保证反应形式的标准化和结果的客观性，但是其所有选择有时可能难以全面反映访谈对象的有关情况，缺乏灵活性。

(三) 回答方式的设计

拟定具体访谈问题之后或同时，还需要考虑让访谈对象以何种方式对每一问题作出反应。对于同一具体访谈问题，可以有不同的反应方式。比如，访谈对象的年龄信息可以通过填空式获得，也可以给访谈对象提供一系列已分好的年龄组，让他们选择最适于自己年龄情况的那一组。其中常用的有填空式、等级排列式、量表式和核对式等几种。表 8-1 列出了每种方式所获取的数据的性质及主要优缺点：

表 8-1 四种反应类别及其特点

反应方式	数据类型	主要优点	主要缺点
填空式	定类测量	误差小,反应灵活性大	难以记分
量表式	定距测量	易于记分	费时,可能产生误差
等级排列式	定序测量	易于记分,强迫区分	难以完成
核对式(分类式)	定类测量(计算总数时可能是定距测量)	易于记分,易于反应	提供数据少,选择可能少

(采自董奇 ,1992)

在访谈设计中，为了选择最佳、最恰当的反应方式，必须综合考虑以下几方面的问题（董奇 ,1992）：

1. 研究变量的性质

选择什么样的反应方式首先取决于变量的性质。比如访谈对象性别方面的信息一定是命名性质的，只能采用填空式或核对式获得，而年龄方面的信息则可以通过不同方式反映出来。

2. 统计处理需要的数据类型

研究变量数据的处理方式决定采用何种反应方式。如果统计分析需要间隔性质的数据，最好选用量表式和核对式。填空式和某些核对式提供的命名数据，只适于进行 χ^2 分析。当事先不能确定数据统计分析的方式时，最好采用量表反

应方式。因为由它获得的间隔数据很容易就可以转化为顺序或命名数据。

3. 反应的灵活性

填空式使访谈对象在作出反应时有最大程度的选择自由，而非式、对错式正好相反，使访谈对象反应时自由选择的程度最低。

4. 完成访谈所需的时间

等级排列式和量表式通常需要大量的时间才能完成。如果一个访谈问卷的内容比较多，已有相当长度，选择对错反应方式也许比较好，这样可以减轻访谈对象的负担，提高访谈效果和质量。

5. 各反应潜在反应误差的大小

一般来说，量表式和核对式的反应误差最大，以这两种方式回答时，访谈对象不仅受社会赞许性的影响，而且还受其他许多因素的影响，如过多使用“对”、“是的”答案的倾向，选择某一固定点作为对每一问题的答案、避免极端评价的倾向等等。等级排列和填空式的潜在反应误差则较小，特别是等级排列强迫访谈对象对多种选择答案进行区分。

6. 计分的难易程度

填空式反应必须先编码，因而是比较难以记分的。而讨论过的其他方式则都是比较容易记分的。

上述讨论的内容主要是针对封闭式问题的，对于开放式问题，则可能需要制定一些代码系统来及时准确地记录访谈对象的谈话内容。

（四）访谈结果的分析

根据研究目的不同，访谈结果的分析方法可以是定性的也可以是定量的。

表 8-2 中列出了几种分析访谈结果的方法。

表 8-2 分析访谈内容的方法

分析方法	评 价
获得总的问卷 / 量表的分数	研究者可能在访谈中完成结构化问卷；可直接进行假设检验
频率统计 / 特定术语的编码 / 主题 / 访谈内容中出现的概念	开放式回答的内容分析的一部分；可检验评价者信度
根据详细的评分系统用明确的标准计分	用于开放式问题，需要培训评价人，可检验评价者信度
定性分析（用编码程序）	如主题分析，理由充分理论，谈话分析技术、演讲分析

（采自艾森克 阎巩固译 2000）

四、评价

如前所述，访谈法的基本特点就是研究者与研究对象进行面对面的交谈。这一特点直接决定了访谈法所具有的如下优点和局限性。

访谈法的主要优点是：

第一，有利于对心理、教育问题进行深入、广泛的研究。访谈法以口头交谈方式进行，包括结构访谈法和非结构访谈法等不同类别，这就使它既可用于收集现时资料，又可收集过去的资料，既可用于定性研究，又可用于定量研究；既可了解客观事实、行为方面的问题，又可了解主观动机、情感、观念方面的问题，既可用于验证某种假设或理论，又可用于提出假设或理论。总之，与其他收集研究资料的方法相比，访谈法可以获得更为丰富、广泛的资料，有助于对人的心理活动、教育规律进行多层次多方面的探索。

第二，能灵活地有针对性地开展资料收集工作。访谈法可以根据访谈对象和访谈过程的具体情况，灵活地采取各种方法，有针对性地、有效地收集研究资料。比如，当访谈对象表现出对问题的不理解或误解时，研究者可针对情况以适当方式重复提问；当访谈对象的回答含糊不清、态度不明确时，研究者可让其详细解释说明；访谈对象说出研究者事先未预料到的、有价值的内容时，研究者可以根据情况进行追问或加以记录。访谈法可以根据访谈对象的知识水平或理解程度灵活变换提问方式，可以在访谈对象自由发表看法时在一定程度上控制谈话方向，充分发挥研究者的主动性和创造性。这是其他方法难以做到的。

第三，可以保证收集到的研究资料具有较高的可靠性。由于访谈是面对面进行的，是研究者与访谈对象相互作用、相互影响的过程，因此，当访谈对象不理解研究者的问题，或者研究者认为访谈对象的回答不完整、不明确时，都可以追问，了解更为确切的信息。研究者可以对访谈环境进行控制，防止干扰，也可以严格控制问题顺序，避免因问卷结构受到破坏造成的误差，此外，在进行访谈时研究者可以在提问或访谈对象回答时观察访谈对象的非言语行为（如表情、姿势、动作等），综合这些非言语信息判断访谈对象回答的可靠性，可以对获得的资料进行信度和效度的评估，保证研究资料的可靠性。

第四，访谈法适用范围广。访谈法是口头进行的，研究者可以对问题进行诠释、说明，它适用于一切具有口头表达能力（思维正常）的不同文化程度的访谈对象，与问卷法相比，访谈法的适用对象更为广泛，如访谈对象可以是文盲、半文盲或因种种原因不能书写的人，而问卷法则难以做到这一点。

此外，访谈法具有较问卷法更高的回收率和有效率，研究者能保证使访谈问卷中的所有问题都得到回答，提高被试回答的有效率。因此，访谈法是心理、教育科学研究中一种重要的研究方法。

当然，访谈法并非一种完善的收集资料的方法，它仍然有着许多缺点和局限性。

首先，访谈结果的准确可靠性可能受到研究者素质的影响。访谈法是由研究者（即访谈人员）进行的，因此，访谈法上述优点的发挥有赖于研究者的素质。如果访谈人员素质较差，能力不强，可能对访谈对象的回答产生误解或在记录时发生错误，如果访谈人员没有掌握必要的访谈技巧，态度生硬，语言不礼貌，可能影响访谈的相互作用过程，造成访谈对象的不合作或提供虚假信息，这样就难以获得可靠的研究资料。访谈结果也易受到访谈人员的主观偏见、价值取向的影响。因此，对访谈人员进行访谈技巧的培训是必要的。此外，访谈人员的性别、性格、年龄、社会阶级、衣着、口音和身材等都可能影响访谈对象的回答。

其次，某些问题不宜进行访谈。访谈法是对面进行的，不具有匿名性。因此，对于访谈对象比较敏感、不愿回答的属于个人隐私的问题，不宜进行访谈。如果强行进行访谈，可能使访谈对象中止访谈或不作真实回答，影响访谈结果。一些无法用语言表达的情感、体验、社会关系变化、动作变化或心理过程等资料不能用访谈法取得，而需要用其他方法（如观察法、社会测量法等）获得。此外，访谈法获得的许多资料需要进一步核实，如对与访谈对象有关系的人进行访谈，或采用其他方法对照检验。

第三，与其他方法相比，访谈法费时费力费财。进行访谈，一般需要聘请访谈人员并对其进行培训，访谈需要抽样，印制各种访谈提纲，准备磁带或记录纸张，还要支付被试费和主试的劳务费、差旅费等，因此，需要花费较多的人力和财力。访谈法一般每次只对一个访谈对象进行，访谈过程和记录整理需要较多的时间。在访谈对象分散的情况下，访谈时间可能更长。访谈法的这一特点限制了它的规模。

第四，访谈法获得的研究资料难以量化。一方面，由于访谈对象文化程度不同，可能对问题的理解不同，而访谈人员对问题的解释也可能不同，这种灵活性造成问题的表述缺少标准化；另一方面，访谈结果缺乏量化的指标，一般只能以某一答案出现的次数、百分比作为指标，访谈对象的回答可能有很大的差异，可能使答案很多，也难以定量计算。

此外，访谈法受环境、时间和访谈对象情绪状态，访谈对象思考问题时间较短等限制，也影响到访谈法的使用。

第二节 问卷研究设计

问卷研究（**questionnaire survey**）是心理学研究中运用最广泛的研究方法之一。访谈研究是通过交谈的方式来收集资料，问卷研究则是通过书面的方式，根

据严格设计的题目或问题向研究对象收集资料和数据的一种方法。问卷是由 questionnaire 这个词翻译而来，它的原意是“一组问题”(a set of questions)或“问题的集合”(a collection of questions)。因此问卷的形式是一份精心设计的问题表格，用来研究人们的心理特征、行为和态度等。

一、问卷调查的类型

(一) 封闭式问卷调查与开放式问卷调查

根据问卷中问题的结构，可将问卷调查分为封闭式问卷调查与开放式问卷调查。

在封闭式问卷调查中，被试只需要从预先给定的答案中根据自己的情况选择一个或几个合适的答案。封闭式问卷的优点主要体现在：回答是标准化的，便于计分和统计分析；由于提供了备择答案，被试不大可能感到困扰而不答或答“不知道”；回答简便，不需花更多的时间来思考，不需要较高的撰写技能。封闭式问卷的缺点是：易产生不可信的回答，特别是有急事或烦恼时；答案是固定的，因此难以发现被试回答上的细微差异；回答者可能会对提供的答案不满或不同意，而拒答。

开放式问卷调查中，问题虽然是统一的，但并没有预先给定可供选择的答案，而是被试根据自己的情况自由回答。开放式问卷的优点是：不受所提供答案范围的限制，被试可以独立、自由地发表看法；适合于研究者对不了解的问题进行探索性的研究；便于赢得被试的合作，他们会把答卷看成是自己的自我表达和一种创造的机会；对那些较复杂的问题，开放式问卷更可取。开放式问卷的缺点是：由于自由作答，可能导致无价值或不相干的资料；各个被试的回答是非标准化的，难以进行对比和统计分析；被试的回答需要有更高语言表达能力，仅适用于教育水平较高的被试。

开放式问卷主要用于探索性研究。正式问卷通常以封闭式问卷为主。考虑到两者的优缺点，在正式问卷中如果在重要问题上加上几个开放式问题，其效果将更好。为了不让被试发挥太多，开放式问题后面的空白，不要预留得太多。在实际的问卷调查中，研究者往往根据研究目的的不同来选用封闭式或开放式或两者的结合。例如，在探索性调查研究阶段通常采用开放式的问题，而在正式调查研究阶段主要采用封闭式问题。

(二) 自填问卷和访问问卷

根据使用方式或填答方式的不同可以将问卷分为自填问卷和访问问卷。自填问卷是由被试本人填写的问卷。而自填问卷又根据发送到被试手中的方式不同分为邮寄问卷和发送问卷。其中，邮寄问卷是通过邮局寄发给被试，被试填写完成后，再通过邮局寄回给研究者。发放问卷则是由研究者或其他人将问卷送

到被试手中，被试填写完后再由研究者或其他人逐一收回。这种自填问卷的优点是可以在较短的时间内收集到大量的资料，并且由于被试在填写问卷的时候不会受到其他人员在场时可能产生的干扰，所收集资料的可靠性更高。缺点是，由于研究者不在现场，无法控制被试的作答过程，如被试对问题的讨论等，可能会影响到研究的效度。

二、调查问卷的基本结构

（一）问卷的标题

问卷的标题也就是问卷的名称或题目，通常置于问卷的起首位置。被试根据问卷的标题，就可以了解到问卷的性质或目的，所以问卷的标题对于了解问卷，具有重要的作用。例如，“中学生压力问卷”，“企业员工工作满意度问卷”等等。有时候也常常把“问卷”称之为“调查表”，事实上，这两者是通用的。例如，“中学生学习障碍调查表”，“重庆市居民生活质量调查表”等等。问卷的标题最好能呈现调查或研究的对象，如前面“中学生”，“重庆市居民”等。但是在某些情况下，研究主题涉及一些比较敏感或个人隐私的内容，为了避免被试对研究主题的防御作用而产生拒绝回答的现象，这时候可将这些敏感的话题改为较中性的词语来表示，如将“大学生的拖延行为调查问卷”改为“大学生生活经验问卷”。

（二）指导语

指导语是问卷结构中必不可少的部分，其作用在于向被试说明与解释问卷的性质、目的以及作答方式。指导语对于问卷的作用就如同仪器的使用说明书。指导语根据说明的内容不同，可分为两个部分。第一部分是说明问卷的性质与目的，通常置于问卷的标题的下方与问卷的内容之前，这一部分通常包括致谢语、为被试保密语以及问卷的编制者。例如，

您好：

我们是西南师范大学心理学研究所的科研人员。这份问卷主要是了解您的日常生活情况，所得资料仅是用于学术研究，因此不需要您填写姓名，请放心回答。问卷题目的答案无所谓对错，只是个人经验的反映。请您仔细阅读各部分的答题要求，并根据自己的实际情况进行回答，不要有遗漏。您对每一题的回答都对本研究具有重要的价值。衷心感谢您的协助与合作。

祝 学习进步

西南师范大学心理学研究所

第二部分指导语说明作答的方式，通常置于每一类问卷内容之间。如果，只有一种类型的话，这一部分的指导语也可以置于第一部分指导语中。例如，

请将你感觉最合适的答案代号写在题后的（ ）中，注意每题仅能填写一

个答案。

以下各种陈述是您对自己的一些看法或态度。请根据您自己的体会，在每个题项后相应的 ☐ 内打“√”。

非 常 同 意	有 些 同 意	不 太 同 意	非 常 不 同 意
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

我对自己充满信心。

总之，问卷中一切可能使被试不清楚、难理解的地方，都应该给予必要的说明和指导，而对于指导语的编写者来说，就是尽可能编写简明易懂的指导语。

（三）问题与答案

问题与答案是问卷的主要内容，也是问卷设计的主题。对于研究对象的特征、行为和态度等资料基本上是通过问题和答案来收集的。问卷中的问题在形式上可分为开放式问题和封闭式问题两类；在内容上可分为有关事实的、有关态度的和有关个人背景的三类。

在问题的形式上，所谓开放式问题，就是要求被试根据自己的情况和感受来自由回答的问题。例如，

你对你目前的学习状况感到满意吗

请你谈谈你对你自己的认识

所谓封闭式问题就是每个问题给出若干个可能的答案，被试可以根据自己的实际情况从中选择一个作为答案。例如，

你对你目前的学习状况感到满意吗？

非 常 满 意 () 满 意 () 无 所 谓 () 不 满 意 () 非 常 不 满 意

你 对 自 己 的 工 作 感 到 满 意 吗 ？ 是 () 否 ()

在问题的内容上主要涉及三个方面的内容：有关实际行为方面的问题，例如：“你参加过心理学实验吗？”；还有有关态度和看法的问题，例如，“你对你这次的考试成绩满意吗”；最后就是有关个人背景的问题，例如要求被试填写有关年龄、性别、职业、文化程度等人口统计学资料。

（四）其他资料

除了上述内容外，问卷还包括一些有关的资料，如问卷的名称、编号、问卷发

放及收回日期、问题的预编码等。

补充讨论 8-1 问卷与测验的比较

问卷与测验都是心理与行为研究中最常用的收集资料的工具。两者在很多地方都有相似之处，甚至有的学者把问卷和测验等同使用（如，Rust & Golombok, 1989）。但是，从问卷与测验的定义来看，两者之间仍具有很大的差异。问卷和测验的差异主要表现在以下几个方面：

（1）编制过程

问卷的编制具有较大的灵活性，根据研究的目的不同，问卷的题目（问题）可采用多种形式；而测验的编制过程则要求非常严谨，必须遵守一定的规则与程序。这两种工具在编制过程上的差异，主要体现在标准化（standardization）的程度上，也就是问卷的编制通常没有标准化的过程，而测验的编制过程一定要遵循标准化的步骤和方法。因此，测验的编制过程中必须进行项目分析（item analysis）、信度（reliability）和效度（validity）检验等过程。而在问卷的编制过程中，则不一定进行这些检验过程。

（2）收集资料的性质

问卷可收集事实性的资料，也可收集如意见、兴趣和态度等心理特征的资料；而测验则基本上是收集心理特征的资料，例如智力、人格、兴趣和态度等。问卷和测验虽然都可以用来收集诸如兴趣和态度等心理特征的资料，但是问卷收集的资料都是针对特定的事件、问题或现象，如，你对将来最关心的事情是什么，你对领导的态度如何等等；测验所收集的资料，则大都具有普遍性与一般性，并不收集具有特殊性的资料。因此，问卷所收集的资料，大都比较具体和客观；测验所收集的资料，则倾向于抽象与主观。

（3）结果的分析

问卷主要收集比较有特殊性的材料，测验则收集具有普遍性的资料，因此问卷比测验的题目所涵盖的范围更广，也更复杂。问卷通常是一个题目测量一个概念，测验则常常是多个题目测量一个概念。所以，问卷的结果分析，常常是逐题分析，而且题目与题目的量化数据通常不具有可加性；测验结果的分析，则常常是把多个题目的量化数据加在一起，再进行统计分析。此外，测验还可以建立常模（norm）以供大规模的解释使用，问卷则通常不建立或无法建立常模，所以问卷的结果就具有局限性。因此，测验可以公开发表使用，而问卷则缺乏这方面的功能。

三、调查问卷的设计

(一) 问卷调查目的确定

一般来讲，问卷调查的目的包括两个方面。一方面是研究者利用问卷来收集资料。研究目的只是为了从总体上了解和描述研究对象某些方面的现状和特征，因此在问卷设计中要尽可能全面的收集有关研究对象这些方面的客观事实。以风笑天编制的武汉市居民生活质量调查问卷为例，该问卷的调查目的是了解武汉市居民日常生活中存在的困难和问题。该问卷调查的内容包括了个人及家庭特征、居住概况、邻里关系、交通状况、家庭生活等 5 项内容，所涉及的内容基本上涵盖居民生活的方方面面；另一方面，研究者进行调查的目的是要对某种心理或行为现象作出解释，即要探讨心理现象之间的相关性或因果关系，那么，所要收集的资料必须是验证某种理论或假设的经验事实。

(二) 问题与答案的设计

研究目的是问卷设计的核心，它决定着问卷的内容和形式。问题与答案是问卷调查设计中的核心问题。

1. 无结构式问卷设计

无结构式问卷设计，也称开放式问卷设计，就是不为被试提供可供选择的答案，让被试用自己的理解和语言来回答问题。例如，

“在职业选择时，你最看重哪些因素？最不看重哪些因素？”

“你与周围的朋友相处得怎样？”

“你对于自己的将来是怎样认识的？”

2. 结构式问卷设计

结构式问卷设计，也称封闭式问卷设计，研究者在给出问题的同时，还提供若干个可供选择的答案，要求被试从中作出选择。结构式问卷设计的问题与答案一般可以采用以下几种方式：

(1) 选择式 (choice method)。即要求被试从二个或三个（多个）备择的答案中选择出最符合自己想法的一项，这些项目必须是互相排斥（只能选一个）并且详尽无遗（每个被试都能回答）。例如：

你觉得你所在单位的领导在工作中采用了哪一种领导作风：

专制 () 民主 () 放任 ()

(2) 查核式 (checking method)。让被试从一群答案中把符合自己想法的答案都选出来。例如，

你通常喜欢收看的电视节目有：

新闻 () 科技 () 娱乐 () 体育 ()

(3) 是或否式 (yes-or-no method)。让被试以是或否，赞成或反对，对问题作

出回答。例如，

你 对 自 己 的 工 作 感 到 满 意 吗 ？ 是 () 否 ()

(4) 等级量表式，也叫评比量表式、Likert 量表式或点式量表：备择答案按一定的标准列出若干个等级，从中做出一个选择。例如，

很重要 重要 无所谓 不重要 很不重要

等级量表式的一种变式是，将答案列成一个连续体，仅在两端上标记号。等级的划分不能太少，也不能太多。太少缺乏区分度；太多令人无法回答。另一变式是，一个问题后面，给出一些答案，要求以一定的标准排列次序，例如 Rokeach 的价值观问卷，要求被试对所列出的诸如求知的需要、归属的需要、权力的需要等 18 种需要，根据其对自己的重要性，最重要的填 1 次重要的填 2 再次重要的填 3 依此类推，一直将 18 种需要的重要性都排出顺序号码。

(5) 等距量表式，一个问题后面给出等距排列的若干答案，让被试从中选择一个。例如，

年龄以 5 年或 10 年为一级 0 - 4/5 - 9/10 - 14 或 0 - 9/10 - 19/20 - 29。等距量表的答案按一般定义是连续的而不是离散的，答案种类过多往往在一个封闭式问题中容纳不下。因而被试往往选择组距两极端的。

(6) 对偶比较法 (method of paired comparison)，问卷中提出的项目，二个成一组，让被试按一定标准（如喜欢与讨厌或赞成与否等）进行比较，所有的各对项目都要进行比较（当然对题项的设计应当是简单、明确、易数量化的）。

例如：对中学生学习目的的调查，你每天在学校中是以什么心情学习的？就下面列出的两种意见组合逐对进行比较，并在符合你心情的项目上划 √ 不能肯定时，就什么也不写。

为了找到好的工作 - - - - - 没有明确的目的，只是无所谓地学习

在每天的生活中想取得成绩 - - - - - 为了培养高尚的品格

为了考上好大学 - - - - - 为了到好地方去工作

(7) 语义差别量表。用一系列描述所研究事物的互为反义的形容词来测试被调查者的态度。

例如：你对比萨饼的看法如何，请在下面每条虚线上的适当位置做个记号。

	1	2	3	4	5	6	7	
热的	-	-	-	-	-	-	-	凉的
味淡的	-	-	-	-	-	-	-	味重的
贵的	-	-	-	-	-	-	-	不贵的

润湿的	-	-	-	-	-	-	-	干燥的
没烤透的	-	-	-	-	-	-	-	脆口的
好的	-	-	-	-	-	-	-	坏的
无吸引力的	-	-	-	-	-	-	-	有吸引力的
新鲜的	-	-	-	-	-	-	-	陈旧的
小的	-	-	-	-	-	-	-	大的
天然的	-	-	-	-	-	-	-	人造的

采用语义法可以评价一个事物在不同维度上的得分，即可以发现事物在哪些方面是优点，哪些方面是缺点，其统计方法同于评比法。

3. 问题顺序的排列

问题的排列顺序影响被试的作答动机及反应。一般来讲，问卷中问题顺序遵循以下几个方面的原则：

(1) 敏感性或隐私性的问题尽量放在后面，放在前面易引起敌对、拒答。

(2) 一般性的问题或容易回答的问题放在前面，特殊性问题或不易回答的问题放在后面。由于特殊性问题或不易回答的问题费时，放在前面，令人望而生畏而导致拒答。

(3) 对后面有启发意义的问题，可以放在前面，但要防止产生反应定势。

(4) 对于时间顺序的问题，应按时间序列依次排列，以免被试的记忆受到干扰。

(5) 有时候为了避免影响被试的情绪，可以将个人基本资料的填写放在问卷的题目之后。

四、问卷调查的评价

问卷调查的优点主要表现在（董奇,1992）：

第一，问卷调查的实施较为经济，节省时间、经费和人力。由于问卷调查可以采用邮寄的方式进行，这样可以在短时间内获得大量的数据，而且也节省人力、物力。此外，问卷调查采用纸笔测验的方式，不需要昂贵的仪器设备，因此简便易行。

第二，问卷内容客观统一，问卷资料便于分析量化。问卷调查特别是封闭式问卷调查，是通过相同的问题和标准化的回答方式来进行，这样就能够在一定程度上避免实施过程中的一些误差因素，而且所获得的数据也比较容易进行定量

分析。

第三，问卷调查匿名性强，回答真实。问卷调查中可以不填写姓名，因此被试能够真实反映自己的观点和态度，特别对于那些涉及有关个人隐私等的敏感问题。

第四，问卷调查具有间接性，相互作用小。问卷调查大都是间接进行的，可以减少研究者和被试之间的相互作用，减少各种可能的心理干扰，有助于提高问卷研究的客观性，减少调查误差。

问卷调查的缺点主要有：

第一，灵活性不够。问卷调查中，为了使收集的资料便于定量处理和分析，常常采用封闭式的问题，因而难以适应每个被试的实际情况。如果被试没有经历过有关的情景，也就不知道自己的体会或感受，可能会出现随意回答的现象。

第二，问卷调查只适用于具有一定文化程度的人，因此使得应用范围受到限制。被试必须具有一定的文化程度才能阅读和理解问题的含义，并能领会作答的要求和方法。因此，问卷调查对于那些文化程度较低的群体就难以进行，从而影响问卷调查的使用范围。

第三，深入性不够。问卷调查通常只能研究一些比较简单、表面的问题，因此无法对复杂的问题进行深入的探索和研究。并且问卷调查的结果和实际的社会情况之间还可能会存在一定的距离。因此开放式问卷调查比封闭式问卷调查具有优势，不仅可以研究“是什么”，还能够进一步研究“为什么”的问题。

专栏 8-1 问卷设计中的常见问题

1. 题项的编写应切合研究假设的需要。例如，如果想要了解大学生的学习动机，问卷中的所有问题都应当与学习活动有关，如果问卷中出现了与环境保护有关的问题，就会让被试不知所措。

2. 题项的内容应明确，避免含混不清。例如，问题太一般化，令人无法回答，如“你经常看报吗？”。语意不清的措辞令人费解，如“你觉得我们的生活够好了吗？”。

3. 一个题项中包含两个或两个以上的概念或事件。例如，“当你遇到挫折时，你是否会努力不懈而且尝试用新的方法去解决。”这个句子就涵盖了“努力不懈”及“尝试用新的方法”两个概念，有时作答者只符合了其中的一个，这对于作答者将会造成困扰。因此像上述的例子最好将其改为两个句子：“当你遇到挫折时，你是否会努力不懈去解决。”及“当你遇到挫折时，你是否会尝试用新的方法去解决。”

4. 题项内容不能用俗语、专有名词或双重否定句来表达。

项目内容不应超出被试的知识水平和能力范围，要适合被试的理解力。例如，对中学生“你对当前教育改革的想法如何？”，像此类的问题已超出中学生作答的能力范围，将会使作答者无法反应。

5. 题项内容应避免社会禁忌和爱好，对于涉及到个人隐私的问题，最好避免直接提问。

6. 题项应避免暗示和诱导作用。如“你想得高分吗？”“专家认为自信心是事业成功的基本保证，你对自信心的看法如何？”

7. 题项内容最好避免容易引起强烈情绪联想的提问。因为强烈的情绪会影响被试的思考，甚至会产生有意的自我防卫，而做出不实的回答。例如，对学生的父母提出“你在孩子的面前是否对学校的老师进行过不好的批评？”这样的提问，就不太可能得到真实的答案。

8. 不要用假设或猜测的语句提问。假设的问题有时可以预测行动的趋向，却不易获得真实的结果。例如，问“你愿意加工资吗？”“假如你有 10 万元，你将如何来使用它？”这类问题往往难以获得真实的结果。猜测往往不准确，因为记忆本身就有误。如问每周看几个钟头的报纸，多少小时电视，都不可能得到真实的结果。

9. 避免不能答或不愿意答的问题。有些问题，虽然不涉及隐私，但却难以回答。例如，“你考试作弊吗？”

10. 是一般的提问还是提有针对性的提问。如果所需要的资料覆盖面大，可把问题提得大些；否则就小些，有针对性地提问。例如对需要的调查，对大学生的提问是“请你根据自己的情况对 18 种需要排序”，对教师的提问是“根据你对某学生的实际情况对 18 种需要进行排序”。前者是一般性的，后者是针对性的。

11 是直接提问还是间接提问。直接提问能得到真实回答的，就应直接提问。但有些问题不能直接问，直接问得不到真实回答，就应采取间接问。例如，关于政治价值观的问卷设计即如此。又如，问个人还是问集体，在措辞上应加以注意。有研究发现，如果先问“你认为，每个人定期 x 光透视胸部，是不是一个好主意？”再问“你是不是照过 x 光？”结果答“是”的，前一问有 96%，后一问只有 54%。实际上两个问题的性质几乎完全相同。但前一问是针对集体，后一问是针对个人的。中国人注重人际关系，这类问题的措辞应特别小心。

12 长句好还是短句好。应当说，只要题项的内容清楚准确，问卷的句子是越短越好，短句易使被试抓住要点，集中思想回答。否则很难进行归纳、回答。如 Morris 生活方式问卷，句子很长，被试很难掌握。又如 Gibson 设计的 Teacher Efficacy Scale(1984) 也是很长的句子，被试很难把握。

13 一张问卷列多少个问题？一般以 30 分钟做完为宜。

14. 处理好相关的问题。例如，由于被试的异质性，有的问题可能只有部分被试有这方面的经历而另一部分被试则可能没有。

第三节 个案研究设计

个案研究 (case study) 最早运用于医学，医生通过对每个病人的病历的详细分析，来判明病理与病因，并提出相应的治疗方案。个案研究也是心理学研究中使用历史最长的手段之一，并在心理学各个领域如发展心理学、教育心理学、社会心理学以及临床心理学中大量应用。例如，个案研究法最常被应用于临床心理学的研究之中。临床心理学家通过个案研究来对某一个人（情绪障碍或社会适应有困难）做深入的研究，以了解引起适应困难的可能原因，为心理治疗提供依据与参考。在教育领域中，对学校中优差生的诊断和辅导等等都可以采用个案研究的方法。

一、个案研究的含义

个案研究的对象是单个的个体，也可以把社会机构或社会组织作为观察单位，通过对个案的深入调查，来研究或探讨与之相关的问题。例如，费斯廷格等 (Festinger, Riechen, & Shachter, 1956) 曾进行过一个有趣的个案研究。他们混入了一个等待世界末日来临的小团伙。该团伙的成员始终期盼着与外星人联系，而且据说有一个成员已经接到了外星人的谕示——地球即将毁灭，因此团伙里的每个成员都等待着灾难降临之前能被外星人的宇宙飞船救走。观察中特别令费斯廷格等感兴趣的是，当预言的灾难没有发生的时候，这个小团伙的表现会怎样呢？通过观察，他们发现，许多成员在预言的灾难日期过去后，妄想程度不是减弱，而是增强。

二、描述型个案研究

描述型个案研究是指搜集单个被试各方面的资料并进行分析的方法。搜集资料的内容虽然根据研究问题而有所侧重，但一般都很重视被试从出生到现在的生活史、家庭关系、生活环境及人际关系的特点等资料。根据需要，也常对被试进行智力和人格测验，从熟悉被试的亲近者那里了解情况，或者从被试的书信、日记、自传或者他人为被试写的资料（如传记、病历等）进行分析。总之，这种方法就是围绕着一一定的目的，搜索单个被试过去和现在各个方面的资料，然后进行周密的分析。

实例学习 8-1 珍妮的信件

美国著名心理学家奥尔波特（Gordon Willard Allport, 1897—1967）认为对人格特质的研究，就是要了解现实的、特定的人，唯一的办法就是研究这个人本身。怎样才能研究某个特定个人的人格呢？他认为行之有效的办法就是通过个案研究，利用个人诸如日记、自传、信件等资料来分析个人的人格特质。下面是奥尔波特通过对一个名叫珍妮给她儿子罗斯的朋友来往信件的个案分析来研究其人格特质的事例。

珍妮 1868 年生于爱尔兰，五岁时移居加拿大。后来同一个美国铁路巡警结婚。在她的儿子罗斯出生前的一个月，丈夫去世。珍妮以艰辛的劳动维持着母子俩人的生活，并把自己的全部爱都倾注在儿子的身上。随着儿子的长大，特别是在罗斯服役结束返回家后，母子两人经常争吵。并且罗斯隐瞒着母亲结了婚。为此，珍妮把罗斯赶出了家门。几年之后珍妮动了恻隐之心，开始去看罗斯并相互通信，直到罗斯过早的去世为止。下面是珍妮写给罗斯两位朋友几封信的一些内容。

珍妮 1928 写的第一封信是罗斯抛弃了他的妻子并与另一女子有了关系时，她感到失望，求助于罗斯的两个朋友，信中解释了她与罗斯之间冲突与对抗的原因。

但是在罗斯和我之间并不是为了钱，绝对不是，是女人，更多是女人。（我字迹难看——我很神经过敏。）有时我见罗斯的时候有点丧失常态——性狂。最初他高兴地谈着积蓄钱，赞扬生意上的特点这类事，我乐得像在天堂那样。他存了钱，在银行里以我们共同的名义存有几百美元。这就是我如此节俭的原因。但他一直与一个女人保持暧昧关系……是很坏的暧昧关系，在他从那里走出来之前，他是那样怕动笔。我帮他解脱了出来……

现在，这个玛丽，是托莱多的屠夫的女儿……在罗斯抛弃他妻子时就给玛丽写信。〔现在〕玛丽说今年冬天来纽约并租一间房子。罗斯和我发生了口角——我拒绝那样做。玛丽永远不能进我住的房子……

我并不打算说这么多，但我心酸、患病了，真的失去了勇气。罗斯绝对不关心我的任何事情——我对他来说是一个巨大的障碍和负担。

在 1929 年的第二封信中，珍妮提到罗斯健康状况很糟，耳朵感染动了手术，医生发现他内耳的大片乳突炎和大脑表面有脓肿。在信中她把罗斯的女友称为娼妇。

罗斯上周说要见我——他仍在医生手中。他看起来身体很差。我请他……与我吃饭，他来了……那是一流的午餐，看起来他很喜欢。

我知道……使罗斯醒悟的方法就是给他想要的东西然后别再管他。

现在我但愿去年五月他把那娼妇带到布朗克斯时我就离开他。但即使现在他娶了她（她可以强迫他那样），他也不会与她共同生活很久。那是我的地方，我为什么要邀请他进午餐——这样他可以不很孤独直至他找到新的娼妇。真的，我知道我对他没有一点用处——他奚落我、鄙视我，但是血缘联系仍然存在，他相信他对此不是没有感情的。

罗斯没有康复，不久便死了。在 1929 年底的信中，珍妮对罗斯女友的敌意爆发了。

虽然这娼妇完全泡在泪水里，当然心碎，但并不因悲伤流泪而忘了物质财富在这个尘世上的重要性。瞧！她说罗斯的衣服和罗斯的车子……[她声称她是罗斯最亲密的亲属]如果她是罗斯最亲近的亲属，那她就会得到救济金，那真是一场足以令人笑死的悲剧。她认识罗斯仅六个月。他们的“伟大罗曼史”是从二月开始的——他们的作为肮脏下流——这下流卑鄙的街头野狗。她在道德上和肉体上杀死了罗斯。

奥尔波特请了 36 位鉴定者来阅读珍妮的信件，一共用了 198 个特质名词来描述珍妮的人格特征，把相同的特质合并之后，共发现了八个主要特质。后来有人(Baldwin,1942)从统计学的角度重新分析了奥尔波特的资料，进一步证实了这八种因素特质。佩奇（ Paige,1966）用计算机对珍妮信件进行复杂的因素分析也得出八种因素特质，见表 8-3。

表 8-3 由两种方法分析出的珍妮的中心特质

常识特质（奥尔波特）	因素特质（佩奇）
1. 好争吵的—多疑的；攻击的	攻击性
2. 自我中心的（占有欲的）	占有欲
3. 感情用事的	归属需要；家庭认可需要
4. 独立的一自主的	自主需要
5 美学的—艺术的	感觉性
6. 自我中心的（自我怜悯的）	殉道性
7. （无比较项）	性欲
8. 怀疑人生价值的一病态的	（无比较项）
9. 戏剧性的一紧张的	（“爱夸张”）

（采自黄希庭，2000）

三、实验型个案研究

个案研究设计也可以与实验研究设计相结合。在个案研究中，根据研究的假设，通过操纵自变量，控制无关变量和对因变量的观测，来推断自变量对因变量的效果。例如在个案研究中有一种叫做变异 - 个案研究 (deviant case study) (Robinson, 1976) 设计。使用这种方法时，研究者需要安排两个既极其相似又有所不同的个案。巴特斯和塞尔马克 (Butters & Cermak, 1986) 曾采用变异 - 个案研究分析了一个世界著名科学家的案例。这个世界著名的科学家 P. Z. 因长期酗酒而患有科尔萨科夫综合症 (Korsakoff's syndrome)，并于 1981 年出现了严重的记忆丧失症状 (遗忘症)。他很难记住新的信息和回忆过去的事情。研究中的另一个个案是与 P. Z. 同年龄的但没有酗酒史的同事 (变异 - 个案分析的比较对象)。巴特斯和塞尔马克把这两个人对于往事的记忆进行了比较，结果发现，对比个案没有表现出像 P. Z. 那样严重的记忆丧失。因此，研究者认为长期酗酒是导致 P. Z. 丧失记忆的一个重要因素。此外 P. Z. 对新信息的记忆丧失的症状与其他科尔萨科夫综合症的患者非常相似。实际上这种把个案的行为和其他人进行对比研究的技术，已经具有实验研究的雏形了。

单个被试研究 (single - subject research) 是在个案研究的基础上，通过严格控制的条件来操纵自变量，从而观察实验处理对行为变化的影响。这是一种兼有实验研究和个案研究特点的研究设计。像实验研究那样，这种设计允许对自变量进行操纵，实验者在控制的条件下对被试的行为进行客观的测量。但是，由于这种方法只处理单个被试，因而又与个案研究相似。这种研究设计有助于研究个体的独特性，同时也能为今后的实验做些有益的探索。

(一) A - B - A 设计

A - B - A 设计是在不同实验处理的条件下，对单个被试的行为进行系统观察的一种研究策略。在 A - B - A 设计中，首先在 A 阶段观测行为发生的频率，作为基线。在 B 阶段，施加实验处理，在实验处理结束后，继续观察并记录行为的变化作为第二个 A 阶段。

下面我们以哈特及其同事 (Hart, et al, 1964) 对幼儿园一个 4 岁儿童的研究为例来进行说明。该研究的对象是一个叫比尔的 4 岁儿童，该儿童看上去很健康，但却过分能哭。研究者认为是成人的注意强化了比尔哭的行为。于是他们采用 A - B - A 设计来验证这个假设。在该设计中，A 表示行为矫正前 10 天比尔每天平均哭 5 ~ 10 次，作为基线状态，B 表示 10 天的行为矫正过程，矫正措施包括：不理他的哭声，用注意来奖励比尔的适应反应 (如，摔倒后不再躺在地上等)。如图 8.2 所示，这一阶段比尔哭的次数逐渐减少。这时，研究者已完成了 A - B - A 设计的 A - B 阶段。但是这个时候还不能确信行为的改善是由实验处

理的效果，也许比尔行为的变化与这段时间内与全体同学能友好相处，或其他什么原因等。

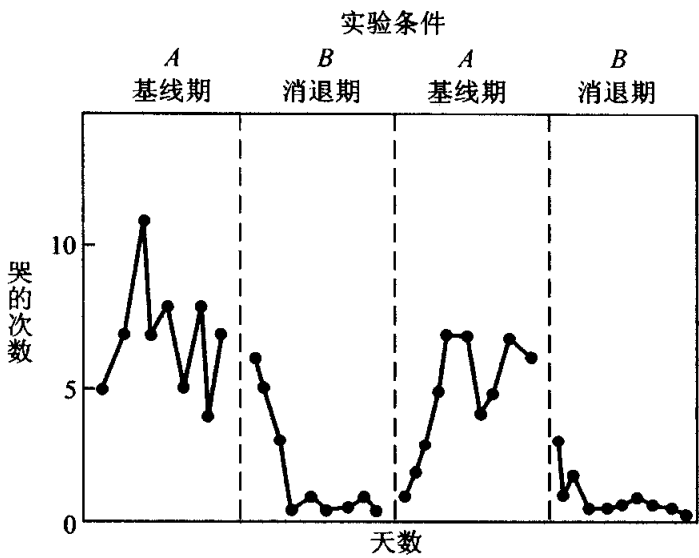


图 8.2 比尔在四个阶段上哭的次数

为了获得更充分的证据，研究者撤掉实验处理，使其行为返回到基线水平，即第二个 A 阶段。由于再次关注比尔哭的行为，所以其行为开始退回到矫正前的状态。如图 8.3 所示，比尔仅用了 4 天就恢复到了原有的水平。因此，研究者可以得出结论，正是由于强化而不是其他的原因，导致了第一个 B 阶段中哭的行为的终止。最后，由于这是一项治疗工作，因此研究者又通过强化适应行为来重新使比尔的问题行为得到控制。

(二) 多重基线设计

A - B - A 设计要求 B 阶段的处理效果不能持续很久。例如，如果一个实验处理导致某种行为保持在一个很高的水平，而不能返回到基线阶段的水平，采用 A - B - A 设计就不太可行。再有某些条件下采用 A - B - A 设计不太符合科学道德标准。例如，通过操作性条件反射原理帮助某个儿童消除了某些不良行为（如酗酒、攻击行为），如果再通过撤除实验处理来观察儿童的行为问题是否能够恢复到基线水平，显然这样研究违背了科学道德。而多重基线设计恰恰可以解决上述问题。

多重基线设计假设，不同的行为（不同的被试）在引入实验处理之前都有不同的基线期。如图 8.3 所示，垂线表示引入实验处理的时刻，在其左边表示基线期，右边表示处理期。当行为 A 被试 A）正在接受实验处理时，行为 B 或 C 被试 B 或 C）仍处在于基线状态在接受实验处理之前仍保持稳定（如图 8.3 中的垂线 A 到垂线 B 或 C 这段时间），并随着实验处理的引入而发生显著的变化，就

可以得出实验处理和行为改变之间的关系。

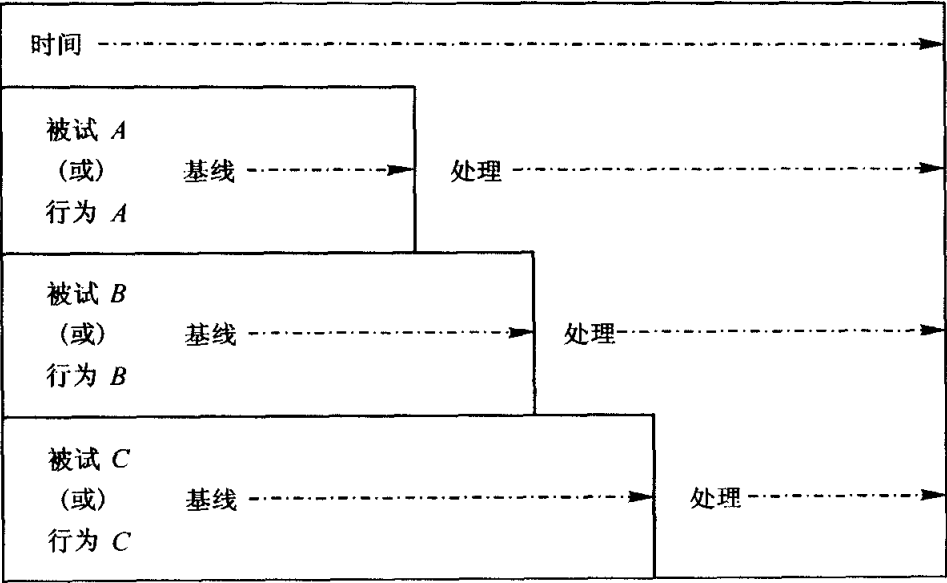


图 8.3 多重基线设计结构

（采自坎特威茨等,1997）

注：不同的被试（或行为）有长短不同的基线期，垂线标志实验处理的引入。

多重基线设计根据研究对象的不同可分为被试内多重基线设计和被试间多重基线设计。如图 8.3 所示，如果在不同的基线阶段比较同一被试的不同行为，就是被试内多重基线设计（行为 A、B、C 分别代表同一被试的三种行为）；如果在不同的基线阶段比较不同被试的同一行为，就是被试间多重基线设计（被试 A、B、C 分别代表三个被试的同一行为）。

一个多重基线设计的理想结果可以用图 8.4 来表示。该研究选用同一被试的 3 种不同行为，并分别建立基线。首先，对行为 1 引入实验处理，但不考虑其他的两种行为。如果实验处理是有效的，则可以观测到行为 1 的变化，但是行为 2 和 3 则保持稳定。其次，对行为 2 引入实验处理，则可以见到行为 2 的变化，而行为 3 保持稳定。最后，对行为 3 引入实验处理。

例如在上例哈特及其同事（Hart,et al,1964）行为矫正比尔不适行为的个案中，如果采用多重基线设计，除了哭的行为之外，研究者还要观察比尔的另一种不适行为，如摔打等。研究者通过设计使这两种行为的基线期不同，先对哭的行为进行矫正，几天后再对摔打行为进行矫正。如果在哭的矫正之后和摔打行为矫正之前，摔打行为保持稳定，并在对摔打行为的矫正后发生显著的变化，就可以认为实验处理具有显著的效果。由于哭和摔打的行为总是联系在一起，因此对一种行为的矫正也会影响到另一种行为的发生。也就是，对比尔哭行为的矫

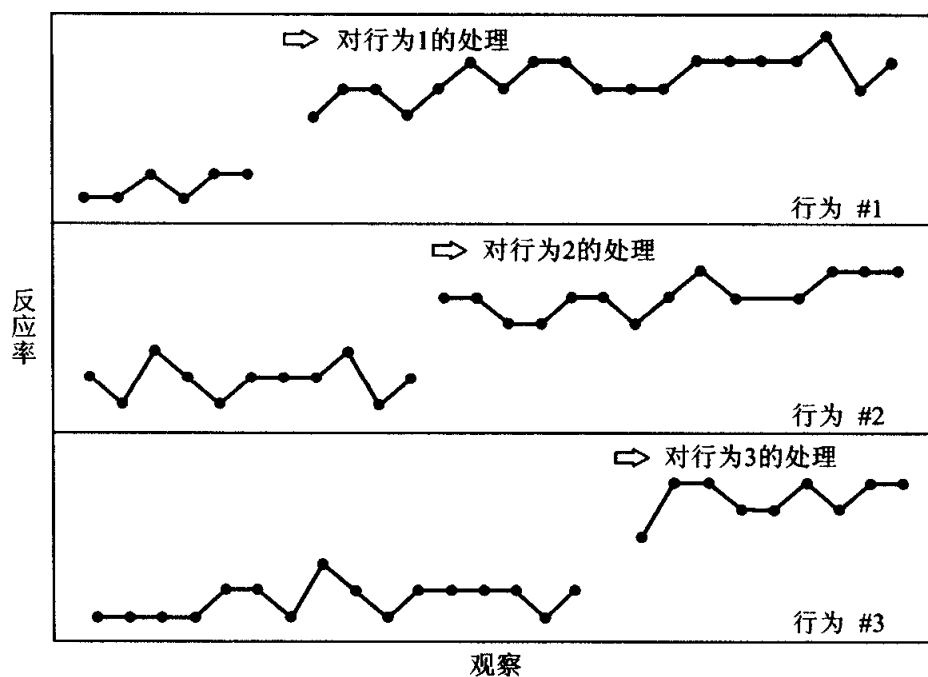


图 8.4 多重基线设计的理想结果

正，摔打行为也会随之减少，这样就不能把摔打行为的改变归因于行为矫正的效果了。因此，研究者采用了被试间多重基线设计。实验前，先找到一个和比尔行为表现十分相似的儿童，然后在一段时间内对比尔哭的行为进行矫正，几天后再开始对另一儿童的摔打行为进行矫正，如果该儿童摔打行为在矫正之前保持稳定，就可以认为矫正行为改变了比尔的问题行为（如图 8.5 所示）。

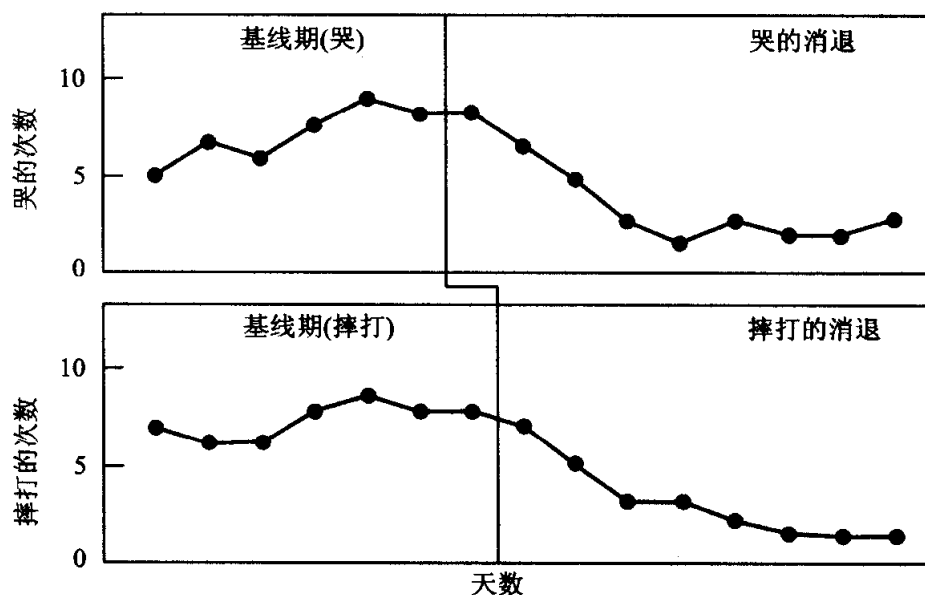


图 8.5 有关比尔问题行为的多重基线实验的假想结果

四、个案研究的评价

个案研究的优点是：

第一，个案研究方法有助于对个体规律的探讨。能加深对特定个人的了解，因而被一些心理学研究者所推崇。例如，对于智力超常儿童的个案研究，由于每个超常儿童的成长背景不同，通过对其个人的发展历程及其生理、心理特点的分析，可以探讨智力发展的规律及遗传与环境的关系等问题。

第二，个案研究方法有助于对事件变化过程的深入探讨。通过与所研究的个案有关的事实深入细致的调查研究，研究者可以积累丰富的资料，从而可以了解事物的变化过程，透过现象探求事物的本质。

第三，个案研究也往往有助于研究者获得某种假设，然后依据这个假设做进一步的研究。精神分析的理论大都是在个案研究的基础上得出的，并且该理论中的某些合理因素也被后来的心理学家的研究所证实。

第四，个案研究方法一般只涉及少数的研究对象，可以节省人力、物力和财力。

个案研究的局限性是，由于个案研究的对象数量较少、代表性差，因此所收集的资料往往缺乏可靠性，也很难得出具有普遍性的规律和结论。很多个案研究基于对回溯式数据（retrospective data）分析，而这种通过对过去经历的回忆所提供的信息带有很大的主观性，例如，个人对过去事件的回忆具有很大的主观性，个人写的日记和自传也往往缺乏准确性，他人为被试写的传记也会因为记忆误差和情感上的好恶而失真，因此影响了个案研究结论的准确性，这也是个案研究的结果遭到质疑的主要原因。

第四节 跨文化研究设计

近年来美国权威心理学刊物《美国心理学家》和其他一些刊物上连续刊发了一些文章，对传统心理科学模式的缺陷和弊端进行了反思，表现出对于文化影响日渐增加的关注。与此同时，心理学的各分支学科也日渐关注文化因素的影响，如认知心理学的生态模式认为认知活动发生于一定的文化背景中，文化因素影响和制约着认知活动；儿童心理学的研究也表明，几乎儿童的每一步成长都印有文化的痕迹；健康心理学也日益重视文化因素对心理疾病的影响。

一、跨文化研究的含义

跨文化研究是运用比较的方法，从文化的普遍性和特殊性的角度来分别考察各种心理产物、理论和行为模式的普遍性和特殊性。因此跨文化方法从一个

新的角度提供了对行为的解释，使心理学可以从文化的角度来看待行为，有助于对行为本质的理解。此外，通过对两种或两种以上文化的分析比较，可以对心理学的概念和理论进行比较性的检验，从而可以验证这些研究结论的效度，改善研究结果的质量。例如，弗洛伊德通过对梦的分析，认为儿童具有仇视父亲、依恋母亲的“恋母情结”。但是对于巴布新几内亚群岛上的土著人的研究表明，儿童在梦境中仇视的对象是舅舅，而不是父亲。因为，在那里舅舅是儿童行为的管教者。为什么土著儿童不仇视自己的父亲，而仇视舅舅呢？这说明儿童潜意识里有一种摆脱权威控制的愿望，谁代表了权威，谁就成为仇视的对象，而与儿童的情欲发展无关。跨文化的研究为纠正弗洛伊德的理论提供了重要的证据。

虽然许多心理学家已经意识到文化与行为的联系，但是在心理学研究中如何把文化因素纳入到研究设计之中一直是个难题。跨文化研究的基本前提是以全世界各种不同文化为样本，对其资料进行比较研究，以验证对人类行为的假设。例如心理学家斯尔（Robert Sears）等人在1950年前后曾研究美国Kansas市儿童教养与人格发展之间的关系，斯尔等人探讨了儿童的断奶时间与儿童情绪之间的关系。在以80个婴儿为观察对象的研究中，发现断奶时间越晚，儿童情绪上的困扰就越大。但是威廷（Whiting）认为，这个结果只能反应一个民族的情况，而不能代表所有文化下的结果。随后，威廷采用相同的研究程序，对75个不同文化背景下的研究进行了比较分析。结果发现，不同的文化背景下，两者之间存在不同的关系。在大多数情况下，婴儿断奶的时间越往后，引起儿童情绪不安的情况就越少。从上述例子可以看出，跨文化研究不但可以使研究者免去单一文化狭窄范围的局限，而且可以发现更多、更广的人类行为的幅度。

通常情况下，凡是把两个或两个以上不同文化来进行比较研究，都可以称作“跨文化研究”。但是，在实际研究过程中，大多数研究者都希望尽可能的选取较多的文化样本，以便能做出有意义的统计结论。

二、跨文化研究的两种取向

文化普遍性（etic）和文化特殊性（emic）是跨文化研究的两种主要研究取向。emic和etic是由语言学家派克（Pike, 1967）从phonemics（音素学）和phonetics（语音学）衍生出来的两个概念。

文化普遍性取向是将一个文化中的行为与另一种文化或所有文化中的行为进行比较，找出所有文化都适用的有效原则，并确定比较不同文化中人类行为的理论框架。从事普遍性研究的研究者站在外部客观的立场上对两种或尽可能多的不同文化群体进行研究，以比较不同民族或文化群体的各种行为，其比较文化行为的标准是绝对的或普遍适用的。普遍性的研究取向的意义在于（1）它可以给人类行为的共同性和差异性提供一种基本的解释，增进人们对人类行为的整

体理解。(2)通过这种研究策略的实施,可以发展出标准的研究程序和方法,提高研究的技术水平。(3)可以作为研究者了解不同文化的基础,因为在研究的初始时,研究者对其他文化没有基本的了解,这种策略可以提供有效的理论假设。(4)普遍性策略在选定的文化之间进行比较,较之深入文化内部进行研究更为节省人力和物力。例如,皮亚杰认知发展阶段理论的跨文化研究发现,虽然不同民族在各个认知发展阶段出现的年龄阶段上存在差异,但各个发展阶段的出现的先后顺序却存在着跨文化的一致性。

文化特殊性的取向只使用某一特定文化的方法、概念和行为解释,运用地方性标准和具有特殊文化意义的理论框架来分析本土文化内的心理问题,因而不具有文化上的普遍性。特殊性研究取向的意义在于,(1)这种研究取向使得研究者可以在文化内部对文化的建构方式、文化与行为的联系进行整体的理解。(2)研究者不是把文化作为一个外在于自己的对象,而是将其本身作为文化的一个部分,因而可以产生对文化的特殊心理体验。(3)它使得研究者可以在日常生活中理解人们的行为,在行为发生的情景中理解行为,在文化生活中理解态度、兴趣、气质、人格、情绪和意志等等。(4)如果承认行为与文化的联系,那么特殊性取向真正把心理学建立在了文化的基础上,因为文化是具体的,行为和文化的联系是与具体文化的联系,而不是与抽象文化的联系。例如,我国台湾学者杨国枢等倡导中国化心理学或中国人的本土心理学,提出心理学要研究中国人的国家民族行为模式,如道德判断、“面子”问题、社会行为、人际交往等就属于文化特殊性的研究取向。

这两种研究取向是相互联系、不可分割的。因此普遍性取向和特殊性取向是从两个不同的角度研究同一种行为,其关系是互补的、共生的,而不是对立。但是,长期以来,跨文化心理学一直是以普遍性策略作为指导思想,特殊性策略至于边缘的地位。(叶浩生,2004)

跨文化研究的最好方法是本国成员与外来成员的合作,前者收集数据并从“局内人”的角度去解释这些材料,而后者可以从“局外人”的位置“向内看”来考察和解释这些材料(Adler,1982)。为了使跨文化研究有意义,第一步是在文化内部进行文化特殊性研究,然后再补充不同文化间的比较研究。

三、跨文化研究的等值性

在跨文化研究中首先要解决的就是可比性的问题。而具有可比性的前提就是要有文化的等值性。玛尔帕斯等(Malpass & Poortinga,1986)曾区分出三种文化等值性:概念等值性、测量的等值性和心理机能等值性。

(一)概念等值性

概念等值性指将某种语言表达的概念用另一种语言准确恰当的表达出来的

程度。所谓概念等值性不是指两句话在语词上的一致，而是指两句话在意义上是等值的。通常情况下，测验项目通常都是陈述句或疑问句，以及两种或两种以上、由编制者准备好的答案选择，供被试者进行选择。因此如何准确地翻译出项目语句的原意，就是概念等值性问题。有时候词语一致意味着概念一致，如将“*I am nervous*”译为“我紧张”。但有时候词语一致但意义有差别，例如 MMPI 中有一项目内容涉及在西方家喻户晓的童话《爱丽丝梦游仙境》，但多数中国人或许不知道这个童话，或许只知其名而不知其内容，如果照搬照译的话，这一项目对中国被试来讲就失去了测量意义。MMPI 中文译者采用《西游记》来代替《爱丽丝梦游仙境》，前者为中国著名的神话小说，从而使该项目内容能为中国人所接受。

提高概念等值性可采用往返翻译 (translation and back-translation) 的策略。通常境况下，先由一个双语者将自陈量表从原本语言翻译为译本语言，然后再由另一个双语者将译本语言翻译回原本语言。此外，也可以采用双语双答的方式，即要求一组双语被试 (熟练掌握两种语言) 对两种语言版本的问卷进行作答，然后检验被试在两种版本上反应的一致性。

(二) 测量的等值性

测量的等值性是指用自陈量表测得的数据在心理测量学指标上呈现出一种可比较的模式 (齐茨、方富熹, 1991)。这里主要指翻译过来后的量表与原量表在信度和效度上的一致程度。不过，测量的等值性首先是样本的等值性，不同文化中的被试应当具有近似的背景。否则，两类群体中的差异就很难归因于文化的差异还是抽样的问题。例如，在美国和非洲撒哈拉群岛上分别抽取大学生，就不能保证被试样本的等值性。因为前者的入学率达到 70.4%，而后者只有 2.1%，因而后者是经过更严格挑选的群体。

一般来讲，一个测验在其主体文化里都具有很好的信度，但转移到客体文化时信度可能会发生很大变化。通常情况下以 α 值的高低来表示量表内部一致性信度高低，但是 α 不能独立于量表所包含的项目数以及被试样本的大小 (Pedhazur & Schmelkin, 1941)。因此项目一总分相关系数 (item-total correlation) 就成为考察内部一致性的重要指标。另外，量表项目间平均交互相关 (average inter-item correlations) 也成为衡量量表信度的重要指标，若量表的项目平均交互相关介于 0.15 至 0.50 间，则该量表有较好的信度 (Clark & Watson, 1995)。近年来，验证性因素分析也成为衡量量表信度高低的的重要方法 (Joreskog & Sorbom, 1993)。

如果要考察两个量表的效度的等值性，可以采用因素分析的方法。首先通过探索性因素分析来考察两个问卷是否具有相同的因素数目，在此基础上考察各个项目是否对应于相同的因素。随后根据原量表中的标准，把因子负荷低于

标准的项目删除。这时所保留下来的项目与原量表具有相同的测量学特征。最后，采用结构方程模型的方法验证保留项目量表的测量模型和结构模型（张力为,2002）。

（三）心理机能等值性

心理机能等值性是指不同文化背景的人在测验、问卷、任务中反映出的行为所表现出基本相同的心理机能的程度。相似的行为可能表现出不同的心理机能或不相似的行为表现出相同的心理机能（齐茨、方富熏,1991）。Dragnes(1986)的抑郁症的例子表明，虽然普遍认为抑郁症的表现有：悲伤、紧张、焦虑，对事物不感兴趣，注意力不集中等，但在不同的文化背景中存在显著的差异。例如，罪恶感 (guilt)不是普遍存在的，在非洲的南撒哈拉人就没有罪恶感的表现。但在西方和日本的患者中罪恶感则是抑郁症的重要成分。日本人谴责自己直接冒犯了其他人，而德国患者则责备自己破坏了道德原则。再如，有研究发现中国被试在 MMPI 临床量表 2(D)及 8(SC)上的得分显著高于美国被试 (Cheung, 1985 ;宋维真,1982)，这可能反映了被试反应倾向上的文化差异，而不能由此结论说，中国人较美国人更抑郁，更倾向于精神分裂症的人格特质。

四、跨文化研究的类型

Vijver 和 Leung(1997) 根据是否具有理论定向和是否考虑背景因素将跨文化研究分为两类 4 种。理论定向可分为假设检验 (hypothesis testing) 研究和探索 (exploration) 研究。假设检验研究根据是否考虑背景因素 (contextual factors) 可分为推广研究 (generalizability) 和理论驱动研究 (theory driven)。同样，探索研究根据是否考虑背景因素 (contextual factors) 可分为心理差异研究 (psychological differences) 和外部效度研究 (external validation) 见表 8-4。

表 8-4 跨文化研究的类型

是否考虑背景因素	理论定向	
	假设检验研究	探索研究
否	推广研究	心理差异研究
是	理论驱动研究	外部效度研究

（采自 Vijver & Leung,1997）

假设检验研究从有关的理论出发形成有关的假设，并通过收集各种不同文化背景中的数据来验证相关的研究假设。其中，推广研究主要是了解一个文化背景中得到的结果能否推广到另外一个文化背景中。这种研究没有考虑文化背景因素的影响。在理论驱动研究中，文化背景因素成为理论假设中的一个成分，

来考察文化变量与其他变量之间的关系。

探索性研究通常是研究者并没有什么预期的假设，因此很多无关的因素可能导致不同文化背景下结果的差异。心理差异研究主要是指通过具有文化等值性的工具来探索不同文化背景中反应的相似性和差异性。由于没有预期的假设，因此如果发现存在不同文化背景中反应的差异，通常对这种文化差异进行事后的解释（post. hoc）。外部效度是通过文化背景因素探讨跨文化差异的原因。

例如，张妙清、宋维真等对中国人人格结构的探讨。由香港与大陆的学者广泛搜集有关华人性格特质的概念，其中既包括由经验表明与中国文化相融合的西方人格概念，如内 - 外向、内 - 外控等，也包括西方人格理论中没有包括却为中国人社会广泛接受和使用的本地化个性概念，如讲人情、讲和谐、讲面子、节俭等。从所搜集到的人格概念中，再归纳、分类出若干种重要的人格构念（Cheung, 1996），形成测验编制的基础。然后，再为每一个人格构念挑选、编制测验项目（其形式是中文陈述句）。项目的来源甚为广泛，有已经发表过的个性研究报告，有中国经典和近代著名小说，还有各类民俗谚语集等等。最后，形成评鉴中国人人格结构的工具。

补充讨论 8-2 你的母亲在你心里吗？ 关于自我概念的一项跨文化比较

自我（self），也称为自我概念（self-concept）被认为是一种编码、储存和提取个体信息的认知结构。中国人和西方人关于自我概念存在显著的差异。中国人的自我概念依赖于同他人的关系，例如，与父母、好友、同事等关系。而西方国家则强调绝对的独立的自我，不包括好友、父母、同事等。朱滢、张力一项研究表明，中国人的自我概念中包含母亲成分，而美国人、英国人则不包含。在美国人的自我意识中，“母亲”和“克林顿”这两个概念都属外界事物。实验中，中国被试分成自我组、母亲组、鲁迅组和语义组 4 种实验条件，要求被试对人格形容词在 4 点量表上进行判断。结果表明，母亲组与自我组无论在总再认率，还是在细节回忆和熟悉感判断上均无显著差异，二者的交互作用亦不显著。而类似的实验程序在西方被试得到的结果与此却有很大的差异。Conway 等人采用英国被试，发现自我组与母亲组之间存在着显著差异，对于细节回忆，自我组成绩高于母亲组，而对于熟悉感，母亲组成绩高于自我组，而且二者之间的交互作用是显著的。Keenan 等人在一个以美国人为被试的研究中也发现自我组的记忆成绩与父母组和语义组都存在显著差异，而父母组和语义组之间没有显著差异，这说明西方被试父母组的加工类似于语义加工。总之，中国被试自我组与母亲组是类似的，西方被试的自我组成绩高于母亲组的成绩，这从行为实验上反映了上述

东西方自我概念的差异。通过上述实验研究者推论，中国人的关于母亲的记忆对应的脑区与自我的记忆对应的脑区可能接近或重合，而西方人的“母亲区”与自我区分别处于大脑的两个不同的半球上。

►本章提要

1. 访谈研究是研究者通过与研究对象的交谈来收集有关对方心理特征与行为的数据资料的研究方法。它是心理学研究中运用最广泛的研究方法之一。访谈研究设计的主要内容有访谈目的的确定、访谈问题形式的设计、回答方式的设计以及访谈结果的分析。

2. 问卷研究是通过书面的方式，根据严格设计的心理测量题目或问题向研究对象收集资料和数据的一种方法。问卷的结构包括有标题、指导语、问题与答案和其他资料。问卷研究设计主要包括问卷调查目的的确定、问题与答案的设计、问题顺序的排列等内容。

3. 个案研究的对象是单个的个体，也可以把社会机构或社会组织作为观察单位，通过对个案的深入调查，来研究或探讨与之相关的问题。

4. 描述型个案研究是指搜集单个被试各方面的资料并进行分析的方法。搜集资料的内容虽然根据研究问题而有所侧重，但一般都很重视被试从出生到现在的生活史、家庭关系、生活环境及人际关系的特点等资料。

5. 个案研究设计也可以是与实验研究设计相结合，在对个案的研究中，根据研究的假设，通过操纵自变量，控制无关变量和对因变量的观测，来推断自变量对因变量的影响。主要包括 $A-B-A$ 设计和多重基线设计。

6. 跨文化研究是运用比较的方法，从文化的普遍性和特殊性的角度来分别考察各种心理产物、理论和行为模式的普遍性和特殊性。

7. 文化普遍性和文化特殊性是跨文化研究的两种主要研究取向。其中，文化普遍性取向是将一个文化中的行为与另一种文化或所有文化中的行为进行比较，找出所有文化都适用的有效原则，并确定比较不同文化的人类行为的理论框架。文化特殊性的取向是用对某一特定文化成员有意义的概念来描述行为，考虑到这一文化成员本身的价值观及其它它们所熟悉的事物。

8. 跨文化研究中存在三种文化等值性：概念等值性、测量的等值性和心理机能等值性。

►本章关键术语

访谈研究 (interview survey)

ABA 设计 (ABA design)

问卷研究 (questionnaire survey)

多重基线设计 (multiple baseline design)

个案研究 (case study)

跨文化研究 (cross-cultural study)

单个被试研究 (single-subject
research)

文化特殊性 (emic)

文化普遍性 (etic)

►复习与练习

1. 心理与行为研究中，为什么要采用访谈、问卷、个案和跨文化的研究设计？
2. 访谈研究设计的主要内容有哪些？
3. 问卷研究设计的主要内容有哪些？
4. 如何进行实验型个案研究？
5. 跨文化研究设计的两种研究取向是什么？

►推荐参考读物

董奇. 心理与教育研究方法. 广州：广东教育出版社，1992

袁方. 社会研究方法教程. 北京：北京大学出版社，1997

王宏印. 跨文化心理学导论. 西安：陕西师范大学出版社，1993

►在线资源

http://www.ac.wvu.edu/~culture/contents_complete.htm

第九章 质的研究设计

很多心理现象无法量化，或者用量化的方法无助于我们理解其内涵。例如，用数字来描述意义、爱、愤怒无助于抓住其本质，而意识、人格、自我、文化等问题也无法通过基因、信息加工得到完全的解释。对于这类问题，如何看到数字之外的另一面呢？又如，有的人担心人类未来会生活在科幻电影所描述的冰冷世界，因为凡是数量化的均可以通过机械来控制。是否真的如此呢？我们如何进行心理学的说明呢？除了数量化，还能科学地研究心理吗？对于这些问题的解答，将用到质的研究方法，在本章中，我们将讨论以下问题：

1. 质的研究的含义
2. 质的研究设计的要素
3. 质的研究的设计模式
4. 常见的几种质的研究类型

第一节 质的研究概述

科学心理学创始人冯特（Wilhelm, Wundt, 1832—1920）在其民族心理学研究中运用的描述和解释就是质的研究。质的研究包含多种具体的研究方法，是一种实证性的研究取向，它具有归纳和解释的优势，最近为更多的心理学研究者所认同，广泛应用于文化、民族、消费、组织、社会、发展与教育、咨询、健康、犯罪等领域的研究中。

一、什么是质的研究

质的研究（**qualitative research**）亦称质化研究，没有确切、公认的定义。研究者从不同的侧面对质的研究作出了界定，例如：

克雷斯韦尔（Creswell, 1994）认为质的研究是一种研究设计范式，是在自然情境中以复杂的、独特的、细致叙述来理解社会和人的过程。巴尼斯特（Banister, 1994）等认为质的研究是试图抓住潜在于行为表述中的意义，以及行为表述结构的意义，捕捉生活于其中、且为人们所诉说的人们所做的结构的意义；是探索、明确和系统化一个已经确认现象的意义；是对一个特定问题或主题意义的清晰表述。这类界定着重于研究目的的说明，强调质的研究在于理解人的行为及其意义。

赫德尔森 (Hudelson, 1994) 认为质的研究的渊源同文化人类学、社会学、心理学、人类语言学相关联, 是站在被研究者的角度来描述和分析文化、人及群体行为特征。此种界定强调质的研究是站在研究对象的角度去理解问题。

柯瑞斯瓦尔 (Krathwohl, 1998) 认为质的研究是用文字来描述现象, 而不是用数字来加以度量。这类界定强调了质的研究的工作语言是非数量化的。

雷纳 (Ratner, 1997) 的定义则突出了该研究取向对文化心理学的重要意义, 认为质的研究是从非普遍性的陈述、个案中获得印象和概括的过程, 是文化心理学方法学的基石。

陈向明 (2000) 从质的研究活动特征的角度认为, 质的研究是以研究者本人作为研究工具, 在自然的情境下采用多种资料搜集方法对社会现象进行整体性探究, 使用归纳法分析资料和形成理论, 通过与研究对象的互动对其行为和意义建构获得解释性理解的一种活动。

从实践的角度来看, 质的研究包括访谈法、观察法、焦点团体法、文献法等资料搜集的方式; 内容分析、动机研究、心理分析、心理传记、个案研究、人种志方法、历史研究、扎根理论方法、叙事分析、行动研究等综合性的研究方法; 符号互动论、批判理论、现象学、诠释学、俗民方法论、女性主义及后现代主义也被认为是质的研究取向。

面对多侧面理解和丰富的实践, 很难用单一定义来界定什么是质的研究。从理解的角度, 可以认为质的研究是一种研究取向, 是研究者试图从参与者的视角, 在研究对象的自然情境中, 以归纳的方式搜集和分析资料, 而达到对心理现象及其意义理解的过程。

二、质的研究的基本特征

(一) 在自然情境中寻求意义和理解

质的研究的目的是理解人们的经验、行为和关于他们世界的意义。不但人的行为受情境的影响, 而且意义也具有情景性, 只有在现场情境中, 通过观察、聆听、或分析实物, 自然地直接接触研究对象的经验世界, 才能理解特定情境和其中的作用过程, 并在情境的变化发展中了解行动的意义。构成质的研究工作环境的自然情景, 可以是任何社会单元, 如个人、家庭、班级、工作组等。在自然情境中研究的特点, 使研究的结果解释也依存于特定的情境, 而不宜简单外推。

(二) 研究者是主要的工具

在质的研究中, 研究者既是研究的主体, 也是研究的主要工具。研究者是质

的研究搜集和解释数据的最理想工具，因为只有人才能适应处于时刻变动中的研究关系，才能胜任复杂的、性质各异资料的分析。研究者不仅可以处理言语信息、结果信息，而且可以处理非言语信息、即时的过程性信息。通过动态性的资料整理总结，不断评价研究的设想、价值观和偏见，探索未曾预料的事件。但作为工具，研究者的价值取舍也可能对研究进程和结果发生不利影响，研究者应敏感地觉察并予以恰当处理。研究者作为研究的主要工具，也说明质的研究对研究者的素质要求较高。

（三）归纳的研究策略

质的研究以归纳的方式搜集和分析资料，从中建立概念、假设和理论。在理解现象意义时，质的研究不是搜集资料来评估已经存在的假设或理论，也不是基于概率进行演绎推论，而是利用直觉和对资料的归纳来建立和发展理论。大多数研究采用质的研究方法的主要原因是现存理论无法解释特定的现象，或是根本就没有相应的理论。归纳策略的运用，要求质的研究遵循弹性的研究计划，不严格限定获取资料的方法和分析问题的角度。

（四）描述性的结果

质的研究的结果是从研究现场关系的描述中自然得出。以过程性描述展示研究者所理解的行为意义，以及依据描述得出结论或提出理论。研究情境中的每件事都可能是理解现象的线索，因此需要搜集各种具体的资料，如现场记录、访谈记录、文献片断、备忘录、照片、图表、音像摘录等。这些资料又以文字或图片等非数字形式，从多角度提供有关情境、人群及其活动的详细的、整体性描述。因此，结果呈现既是描述性的，又是脉络性的、过程性的。

三、质的研究与量的研究

量的研究是运用一定的数学方法，用数值来判断研究对象诸因素的关联，表示、分析研究的结果。质的研究是研究者在与研究对象的互动中，通过深入、细致、长期的体验，然后对事物的“质”得到较整体性、解释性的理解。两种研究取向在理论基础、资料获得以及解释等方面都存在区别。

（一）理论基础

量的研究的哲学基础是逻辑实证主义，主张实证主义方法是认识人和社会的唯一途径。心理学的研究可以运用假说验证来建立法则，通过对有代表性的样本的研究结果推论、演绎，就可以解释行为，充分理解心智与行为的关系，进而控制、预测行为。

质的研究理论基础是现象学、符号互动论等，认为人对世界的理解是经过个人“知觉”的过滤选择，再加上个人诠释建构而成，意义是建构出来的。心理学的研究应该以整体性的观点为基础，不预设立场，从整体脉络中了解事件、现象

的意义，探索个人的意义及其社会文化的意义。

（二）适用对象

量的研究注重研究对象、研究问题的普遍性、代表性，追求研究资料、研究结论的精确性，因此比较适合变量关系的确认、宏观层面问题的调查与预测。

质的研究注重研究对象、研究问题的个别性、特殊性，注重提出新问题或发现问题的新角度，把自然情境作为资料的直接源泉，因此比较适合在微观层面对个别事物进行细致、动态的描述和分析。

（三）研究过程

量的研究是静态化研究，重视对研究情境与程序的控制，从操作化概念与假设出发，以实验、调查等方法获得数据来验证假设。研究强调严密设计以保证结论的客观性及可重复性，要求研究者价值中立。

质的研究具有动态性，不一定事先设定理论假设，研究倾向于运用无结构、弹性的程序，用参与观察、访谈、实物分析等方法来搜集、分析数据并归纳结果。研究者对研究的影响不可避免，要求研究者对自己与研究对象之间的关系进行反思。

（四）研究呈现

量的研究使用确定化的程序，统计应用较普遍，数据或结果往往以数字与图表来表达，写作通常是第三人称。质的研究程序灵活，结果呈现方式多样，多以叙事的方式，通过文字和音像资料展示研究结果的内涵，写作通常是第一人称。

质的研究与量的研究是两个极端，质的研究倾向于理论的建立，量的研究则倾向于理论的测试，理论的建立需要广泛而无偏见地搜集资料，理论的测试需要客观精密的验证。两种研究又反映了事物质与量的辩证关系，两者在心理学研究上相辅相成，都是必需的。

四、质的研究的应用

在研究实践中，许多主题既可用量的方法研究，也可以从质的角度描述，二者经常互为补充。在定量研究中，研究起始阶段为发现假说而开展的“实地观察”，验证假说中为丰富验证数据、增添研究结果的说服力而进行的描述，因果推论研究中对概率描述的补充性说明，都是质的研究的体现。而有的研究问题，如了解历程、探求意义、情景描述等，则更适合使用质的研究方法。表 9-1 以组织发展变革为例展示了质的研究方法在不同的研究类型中的运用。

表 9-1 质的研究在不同研究类型中的运用

研究类型	问 题	方 法
探索性研究		
认识鲜为人知的事件、情况或情景；识别或发现一些重要的变化；提出假设	这个组织里正在发生什么？组织里的工作模式和核心是什么？模式和核心是如何组织联系的？	个案、观察、深度访谈
描述性研究		
记录事件、情况或情景	正在发生的重要的事件、行为、态度、过程和结构是什么？	个案、人种志、观察、深度访谈、问卷调查
解释性研究		
解释事件、情况或情景的形成原因；识别影响事件、情况或情景形成的原因	什么事件、价值观和政策正在影响组织而形成现在的特征？这些力量是如何相互作用而影响这个组织的？	个案、现场调查、人种志、深度访谈、内容分析
预测性研究		
预测事件、情况或情景的结果；预测可能导致的行为或行动	该组织现在的政策可能对未来产生什么影响？谁会受到影响？会受到什么影响？	观察、深度访谈、问卷调查

（一）了解历程

从时间维度说明事物发生发展的来龙去脉以及阶段性，也就是从过程的角度，而不是单从结果的角度描述问题。例如，研究一个抑郁症患者的人际概念的发展，就不是仅仅描述此人当前的抑郁状况和人际概念状况，而是从时间的维度来探索其抑郁的发展过程以及在这个过程中人际观念的发展与改变。再比如，当研究者从习惯使用量的研究方式转为使用质的研究方法时，研究者往往会有许多不适应或内心冲突，如何界定这种冲突、使之服务于该方法的研究，就需要从过程的角度加以描述。

（二）探求意义

了解事情、事件和行动的意义。探求意义是阐释人们如何理解事物、主观意图和行为及其影响，这种理解又如何影响其行为的过程。例如，儿童顺从他人的主观意义是什么，这种理解对于其个体的发展有什么价值。对于这样的研究目标，其对象的性质不适宜量化，或了解意义比了解其数量关系更重要，往往只有

质的研究方法才可以胜任。

（三）情景描述

当新的现象出现或挖掘已有现象的新内涵时，往往需要通过情景的细致描述来加以界定。通过对特定的情境描述，可以了解事件、行动、观念以及意义是如何在特定的情形下产生出来的，了解某些现象的影响因素。情景描述因此成为提出研究假设、发展理论的研究基础。例如，文化冲突模式下组织人际互动模式是什么？与一般的组织有什么不同？就需要对组织情景中的人际互动进行细致的描述，从而可能发现其中的特点，归纳出其可能独有的互动模式。

总的来看，凡是从被研究者的视角、采用归纳方式进行研究的课题均可以使用质的研究方法。在具体的研究中，质的研究对于处理具有特殊性、意义性、陌生性等特点的问题更具有优势。

补充讨论 9-1 关于质的研究的地位

为什么质的研究方法处于劣势地位？从历史来看，心理学独立的历史短暂，其发展需要解决的首要问题是谋求科学的地位，而实验和测量要比质的研究方法看起来更科学；从现代西方心理学的实践来看，心理学研究需要使用科学的仪器和资源进行研究，以获得最大限度的支持。

为什么用质的研究方法？质的研究为什么越来越受到重视？大致有以下几种原因。

一、实验和测量的局限性

实验研究的伦理问题。实验要使研究对象接受特定的情景，说服他们暴露于各种情景下……这些操纵可能会伤害其自尊、同一性、自我决定的感觉，乃至身体健康。

依靠测量是低效的。将心理简化为数字会忽略掉一些非数量维度的信息。例如，我们如何量化意义、爱、愤怒或疑惑？可以用数字来描述它们，但是我们可能无法以此抓住其本质。

控制的问题。为了发现变量之间的关系，其他的变量必须得到控制。且不说我们无法控制个人带到实验中的经验，控制在复杂的物理系统中也是成问题的，更不用说人类。

化约主义（还原论）的困扰。研究者用生理和信息处理等因素对人类行为进行解释、分析，那么这些解释从根本上回避了研究最初想要回答的问题——例如意识、意义、人格、自我等。

实验及其相关的方法本质上是决定论的。一方面寻找到建立因果关系的基点是困难的：另一方面，人类的心理和行为存在某种程度的自由，例如决定于外

界强制的道德行为就没有意义。

二、研究主题决定研究方法的选择

当研究者不是寻求因果关系、不是去发现普适的定律、不是去了解个体记忆系统中的认知过程，而是去理解现实和经验，那么质的研究方法就是恰当的。一般来说，传统的量的研究方法无法捕捉鲜活的生命，质的研究方法则是试图捕捉生活。

另外，质的研究方法为心理学研究者所认可，并不是改变心理学来适应质的研究方法，而是质的研究方法适合于一部分心理学所研究的问题。

第二节 质的研究设计要素

质的研究设计中必须考虑五个要素是：研究目的、观念架构、研究的问题、研究方法和效度（陈向明 2000；Maxwell, 1996）。这五个核心要素紧密连接形成互动的整体，研究目的是研究的根本和出发点，是确定其他研究要素的基础；观念架构与研究目的紧密相连，并影响研究问题的形成；方法则是针对回答的问题汇集和分析资料，是研究的可行性和效度的保障。

一、研究目的

明确研究的目的（purpose of a study）是研究设计的主要部分。如果不明确通过研究可以获得什么、做什么，就无法确定研究工作的方向和重点，无法指引研究设计的其他部分，无法确保研究的价值。

研究目的分为个人的目的、务实的目的和科学的目的。

个人目的指研究者个人选择质化研究的动机：即研究者为什么选择质的研究方法？质的研究并不容易实现，当决定使用这种研究方式时，就必须考虑使用的理由与研究问题、研究条件之间是否存在矛盾。

务实目的是研究本身对于完成某事，改变某个情况，或达成某些目标的功效。务实的目的是推动进行质的研究的重要部分，一方面任何研究的最终目的之一就是应用于实践；另一方面，许多研究要进行长期的实地调查或追踪研究，付出大量的努力，明确的务实目的是研究的一个有力支撑。

科学目的着眼于了解某些事，对于发生什么，以及为何这件事情会发生等方面深入的理解，是为认识世界、追求真理而进行的研究。

任何研究都可能有一种或多种目的。有些目的要研究进展到相当程度才会明显，有些目的随着研究进行还可能发生改变。必须考虑各种目的对于效度、资料搜集、分析的潜在影响，混淆各种目的会影响研究设计的严谨性和可行性。务实和科学是研究目的设计中的最主要部分。

二、观念架构

观念架构(conceptual framework)也称为概念性架构、研究的理论构架,是质的研究的基点,是指支持和贯穿于研究中、解释研究对象的关键假设、期望、信念、概念、变量和理论,以及它们之间的假设关系。

形成观念架构有三个主要的渠道:分别是研究者的经验性知识、现存的理论和研究,以及测试性或探索性的研究。

研究者的经验性知识是研究者与研究问题有关的个人经历,以及自己对该问题的理解和看法。研究过程很难和研究者的生活截然分开,在研究各阶段中,研究者的期望、信念和假设是提出新概念或观点的素材。但经验性知识也会影响研究的深度和有效性,因此需要用备忘录记录各种经验性知识,以提供检视它们对研究影响的依据。

现存的理论和研究是指在研究主题内,已提出的有关理论和完成的研究。理论的功能在于展示研究资料之间的内在联系,特别是突出那些不易注意到或易被误导的关系,有助于研究者把握研究的方向。如果没有充分运用现有理论,则可能遗漏了只有理论才具有的洞察力。但过分依赖理论,甚至将理论强加于研究之上,会阻碍研究者看到那些不适应理论的事件和关系。已有的研究可以用来评价和发展当前的研究,例如,文献中是否提出了未被回答的问题,先前研究的方案和方法有何问题,本研究的理论构想能否回答先前研究获得的资料等等。

补充讨论 9-2 已有理论知识对观念架构的阻碍

当我开始研究大麻的使用的时候,意识上最主宰的问题,也是唯一值得去看的问题是“为什么人们会做那样奇异的事情”?而且意识中偏好的回答是去找一个心理特质或社会属性,来区别使用 and 没有使用大麻的人……。我偶然发现,一个更大更有趣的问题是:人们如何学习界定他们自己内在的经验?(Becker,1986)

测试性的研究是将自己的设想付诸实践,检验或探测它们的内涵。测试性的研究可以使研究者发现既有构想的不足,更精确地着眼于研究的焦点,最终有助于建构更成熟的理论。

三、研究问题

研究问题(research question)是质的研究设计的核心,直接关系到研究设计的其他要素,是决定研究成功的重要方面。

（一）研究问题的类型

研究问题可以进行多种划分，例如概括性问题和特定性问题；探讨变量关系、类别差异的结构性问题和探索事物动态变化的过程性问题；进行现象说明的描述性问题、从研究对象的角度进行说明的阐释性问题、探讨概念关系的理论性问题、探讨研究适用情景的推论性问题和进行价值确定的评价性问题等。

质的研究适合处理特定性问题和过程性问题，以及描述、阐释、理论水平的问题。推论类问题会受取样的影响、评价类问题会受价值取向的影响，这两类问题在研究过程中需要对相应的影响特别探究和说明。

（二）研究问题的确定

研究问题是充分考虑问题与研究目的、已知事实、当前的观念架构之间的相互影响，经历反复斟酌而形成。研究问题要清晰，要避免常见的错误，要能够通过研究得到回答。例如，“大学生心理健康研究”就失之宽泛而难以通过单个研究来完成；“员工如何处理同事的排斥”包含未曾检验的假设，也不可以作为研究的问题；缺乏概念或经验支持的问题也不是好问题。

问题确定后就需要提出研究假设（hypothesis）。质的研究假设具有动态特性，不是仅仅根据既有的理论和研究发现而提出的，而是在与资料互动中形成、发展和检验。研究假设表达研究的指向而不是研究的具体内容，一般采用“×××的行为与观念研究”、“×××的跨文化研究”、“×××个案研究”等形式，而不是诸如“××在××中的效应”、“××对××的影响”等方式。

四、方法

在方法设计部分，主要考虑如何与研究对象建立研究关系，确定取样、资料搜集和分析的具体方法，以及研究结果呈现的方式。

（一）建立研究关系

研究关系（research relationship）是指在研究过程中研究者与研究对象之间的关系。

建立符合伦理的研究关系。在质的研究中，如何进入研究的现场，与研究对象建立起符合研究伦理的关系较之于量的研究更重要。一方面，研究离不开良好的互动关系；另一方面，研究往往要深入研究对象的生活，不可避免地会产生价值冲突、侵犯等对研究对象不利的事情。

确立研究者的角色和研究关系的亲疏度。在研究互动过程中，研究者不可能采用完全中立的态度，必须扮演从“局外人”到“自己人”之间的某种角色。人们对陌生人和熟人的反应不同，因此，要根据研究需要确立研究者表演的角色和所建立研究关系的亲疏程度。

在研究关系建立和持续过程中，应吸取相关文献中成功的经验，发挥研究者

建立关系的能力、合理运用沟通技术，不断协调与研究对象的关系。

(二) 取样

任何研究都需要取样决策，质的研究也不例外。在研究过程中，要不断选择时间、场景、观察或访谈对象以及内容，思考它们对于回答研究问题的代表性以及效率。质的研究有结构化和非结构化两种取样方式可供选择，见表 9-2。

表 9-2 结构化与非结构化取样的比较

	非结构化取样	结构化取样
总体的外延和特征	事先均不知道	事先知道,分布特性可以估计
取样次数	逐步确定标准来重复取样	根据预定计划一次完成
样本大小	事先无法确定	事先可以确定
取样结束的标准	达到理论饱和	样本被研究

结构化取样 (structured sampling) 是一种统计取样，是在具体取样之前就确定样本的结构。例如，计算机技术在人们心目中的形象（社会心理表征）要受职业和文化的影响，因此，取样的职业方面就可考虑设计者、职业应用者和日常应用者；再选择不同地区的群体作为文化因子的考虑。结构化取样结构清晰，但可能忽略有关的取样维度，例如，不同性别、文化程度的人对计算机技术的认识也可能会不同。结构化取样适合于进一步分析、区分和检验已知取样维度的异同，在质的研究中使用相对较少。

非结构化取样 (unstructured sampling) 中最重要的是理论取样。理论取样 (theoretical sampling) 通常用来发展或建立理论，是从经验或理论出发，以提供深度、多元化资料为标准，随数据的搜集和解释逐步选择特定的情景、人物、时间以及事件，直到达到理论饱和 (theoretical saturation)，即没有新信息可以用来进一步发展理论的状态。

补充讨论 9-3 马斯洛“自我实现的人”的研究取样

“自我实现的人”是人本主义心理学家马斯洛 (Maslow, A) 对心理健康者的称谓。自我实现的人的提出对于心理学、哲学、管理学等领域都产生了极大的影响。现实中并不存在一个完美的人作为研究的对象，在“自我实现的人”内涵确定之前，也不存在筛选被试的完整标准。在此情况下，马斯洛如何选择研究对象来确定自我实现的人内涵的呢？

马斯洛采用一种重复技术，“以个人或文化的非技术的信仰



作为开始，对自我实现症候群的各种扩大用法和定义进行比较，然后再仔细地给它下定义”，“根据修正过的通俗定义，第一批研究对象小组选出来了，其中一组质量高，一组质量低。以临床风格对这些人进行尽可能仔细的研究，在经验研究的基础上，最初修正过的通俗定义又按照现在手中的材料进一步修改。这样就得出了第一个临床定义。按照这一新定义，对最初的研究对象进行重新筛选，一些人保留，一些人淘汰，一些新的成员补充进来。然后，又继续对第二种水平的研究对象小组进行临床研究，如果可能，进行试验和统计研究。这又导致了对第一个临床定义的修改、订正和补充。然后，又根据这一个新的定义进行再筛选。经过这样不断重复，一个最初模糊、不科学的通俗概念就变得越来越精确，在特性上越来越便于操作，因而也越来越科学。”

（采自许金声等译，1987）

目的取样（purposive sampling）也是质的研究常用的一种非结构取样方式。目的取样与理论取样最大的不同是它适合已有理论的情形。研究者基于当前研究已经确认的变量，主观地确定何者被确定为样本，其目标是确定信息丰富（由此及彼）的个体以进行深入的研究。目的取样的过程受研究者对当前研究状态的判断制约，确定“什么是与研究有关的”至关重要，正是在“什么是与研究有关”的见解启发下，研究者有意识地去寻求典型的和趋异的数据。目的取样对于发现可能存在的结构变异、说明所研究的现象十分重要。

取样决策应该充分考虑质的研究的动态过程性，根据需要确定和转换各种取样方法。除了理论取样和目的取样外，还有滚雪球取样（snowball sampling）、极端个案取样（extreme case sampling）、最大变异取样（maximal variation sampling）强度取样（intensity sampling）、典型个案取样（typical case sampling）等方式。

（三）资料搜集

参与式观察、深度访谈和文献是质的研究资料的三种最主要来源。随着互联网的兴起，在线资源也成为许多研究的资料来源（Merriam, 2002）。

确定资料搜集的方法。方法没有优劣之分，重要的是运用恰当、获得研究所需要的资料。在充分了解各种方法特点的基础上，结合在真实情景中、特定的时空条件之下的可行性确定运用何种方法搜集资料。必要时要通过试验来确定方法的可行性和有效性。

使用三角验证法是保证所搜集资料有效的重要策略。资料的三角验证（triangulation）是使用不同的方法、从不同的研究对象或情景中搜集资料，以排除特定方法的限制和系统偏差（如图 9.1 所示）。

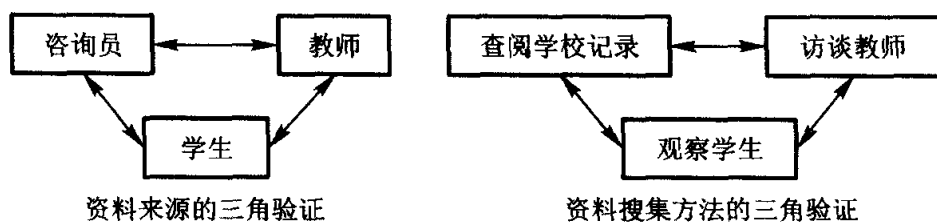


图 9.1 三角验证法

(四) 资料分析

资料分析的原则是及时、自省、比较和创新。及时就是不能待资料搜集整理完毕后才进行分析，资料分析和搜集一般是同步进行的。自省就是研究者不断反省自身与资料，以体悟出新的思路，使数据呈现出更合适而完整的主题。比较就是随时将分析结果与现存理论或他人经验进行比较，判断何者相似，何处不同。创新就是放弃预设成见，开启思考，获得更深层次的体会和见解。

资料分析主要用归纳法。归纳是通过将数据分门别类加以组织，并找出类别间的模式与关系，归纳出的模式又进一步调整分析的过程，直到形成理论或者得到结论，回答所研究的问题。

分析包括预处理和正式分析两步。预处理包括资料的整理、审核、将录音逐字誊写，给材料标页码等。经过预处理，使研究者能够了解资料的整体性质和范围，初步形成编码设想。正式分析包括细读数据、了解情境的相关性、确认数据中的意义单元，并找出信息单元中隐藏的意义、以归纳法寻找各种类别，以及用归类和比较法使主题呈现出来。

资料编码是分析的基础和核心。编码（coding）是依据一定标准将资料归入相应的类别中的过程。与量的研究结构性编码不同，质的研究编码的目的是协助发现概念、诠释资料的意义，是通过资料之间的比较，将资料中的词汇、句子、段落等意义单元用编码系统表现出来，是依据资料而实现对资料的统整。按照概括程度，编码可分为开放式编码、主轴编码和主题编码，或一级码、二级码和三级码等。各种编码单元互相联系形成编码系统，编码系统的层次和数量则依据具体资料确定。

补充讨论 9-4 研究的编码

为什么美国的少数民族学生不愿意教书？通过不同种族教师对“为什么你认为有色人种学生会去教书？”这个问题的回答，研究者 Gordon 总结归纳出了 17 个主题，并将它们分成教育经历、文化和群体心态、社会和经济障碍三类：

教育经历

高中没有毕业	缺乏足够的准备
在学校中的负面经历	不良的学生纪律 / 得不到尊重
教师缺少面对多样化学生群体的准备	缺乏大学的支持

文化和群体心态

缺乏学业方面的鼓励	无种族意识
缺乏角色榜样	社会地位低下
要获得回报，需要太多的教育	

社会和经济保障

低收入	消极印象
不良的学校条件	其他方面更多的机会
种族主义	

(采自弗林克尔、瓦伦，2004)

已经有多种计算机软件用于质的研究分析中，如 ATLAS/ti、NUD. IST、NVivo 等。软件可以帮助资料整理、编码、发现规则和建立概念网络等，但无法全面替代研究者对资料意义的挖掘。

(五) 研究结果呈现

结果呈现是非常具有挑战性、创造性的过程。它不但与资料搜集、资料分析、发展理论等结合在一起，而且呈现方式灵活多样，研究报告、论文、专著，乃至戏剧、电影等都在可选之列。

研究报告 (research report) 是研究结果呈现的主要形式。报告的风格、书写角度要根据读者、研究问题、研究者与被研究者的关系等来确定。例如，公众的阅读兴趣是研究发现，而研究者则关注研究过程中与主题有关的各种现象、所使用的研究方法、对研究关系的反省等，以判别研究的质量。

叙述是质的研究呈现的关键。成功的研究报告具有以下一种或多种突出特征：

- ✧ 用第一人称的方式叙述，将资料引述与解释交融在一起
- ✧ 坦诚介绍研究者使用的方法和研究过程中所做的反省
- ✧ 用当事人的语言来描述、真实再现当事人的具体观点
- ✧ 详细描述事件发生的情形和当事人的行为、情绪反应
- ✧ 根据当事人所在的文化背景对研究结果进行检讨、思考
- ✧ 用研究结果检验某一理论模式或者观念架构

五、效度

效度是研究的描述、解释、诠释、结论，以及任何其他说法的正确性。将效度作为研究设计的要素，可以避免无法证明研究自身是否可信，是否有效的问题。

（一）影响因素

影响效度的因素很多，如遗忘、间接资料、反应性行为、研究者的偏见等。其中研究者的偏见和研究对象的反应性行为是影响效度的最主要因素。不同效度的敏感因素不同，例如，影响描述效度的主要是资料的错漏，可能是观察失误或记录不全面等原因造成的；影响阐释效度的主要是以研究者的观念去解释而忽略了研究对象，无法从研究对象的角度表明资料的意义；影响理论效度的主要是忽略或不去搜集矛盾的资料，没有考虑替代性解释，使理论无法真实反映所研究的现象。

（二）效度的测试

效度的测试是通过找出威胁效度（validity threat）的证据来测试效度。效度测试无一般化程序，测试方法依研究而不同，常用的有证伪法、三角验证等。

证伪法 找出矛盾的证据以及负面的案例来判别结论有无错误。由于资料本身也可能存在错误，因此，使用证伪法时要严格检查支持性与矛盾性的资料。

三角验证 三角验证是质的研究中保证研究效度的重要方法。Denzin (1989) 区分了四种三角验证类型，即不同的数据来源、不同的研究者、多种理论视角，以及不同的研究方法。另外，也可以进行不同学科的三角验证，即从其他学科观点和视角来辅助说明心理学研究。

例如，要保证访谈资料的可信度，三角验证法可能的措施包括：用不同的方式重复询问被访谈者，对回答问题和自发谈话进行比较，研究者观看被访谈者描述的事，访问现场中的其他人，以及用不同性别访谈者来排除可能的性别差异反应。

反馈法 反馈对于辨识影响效度的因素、研究者偏见、研究方法或逻辑缺陷等都很有用。研究对象熟悉的人、不熟悉的人以及研究对象本人都是反馈信息的来源，他们的有关资料和结论的意见，可以用来排除误解或检查是否忽略有关的意见或观点。

丰富的资料 丰富的资料指对于事物或现象的细节描述详尽而完整，能够呈现其真实状况，如访谈中逐字逐句的记录，观察中特殊、具体事件的详细描述。丰富的资料可为研究有效性提供证据。

方法提供不同角度的效度证据，但并不必然保证研究有效。例如，三角验证

中，不同方法本身可能就是无效资料的来源，研究者也可能不自觉地选择支持预期结论的研究方法。因此，应该思考每种偏差的可能来源，并寻找可行的解决办法。

第三节 研究设计的模式与论证

研究设计模式是研究的组成要素和这些要素的内部逻辑与结合方式，是从最初设定要探究的问题到这些问题最后答案的规划。研究设计论证是对研究设计作出的说明，是研究设计的总结，也是一份向特定听众传达并论证研究的文件。

一、研究设计的模式

（一）线性设计模式

线性设计模式（linear design）是量的研究的基本设计模式，各个要素按先后次序逐步进行。如表 9-3 所示，研究基于理论知识和经验产生特定的假设，经过操作化、取样、搜集数据、解释数据、以及效度而完成。

线性设计模式步骤分明，但不能够妥帖地描述质的研究逻辑与过程。质的研究是开放的、具有相当弹性的历程，各部分相互依赖，无法一次性完成。随着每个元素的发展，其他要素需要重新思考和修正。

线性设计模式并非与质的研究格格不入，如果把质的研究理解为由多个循环往复的线性过程组成的，就有助于设计和实施研究。

表 9-3 质与量设计模式的线性比较

步	骤	量的研究	质的研究
1	心理学理论		确认感兴趣的问题或主题
2	从理论产生特定的假设		提出探索性的研究问题
3	计划特定的程序、方法和工具		搜集、解释最初的数据
4	搜集数据		提出尝试性的假设
5	分析、解释数据		进一步搜集、解释数据
6	决定研究假设是否被证实		将尝试性的假设精确化
7			进一步搜集、解释数据
8			提出更确定性的假设
9			产生理论

（ 采 自 Ponterotto & Grieger,1999）

(二) 互动模式

马克斯韦尔(Maxwell, 1996)提出的互动模式(interactive model)是现在广泛认同的一种设计模式。互动设计模式包括确定研究的目的、探讨观念架构、定义研究的问题、决定研究方法和保障效度五种设计要素。互动模式认为设计的各个组成部分之间是同时发生、相互作用的关系,设计是一个互动的过程。汇集和分析资料、发展理论、效度威胁的消除,通常是同时进行的。任何一个部分的变动都会受到其他部分的牵涉和影响,每一项工作都影响其他的或全部的工作。

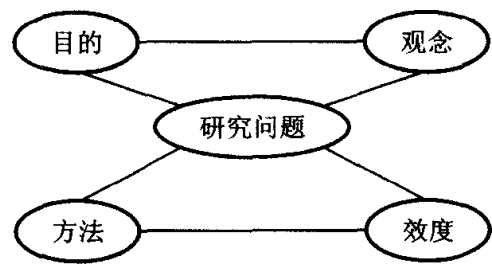


图 9.2 互动模式图
(采自 Maxwell,1996)

虽然各种要素之间都相互影响,但有的关系是相对重要的,在图 9.2 中用连线来表示,相对次要的关系则没有用连线表示。要素及其关联构成了两个重要的三角式关系,由研究目的、研究的观念和研究的的问题所组成的三角关系,是研究设计的外部成分;由研究的的问题、方法和效度组成的三角关系,所表达的是研究设计的内部成分;研究的的问题处于这两个三角关系连接处,是研究的外部成分和内部成分的纽带。

(三) 立体两维互动模式

立体两维互动模式是陈向明(2000)基于互动设计模式提出的,见图 9.3。认为质的研究过程像不断向下转动的螺旋圆锥体,锥体的任何横切面都可以由互动模式的五个成分组成,这五个成分之间也同时在水平面上发生着互动,即平面的“两维”。研究中的每一个部分都随时间的流逝而不断地缩小聚焦范围,以螺旋转动的方式不断地往下旋转,同时带动圆锥体的纵切面往下运动,随即构成了“立体”。

立体两维互动模式在互动研究设计模式上加入了时间成分,既说明了各个成分之间的逻辑互动,也说明各个成分随着研究的深入而发生的螺旋转动可能是不同步的,表达研究设计各要素变动复杂多样的特点。这个模型的螺旋转动也可以反转理解,随着研究的进行而螺旋向上转动。在锥体的各个横切面上,原先处于底部混沌状态的各个要素及其关系随着螺旋上升而愈加分化、彰显,所研

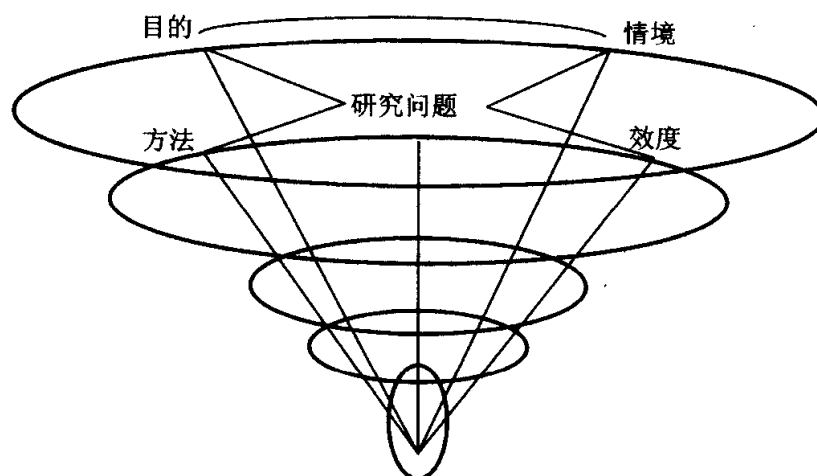


图 9.3 立体两维互动模式图

（采自陈向明，2000）

究的问题、目的也愈加明了，观念架构也更加完整，影响效度的因素以及验证效度的方法也更具有针对性。

二、质的研究设计论证

（一 研究设计与研究设计论证

研究设计论证（research proposal）也即研究计划报告，是解释研究的目的、主题和意义，以及对研究采用方法及其适当性做出的说明。

研究设计论证依赖于研究设计。研究设计中各个要素的关系是论证的主要因素，研究设计的质量决定了论证的水平。在研究设计尚未清晰、成熟之前就急于撰写论证，很可能事倍功半。

研究设计论证对于完善研究设计具有促进作用。虽然论证无法详尽地表示所要进行的研究工作，也不是所有的研究都需要向他人说明研究设计，但设计论证可以帮助研究者整理研究思路、将研究设计付诸实践。成熟的研究者可能不需要将每个研究的论证写出，但互动式渐进的过程往往超出初学者的驾驭能力，撰写详细的设计论证对完善研究设计不可或缺。

（二）研究设计论证的撰写

研究设计论证的重点是表达研究设计本身的严谨性、可行性。设计论证不是简要说明研究的内容、文献、研究的主题或方法，论证的每一部分都应该对研究中可能遇到的、各个重要的问题以及解决策略等提供清楚的答案。如，为什么这些事情值得去研究，为什么使用某种研究方法，通过研究可以学习到什么，我们如何知道结论是有效的等。

研究设计论证没有标准模式，一般包括摘要、前言、观念架构、研究问题、目的、方法、效度、结论、参考文献、附录等要素。

摘要：摘要是研究设计的概要。摘要的内容包括研究目的、研究问题及研究方法等。摘要的表达应该扼要，让读者迅速了解研究的内容。摘要是设计论证书的选用内容。

前言：本部分表达研究的目的、研究的概要、研究的类型。在表述上不要太长。

观念架构：观念架构是研究设计论证的重要组成部分。本部分通常是通过述评已有的研究而建立本研究的理论基础，说明针对研究现象所要使用的理论是什么，表现研究与已有的理论、研究之间的关系，表明研究的主题。

观念架构的论述需要文献的支持，因而本部分可以使用“文献综述”作为标题。在使用文献时，要根据文献所介绍的理论、发现与本研究设计的关系取舍，不能堆砌无关的文献。

如果已进行实验性研究或前期研究，可以在此说明，并解释它对研究的意义或影响。

研究者的资历也是观念架构的一部分。对于寻求研究支持的论证，例如学位论文的研究设计、申请基金资助的设计，就要介绍研究者擅长的相关领域与经验，为他人信赖本研究的价值提供佐证。

研究问题：研究的问题是研究的焦点，是论证的中心。研究问题通常在陈述研究的观念架构之后，在先行的研究、理论与个人经验的铺垫下清楚地提出。

本部分既要明确地界定问题范围，凸显研究问题对学术、对实践的重要价值，也要处理好研究问题与已有的理论和研究、研究者的个人经验、探索性的研究，以及研究目的的关系。

◇ 研究问题的焦点是什么？

◇ 是否适合于用质的研究方法，问题是否属于意义、理解或过程类问题？

◇ 问题是否已经陈述清楚？

◇ 文献中有无该问题的论述？

◇ 研究的问题与以往的研究有何关系？

◇ 研究者的个人视角与本问题之间有何关系？

◇ 研究该问题的基本条件是否具备？

研究目的：明确地表达本研究的方向，表明本研究对知识或实践的价值或贡献。研究目的是判断后续研究可行性、预期结果价值的重要依据。

研究方法：研究方法部分主要说明研究的性质、如何建立研究关系、取样的策略、资料搜集策略、资料分析方法等。

研究方法的确定依赖于研究问题和研究的环境。在论述研究方法时，要结

合研究场景以及社会情景的描述说明使用某种方法的正确性、恰当性与价值。

由于各种研究方法来自不同的传统，相同名称的方法内涵上可能有差异，因此设计论证中对于有关方法需要准确地描述。例如，访谈是如何安排的，将在什么地方、用多长时间，用的是什么类型的访谈计划，与访谈问题有关的知识或领域是什么；观察的类型是什么，在什么时间和地点；用到了哪些类型的文献，是如何得到的，官方的或私人的文件的最初用途是什么。

一般不需要清楚地说明所用方法的理由。若采用一种新方法或可能存在对该方法的偏见，就要说明使用这种方法的理由，解释选择的依据。

效度：如何保证研究的效度，可以放在研究方法中论述，但独立论述更好。专门讨论研究的效度，一方面是为了说明如何处理效度的威胁，另一方面是表明对效度的重视。

本部分主要说明威胁研究效度的因素，以及相应的测试效度的方法。应针对每一种重要的威胁效度的因素阐述相应的效度保障策略。

初步的结果：如果已经进行前期研究，可以说明到目前为止，研究的可行性，以及暂时的研究结果。初步研究不但影响概念架构，而且通过对初步研究结果的评估，有利于说明研究设计的问题、搜集、分析资料的方法、效度策略的恰当性、可行性。

结论：简要总结前面的内容，重申研究的目的与问题，说明能够取得何种资料、预期的结果，以及研究可能对本领域的影响等。

参考文献：列出确实引用的、重要的文献。

附录：不便于在正文中表达，但对于研究设计论证较重要的内容，都可以列入附录中。包含：

- ✧ 研究的时间进度表
- ✧ 介绍信、准许研究的信
- ✧ 问卷、访谈提纲以及其他的研究工具
- ✧ 受访者列表
- ✧ 观察的时间程序
- ✧ 分析工具的说明
- ✧ 研究问题、研究方法、资料分析策略的关系矩阵
- ✧ 先行研究已经完成的部分观察笔记和访谈记录范例
- ✧ 某些特殊的资料和分析技巧
- ✧ 研究对象的背景资料等内容

根据不同的目的，研究设计论证还可以包括研究的预算、管理计划以及设备与特殊资源等内容。

论证应做到思考严谨、避免漏洞。用于沟通的设计论证，需要考虑读者特

征，避免不必要的复杂格式、华丽的辞藻，向读者清楚地说明要做什么，为什么这样做，力求达到最好的沟通效果。

补充讨论 9 – 5 研究基金的申请

在人们的印象中，成功的心理学研究者是专业文章加书籍堆积起来的。如果真的如此，许多研究者就会自认幸福。然而，学术研究是消耗资源的，除非进行问卷调查或文献研究，即使复印费超支研究者也能承受得起，大量的研究没有外部的资助是很难进行的。研究者们需要花大量的时间申请研究基金。研究者所在机构的研究资金通常有限，使用这些资金也需要申请。申请外部基金资助通常会有一些困难，并不是所有的申请都会得到资助。

成功的申请报告具有以下特征：

清晰地陈述问题，强调其重要性	充分应用文献，说明研究的起点
呈现探索研究的结果	论述清晰简明
说明分析数据的方法	陈述预期的成果
说明研究经验	资金数额申请恰当
合乎专业规范	有说服力

研究基金申请取决于研究的价值和设计的科学性，设计论证在其中扮演重要的沟通角色。

(采自 Giles,2002)

第四节 质的研究表现形式

心理传记、人种志和行动研究是三种质的研究，它们所体现出的质的研究特质不尽相同。

一、心理传记

(一) 概念

心理传记 **psychobiography**) 是个案研究的一种，起源于弗洛伊德 Sigmund Freud,1856—1939)对达·芬奇、摩西等人的研究 该方法是把人当作“ 故事 ” 是将正式的或系统的心理学概念或理论应用于传记研究，通过分析个人的生活资料 and 叙说来了解人们如何建构他们的世界，是凭借个人世界观的重建而寻求宏观与微观观点相结合的方法。该方法适用于寻求个人生命意义的理解，理解重要的个案，以及产生新的假设和理论。

补充讨论 9-6 弗洛伊德和达·芬奇



弗洛伊德以文艺复兴时期画家达·芬奇为研究对象，以精神分析的观点和传记的研究方法，阐释了达·芬奇艺术与科学活动的心理起源。通过研究，弗洛伊德认为童年经验和“恋亲丛”对达·芬奇的艺术创作以及其命运具有决定性的意义；他是一个特殊类型的同性恋者和强迫性神经症者，把想象特别是幻想视为创作的出发点和主要形式；认为艺术家的创作是一种性欲的宣泄，主张艺术和科学成就是性本能升华的两种表征；并提出以“性压抑”为本源的科学研究类型。

关于达·芬奇的传记著作遭到了弗洛伊德已经预料到的谴责和诽谤，因而在该书中他就为自己辩解，这种辩解至今仍然普遍适用。他写道：

我像其他人一样，受这位神秘的伟大人物的吸引力驱使，在他的天性中人们发现了只能用如此鲜明的抑制来表达自己的强大的本能激情。然而，无论列奥纳多的生活真相怎样，我们不能停止对它做精神分析解释的努力，甚至我们能够完成一种新的突破。我们必须用通常的方法在传记领域中界定出精神分析学所能取得成绩的限度；否则，每一种将来可能出现的解释都会成为一种失败摆在我们眼前。在精神分析研究控制下的材料是由一个人生活的历史组成的：一方面是事件偶然情况和背景的影响；另一方面是这一对象被记述过的反应。在心理机制知识的支持下，努力为本性建立一种在他的反应力量上的动力学基础，揭示他的心理原始动机的力量以及它们以后的转变与发展。如果这样是成功的，在他生活过程中的个性行为就能在性格和命运、内外力量的结合中得到解释。假如这样不能提供任何的结果，这可能——……在于与列奥纳多相关的资料不准确、不完整，……即使史料非常丰富，我们对心理机制的论述有最大把握……精神分析研究仍然不能使我们理解一个人怎么会必然成了这样的人，而不是另外的一种人。

（采自弗洛伊德，1910/1998）

传记研究的出发点是认为人类社会是由符号所连接而成的意义网络，研究的主题是个人所建构的意义，叙说者的传记性表达则是理解个人如何建构、维持、理解和接受意义的主要途径。获得个人叙述的最重要方法是叙述访谈（narrative interview），是邀请叙说者谈论自己的生活，叙说者自己决定说的内容与方式，访谈者不能打扰叙说者。

传记研究有因果式和连贯整体式两种取向。因果式研究着重于行为原因的

探讨，例如，寻找导致某种成人行为的儿童期的事件或者原因，通常具有精神分析的理论传统。连贯整体取向旨在寻求行为重复出现的规则和模式，是通过提出概念或解释使我们能够理解个人在特定情境下的行为。

(二) 研究设计

心理传记研究的核心是选择案主、搜集资料和写心理传记。

在选择研究对象，也就是感兴趣的案主时，首先要做一些整体性的了解，避免特别崇拜和讨厌的人物，以免过多的情感卷入。

搜集资料包括现存的资料和访谈资料。现存的资料包括已经出版的书信、日记、自传、回忆录、手稿、人事档案等。通过访谈案主及其亲朋好友获得的访谈资料是最重要的，是研究创新的基础。对于访谈资料，还应该将论证性的言论从叙说性的言论中分解出来，叙说性的言辞才是传记的基础。这些资料能够根据一定的线索对个体形成在一定的社会历史背景下，从出生、青少年直到晚年的人生各个阶段，和家庭、教育、婚姻、工作、业余生活等各个生活主要方面的整合。

在形成传记的时候，不但要保证所用资料的有效性、资料与研究结果的一致性，还要注意避免用自己预设的范畴来理解研究对象，避免将研究对象极端化，避免用不足的资料得出极端的结论。

Runyan(1988)提出了一种心理传记的研究方法，包括以下8个环节。

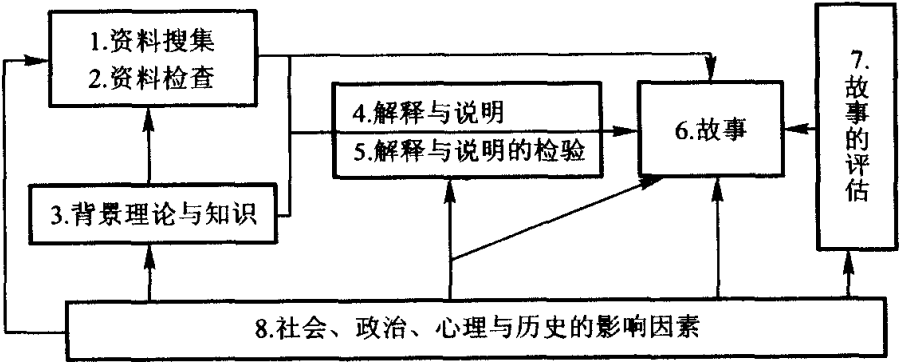


图 9.4 对个体生命的了解与知识的进展循环图

(采 自 Runyan,1988)

1. 证据与资料的汇集，包括信件、日记、访谈、档案等。
2. 证据与资料的批判性检查，包括在既有证据基础上检测错漏，找出不同证据的相对比重。
3. 背景的理论知识，包括心理学的发展理论，相关的文化与历史背景，及相关的生物学知识。
4. 产生对个案新的理解和说明。

5. 对于所提出的解释予以批判性的评估并试图证伪。
 6. 人生故事的产生。结合一些特殊的解释或说明，用理论与背景知识来建构出具体的故事。
 7. 对于所叙说的故事予以批判性的评估。从个案的讨论中发现资料的适当性，理论的恰当性以及解释的可信度。
 8. 上述的各个环节均受到社会、政治、心理与历史因素的影响。
- 心理传记研究具有互动性质，以上任何要素的变动都会给研究带来新的发展。例如 新资料的发现 理论的进展 新的故事的提出 社会背景因素的变化等。

二、人种志

(一) 概念

人种志(**ethnography**)方法亦称文化人类学方法，是典型的质的研究方法，是通过在自然情境中以局内人的立场去观察、描述、判断和解释，而获得对一定社会文化背景中个体和群体的心理现象、心理过程、行为及其意义的整体理解。人种志研究源于人类学和社会学对特定族群社会和文化生活的详细描述和分析，是社会心理学研究的重要模式，主要应用于健康心理学、临床心理学和职业心理学的研究中。

人种志强调现象在行为者眼睛中意味着什么，强调被研究者的视角，广泛运用参与观察。研究者参与到研究对象群体的自然情境中，记录、描述该群体的文化、行为与价值。通过整体性和全局性的描述，人种志研究揭示出平时不易被常人所察觉的研究对象的实质内涵，发现那些习以为常但非常重要的因素，提出重要的假设。

补充讨论 9-7 罗森汉对“疯子”的研究

如何区别正常和异常的行为是心理学的基本问题。然而心理卫生专业人员真的能区别精神疾病和精神健康吗？对病人进行心理诊断时是依据病人本身的特点，还是诊断当时病人所处的情境？确定该问题答案的方法之一就是让正常人进入精神病院，看他们是否被发现是心理健康的。

罗森汉(**Rosenhan**)招募了 8 个被试(包括他自己)作为假病人，被试的任务是想办法进入美国 5 个州的 12 家心理医院去。所有假病人都遵从同样的指示，先打电话与医院预约，到达医院后都声称听到一些声音，说“空的”、“轰”和“砰”等声音。除此以外，所有被试的表现完全正常，所提供的信息除了化名和改变职业身份外也都是真实的。



所有被试都获准进入医院，7 人都被诊断为“精神分裂症”。入院以后，这些被试的行为表现完全正常。所有被试一开始偷偷记录经历，后来这种“做笔记的行为”居然也被认为是一种症状表现，他们就索性公开做笔记了。为了尽快离开医院，他们都表现得像个模范病人，积极与工作人员配合，接受所有的药（并没有吃）。

假病人在医院的时间从 7 天到 52 天不等，平均 19 天。这项研究的主要发现是，这些假病人没有一个被医院工作人员查出是假的。更有趣的是，有 3 个假病人在住院期间，医院里的 118 个真病人中有 35 人认为他们不是真正的精神病人，而认为他们是新闻记者，正在对医院进行调查。

病人和医院工作人员之间的接触很少，即使有接触，也显得很奇怪。被试试图接近各种各样的工作人员，向他们询问一些很普通、正常的问题，如“我什么时候可以出院？”结果 71% 的医生，88% 的护士和服务员没有停下来回答，径直走开；23% 的医生，10% 的护士和服务员只与他们有目光接触；真正停下来与他们说话的医生只有 4%，护士和服务员只有 0.5%。与此形成鲜明对比的是，他们却得到了大量的药片。这 8 个被试总共得到 2 100 粒药丸。

罗森汉认为这是因为精神病医院这个情境对诊断行为产生了极大影响。对精神病的诊断是依据情境而不是病人本身的特点。一旦入院，就被认为是精神不正常的。更重要的是，罗森汉指出了诊断标签的影响。当一个人被打上了某种标签以后，工作人员将认为病人的所有行为都是来自于那个标签，而排除病人所有其他的特性。所以，这些被试做笔记的行为才被认为是另一种症状表现。

贴标签会带来相当大的危害性。标签不但使医院工作人员囿于这个标签来看待、解释病人的所有行为；病人也会以诊断标签来看待自己的一切；更重要的是，被贴上标签的人还会受到周围人的歧视，嘲笑，冷落等不平等的对待。这对于心理问题的治疗和康复极为不利。因此，尽管现在人们仍在继续使用心理诊断和诊断标签，但认识到了上述问题后，在使用心理诊断和诊断标签时就更慎重。

（采自 Rosenhan, 1973; Hock, 2004）

（二）研究设计

人种志研究设计的核心内容是确定如何进入研究现场、以长时间的参与观察搜集研究资料，形成描述性分析。

人种志研究是在日常生活中进行的，研究假设通常是随着资料的收集而产生，探究的主题也是在逐步分析中产生的。因此，做好如何进入现场，为被研究者接纳等准备是该研究要解决的首要内容。为了研究的方便，人种志研究通常

寻找一个研究现场中的人作为“联络人”的角色，他是研究者和研究对象沟通的“翻译”和中介。

搜集资料用的是非结构方式，通过广泛地做笔录而详细描述他们所观察到的所有细节。由于单纯的观察无法获得心理现象的深入理解，应该采用三角验证的方法，结合访谈、晤谈、文献等方法来获得有效的资料。

形成研究结果的过程通常也是理论建构的过程，人种志是长于理论建构的方法。形成理论的方法通常称为扎根理论。扎根理论（grounded theory）强调资料优先于理论，是基于所搜集的资料，发现、归纳出其中的核心概念而形成理论，并不断地将理论与资料进行对比，直到归纳出的“理论”能够解释全部数据。

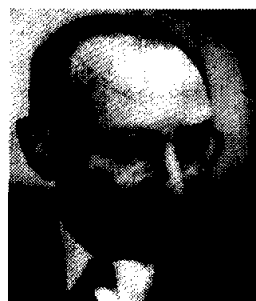
三、行动研究

（一）概念

行动研究（action research）在 20 世纪 40 年代从勒温（Lewin, K）的工作中发展起来 70 年代末期在心理学的研究中受到重视（巴尼斯特，1994）。行动研究不是单纯质的研究方法，而是一种偏重于质的研究方法的综合研究取向。行动研究是从实际需要中寻找课题，在实际工作过程中由实际工作者与研究者的共同参与，通过全程干预，使研究成果为实际工作者理解、掌握和应用，从而达到解决实际问题、改变社会行为的研究方法。行动研究是适合于教育、咨询、健康心理学、职业心理学等领域广大实际工作者的研究方法。

补充讨论 9-8 勒温与行动研究

美国参加二战期间，需要将好的肉食大量供应前方战士，因而导致后方肉食的短缺。说服居民食用内脏，改变其食精肉或排骨的习惯，是当时紧迫的社会问题。社会心理学家勒温与玛格丽特·米德一起进行了关于“食物习惯”的研究，这是一种经典的源于生活并且作用于生活的研究。由于研究卓有成效，他们不但通过研究解决了当时紧迫的社会问题，而且在这种研究中，提出了关于团体决策的心理学理论，以及关于心理生态学的心理学理论等。



在另一项勒温和其学生与犹太人、黑人进行合作研究中，实践者均以研究者的身份参与到研究之中，积极地对自己的境遇进行反思，力图改变现状。勒温将这种结合了实践者智慧和能力的研究称为“行动研究”。影响广泛的行动研究法，就是在这样的研究中得到体现并逐渐成熟的。

行动研究的主要特征是为行动而研究、由行动者研究和在行动中研究。为行动研究指出了研究的目的，行动研究从亟需解决的问题开始，通过行动与研究的结合，理解和解决实践工作者遇到的问题。在这里，因果关系并不是最重要的，相反，“今后该怎么做”却是迫切需要的。行动者研究表明研究的主体是实践工作者，专家、学者提供意见与咨询，是协作者。因此，行动研究也是促发行动者学会反省、扩展视野，提高探究问题与解决问题的能力，促进发展的研究。在行动中研究指出了研究的情境和研究的方式，研究的环境就是现实的工作情境。情境的特定性使行动研究结果不宜作情境推论。

（二）行动研究的设计

行动研究的设计包括计划、行动、观察和反省四个环节，如图 9.5 所示。研究以对问题的界定与分析为起点，包含对计划及其实施情况的评价，并在这种评价的基础上加以改进。

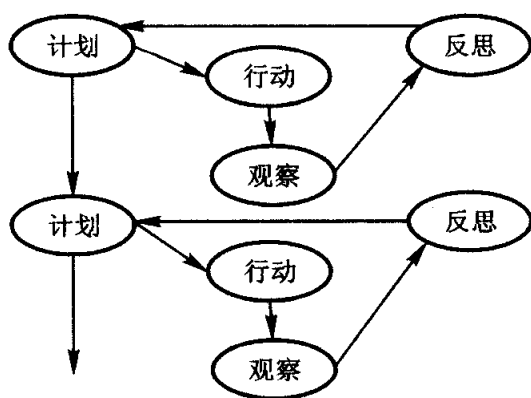


图 9.5 行动研究的螺旋循环示意图

计划是行动研究的第一环节。行动研究从诊断需要解决问题入手，形成假设来解释该情景的事实，并设想改变当前状态的行动方案。计划包括总体设想和具体的行动步骤，既包括存在的问题，受制约的因素，又包括解决的设想，即具备可行性。计划又必须具有充分的灵活性、开放性，能随需要把始料不及、未曾认识、在行动中发现的情况、因素容纳进去。

行动是指计划的实施，是行动者有目的、按计划的实际行动过程。行动促进行动者认识的改进和行动所在环境的改进。在行动中，要协调各方面的力量，随着对行动及背景认识的逐步加深，及各方参与者的观察和评价建议，考虑实际情况的变化，进行不断的行动调整。

观察是指对行动的过程、结果、背景以及行动者的特点的考察，是数据搜集的过程。搜集数据是行动研究的中心，也是反思、修订计划和进行下一步行动的

前提条件。行动研究需要搜集质性和量化两类数据，因此需要灵活运用各种数据、资料的采集分析技术，系统收集行动背景、行动过程和行动结果的资料。背景资料是分析计划有效性的基础；过程资料是判断方案效用及其模式的依据，主要记录计划实施者的参与方式，材料的使用，主要活动的安排，意外和干扰排除等；结果资料是确定行动方案效果的依据，主要记录各种预期的与非预期的、积极的和消极的结果。

反思是一个行动循环的终结，又是过渡到另一个循环的中介，包括整理和描述数据、做出解释和评价。整理和描述是对观察感受到的、与制定、实施计划有关的各种现象归纳整理，描述出本循环过程和结果。评价解释是对比计划与结果而作出判断评价，提供下一步行动计划的判断和构想。

►本章提要

1. 质的研究是逐渐为众多的心理学研究者所认可的研究取向，广泛运用于心理学的各个领域。

2. 质的研究是研究者试图从参与者的视角，在研究对象的自然情境中、以归纳的方式搜集和分析资料，而达到对心理现象及其意义理解的过程。

3. 质的研究的基本特征是以寻求意义和理解为目的，在自然的情境中、以研究者作为研究的主要工具、注重描述性的资料、运用归纳的方法、形成描述性的研究结果。

4. 质的研究在理论基础、适用对象、研究过程和研究呈现方面与定量研究都存在差异。在研究实践领域，两种方法通常是交错使用、相得益彰。质的研究方法更适合从被研究者的视角、采用归纳方式进行研究的课题，对于处理具有特殊性、意义性、陌生性等特点的问题更具有优势。

5. 质的研究设计中必须考虑五个因素是研究目的、观念架构、研究的问题、研究方法和效度，其中研究目的是研究的根本和出发点，是确定其他研究要素的基础，概念情境与研究目的紧密相连，并影响研究问题的形成，方法则是针对回答的问题汇集和分析资料，是研究的可行性和效度的保障。

6. 研究设计论证是解释说明研究的目的、主题和意义，以及当前研究将采用的方法及其适当性做出的详细说明的文本，是研究设计的总结。可用于向特定听众传达并论证研究。

7. 心理传记研究、人种志和行动研究三种具体的质的研究形式，分别适用于不同的研究领域。

►本章关键术语

编码(coding)

非结构化抽样(unstructured sampling)

观念架构(conceptual framework)	行动研究(action research)
互动模式(interactive model)	叙述访谈(narrative interview)
结构化抽样(structured sampling)	研究报告(research report)
理论饱和(theoretical saturation)	研究的目的(purpose of a study)
理论抽样(theoretical sampling)	研究关系(research relationship)
目的取样(purposive sampling)	研究假设(hypothesis)
人种志(ethnography)	研究设计论证(research proposal)
三角验证(triangulation)	研究问题(research question)
效度(validity)	扎根理论(grounded theory)
心理传记(psychobiography)	质的研究(qualitative research)

► 复习与练习

1. 质的研究的基本特征是什么？
2. 选择下列“研究题目”中的一个，讨论使用质的研究方法的可行性，并提出自己的研究设计。

中国人自尊结构的探索	公务员培训的需求研究
一个成功者的心路历程	农村高中决策研究
外出民工打工动机的动向研究	私立预备学校中教师互动模式研究
大学生未来时间定向探索	自由职业者工作特点研究
溪口中学推进人格养成教育的研究	对热衷于探险行为的人的研究
企业员工激励与约束机制探索	对×××志愿者的研究

► 推荐参考读物

车文博. 弗洛伊德. 列奥纳多·达·芬奇和他对童年的一个记忆. 弗洛伊德文集第四卷, 长春出版社, 1998

马斯洛. 自我实现的人. 许金声等译. 北京: 三联书店, 1987

胡寄南. 现代中国优秀人才的个性心理品质. 胡寄南心理学论文选, 上海: 学林出版社, 1985

玛格丽特·米德. 萨摩亚人的成年—为西方文明所作的原始人类的青年心理研究. 周晓红, 李姚军译, 杭州: 浙江人民出版社, 1988

► 在线资源

ATLAS/ti 用于图形, 声音, 影像, 文字的分析处理, 见 www.atlasti.com/

NUD.IST 用于团队协作研究中处理大量数据; NVivo 用于定性合并, 探索、

修饰模型；适合处理多媒体等复杂的数据以及用于深层的分析，见 www.qsrinternational.com/

关于质的研究的一般资料，可以浏览在线杂志 *The Qualitative Report* ,网 址 www.nova.edu/ssss/QR/

第三编 资料的统计分析 与 SPSS 应用

科学研究离不开资料分析——特别是统计分析,从科学研究的步骤上说,对资料的统计分析决定于研究者的先期假设以及相应的研究设计,发生于观察、实验、测量等收集到数据之后。因此,逻辑上,对资料的统计分析应当在研究设计(本书第二编)之后。在这个意义上,统计分析似乎不是科学研究中最重要的一环。但在本编中,读者会发现统计分析的重要意义体现在它同样制约着其他的研究环节。正如 David Salsburg 所说:“一流的科学家可以做出很有价值的实验,产生新知识;二流科学家只是忙于各种实验,收集大量数据,但对知识的累积没什么用处。”统计方法正是帮助一流研究者通达真理发现的利器!

本编“资料的统计分析与 SPSS 应用”将按照资料分析的逻辑顺序并结合 SPSS for Windows 统计软件介绍从数据整理到描述统计再到推断统计,从初等统计到高级统计的线索,教会读者如何掌握和妥善利用统计知识和技能为证实研究假设和科学发现服务。本编包括四章,第十章主要介绍资料的预处理过程,包括如何对资料进行量化、编码、录入并进行初步的描述统计,在补充材料中将对数据的探索性分析这一新近的统计分支作一介绍。第十一章主要介绍推断统计的有关知识和 SPSS 应用,包括抽样分布和假设检验原理以及 SPSS 中实现各种统计检验的菜单功能。第十二章围绕第二编涉及到的各种研究设计,通过实例分析着重介绍其统计分析的 SPSS 操作过程,便于读者模仿学习。最后,作为一种常用的高等统计方法,第十三章将介绍因素分析原理及其 SPSS 应用。

第十章 数据的整理和描述

统计

采用是 / 否式的“迫选式”回答方式调查人们对于某个事件的态度，或者采用结构化问卷调查大学生的自尊水平，又或者用点击率来反映不同性别用户对不同类型互联网网页的偏好程度……以上这些“研究”得到的资料或数据具有相同的特征吗？可以使用相同或类似的方法进行统计分析吗？如何将这些资料变化、录入到 SPSS 统计软件中去？

日常生活中，我们常用平均数这样的概念来反映一组数据的大致特征，例如，你会盘算这个月你平均每天花了多少生活费。如何才能准确地、快速地反映数据的关键特征？统计学用什么办法来解决这些问题？如何利用 SPSS 帮助达成这些目的？学习本章，你将了解到：

1. 资料如何进行编码、量化。
2. 变量或数据的基本类型及其特征。
3. 如何利用 SPSS 快速读取数据和对数据进行初步整理。
4. 如何利用 SPSS 绘制常用的统计图、统计表。
5. 如何对数据的关键特征进行描述和探索。

第一节 资料的量化和录入

任何研究设计最终都会得到一些数据资料，然而杂乱无章的数据须经过整理才会显现出有用的信息。初学者往往将统计学仅视为数据分析和呈现数字结果的工具，事实上，数据如何被收集以及为什么要这样收集和具体的数据分析过程一样重要。本节首先介绍如何将原始研究资料进行编码、量化，转变成可以进行统计操作的数据和变量类型，接着结合 SPSS 统计软件介绍如何录入数据。

一、资料的量化

资料量化的目的是将原始研究资料转换成电脑或类似机器能够判读和处理的信息——机读格式 (machine-readable form)，以便于进一步的量化分析 (quantitative analysis)(艾尔·巴比, 2000)，它是伴随着电脑在统计分析中的深入应用不断发展的。这里所谈的资料量化主要即指资料的编码 (coding)。

（一）资料的编码

原始资料要转换成便于进行统计分析的数据，必须借助于编码，即将研究资料加以类别化并转换成计算机可识别的符号的技术性程序和过程。一方面，目前统计分析普遍使用计算机和专业统计软件，无法对自然语言直接进行识别和分析；另一方面，为了方便研究各环节的参与者交流，编码通常是遵循一定规则的系统化、结构化的过程，不是随意地用某个符号代表。

1. 编码和编码框架

通常，编码的码值可以是数字或代码（字符串）。例如，问题询问您的性别，男/女这两种答案本身就可作为码值，更习惯化的方式是将其编为 1 和 2 这样的数字型码值。编码可以在研究设计的同时进行，便于研究者直接将被试的反应记录和输入为相应的码值，可称为事前编码；也可在获取被试反应的同时进行，称立即编码；还可在研究之后根据研究目的和所记录的反应，构建编码系统，称事后编码。

编码的关键环节是类别化（categorization）和制订编码框架（coding frame）。类别化过程应当遵循目的性（类别基于研究问题和目的）、完备性（类别无遗漏，能涵盖所有可能反应）、互斥性（任意反应能够且仅能分配到一个类别）和单维性（各类别处在同一维度或水平上）等原则。例如，某研究者调查中学生利用网络的现状，设计了这样一个题目：

当你上网时，你通常是（请将你选择的项按所用时间从多到少的顺序写在横线上）：

看新闻 ②玩游戏 ③看电影 ④查资料 ⑤学习（如学习英语等） ⑥其他_____

在其提供的前 5 个选项中，“查资料”和“学习”违反了互斥性原则，可能会造成被试回答时的迷惑。其次，也违反了单维性原则，题项的编码框架混乱，容易导致被试反应时信息遗失。研究者实际上试图收集两种信息，一是各种网络利用方式的频数分布（适用计数数据分析方法）；二是利用各种方式的偏好（适用等级资料分析方法）。如果被试将所有 5 个选项都选中并进行主观排序，则会损失第一种信息（各类别频数相同），读者会发现，此时的编码（假设为 1~5）并不在同一个维度上。这里，应当将评价选择和排序回答分开。

对年龄这类测量型数字信息，可以原封不动地用于分析。但许多时候要根据实际情况和统计分析的需要对原始数据重新分类，比如月收入可分为 500 元、500~1 000 元、1 000~2 000 元、2 000~5 000 元、5 000 元以上等；受教育水平可分为高中及以下、本科、研究生等。文字型信息则属于名义型变量，可用数字或符号作为代表，如上述的性别，再如职业类型，可以归类为工人、农民、教师、机关干部、企业员工、商人等，分别用 1、2、3、4、5、6……表示。此外，编码有时还需要

考虑未作回答或极端回答的情况，如通常都采用一个特殊的码值。

编码框架亦称分类框架（classification scheme），通常是包括了所有问题的所有可能回答和相应的代码的表格，研究者可依此对所收集资料进行编码。通常，它和指导说明一起组成编码簿（code book）对研究测试者、资料登录者、数据分析者及报告撰写者提供指引，对于牵涉到大规模数据资料的研究（例如编制常模）而言，准确而完善的分类框架和编码簿至关重要。如表 10 - 1 所示，编码框架的内容通常包括变量名及说明、变量值编码及说明、变量对应数据列及问题、被试代码等信息。由此每个被试的原始信息就全部成为一串数字，便于计算机录入（参见表中的编码示例）。

表 10 - 1 编码框架示意

问题 号码	变量名	变量含义	编码范围 /占用列数	编码示例	编码含义
1	ID	被试编号	3 列	150	第 150 号被试
2	day	资料收集时间	6 列	031105	2003 年 11 月 5 日
3	gender	性别	1 或 2	1	男性
4	grade	年级	1 ~ 10	1	初中一年级
				6	高中三年级
				8	大学二年级
5	timepd	每日上网时间	1 ~ 5	1	< 半小时
				2	半小时 ~ 1 小时
				3	1 小时 ~ 2 小时
				4	2 小时 ~ 4 小时
				5	4 小时以上
6	active	在网上主要活动	1 ~ 8	1	浏览新闻
				2	玩游戏
				3	看电影
				4	听音乐
				5	聊天交友
				6	搜索或下载资料
				7	远程学习
				8	其他
...

2. 几种常见的资料编码方式

不同的研究设计和资料收集方法，编码方式可能差异很大，这里列出几种常

见方式供读者参考。

• 封闭式题项：这种题项的回答通常只包含有限个确定的类别，比较易于编码，只需注意各答案和编码的内在逻辑顺序即可。如：

(1) (单选) 您目前的婚姻状况：反应编码

1 ☐ 未婚 2 ☐ 已婚 3 ☐ 离婚后未再婚 4 ☐ 离婚后再婚 5
5 ☐ 丧偶后未再婚 6 ☐ 丧偶后再婚 7 ☐ 未婚同居

(2) (选项固定，可多选) 您的学习目的：反 应 编 码

1 ☐ 考高分 2 ☐ 增长知识 3 ☐ 让父母高兴 4 ☐ 将来找个好工作 1356000 or
5 ☐ 让老师喜欢我 6 ☐ 让同学羡慕我 7 ☐ 其他 1010110

(3) (多选排序题) 您择业中考虑的主要因素有 (依据重要性先后排列，限选三项)

1 ☐ 收入 2 ☐ 福利 3 ☐ 专业 4 ☐ 发展 5 ☐ 地区 6 ☐ 地位 反应编码
7 ☐ 风险 8 ☐ 劳动强度 9 ☐ 个人爱好 10 ☐ 其他 148

(4) (不限定排序题) 您择业中考虑的主要因素有 (依据重要性先后排列)

1 ☐ 收入 2 ☐ 福利 3 ☐ 专业 4 ☐ 发展 5 ☐ 地区 6 ☐ 地位 反应编码
7 ☐ 风险 8 ☐ 劳动强度 9 ☐ 个人爱好 10 ☐ 其他 1480000000

• 内容分析中的编码：内容分析法 (method of content analysis) 是收集和分析文本中专项内容 (包括语词、意义、画面、符号、概念、主题等) 的技术，这里，“文本”可指任何口头和书面的语言材料，也可指非语言和形象材料。内容分析往往需运用客观、系统的计数和记录程序提取所需的信息，由于往往牵涉到大量文本，研究者越来越倾向于使用计算机进行资料分析和信息提取。内容分析中的编码除了要确定分析或记录单元，如字词、主题等，还要区分情景单元，即分析或记录单元出现的上下文，例如要研究小说中的冲突主题，表明冲突的各种行为是分析和记录单元，冲突发生的背景或性质即是情境单元，如发生在配偶、同事、路人之间，是口角冲突或械斗等。

内容分析编码的特殊之处在于某些隐含信息在文本中无法直接找到，只能通过上下文的语义分析和一定的操作性定义加以确定，这可称为隐性编码 (latent coding)，例如确定小说中人物的情绪状态和人格特征等。再如要分析新闻媒体对名人社会效应的描述，在设定变量的编码时就要考虑：何谓“名人”？影视明星、歌星算不算？贪官污吏算不算？“社会效应”的操作性定义是什么？这种效应可能有哪些类型？如果名人仅限于正面形象，则诸如“杰出的”、“伟大的”、“成功的”等褒义语词可用来确定显性编码 (manifest coding) 而如果要讨论反面形象，则丑闻、获罪等可用来确定隐性编码。类似地，也可像表 10-2 那样编制一个编码框架。

表 10-2 内容分析编码框架示意

变量名	变量含义	编码示例及含义
medname	媒体名称	中华网(字符型变量)
pressday	新闻发表日期	031105—2003 年 11 月 5 日
title	新闻主题	经济(字符型变量)
celeb	是否名人	1—名人 2—非名人
career	行业	1—官员;2—社会活动家; 3—艺人;4—企业/商业人士; 5—自由职业者;6—其他
assess	形象评价	1—极为反面;2—比较反面; 3—中性;4—比较正面; 5—极为正面;0—无法确定
...

(二) 资料的分类

资料的分类指将实验或调查中收集的数据信息按照某个框架或标准划分成方便电脑进行分析处理的不同类型,是对资料编码、量化的自然结果。决定资料类别的主要标准是变量的测量尺度(measurement scale),或称为变量的测量水平、等级等。这个问题和前面所讲的编码密不可分。

关于变量的类别参见第一章“研究类型与变量”,从统计分析的角度看,需要强调变量和数据之间的关系。前面提到通常区分变量为四种测量类型,即名义变量、等级变量、等距变量和比率变量。同一个变量,测量的尺度不同,其观测数据所含有的数学特性和信息类型也是不同的。比如,同样是某个测验的得分,如果题项是一个是非题,则得到的数据就只能是“是”和“否”,即便给它们分别编码 1 和 0,其表达的意义只能是一个“标识”;如果是一个主观题(例如作文题),则观测的数据可以是 0 到满分之间的任意值。因此,即便同样的编码,数据类型会因为测量尺度而不同。与其说研究者关心的是变量的测量尺度,不如说是关心变量观测值(即数据)的测量尺度。因此就测量收集到的资料类型而言,名义尺度也可视为一种测量尺度。

数据的测量尺度直接关系到它适用于何种统计方法。通常,研究的目的、涉及到的变量个数、研究设计类型、样本间的关系以及观测指标的测量尺度等条件决定了应当选取何种统计方法(Hansen,2000)。此外,不同的测量尺度也决定了测量具有不同程度的准确性和敏感性,例如,编制问卷的计分标准或选择反应

量表时，应尽可能选取最敏感、最具区分度的量尺。所以要使某个行为被准确地量化，最好采用连续型的（等距或比率）量尺对其加以测量。

特别要指出的是，SPSS 软件对变量类型的规定有所不同。在 SPSS 中 数据资料的基本形式是变量和记录（case）交叉形成的矩阵式数据集，输入变量或其观测值时都需要确定变量类型。表 10 - 3 所示为 SPSS 中的主要变量类型及其属性。

表10 - 3 SPSS 中各变量类型及其属性

SPSS 变 量类型	系统默 认位数	小数 位数	输入方式	显示方式	示 例	
					输 入	显 示
Numeric	8	2	标准格式或 科学记数法	标准格式数值， 圆点表示小数点	65.789	65.79
Comma	8	2	带逗号的数值 或科学记数法	圆点为小数点，逗 点为三位分割符	1,343,438.10	1,343,438.1
Dot	8	2	带圆点的数值 或科学记数法	逗号为小数点，圆点 为三位分割符	34,3434E2	3.434,3.4
Scientific notation	8	2	标准格式或 科学记数法	科学记数法	457.8E4	457.8E4
Date			格式繁多	格式繁多		
Dollar	8	2	带 \$ 或不带 \$ 或 科学记数法	有效数值前带 \$ 以逗 点为分割符	\$12343	\$12343
Custom Currency			用户自定义型			
String	8	/	一串字符串		sex	sex

二、资料的录入

数据经过整理、“量化”成数字或符号后需要转入读卡机或计算机中以便进行分析。早期的资料需要在编码之后由专业人员打上卡片或录入磁带，再由读卡机或计算机录入资料。新近发展的光学标记阅读机（optical marker reader）则更加快捷、方便，被试只需要在特制的答卷纸上按规定格式在相应问题的答案上做上记号，之后就可以通过阅读机很快地实现录入及同计算机的对接和传输。然而这种录入法也有缺陷。首先，由于扫描卡只适用于选择性题目，问卷或其他

收集数据的工具，题型受到限制，如无法使用开放型问题。其次，问卷格式与扫描卡格式必须一致，如有改变，扫描卡就需要调整，在技术上要求高，成本大，不适用于小型研究。第三，填写扫描卡有一定的难度，有些文化程度低的人可能不明白要求，容易出错。以下主要介绍如何在 SPSS 软件中实现资料录入。

（一）SPSS 操作界面简介

SPSS 即“统计产品和服务解决方案”（Statistical Product & Service Solutions）（2000 年前称“社会科学统计软件包”，Statistical Package (for) Social Sciences），是社会科学领域应用比较广泛的统计软件之一，目前已发行至 13.0 版本。教材主要结合 10.0 版本进行讲解。

图 10.1 所示即 SPSS 的主界面，也是其数据编辑窗口，类似于 MS Office 办公软件中的 Excel 界面，它包括标题栏、菜单栏 10.0 中共有 10 个下拉菜单）、工具栏，工具栏下方是数据输入区，其下方则是一个矩阵式的数据表。该表由若干行和列组成，每行对应了一条记录或案例，每列则对应了一个变量。窗口最下方是状态栏。

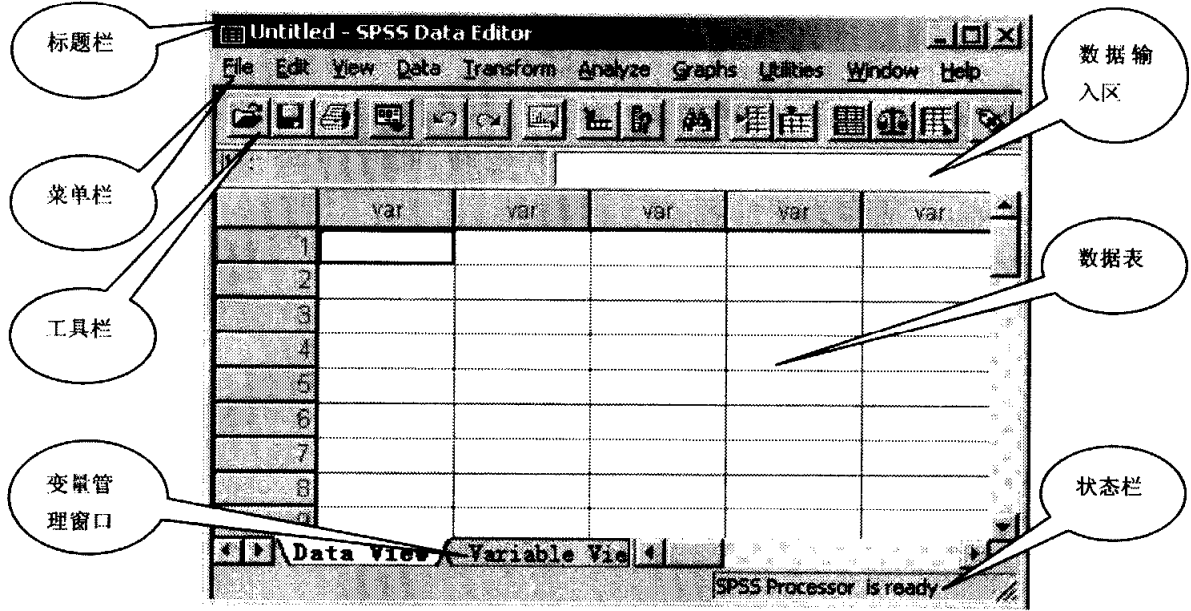


图 10.1 SPSS 主界面及数据编辑窗口

SPSS10.0 相比以前的版本最大的改进之一是在主界面中增加了一个变量管理页面，点击 Data View 旁边的 Variable View（或双击变量名）进入变量管理 / 定义窗口（图 10.2）。通过它可以方便地实现变量管理。这个新窗口形式上唯一的不同是中间的变量集替代了原来的数据集，注意变量集的构成方式：此时变量为行（case），而其属性为列，可以分别编辑或修改这些属性以帮助正确分析，便于观看分析结果。而在以前的版本中管理变量往往需要通过 Data 菜单中的

Define Variable 过程实现。变量包含的属性和定义方法如下：

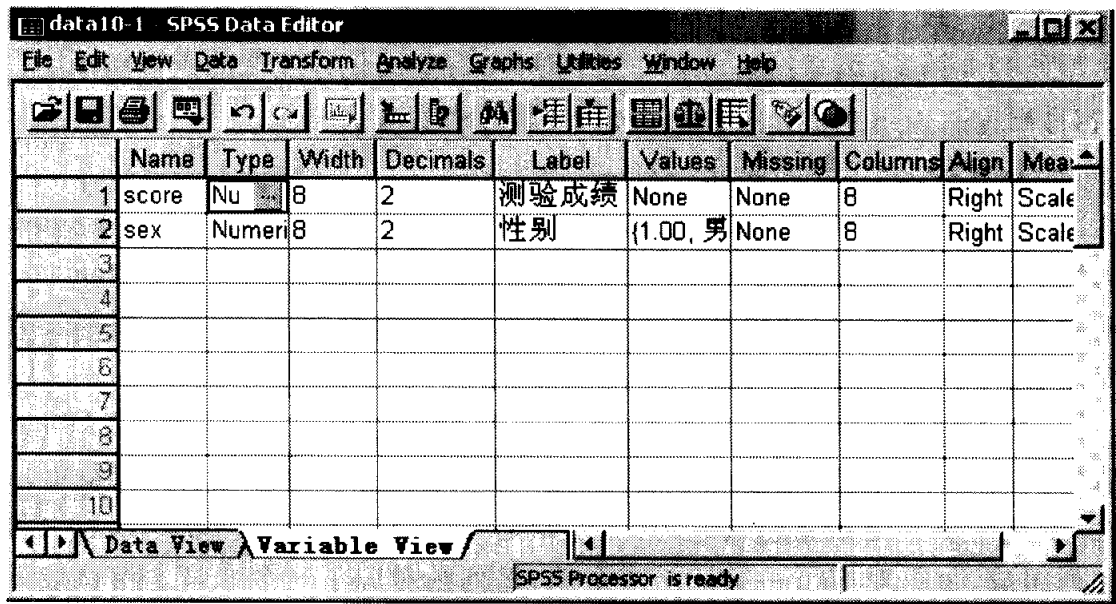



图 10.2 SPSS 主界面变量管理窗口

- 变量名(name) 如 gender (表示性别)，长度不超过 8 字符，通常为字母或数字，不区分大小写，变量名中字符使用有限制，同时 SPSS 保留字（如 ALL、NOT、BY、OR 等）不能作为变量名。
- 变量类型（type）如 gender 可定义为字符型（需输入字符），也可默认为数值型（只需输入 1、2...），单击该列单元格右侧，弹出变量类型对话框，可选择不同类型和改变长度。参见表 10-3。
- 变量长度(width) 默认为 8 位（字符型变量 8 字节），可更改，但显示效果受显示列宽（column）限制。单击单元格右侧  可改变。
- 小数位数（decimals）默认为 2 位（包含在变量长度中），输入超过限制四舍五入。
- 变量标签（label） 变量名的附加说明 因为 8 位的变量名往往不足以说明变量含义。可输入中文，如将 gender 标签输入为“性别”。定义标签后，鼠标停留在数据窗口中变量名位置可显示标签内容，同时输出结果用标签代替变量名呈现。变量标签要和下面的变量值标签严格区分。
- 变量值标签（value label），对变量取值的附加说明，因为变量值通常为数字或字符，某些时候含义并不明晰。单击单元格右侧，弹出值标签对话框（如图 10.3）上、下方各输入变量值和标签 单击 Add，最下方框中显示出值和标签对应关系。如 gender 取值为 1、2，定义值标签分别为男性和女性。选中某个已定义的标签，更改后按 Change 接受，也可以去除（Remove）标签。可见，变量值标

签主要对分类变量有意义，因其取值的数字往往没有意义；连续变量观测值只要清楚单位，其意义一目了然。和变量标签一样，定义变量值标签的好处是可以在结果文件中直接看到标签。此外，在数据窗口中选中 View => Value Labels 数据以标签形式显示。

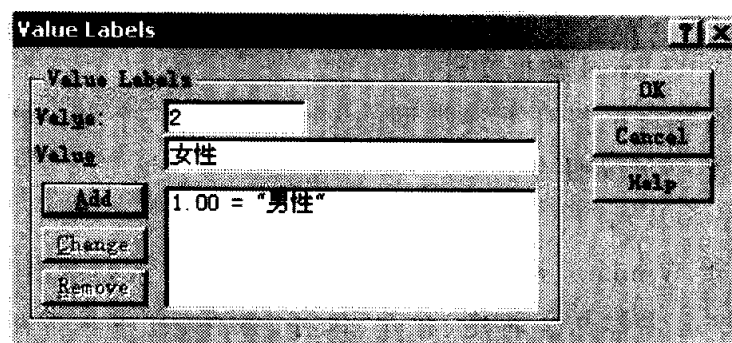



图 10.3 变量值标签 Value label 对话框

- 缺失值 (missing value)，定义缺失值以便于进行缺失值分析 (missing value analysis)。通常缺失值被视为错误而忽略，但产生缺失值的原因很复杂，比如被试可能不愿意回答某些问题，或者不知道答案，或者答案出乎意料，如果不对缺失值进行过滤或处理就会影响分析的结果。缺失值的处理参见数据的整理。

- 变量显示宽度 (columns) 默认为 8 字符，通常应大于变量长度和变量名长度，单击单元格右侧  可改变，也可以直接在数据管理窗口拖动单元格间栅格线 (grid line) 调整。

- 变量对齐方式 (align)，分三种情况，数值变量默认右对齐，字符变量默认左对齐。

- 变量量尺 (measure)，分三种情况：等距 (scale)、等级 (ordinal)、名义 (nominal)，含义不再解释。

(二) 直接打开数据文件

用户可以像利用 Excel 输入数据一样直接在数据窗口输入数据，可利用小键盘、Page UP、Page Down、Home、End 键切换单元格。注意受到前面所讲的数据格式的制约，已输入数据的单元格中数据的显示值可能和输入值不同。不过通常不采用直接输入方式。

SPSS 现在可以直接读入许多格式的数据文件，其中就包括 Excel 各个版本的数据文件。选择菜单 File ==> Open ==> Data，系统就会弹出 Open File 对话框，单击“文件类型”列表框，在里面能看到直接打开的数据文件格式，分别是：

SPSS(*.sav)	SPSS 数据文件	Lotus(*.w*) Lotus 数据文件
SPSS/PC+(*.sys)	SPSS4.0 版数据文件	SYLK(*.slk) SYLK 数据文件
Systat(*.syd)	Systat 数据文件	dBase(*.dbf) dBase 系列数据文件
Systat(*.sys)	Systat 数据文件	Text(*.txt) 纯文本文件
SPSS portable(*.por)	SPSS 便携格式数据文件	data(*.dat) 纯文本文件
Excel(*.xls)	Excel 数据文件	

注意,SPSS 能够直接打开上述类型的文件,同样,一个新的数据集也基本上可以保存为同样类型的文件。

☆ SPSS 操作提示和技巧

和大多数基于 Windows 系统的软件(如 MS Office 系列软件)一样,SPSS 的文件菜单(File)中通常都包含新建、打开、保存、另存、打印等常用命令并有相应的工具条。从新建和打开命令中可以看出 SPSS 中常见的文件类型,和一般的软件不同的是,这些文件类型往往联系密切,比如 *.sav 文件是 SPSS 的数据文件(data file),*.spo 文件是数据经过统计分析产生的结果文件。SPSS 采用的是窗口管理模式,方便使用者打开、编辑、存储文件。SPSS 主要包括 5 种窗口:数据管理窗口(data)、数据输出窗口(output)、语句窗口(syntax)、图表管理窗口(chart)、帮助窗口(help)。各窗口的结构和特点将陆续介绍,详细内容读者也可参阅推荐读物。值得提醒的几点是:

① SPSS 一次只能打开一个数据文件,所以经常存盘或另存是个好习惯!对当前数据执行的一个命令没有结束前,不能打开新的数据文件。

结果和语法文件虽然可以打开多个,但新产生的结果和语句默认地添加在最后打开的文件中。

妥善利用好语句,特别是对数据调用重复分析过程时,小小的修改可以事半功倍。SPSS 可以进行完全窗口菜单、程序和混合三种运行管理方式(蔡程辉,庄永龙 2001),第一种方式适用于初学者,是本书的主体内容,但可能无法满足一些精细要求;而一些高级统计功能则只能借助语法运行。

妥善利用好输出结果的格式设置 因为结果文件中的许多内容如表、图等可以拷贝到 Word 文档中直接成为研究报告的内容,设置恰当可以省掉很多编辑工作。

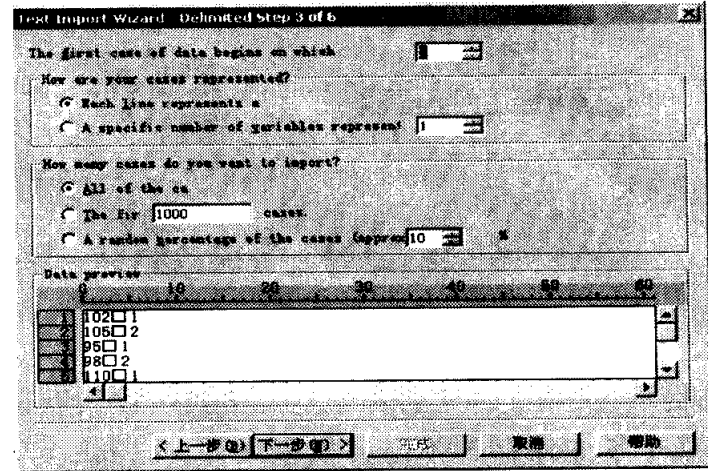
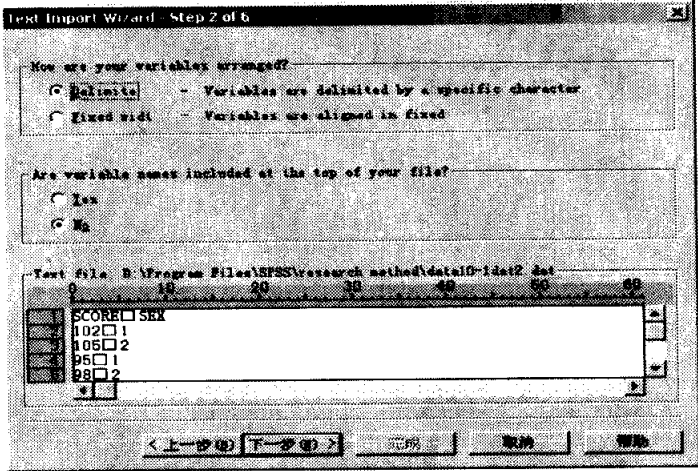
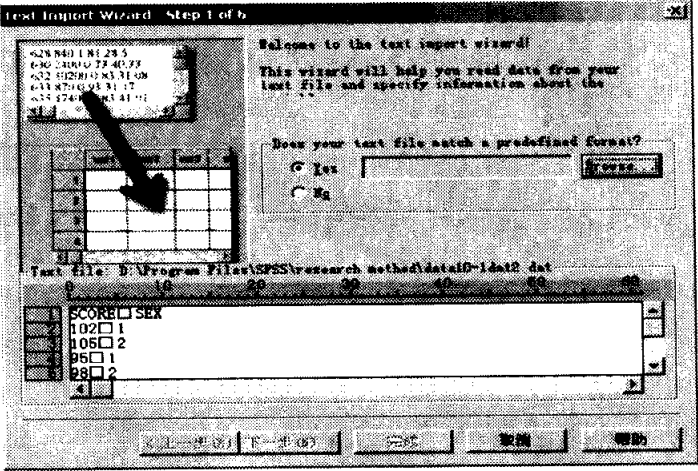
(三) 使用文本导入向导读入文本文件

这里重点介绍文本导入向导(text import wizard)读入文本文件。对于初学者来说,该向导是输入数据的最好帮手,也是最通用的数据读入方式。大多数研究者习惯于先将大量数据按照编码框架用计算机行列对齐输入纯文本文件(如 Windows 自带的写字板、留言簿等程序 后缀名 txt 或 dat)。调用前面提到的 File ==> Open ==> Data 可以直接读取纯文本文件,不过一旦执行,将自动进入该向导。

选择菜单 File ==> Read Text Data ,系统会弹出和上面一样的 Open File 对话

框，文件类型自动跳到 Text (*.txt)。选择文件后，可以看到向导共分 6 步，分别对读入格式做出设置。界面上方通常有一些格式选择框，下方则是做出相应步骤后的数据读入效果预览。举例说明其具体操作。

例 10-1 读入 40 名被试智力测验分数和性别的数据，变量名包含在第一行，两个变量间用空格分隔开。各步骤界面如图 10.4。



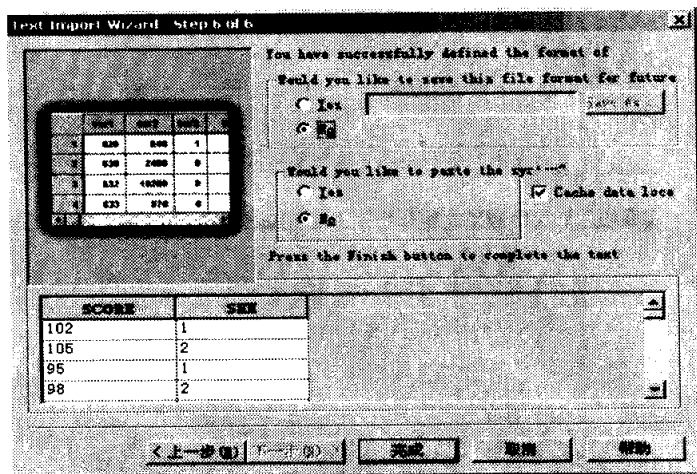
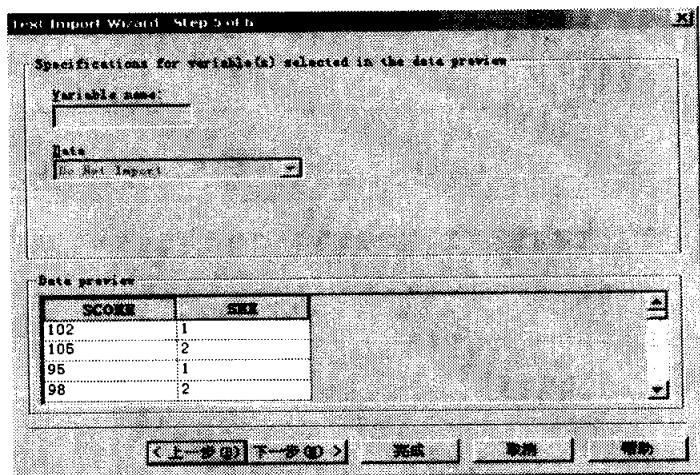
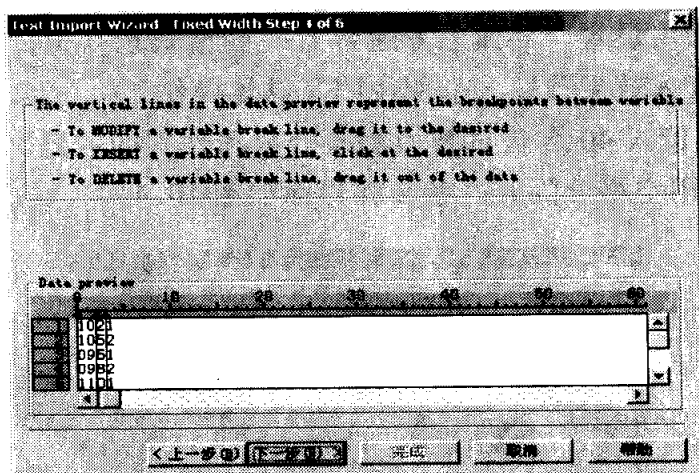


图 10.4 文本导入向导 6 步骤示意

第一步首先提示“文本文件是否和预定义格式一致”，默认选择“No”并单击“下一步”，若有某个事先保存的预定义模板文件（*.tpf）可选择“Yes”。第二

步设定变量按 Delimited (用分隔符区分) 或 Fixed Width (固定宽度) 排列以及是否包含变量名; 后者表示固定宽度, 用于各变量观测值码位相等情况, 前者表示按分割符区分, 这里选择前者。文件中确实包含变量名, 选择 “Yes” 然后进入下一步。第三步设定 “第一条记录从第 n 行开始” 选择 “一行代表一条记录” 和 “导入所有记录”。第四步选择变量的分隔符, 可选择制表符 (tab)、空格 (space)、逗号、分号 (semicolon) 或自定义符号。这里系统自动识别了空格, 从预览中可以看出变量分割正确。第五步已经可以看到初步的读入结果, 可进一步设定变量属性。这里直接单击下一步。最后选择 “是否保存文件格式以备下次使用” 以及 “是否将以上操作粘贴为 SPSS 语句”, 通常可以不管, 单击完成后数据被导入。

若在 Step2 中选择固定宽度, 则在 Step4 中会出现另一个界面, 提示用带箭头的竖直线在变量间进行分割, 如前两列为第一个变量, 则在 2、3 列位置单击插入分割线, 也可以通过拖拉分割线修改或删除变量分割, 不再详述。此外, 若存在多个格式相同的文件需要导入, 则可以利用文件菜单中的 “应用数据字典” 套用已经导入并设置好的文件的格式和标签等属性。

(四) 数据库查询转录

SPSS 不仅可以直接打开许多类型的数据文件, 还提供了另一个适用范围更广、但使用上较为专业的数据转录方式——数据库查询 (Database Query)。不过此种方式的使用与 Windows 系统中安装的 ODBC (Open Database Capture) 引擎有关, 而且由于 SPSS 10.0 可以直接打开 Excel 所有系列的数据文件, 因此一般用处不大。

第二节 资料的初步整理

资料的整理也叫资料净化, 包括无效资料的剔除, 缺失值、异常值以及诸如编码错误、录入错误、数据读入错误等的审核、筛查和修正。另一个重要方面是对数据进行排序、变换、重编码、重分类、分类汇总、加权, 数据文件的合并、分割, 对缺失值的处理等预分析, 目的是使数据条理化、系统化, 更便于统计分析和得出可靠结论。以下分别介绍 SPSS 中进行数据整理的常用过程, 这些命令包含在数据 (编辑) 菜单 (Data) 和 (数据) 变换菜单 (Transform) 下。

一、数据的排序和选择

(一) 记录排序和异常值筛选

一个已读入的数据文件, 调用 Data => Sort Cases, 打开观测量排序对话框,

如图 10.5 左，这是我们第一次认识一个标准的 SPSS 命令过程主对话框（与文件操作对话框和向导对话框等不同）。通常，一般的对话框的结构为：

- 备选变量框：位于最左侧，数据中所有变量和标签均按顺序显示。这里总共有两个变量。

- 分析或处理变量框：位于右侧或右侧上端，有时不止一个，可以从备选变量框选择适当的变量进入相应分析框，每个分析框和备选框间有一个移动变量按钮，点中变量后按钮中央方向箭头呈黑色，显示该变量的去向，单击按钮或双击变量实现移动。这里选择按照“性别”变量排序。

- 单选项：多个带有圆形旋钮的互斥性并列选项，如这里的升序和降序（ascending vs. descending），选择必居且只居其一，选择后圆形旋钮中心有黑点。这里选择降序排列。

- 复选框：一个或多个带有方框的选项（见图 10.6），可多选或不选。复选框和单选项通常对应一些统计分析或处理。

- 下级对话框功能按钮：通常分布在右侧或底端边缘，允许打开新的下级对话框设定统计分析的参数或其他选项，通常在名称后有省略号（见图 10.6）。

- 执行功能按钮：通常分布在右侧或底端边缘，如果是主对话框，则通常为并排的 5 个方框，分别为运行（OK）、粘贴（Paste）、重置、取消和帮助。如果是下级对话框，通常只有返回（Continue）、取消和帮助三个。当选择或指定的变量、参数不合要求时，运行和粘贴按钮显灰不起作用，这两个按钮实际上代表了前面提到的两种运行管理方式。

例如 按 OK 立即运行排序；按 Paste 则暂不执行，而将命令设置粘贴到语句窗口中去。通常，熟练用户都是先利用对话框选择分析过程和参数，然后 Paste 到语句窗口中进行修改，最后将程序提交系统执行，实现混合运行管理，见图 10.5 上、下。

上例中，如果直接进行排序，从数据表中显示可看出，数据排序按照 sex 变量降序排列，每个记录整行移动（保证同一个记录中的不同变量观测值仍在同一行）。此时，由于性别变量排列有序，显然如果存在异常值（包括错误回答、输入和缺失值）就很容易发现并进行处理。如果将命令粘贴到 Syntax 窗口（如图 10.5 下），然后仿照上面的设置复制一个语句，并将变量和排序方法分别改成“score”和“A”。将鼠标移动到第二个语句中（以点号为分隔符）并执行（Run）该语句，则数据集会按照测验成绩变量升序重新排列。选择多个排序变量，按照选择变量顺序嵌套排序，即首先按第一个变量排序，若有等值出现则按第二个变量排序，以此类推，但仍然保证同一个记录中数据不变。

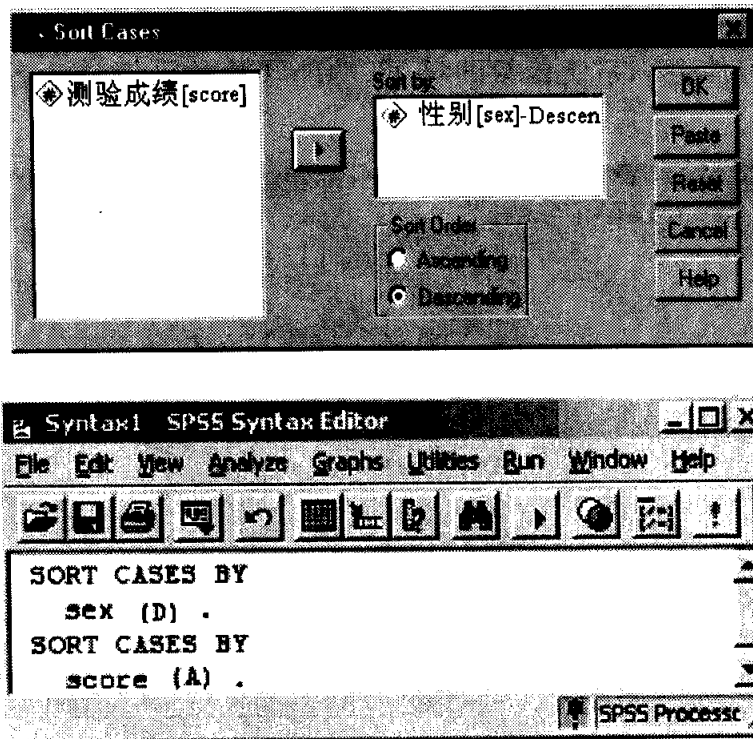


图 10.5 记录排序主对话框和排序语句

(二) 记录的选择

当需要对特定的记录进行分析时需要利用 Data 菜单下的 Select Cases 过程设置记录选择规则，其对话框如图 10.6 上，默认的是选择所有记录，此外规则自上而下为：条件选择（包括一个条件设置按钮），随机抽样选择、特定编号范围选择或使用过滤变量（filter variable）（上一次条件选择产生）选择，还可以选择设置将未选择记录过滤或去除（通常都选择过滤）。不论何种规则，确定后被选择的变量参加统计分析，未选择的变量记录编号上被划上斜线表示其被过滤，若按 Reset 按钮则取消选择，回复到全数据集。此外，打开条件选择中的条件设置按钮对话框如图 10.6 下：实际上许多统计分析过程中都包含这样一个子过程，包括左侧的备选变量框和右侧的条件表达式框，右下方包括一个常用数字 / 运算符输入板和一个函数选择列单。SPSS 中提供了 70 种系统函数，包括算术函数、统计函数、缺失值函数、字符函数、时间函数（DATE./XDATE.）、分布函数（CDF.）、反分布函数（IDF.）、逻辑函数等。使用者可以自定义一个基于选择变量的条件表达式，也可以调用函数（从 Functions 列表拖动右侧滚动条双击选择）并按该函数参数要求选择变量或写入表达式及其他参数规定条件表达式（如 `score <= 100`）。执行条件选择的同时会产生一个新的过滤变量（变量名默认为 `filter_$`），选中的记录对应值为 1 未选中为 0。注意运算符输入板中的逻辑运

算符可以帮助进行多变量组合条件选择，如“和”（&）“非”（~）或（!）等，读者可以自行练习体会。系统函数的符号、功能和参数设置请读者自行参考有关文献。

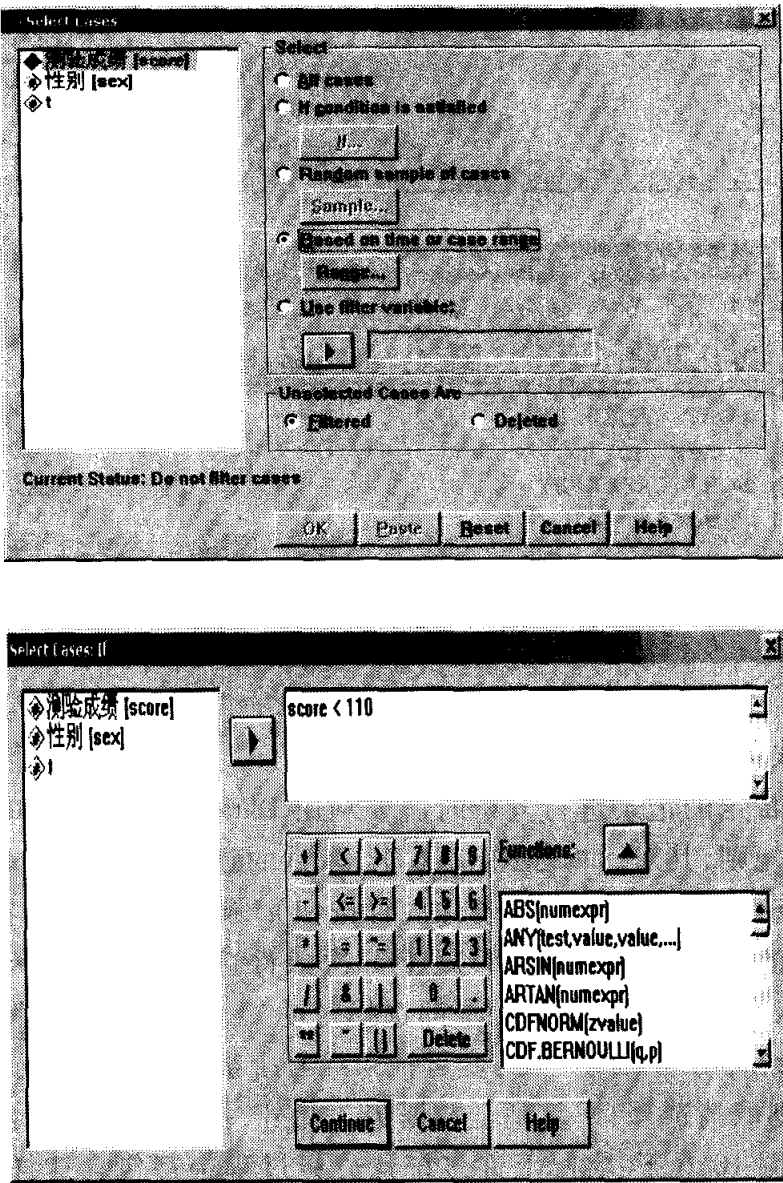


图 10.6 Select Cases 主对话框和 If 条件选择子对话框

二、数据的排秩和重编码

一些统计分析要求对原始数据先进行变形才能使用，比如求等级相关有时需要先求出原始数据的秩 / 等级（rank）；又如简化变量取值的类别，将一些无差异的类别归为一类而需要重编码。这些变换涉及到 Transform 菜单中的 Rank

Cases 和 Recode 命令。

(一) 变量排秩

对连续型变量按照取值大小求秩，并建立一个新变量保存。Rank Cases 主对话框如图 10.7，这里可以看到备选变量有两个去处，可以分别选择进入排秩变量框和分组变量框（将变量按照分组变量各水平分别排秩），在后面还有许多统计过程对话框中都包含这种标题为 By 的分组变量框。左下角的单选框可选择将等级 1 赋给最小值或最大值；分组变量框下方是一个复选框，用以在结果中报告排秩汇总表。下方还有两个功能按钮，可以分别选择排秩的其他方法和重复值 (ties) 处理方法，如按指数分 (savage score)、百分位等级 / 相对等级、百分位等分 (Ntile)、标准正态 Z 分数等排秩；用重复值应占等级的最大值或最小值等赋重复数据秩。具体请读者自行练习体会。

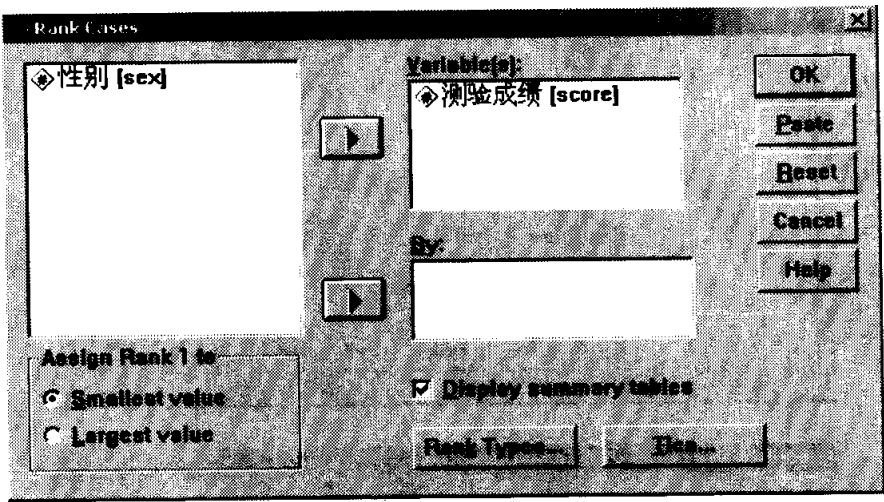


图 10.7 Rank Cases 主对话框

(二) 变量重编码

重编码常用于变量的“降阶”处理，即将高量尺变量转化为低量尺变量。Recode 菜单包括两个过程 Into Same Variables 和 Into Different Variables 为避免原始数据丢失，建议使用将重编码保存为新变量。调用 Into Different Variables，其功能实际上包含了 Into Same Variables 的功能 (如图 10.8 上)。主对话框中的重编码变量框提示将数值型变量转换为新的输出变量，右侧可为该新变量命名（输入后点 Change 钮），编码规则由下方的 Old and New Values 下级对话框规定 (如图 10.8 下)：对话框左、右分别为旧编码和新编码，编码方式包括“点对点”式 (如将 -2、-1、0、1、2 的语义量表数据转换成 1、2、3、4、5)、系统或用户缺失值处理 (如将缺失值编码为 999) 和“段对点”式 (一段范围的数值转换为点值的降阶处理) 三种。此外，注意在重编码变量框下方也有一个 If 按钮 其作用是设

置变量重编码的条件。

实际上 Transform 菜单中还包括三个重编码或标识过程：一是计数（Count）过程，用于将连续或多类别分类变量降阶转换为二分类计数数据，不过这个功能通常可以由其他更有力的过程代替，不再详述。二是自动重编码过程（Automatic Recode），自动将变量按照从小到大或从大到小赋予连续出现顺序并生成新变量，如 10、11、12、12、13、13、14 这样的序列自动重编码成为 1、2、3、3、4、4、5（12 为第三个出现的数），这实际也是一种特殊的排秩方法。三是变量类别化（Categorize Variables）过程，相当于 Rank Cases 中的按照百分位等分（Ntile），可用于将连续型变量按指定百分位转换为分类变量。

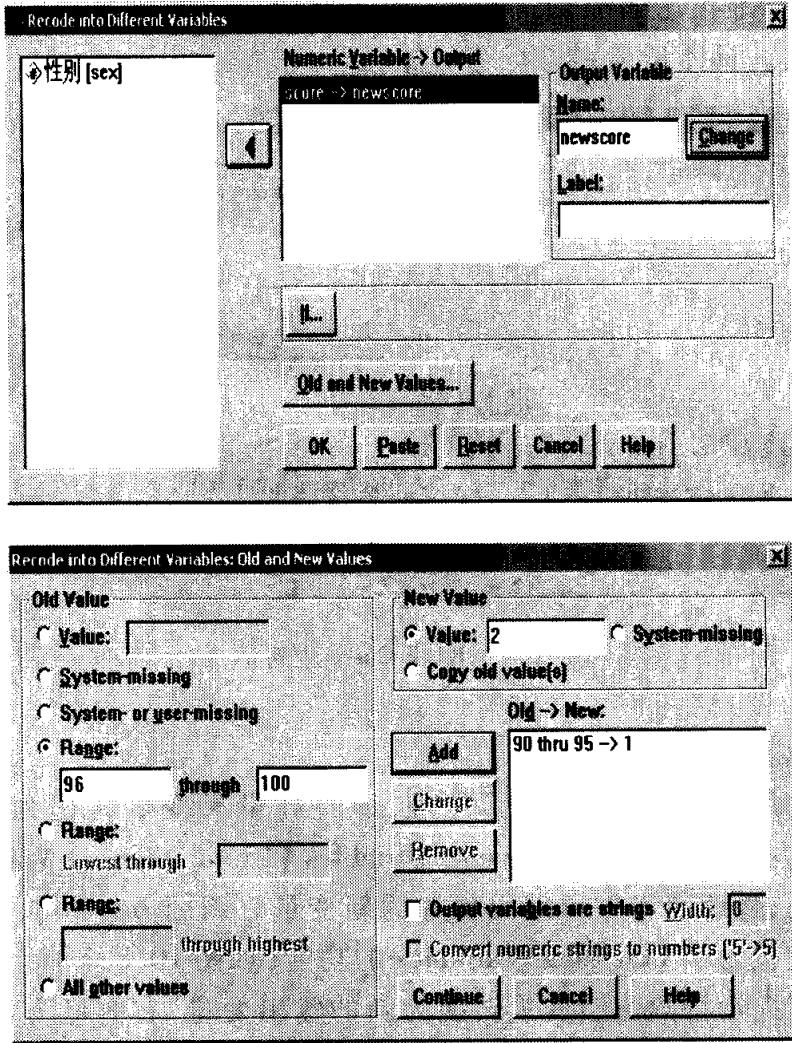


图 10.8 Recode主对话框和重编码规则子对话框

三、数据的其他变换

Transform 菜单中的计算（Compute）过程可以提供更丰富的变量变换功能，

可计算数值或字符型变量的值、同类变量间进行函数运算或利用函数进行变量的计算和转换。打开主对话框（如图 10.9），其界面类似于记录选择，左上角是新的目标变量（target variable）名及属性定义按钮（如果目标变量和已有变量同名，则覆盖原变量），左边为被选变量框，右上角为数值表达式编写框，下方是数字 / 运算符输入板和函数列单。该过程除了可以直接在原变量基础上（就像 If 中设定选择变量的条件一样）进行运算，还可利用执行按钮上方的另一个 If 条件按钮设置更复杂的条件，指定新变量中符合不同条件的记录以不同的取值。该过程功能十分强大，读者不妨多练习体会。例如，使用者可以：

- 直接定义一个新常量的取值（常作为标志变量，如 $s = 100$ ）直接在数值表达式框中输入数值即可。

- 定义新变量为某些已有变量的函数，若为自编表达式，则利用输入板写入表达式，并选择相应的变量进入表达式（如 $total = score / 100$ ）或调用系统函数，选择相应函数并设置好参数（如 $abs = SQRT[score]$ ）、 $best = MAX(test1, test2, test3)$ 等。

- 定义条件运算，即在 If 中设定条件，赋值给符合条件的记录（比如将测验成绩间隔 10 分分组，可以先令新变量 $newscore = 1$ ，然后分别设置 “ If ... $score > 100 \cdots newscore = 2$ ”、“ If ... $score > 110 \cdots newscore = 3$ ”、“ If ... $score > 120 \cdots newscore = 4$ ”完成转换运算。当然，这种情况下同样可以利用系统函数。

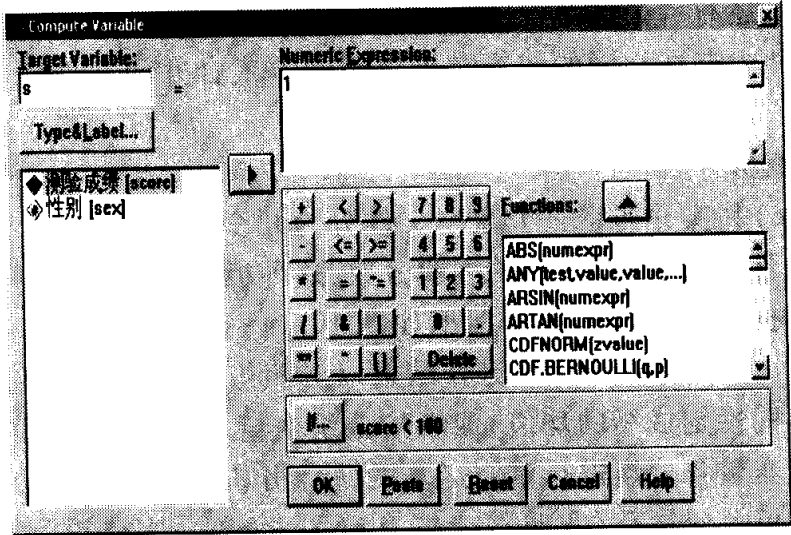


图 10.9 Compute 主对话框

Data 中的 Weight Cases 可用于对变量进行加权，例如根据某个频数变量对

其他变量进行加权。注意：变量中含有零、负数或缺失值的记录将被排除在处理之外，且在后续分析中加权处理一直有效直到关闭加权处理。该过程比较简单，请读者自行练习。

四、数据的合并和拆分

用于合并多个具有相同或相似组成的数据集（如研究的多个样本数据）或将数据按照某个变量的水平进行分组处理。主要牵涉到 Data 菜单下的 Merge files 和 Split File 命令。

（一）数据合并

合并数据总是基于当前数据文件增加变量列或增加记录行。增加记录行时调用 Merge files => Add Cases，选择要合并的文件，系统首先对新旧文件中的变量进行匹配（如图 10.10 上），名称相同的变量自动进入右侧变量框作为新数据中保留变量，名称不相同的变量则用不同的记号标出，如果两文件中使用的某些变量名称不同，则需要手工配对（选中两变量后按 Pair）。注意，为防止数据合并发生错误，最好标明合并后数据源，即勾中最终变量框下方的 Indicate case source as variable，用数字表示各记录来源的文件（默认用 0、1…表示合并操作次数）作为一个数据源变量（默认变量名为 source01）。

增加变量列时调用 Merge files => Add Variables，选择要合并的文件，仍然首先对新旧文件中的变量进行匹配（如图 10.10 下），注意，这时名称相同的变量反而被排除在合并变量之外，因为这时是增加新的变量。主对话框中除包含变量源复选项外，还有一个按关键变量（key variable）记录匹配进行合并的功能按钮，当两个文件中存在相同的变量时可以不同方式合并两文件中的变量。这里不再详述。

（二）数据拆分

（Split File）其实并非将一个数据文件分为几个文件，而是在同一个数据文件中按某个条件分组，如分性别的测验分数，这种拆分在后续运算中一直有效直到取消或更改拆分设置。其主对话框如图 10.11，选择分组变量进入分组变量框（最多 8 个），可以选择将后续运算结果以对照组形式呈现（如在同一个表格中分别呈现不同性别成绩的均值）或者按分组变量水平逐个呈现结果。该过程对于需要按某变量各分组进行重复运算的统计分析特别有用，例如多因素方差分析中进行简单效应检验（参见第十二章有关内容）。需要提醒的是拆分之后别忘了取消。

五、数据的其他整理方式

有时需要对数据集进行转置（如将原来的变量视为记录，对不同变量在同

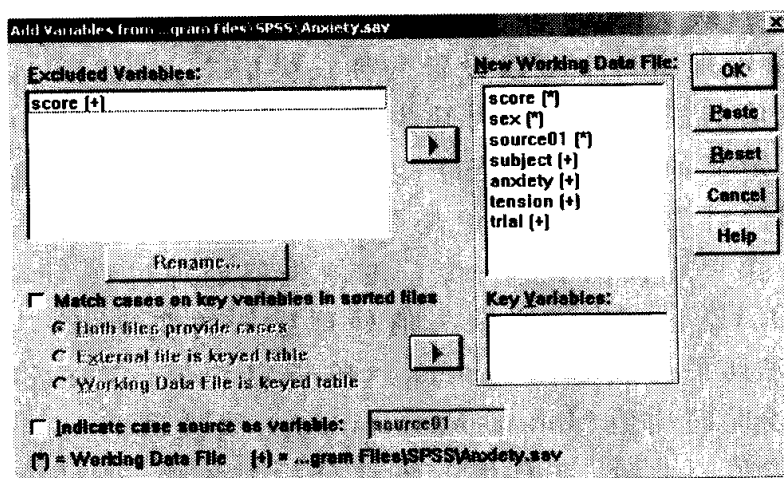
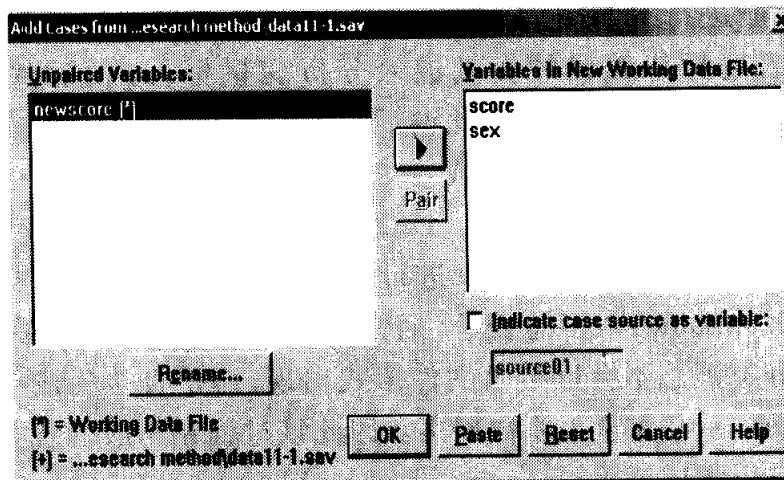


图 10.10 Merge Files 两个过程主对话框

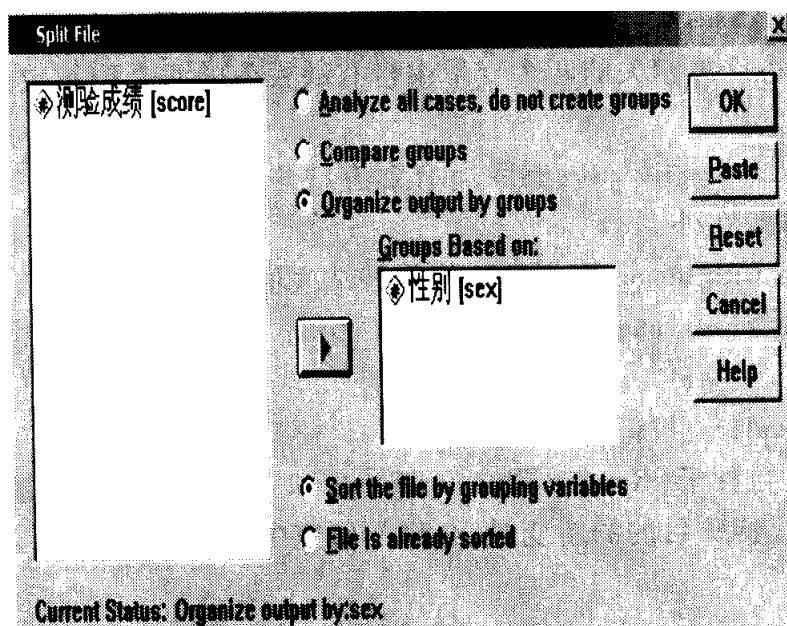


图 10.11 Split File 主对话框

一个案上的特征进行分析)，需要用到 Data 菜单中的 Transpose 过程，其操作很简单，只需选择待转置的变量进入分析变量框，运行后生成新的转置数据文件并自动成为当前文件，此时，新数据集中原变量变为记录，原记录变为变量，通常还会在开头生成一个新的原变量名变量（名称默认为 case_lbl）。

调用 Data => Aggregate 过程可以进行变量的分类汇总，即根据一个或多个分类变量（break variable）对汇总变量（通常为连续变量）进行简单的描述统计。不过其特点是分类汇总不会以结果文件形式出现，而是直接生成一个新数据文件（默认名为 aggr. sav）或覆盖原数据集。因此该过程在一般的统计分析中并不常用。请读者自行尝试练习。

前面谈到了缺失值的可能来源，SPSS 数据集中缺失值默认为圆点“.”，在变量管理窗口中可以设置两种缺失值定义（如图 10.12 上），分别指定变量取值的三个离散值为缺失值，或指定一个范围加上一个离散值为缺失值。定义的缺失值相当于空个案，通常它所在的整个记录都会被排除在后续分析之外，因此不论是否定义缺失值，必须对空数据进行处理。Transform 菜单中提供了 5 种专门的缺失值处理过程“Replace Missing Values”（图 10.12 下），先定义缺失值处理后新的变量，然后选择处理方法：以变量的序列平均数（series mean）、临近点均值（定义临近点范围）、临近点中数、线性内插值（linear interpolation）、线性回归值（linear trend at point）代替缺失值。此外，在具体的统计分析方法中，还可能包括以下几种缺失值处理方式：

- Exclude cases pairwise：成对删除参与分析的变量中含缺失值的记录，例如相关分析时将计算相关的两个变量中含有缺失值的两个配对数据全部排除在计算之外。

- Exclude cases listwise：删除所有含有缺失值的记录，凡整行中含有缺失值，该行全部排除在外。

- Replace with mean：用变量均值代替缺失值。

- Exclude cases analysis by analysis：统计检验时，删除含有缺失值的记录。

- Report values：报值。

此外，随机数种子（Random Number Seed）和生成时间序列变量（Create Time Series）用于生成特殊目的的新变量，心理学研究中应用较少，不再赘述。

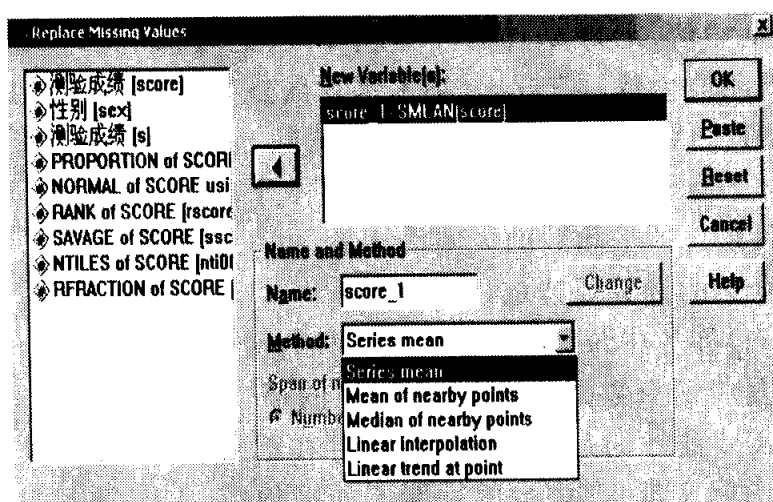
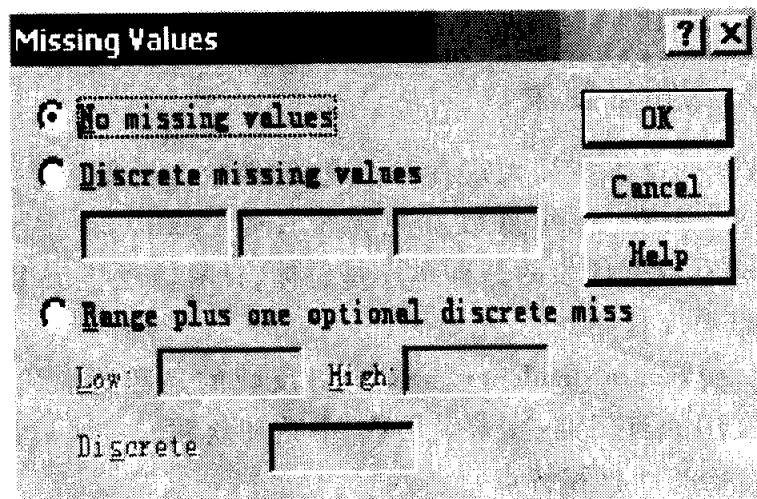


图 10.12 缺失值定义方式和处理方式

第三节 数据的描述统计

描述统计(**descriptive statistics**)主要研究如何描述一组数据的全貌，概括事物的某方面属性，其本质上是为由局部（或样本）信息推断全局（或总体）服务的，因此从统计的目标看，它和前面所讲的数据整理存在着密切联系，这从 SPSS 软件提供的种类丰富的描述统计或数据整理过程也可看出：许多统计过程具体功能可能略有差异，但并没有本质区别甚至往往难以区分。因此，诸如描述统计、概括统计、基本统计分析、单变量统计大体上没有什么区别。本节将结合统计教材的习惯处理和 SPSS 软件的统计功能介绍数据分布特征度量、数据地位度量和相关度量的基本知识，以及如何利用 SPSS 进行数据特征的描述、汇总统计

和有关的统计图表制作。

一、数据分布特征度量

描述一组数据的基本特征大致有三种途径，即数据的集中趋势、离中趋势和数据的分布趋势。通常，这些量数被视为统计量（statistic），同时又都是参数（parameter），关键看数据代表的是样本还是总体，计算方法也大多相同。

（一）集中趋势度量

集中趋势度量（central tendency measures）反映的是数据的代表值或相对集中的中心位置，常用的有三类：算术平均数、中数和众数。

1. 算术平均数的计算和特点

算术平均数（arithmetic mean or mean）亦称均值或均数，记为 M 作为总体参数常指总体的期望值，记为 μ 作为样本统计量，记为 \bar{x} 。其计算就是将所有观测值之和除以观测值个数。实际上，算术平均数是另一类所谓“加权平均数”的特例，即各观测值权值相等的特殊情况。平均数是最常用的集中量，具有稳定性、样本一致性，适合于后续的统计分析。样本平均数（ \bar{x} ）常作为总体平均数（ μ ）的估计量，且是总体平均数的无偏估计量。

2. 中数和众数的计算及特点

中数（median）亦称中位数，本质上是一种百分位数（你能猜出它是多少位数吗？），其含义是将所有数据分成大、小各半的那个数，记为 M_d ，据此中数的计算也比较简单，需要注意的是中数位置附近存在重复数时，其计算又有不同的规定（详细算法请参阅文献）。中数的性能不如平均数，它实际上主要是考虑“中间”位置的观测值，因此无法保证原始数据的信息得到保留，主要用于对数据集中趋势的快速估计。但当所获得的数据资料呈偏态分布时，中数的代表性优于算术平均数。此外，中数可用于等级尺度数据集中趋势度量。

众数（mode）（或范数）指数据中出现次数最多的那个观测值或次数分布中频数最多组的组中值，记为 M_o 。显然，众数的计算最为简便，同时损失的信息也最多。它主要用于对数据集中趋势的快速估计和名义尺度数据集中趋势度量。

（二）离中趋势度量

离中趋势度量（或离散趋势度量）（dispersion measures）反映的是数据相对于集中量数的离散、分散程度，因此，它不仅可以反映数据的离散程度、变异程度，而且可以衡量集中趋势度量对数据的代表性。常用的离中趋势度量有全距、平均差、方差和标准差、差异系数、四分差等。

1. 全距和平均差

全距（range）又称极差，即数据中最大值和最小值的差值，是最简单的离中量。全距通常不能全面衡量各观测值的离中趋势，为此，需要考虑“离（均）差”，

即各观测值与平均数的差 ($x - \bar{x}$)。很容易证明,所有数据的离差和为 0。

平均差 (average deviation) 用平均绝对离差 即 $\sum |x - \bar{x}|/n$ 来度量数据离中趋势。其大小能够反映数据的离散程度。不过其缺点是牵涉到绝对值计算,不利于后继的四则运算。

2. 方差、标准差

方差 (variance), 亦称变异, 是各观测值与平均数离差平方的算术平均数。总体方差记为 σ^2 ; 作为总体方差的无偏估计量, 样本方差 (亦称为均方, mean square) 的计算公式为:

$$\sum (x - \bar{x})^2 / (n - 1) \quad (10-1)$$

方差的平方根称标准差 (standard deviation) 记为 S , 即:

$$S = \sqrt{\frac{\sum (x - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum x^2 - (\sum x)^2 / n}{n - 1}} \quad (10-2)$$

方差或标准差是最常用的离中量, 它不仅反映了数据的离中趋势大小, 可直接比较几组“同质”数据离散程度, 而且反映了数据偏离集中量的大小程度。依照经典的“真分数”(true score) 测量理论, 观测指标上的方差(变异)大致有两个来源: 一是数据的随机变异, 即各种误差; 另一种即是由于观测指标所来源的不同对象被施加了不同处理, 即处理效应 (treatment effect)。将这两类变异进行差异比较, 就可以判定处理是否有效, 这正是方差分析 (analysis of variance, ANOVA) 的基本思路。方差通常也用于异常数据删减, 即读者熟知的“ 3σ 法则”, 其原理是正态分布数据绝大部分 (99.73%) 落在平均数左右三个标准差范围内。方差还可用于求解数据的标准分数 (standard score)。

3. 其他离中量数

当两组数据非“同质”时不能直接使用方差来比较数据的相对离散程度, 而需要利用一个相对离中量, 即差异系数 (coefficient of variation) 来统一单位和观测值的水平, 即数据标准差比平均数的百分比:

$$CV = \frac{S}{\bar{x}} \times 100\% \quad (10-3)$$

除了上述主要的离中量外, 四分差 (semi-interquartile range) 有时也用于度量等级尺度数据或尾端数据不明时的离中趋势。四分差指数据中间两个四分位数 (quartile) 差值的一半 即 $(Q_3 - Q_1)/2$ 。

(三) 分布趋势度量

仅了解数据分布的集中趋势和离中趋势还不够, 有时需要对数据分布的形态 (如是否对称、偏斜程度等) 作度量和判定。分布趋势主要包括偏度 (skewness) 和峰度 (kurtosis) 两方面。

1. 偏度

是对分布偏斜方向和程度的描述，若分数较多地集中在大数端，称正偏态（positive skew）分布，反之称负偏态（negative skew）分布，而正态分布的基本特征是钟型、中心对称，有关其介绍参见第十一章。图 10.13 所示的是三种分布形态图示和利用平均数、中数、众数三种集中量间的关系判别分布形态的经验方法。若 $M > Md > Mo$ 则为负偏态，分布曲线的“小尾巴”朝向数轴负方向，反之为正偏态（中数始终在中间）。正态分布中，三种集中量数相等。

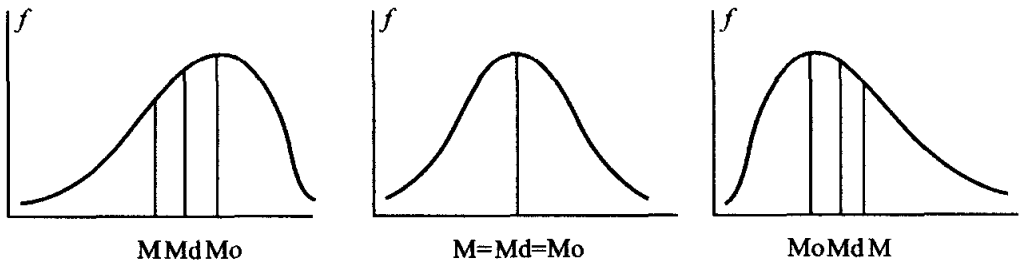


图 10.13 正态分布、偏态分布及其特征示意
(自左至右依次为：负偏态、正态、正偏态)

度量偏度的指标是偏度系数，其公式为（10-4）式。当分布为对称分布时，偏度系数为 0，偏度系数为正数，分布为正偏态，反之则为负偏态，偏度系数绝对值越大，分布越趋于偏斜。

$$sk = \frac{n}{(n-1)(n-2)} \sum \left(\frac{X - \bar{X}}{S} \right)^3 \quad (10-4)$$

$$kt = \left[\frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{X - \bar{X}}{S} \right)^4 \right] - \frac{3(n-1)^2}{(n-2)(n-3)} \quad (10-5)$$

2. 峰度

是对分布高狭程度的描述，与正态分布中心点的峰度相比较，尖而高的称尖峰（leptokurtic）分布，反之称为低峰（platykurtic）分布。度量峰度的指标是峰度系数，其公式为（10-5）式。当峰度系数为正值时，表明数据趋于尖峰分布；为负值则趋于低峰，正态分布的峰度为 0。

二、数据地位度量和相关度量

（一）地位量数

地位量数亦称相对位置量数，用来表示个体分数在团体中所处的相对位置。其特殊之处在于：它们不是用“一个”量反映数据的某种分布特征，而是和个体分数相对应，常用于测验的常模及个体分数评价。

1. 百分位数和百分等级

百分位数 (percentile) 亦称百分点 (percentile point) 常记作 P_m 其中的 m 表示百分比, P_m 的含义就是次数分布中相对于 m 百分点的原始分数, 即有 $m\%$ 的数据低于该数。百分等级 (percentile rank) 常记作 PR , 和百分位数刚好相反, 它表示某原始数据在分布中的百分位, 即次数分布中低于该原始分数的数据百分比, 常用在测量中解释测验分数。未分组情况下百分等级分数的计算公式如 (10-6) 所示 其中 R 为数据按从大到小排列的等级, N 为数据总数。

$$PR = 100 - \frac{100R - 50}{N} \quad (10-6)$$

2. 标准分数

标准分数亦称 Z 分数、基分数, 是以标准差为单位度量个体分数偏离数据平均数大小的量, 反映的是个体分数在团体分数中的相对位置高低。另外, 以 Z 分数为基础转换得到的测量中的其他导出分数也可统称为标准分数, 如 T 分数、标准九分数、标准十分数等。 Z 分数的计算公式为:

$$Z = \frac{X - \bar{X}}{S} \quad (10-7)$$

显然, Z 分数的正负号和相对大小就可以表明个体分数在团体中的相对位置高低。因此, Z 分数常用来进行测验中分数的合成和评价。若原始分数服从正态分布, Z 分数可进一步表示个体分数在团体中的相对“位置”, 这是因为当正态数据标准化后成为标准正态分布, 其 Z 分数 (横坐标值) 和正态曲线下的面积 (概率) 存在着严格的对应关系。这时通过查找标准正态分布表就可以根据 Z 分数换算得原始分数的百分等级。根据这个特性, Z 分数常用在测验或考试中确定分数线或录取率。此外, Z 分数还可以用于表示测验的常模分数和剔除极端数据。其有关的应用参见第十一章正态分布内容。

(二) 相关量数

相关量数通常是对两组数据间关系的度量, 这和前面的量数有所不同。研究变量间的关系是心理科学研究的重要任务, 由于心理和社会现象的极端复杂性, 研究者通常更多的是假设变量间存在着相互关联、相互影响的“不严格确定的依存关系” (黄希庭 2004), 而不是简单地确定因、果来考察完全确定性的关系。例如, 智商与学业成绩, 运动员的赛前焦虑与比赛成绩, 动机强度与工作绩效, 薪酬和员工工作满意度等之间的关系都属于相关关系。

两变量相关在取值上的表现是给定一个变量的值, 另一个变量的值在一定范围内波动, 统计学上常用相关系数 (correlation coefficient) 来度量变量间的相关程度, 用相关分析 (correlation analysis) 来研究变量间的相关关系, 它主要包括相关系数统计量的计算和相关系数的显著性检验。这里主要介绍相关系数及其含义。

1. 相关系数的含义及相关类型

相关系数作为样本统计量常用 r 表示，作为总体参数用 ρ 表示，其取值通常在 $[-1, 1]$ 区间内。根据相关系数的取值可以区分出不同类型的相关关系（见表 10-4）。注意相关系数只是一个顺序尺度数据，只能比较大小。和任何其他统计分析方法一样，不同类型的相关有不同的适用场合，对于变量或数据也有不同的要求，这是统计分析者应当着重掌握的。

表 10-4 相关的各种分类标准及分类

相关系数取值	相关分类	数据量尺	相关分类
$0 < r \leq 1$	正相关	两列连续等距正态数据	积差相关
0	零相关	连续等距非正态或等级	等级相关
$-1 \leq r < 0$	负相关	—列连续正态和—列名义	质量相关
$r = \pm 1$	完全相关	两列名义尺度数据	品质相关
函数形式	相关分类	变量个数	相关分类
直线型	线性相关	两变量间	简单相关
曲线型	非线性相关	—组变量和另一变量间	复相关
		—组变量和另一组变量间	典型相关

2. 皮尔逊相关和斯皮尔曼等级相关

英国统计学家皮尔逊（Carl Pearson, 1857—1936）最早提出一种相关的度量方法，称积差相关（product moment correlation）（习惯也用他的名字称呼）。该法适用于两列等距（及以上）数据，且要求数据总体呈正态分布，是考察线性相关关系的最常用方法。计算公式为：

$$r = \frac{S_{xy}^2}{S_x S_y} = \frac{\sum Z_x Z_y}{n} \quad (10-8)$$

其中 $S_{xy}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{n}$ 即 x 与 y 的协方差（covariance）。

积差相关的条件较为严苛，在其基础上发展的斯皮尔曼等级相关（Spearman's rank correlation）适用场合更宽泛，可用于两列总体不服从正态的连续等距数据或等级数据，用于等距以上数据时要先转换成连续编号的等级数据。无重复等级情况下斯皮尔曼等级相关系数计算公式为：

$$r_R = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)} \quad (10-9)$$

其中 n 为等级数目，也即数据对数， D 为对偶等级之差。

此外，等级相关中还有一类肯德尔等级相关（Kendall's rank-order correlation）包括 W 系数、 U 系数等，多用于特定的量表和题项分析。肯德尔 W 系数亦称和谐系数（coefficient of concordance），专用于求多列等级变量间相关程度，常用来对 N 件事物的 K 次评定（或 K 个评价者，或 K 次评价）分数间的一致性进行评价。与之匹配，肯德尔 U 系数亦称一致性系数（coefficient of consistency），专用于对 N 件事物的对偶比较评定（将 N 件事物两两组合后评价，总评价次数仍为 K 次）间的一致性评价，较少用。注意这两种系数的取值范围和意义不同于一般的相关系数。下列是无重复等级情况下 W 系数的计算公式：

$$W = \frac{\sum_{i=1}^N R_i^2 - \frac{(\sum_{i=1}^N R_i)^2}{N}}{1/12 K^2 (N^3 - N)}; R_i \text{ 第 } i \text{ 件事物的 } k \text{ 个等级之和} \quad (10-10)$$

3. 质量相关

指一列变量为等比或等距的测量型数据，另一列变量是二分类性质变量，欲求两者之间的直线相关，多用于量表和题项评价。按照二分类变量的来源不同又包括点二列相关（point-biserial correlation）和二列相关（biserial correlation）两种。

点二列相关（ r_{pb} ）适用于一列数据为等距或等比（单峰对称分布）变量和一列标准的二分变量（如文盲与非文盲），实际上可以看成是积差相关的一种特例（陈世平，2003）；二列相关（ r_b ）适用于两列等距或等比变量，不过其中一列被人为划分为两个类别，如测验中按照某个分数线划分为及格和不及格。两种相关系数的计算公式分别为：

$$r_{pb} = \frac{\bar{x}_p - \bar{x}_q}{S_t} \cdot \sqrt{pq} = \frac{\bar{x}_p - \bar{x}_t}{S_t} \cdot \sqrt{\frac{p}{q}} \quad (10-11)$$

$$r_b = \frac{\bar{x}_p - \bar{x}_q}{S_t} \cdot \frac{pq}{y} = \frac{\bar{x}_p - \bar{x}_t}{S_t} \cdot \frac{p}{y} \quad (10-12)$$

公式中 p, q 为二分变量两个类别的比例（可任意指定，但相关系数的符号会发生变化，解释相关方向时略有不同）， S_t 指连续变量标准差， $\bar{x}_p, \bar{x}_q, \bar{x}_t$ 指二分变量两个分类对应连续变量的分组平均数和总平均数，（10-12）式中的 y 表示 p 或 q 对应的标准正态曲线纵高（ p/q 较小者作为概率查出的正态概率密度曲线高度值）。

4. 品质相关

品质相关系数度量的是两个性质变量间次数分布的关联程度，专用于计数数据间相关程度度量。这是品质相关和其他相关间的不同之处。但和所有“相关”的含义一样，两个性质变量的次数分布间存在关联意指“一个变量的取值（此处为变量的不同分类）发生变化则另一个变量不同分类的人数分布发生变化”。例如，下方左边表格中的数据是 100 名男、女大学生的英语四级考试的成

绩（及格或不及格）人数分布。表中数据是典型的两个二分变量交叉形成四个单元格，称“四格表”。男性中及格和不及格人数分布与女性中两类人数分布完全一样，说明当“性别”变量取值发生变化时（从 1 到 2），“成绩”变量两个分类上的观察指标（人数）并未变化，即性别和成绩分类两个变量（的次数分布）间不存在关联。再来看右边表格中的数据，调查了 160 对青少年及其家长的吸烟行为，考察父母吸烟是否与子女吸烟存在关联。从四格表人数分布看出，父母吸烟的子女吸烟人数“多于”不吸烟人数，父母不吸烟的子女吸烟人数“少于”不吸烟人数，即“父母是否吸烟”变量的不同分类对应着“子女是否吸烟”变量不同分类上的人数变化，或者说父母是否吸烟的不同分类在子女是否吸烟不同分类上的人数分布不同。这都说明直观上看，“父母是否吸烟”与“子女是否吸烟”之间存在关联。

性别 \ 成绩	1. 及格	2. 不及格	合计
1. 男性	20	30	50
2. 女性	20	30	50
合计	40	60	100

子女 \ 父母	1. 吸烟	2. 不吸烟	合计
1. 吸烟	65	15	80
2. 不吸烟	25	55	80
合计	90	70	160

当然，这种四格表的直观分析还不能真正说明两个变量间是否存在相关，和所有的“相关”一样，这种关联的真实性或显著性还需借助假设检验考察，主要是用到 χ^2 检验。所谓“品质相关系数”是度量这些样本计数数据间相关或关联程度大小的一类统计量，绝大多数是由假设检验中的卡方统计量变换而来。例如用于四格表数据的 ϕ 系数（Phi coefficient）和用于 $R \times C$ 表（一个变量为 R 个类别另一个变量为 C 个类别）的列联系数（contingency coefficient）等。下列是这两种相关系数的计算公式，其中 N 为总次数。

$$r_{\phi} = \frac{(ad - bc)}{\sqrt{(a + b)(c + d)(a + c)(b + d)}} = \pm \sqrt{\frac{\chi^2}{N}} \quad (10 - 13)$$

$$C = \sqrt{\frac{\chi^2}{N + \chi^2}} \quad (10 - 14)$$

三、SPSS 进行描述统计的若干途径

SPSS 中的许多过程都包含集中量、离中量及相关量三种统计量数的计算功

能，这些描述统计通常可以由 Analyze 菜单下的 Reports Descriptive Statistics 的某些过程以及 Means、Correlate 等过程来专门进行。下面分别对其构成和操作做简单介绍。

(一) Reports

报表子菜单包括 OLAP Cubes (在线数据处理)、Case Summaries (记录汇总)、Report Summaries in Rows (分行汇总报告) 和 Report Summaries in Columns (分列汇总报告) 四个过程，主要功能是对等距尺度的观测指标进行分组集中量和离中量描述，还可以报告形式显示或打印出符合研究者特定要求的记录。该子菜单侧重于定制报表，对于完整研究的统计分析则无太大价值。

OLAP 主界面如图 10.14 左，分别选择自变量和因变量进入分组变量和汇总报告变量框，其两个功能按钮分别可以选择报告描述统计量类型（和下一章要讲的 Means 过程类似）以及标题和注解定制。其中包含的描述统计量有：总和、记录数、均值、标准差、占总和的百分比 (percent of total sum)、占观察量总数的百分比 (percent of total N)、中数、分组中位数 (grouping median)、均值标准误 (std. error of mean)、最小值、最大值、全距、首值、末值、方差、偏度 (skewness)、偏度标准误 (std. error of skewness)、峰度 (kurtosis)、峰度标准误 (std. error of kurtosis)、调和平均数、几何平均数等 20 余种。OLAP 的主要特点是提供了很好的交互方式，用户可按自己的喜好定制报表的格式和内容。例如，对 y_1 、 y_2 、 y_3 、 y_4 四个因变量在 f_1 、 f_2 两个自变量上的分组描述统计，给出的基本结果是不考虑 f_1 、 f_2 的分组的统计量（默认的 total）双击该报表， f_1 、 f_2 两分组变量右侧会分别出现一个下拉列表，可以分别选择两变量的各个分组，则下面表格中显示的是该组相应统计量。分别选择该列表，用户就可以得到两变量任意水平组合下的描述统计量，比如可以要求显示不考虑 f_1 的分组时， f_2 第二水平（编码为 1）下的描述统计，或者显示 $f_1(0) \times f_2(1)$ 处理下的描述统计，等等。图 10.14 右说明了该交互过程。

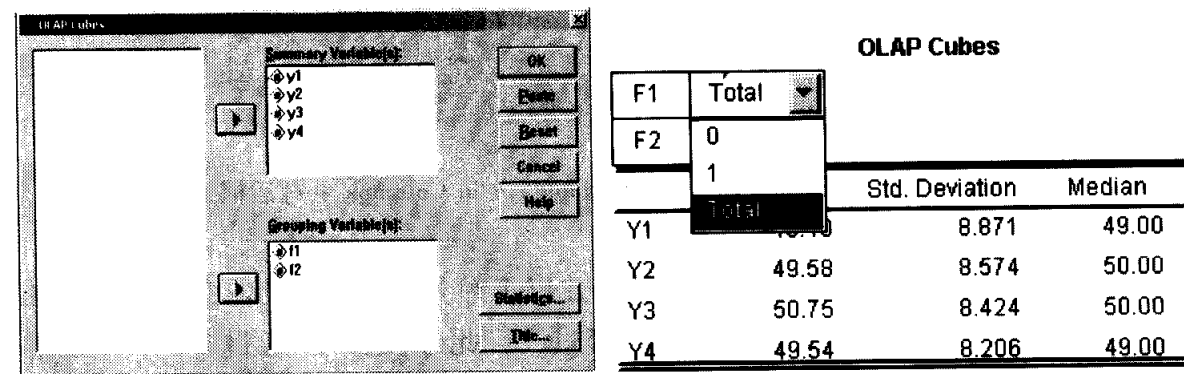


图 10.14 OLAP Cubes 主对话框及结果交互显示方式示意

Case Summaries 用于计算指定变量的分组统计量，若有多个分组变量则报告水平交叉组合后的单元格描述统计量。其主界面、变量选择和两个功能按钮类似于 OLAP;勾中 Display Cases 复选框，后可再选择分析指定数量记录及缺失值、记录号处理方式，同时在结果中给出指定数量记录的列单和描述统计。除了对记录的“穷举”显示外，其功能可以由其他统计过程替代（如 Means 过程），不再详述。

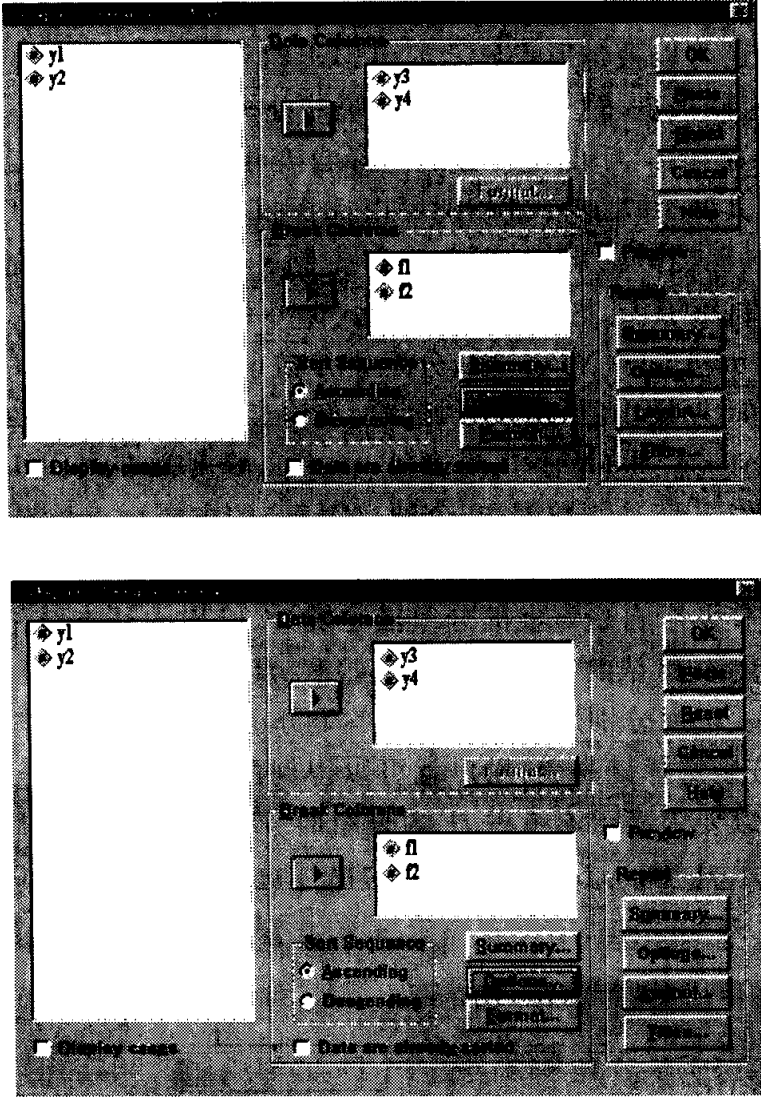


图 10.15 Summary in Rows 和 Summary in Columns 主对话框

分行汇总报告也可以罗列原始数据，并以汇总量为行，以报告变量为列给出报表，其报表格式设置更精细（如图 10.15 上）。此外，还可以通过 Summary、Options、Layout、Titles、Format 等功能按钮设定报告何种汇总量（复选，分行报告）、报告页面设置和布局、标题等报表格式和内容。分列汇总报告的不同在于，它不

能罗列数据，而是以分组变量各分组为行，以报告变量为列给出汇总报表（如图 10.15 下），也可以通过 Summary（单选，分列给出各报告变量汇总量）、Options、Layout、Titles、Format 等按钮设置报表格式和内容。若有多个分组变量，则两种方式都能够提供各分组（交叉处理）的汇总量。这些设置比较繁琐，读者可参考文献练习体会。以下是两种方式的报表示例：

表 10-5 分行汇总报告（左）和分列汇总报告（右）结果示例

Page 1				Page 1			
F1	F2	Y3	Y4			Y3	Y4
				F1	F2	Mean	Sum
0	0						
	Mean	51	50	0	0	51	1259
	1				1	53	1348
	Mean	53	54				
1	0			Subtotal	f1	52	2607
	Mean	53	48				
	1			1	0	53	1212
	Mean	47	45		1	47	1135
Grand Total							
	Mean	51	50	Subtotal	f1	50	2347

（二）次数分布表和次数分布图

数据整理的直接目的就是编制数据的次数分布（frequency distribution），即实验或调查中观测的样本究竟取哪些值？这些值各自出现了多少次？它比前面所讲的分布趋势度量更加具体、形象，便于对数据的分布有直观的了解。不严格地，根据原始数据进行的任何形式的数据分组或分类都可以叫做次数分布，统计教材上所谓的统计分组（grouping）、统计汇总（summary）本质上就是编制次数分布，其结果的形式通常可以用表格或图形来表达，即次数分布表和次数分布图。利用 SPSS 软件可以很方便地作出变量的次数分布表和次数分布图。

1. 离散变量的次数分布

离散变量通常可以是类别变量或名义变量，也可能是等级变量，其取值通常是离散的，其分组的标准（标志）很明确，一个取值就是一个分组，因此其次数分布就是各个类别或等级（位次）观测值发生的次数。举例说明 Frequencies 过程的用法：

例 10-2 假设调查了 26 名被试的职业类型，试作出其次数分布表和次数

作者注：某些比率变量取值也有限，这时要注意其观测值的跨度范围，如社会统计中要了解居民家庭或户规模（包含人口数），其取值大多集中在 1、2、3、4、5……，数字越大次数越少。

分布图。

解答：第一个变量 cat 即职业类型，读者可以从变量属性中看出 26 个观测值的含义：1—工人；2—农民；3—商人；4—知识分子；5—其他。调用 Analyze => Descriptive Statistics => Frequencies 其对话框如图 10.16 上所示，除了备选变量框和分析变量框外，左下方有一个“显示次数分布表”复选项；下方并排有三个功能按钮，分别为统计量（如图 10.16 下）、次数分布图（Charts）和格式设置选项。格式设置选项可以设定各种次数分布表显示方式，如按照各观测值次数多少排列，多变量同时统计时表格输出方式（选择 compare variables 多变量并列，选择 organize output by variables 各变量单列）以及压缩（suppress table 输出表组数。主要看一下统计量对话框，它包括四块内容，分别是：

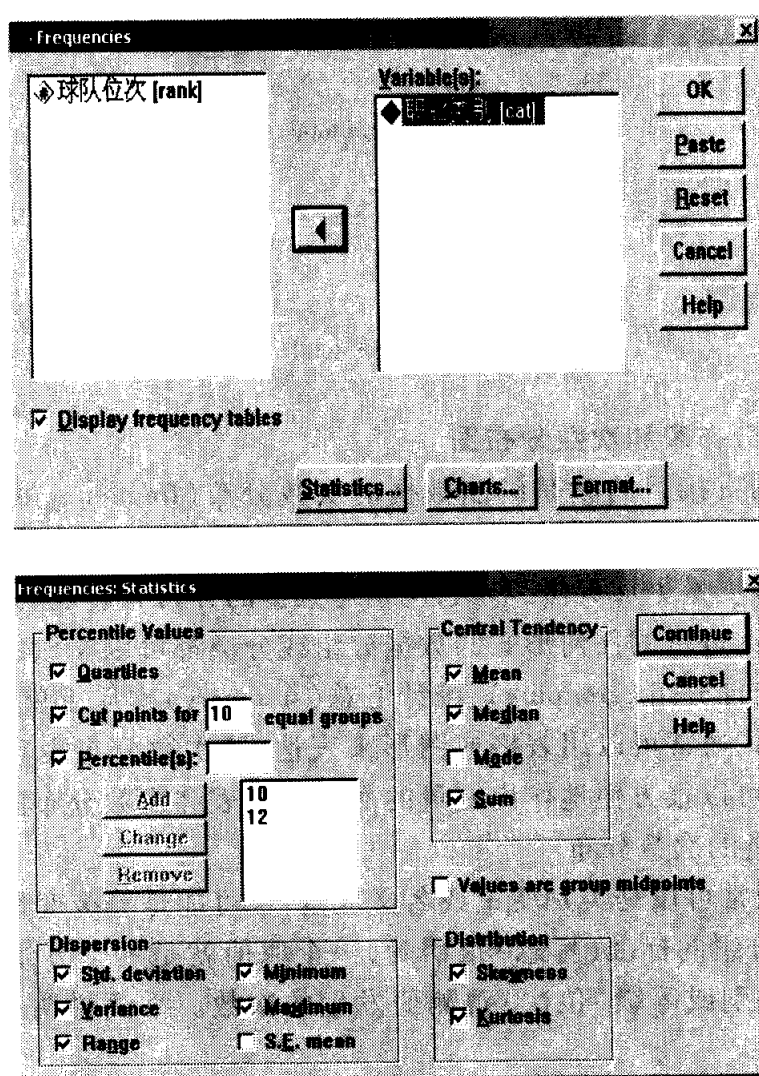


图 10.16 Frequencies 过程主对话框和 Statistics 按钮界面

- 百分位数 (percentile values)，包括三个复选框：四分位数，显示 25%、

50%、75%的百分位数；平均分组，按百分比将数据平均分为所设定（2 ~ 100）的相等等份，并相应输出各百分位数。如输入 10 则输出 10%、20%、30%等百分位。自定义百分位数 输入 0 ~ 100 整数后按 Add 增加输出。

- 集中趋势量，包括均值、中数、众数和总和。
- 离中趋势量，包括最值、全距、方差和标准差、均值标准误（S. E. mean）等。
- 分布趋势量，包括峰度系数和偏度系数。

例中先勾中显示次数分布表并按默认格式输出职业类别变量的次数分布表（表 10 - 6），可看出 SPSS 自动按照累加次数分布表的形式给出次数分布，第一列是有效的变量类别值，第二至第五列依次是次数、百分比、有效百分比和累积百分比。当然，由于是离散变量，有关分布的其他描述统计并没有多少意义。

表 10 - 6 26名被调查者的职业类型

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	工人	8	30.8	30.8	30.8
	农民	3	11.5	11.5	42.3
	商人	6	23.1	23.1	65.4
	知识分子	6	23.1	23.1	88.5
	其他	3	11.5	11.5	100.0
	Total	26	100.0	100.0	

SPSS 中的次数分布图制作通常有两种方式，除了在 Frequencies 过程中直接制作外，还可单独使用主菜单中 Graphs 的相应统计图功能完成，其中的设置和选择更为复杂，不赘述。打开 Charts 按钮，可以单选制作的次数分布图有条图（bar charts）、圆形图或饼图（pie charts）和直方图（histograms）。三者各自适用的数据类型是：前两种主要用于本例中的离散变量，而直方图通常用于连续变量。图 10.17 左即本例中做出的职业类别人数分布条图，条图的坐标系中，横坐标没有实际意义，只表示变量各个类别，纵坐标表示次数或相对次数（作图时选择 percentages），从各直条的高度可以直观看出各类别人数分布多少，如这里工人最多，农民和其他职业最少。如改用饼图，则不仅可以直观看出各类别次数，还可以看出各类别次数占全体的相对比例，图 10.17 右即编辑后的饼图，最多的工人类别被突出（explode）显示。

2. 连续变量的次数分布

连续变量通常是等距或等比变量，其取值理论上可以取一段范围中的任意

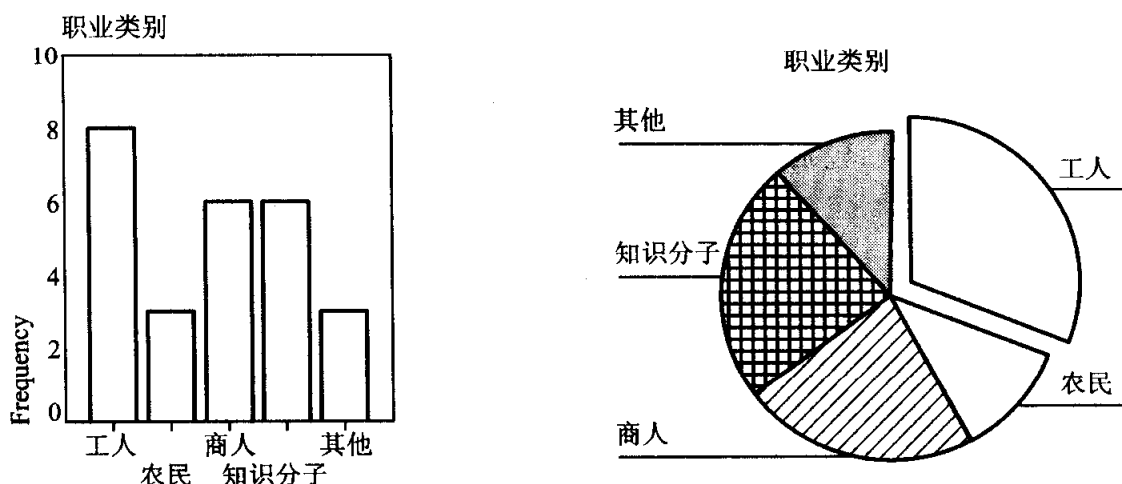


图 10.17 职业类别变量次数分布条图和饼图示意

值，其次数分布的编制通常采用分段分组的方法，关键是选取适当的分组。关于手工制作连续变量次数分布表的方法和某些具体规定（如关于分组界限、分组代表值等）不再赘述。需要注意的是 SPSS 不能自动完成分组后的次数分布表，这里所列出的次数分布实际上是一种“穷举式”次数分布，即将所有不同数值的次数统计并显示出来。不过借助描述统计特别是次数分布图了解数据的分布规律后可以使用其他的 SPSS 过程（如数据转换过程）重新建立分组后次数分布表（见陈世平, 2003）。下面举例说明连续变量次数分布表的制作及特点。

例 10-3 假设测查了 100 名企业员工的焦虑程度（总分 20 分）作其次数分布。

对于连续变量，描述统计量比较重要，可以不选择输出次数分布表。在描述统计子对话框中选择平均分成 10 组 统计方差、均值、中数、峰度、偏度、最值等。在绘图子对话框中选择作直方图并添加正态曲线（with normal curve）输出结果包括描述统计和次数分布图见图 10.18 左、右。左图中罗列出各描述统计量以及按 10 个平均分组给出的 9 个百分位数；右图中自动将数据分成了 9 组（双击编辑该图横坐标，可改变组距大小、呈现方式等），横坐标显示的是各组组中值。从直方图中次数分布和正态曲线的拟合程度以及描述统计中的峰度和偏度数值可以看出，该数据分布大致呈正态分布。

此外，SPSS 中进行单变量描述统计还可以通过 Descriptive Statistics 菜单下的 Descriptives 过程来实现，其对话框构成类似于 Frequencies 过程，只有一个 Options 按钮，其包含的描述统计功能也远不如 Reports 或 Frequencies 过程丰富 不过其最大的功用是有一个 Save standardized values as variable 复选框可以直接将变量 Z 分数求出并保存在数据集中供后续分析用。最后，若要计算数据的百分等级分，可以直接利用其公式（如公式 10-6）调用 Compute 过程直接计算。

Statistics		
焦虑得分		
N	Valid	100
	Missing	0
Mean		10.7304
Median		10.7000
Variance		.62081
Skewness		.066
Std. Error of Skewness		.241
Kurtosis		.036
Std. Error of Kurtosis		.478
Minimum		8.86
Maximum		12.86
Percentiles	10	9.6160
	20	10.0800
	30	10.4000
	40	10.5400
	50	10.7000
	60	10.8600
	70	11.1080
	80	11.3000
	90	11.8860

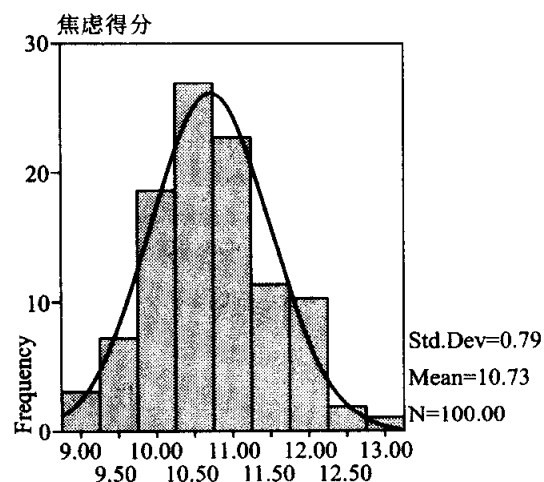


图 10.18 焦虑得分次数分布统计表和直方图

(三) 相关度量计算过程

作为描述统计量的相关系数计算，在 SPSS 中主要通过 Analyze 菜单下的 Correlate => Bivariate 和 Descriptive Statistics => Crosstabs 来实现。不过这些 SPSS 命令包含整个相关分析过程，在计算相关系数的同时也会给出假设检验结果，所以这里只对各过程可以统计的相关系数类型作简单介绍，其余内容留待第十二章介绍。

1. Bivariate

该过程主要计算多个变量间“两两”直线型相关同时进行相关系数的显著性检验，结果显示在一表格中。该过程的主对话框（如图 10.19）除了两个变量框外，只有 Options 功能按钮，可以选择输出每个变量的均值、标准差及交叉积和协方差以及缺失值处理方式。后两种矩阵主要用于其他高级统计过程（如探索性因素分析、验证性因素分析等）的分析。注意，Bivariate 过程提供的相关系数计算只有三种：Pearson、Spearman、Kendall's tau-b，前两种已经述及，主要用于计算连续等距正态数据间相关以及等级数据或非正态连续等距数据；后一种是 Kendall 等级相关系数之一，和 Spearman 相关系数类似。在显著性检验中可以选择单侧或双侧检验和显著性标定方式，在结果中用上标星号（*）表示 0.05 水平上显著，用双星号表示 0.01 水平上显著。

2. Crosstabs

该过程主要是用来进行列联表或交互表分析（cross tabulation）所谓列联表指由两个或多个分类变量（可以是名义变量或等级变量）组成的交叉次数分布

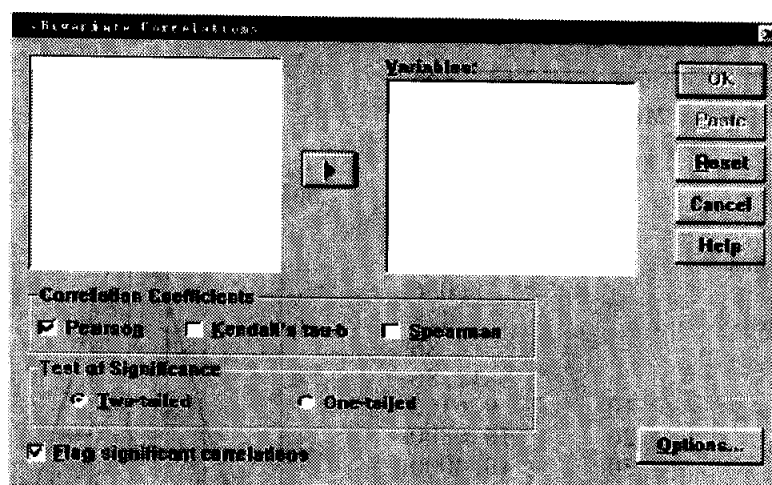


图 10.19 Bivariate 主对话框

表，亦称交叉表。其单元格相当于变量交叉后形成的“处理”组合，例如前面品质相关中提到的四格表就是最简单的列联表。二维或双向（two-way）列联表由一个行变量和一个列变量构成；如再加一个控制变量则构成一个三维列联表；多个行、列、控制变量构成多维（multiway）列联表。列联表分析提供了对多维列联表的描述统计和 χ^2 检验，可进行相应的期望值估计。这里只介绍它提供的主要相关量数及其适用条件， χ^2 检验原理和应用留待十一章、十二章介绍。

Crosstabs 过程提供的相关量数主要是在 Statistics 功能按钮中，如图 10.20 所示，除了卡方统计量检验和相关系数检验外，主要提供了适用于名义变量、等级

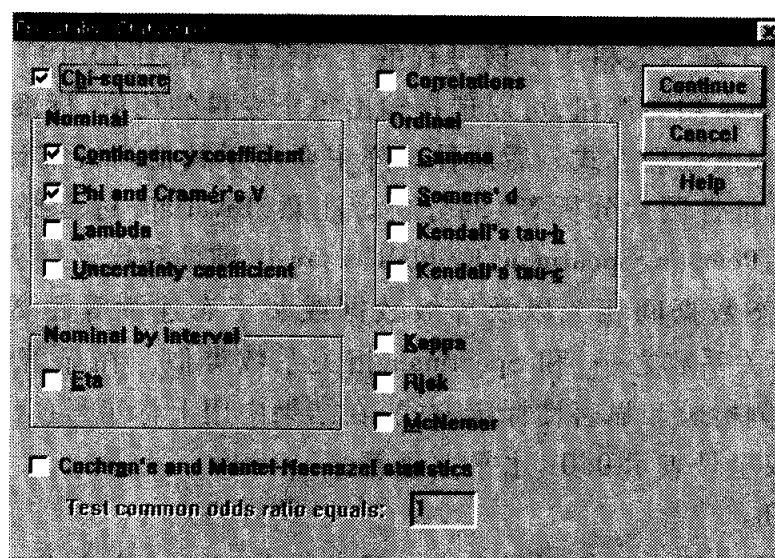


图 10.20 Crosstabs 过程 Statistics 子对话框

变量、名义加等距变量等类别的一些相关量数，包括一些广义的相关量数（通常不称为“相关系数”，但仍反映变量间的关联关系）。其含义和适用条件见表 10-7。

表 10-7 各类关联量数含义及适用条件说明

分类	名称	说明
名义	Contingency coefficient(C, 列联系数)	基于 χ^2 统计量, 取值在 0 ~ 1 间, 不可能达到 1, 越大则表示关联性越强
类别	Phi & Cramér's V(ϕ/V , 校正列联系数)	都是对列联系数的校正, ϕ 在四格表中取值在 0 ~ 1 间(包括 1), V 对 ϕ 进行调整, 取值在 0 ~ 1 间, 越大表示关联性越强
	Lambda(λ)	当用自变量预测因变量时误差降低的比率, 取值[0,1], 越大表示自变量越能很好预测因变量, 取值为 0 表示两变量独立
	Uncertainty coefficient(不确定系数)	类似于 Lambda, 取值接近 1 表明第一个变量反映第二个变量信息越多
等级	Gamma	反映两变量间的等级对称关联性, 取值[-1, 1], 为 1 时表示所有数据集中在表格左上角到右下角的对角线上, 0 表示两变量相互独立。该类最基本的指标
	Somers' d	tau-b 的非对称扩展, 表明同序对超过异序对的比例
	Kendall's tau-b(τ_b)	等级尺度的非参数相关检验法, 是对 tau-a 的调整(tau-a 表明同序对超过异序对的比例); 取值[-1,1], 适用于行、列分类相同的正方形表
	Kendall's tau-c(τ_c)	类似于 tau-b 系数, 适用于非正方形表
名义 × 等级	Eta(η)	自变量和因变量分别为名义、等距型, 反映自变量组间变异占因变量总变异的比例; 常用平方表示关联性, ANOVA 中常用, 类似于 R^2
	Kappa	评价一致性度量(measurement of agreement), 一般要求行、列分类相同, 取值[-1,1], 其解释同积差相关系数
	Risk(相对危险度)	只适用于四格表数据, 给出相对危险度(RR)和优势比(Odds ratio, OR)作为估计, 置信区间 ≤ 1 表示两变量无关联, 否则表示有关联
	McNemar	两相关样本二分类变量间关联程度的非参数检验统计量
	Cochran's & Mantel-Haenszel statistics	包含控制变量情况下四格表(特指 SPSS 中)数据的分层关联性度量, 给出公共优势比值和检验

补充讨论 10-1 探索性数据分析和 Explore 过程简介

探索性数据分析 (exploratory data analysis, EDA) 是上世纪 60 年代普林斯顿大学统计学家图基 (John W Tukey) 发展的一个新的统计学分支, 1962 年他发表了《The Future of Data Analysis》成为该领域的奠基之作, 1970 年他出版该领域第一本专著《Exploratory Data Analysis》。80 年代陈忠琰等将 John Wiley & Sons 公司 1983 年出版的《理解稳健性和探索性数据分析》翻译介绍到国内。所谓“探索性”主要是和假设检验等传统统计学的“验证性”数据分析 (CDA) 方法相对, 和通常讲的描述性统计一样, 探索性数据分析也是对数据的一种预先考察, 是建立正式的统计分析模型的基础, 因此严格来讲它是完整、科学的统计分析过程的环节之一。其思想古来有之, 例如皮尔逊 (Karl Pearson, 1857—1936) 对于正态分布理论的怀疑和对描述统计学的推崇。

传统统计方法通常是基于一定的数据分布假设 (如正态分布) 采用相应的数学模型进行统计分析, 但客观上多数数据并不满足这种假设 (实际应用者往往省略了这一假设检验过程), 从而影响了统计方法的效度; 探索性数据分析则从客观原始数据出发去探索数据的内在规律。其次, 传统的统计分析以概率论为理论基础, 对估计、检验和预测等方法都会给出基于一定置信度 (精确度) 的度量方法和度量值, 有时显得过于刻板 (如假设检验中小概率事件原理的应用); 而探索性数据分析则相对放松了对概率论理论和精确度的刻意追求, 更加强调数据特征值的稳健性 (robustness)。第三, 传统统计分析方法运用的数学工具越来越艰深, 阻挡了相当一部分实际应用者; 而探索性数据分析运用简单直观的茎叶图 (stem and leaf plot)、箱图 (boxplot)、残差图 (residual plot) 和字母值 (letter values)、数据变换、中位数平滑 (median polish) 等方法, 极大地扩大了统计分析的用户群, 可以说使传统统计学的面貌焕然一新。 (袁卫, 1999)

探索性数据分析的主要思想和内容包括:

1. 耐抗性 (resistance) : 即对数据中的奇异值 (outliers) 或极端值 (extreme values) 不敏感, 统计分析或概括的结果对于数据小部分的改变保持稳定。例如, 在集中趋势度量中, 中数是耐抗性较好的, 而算术平均数的灵敏性就比较高。另一个类似的概念是稳健性, 指对于概率模型假设的违背不敏感, 二者强调的内容不同。相比众数而言, 中数不算是耐抗性高的, 而算术平均数则既缺乏耐抗性又缺乏稳健性。

2. 残差 (residuals) : 即从原始数据中减去概括统计量或所拟合模型的期望值后剩余的部分:

$$\text{残差} = \text{数据} - \text{拟合}$$

残差可能意味着未加考虑的系统变异，因此残差分析在许多统计分析方法中是至关重要的环节，例如回归分析中的残差诊断。探索性数据分析认为应当“利用耐抗分析将数据中的主导行为和反常行为分离开来”，当数据的主体“遵循某个一致的模式，这个模式决定一个耐抗拟合”（陈忠琰等，1998）。

3. 重新表达 (re-expression)：相当于数据转换，探索性数据分析要求尽早考虑数据的原始测量尺度是否合适和什么样的尺度能够帮助简化数据分析。若原始尺度不合适，则将数据转换后或许有助于促进对称性、变异的恒定性、关系直线性或效应可加性。例如数据的幂变化等。

4. 图示 (graphical displays)。图示的最大特点是直观性，也是探索分析的主要特色。探索性数据分析中主要用到的统计图示有茎叶图、直方图、箱图、正态概率 P-P 图和正态概率单位 Q-Q 图等。

总之，探索性数据分析强调“灵活探求线索和证据”；而验证性数据分析则着重“评估现有证据”（陈忠琰等，1998）。两种方法在统计分析过程中相辅相成，交替使用才能得到满意的结果。

Descriptive Statistics 菜单下的 Explore 过程是专用于进行探索性数据分析的，该过程可用来进行奇异值或极端值的筛查，考察样本分布特征有关分布的前提假设条件进行检验，及探索组间差异的特征。探索的结果给出分组和不分组条件下的描述统计量与诊断性统计图，包括茎叶图、直方图和箱图等。

以具体例子简要说明 SPSS 探索数据分析的流程和结果，数据见 SPSS 自带的 Employee data. Sav，这是一个员工薪水及其影响因素的研究原始数据。打开 Explore 主对话框，操作流程如下：

1. 分别选择变量 salary、gender、id 进入因变量、因素和标识变量框；
2. Display 单选框选择同时输出统计表和统计图；
3. 打开统计量按钮，选择描述、**M** 估计量 (**M** estimator)、奇异值三个复选项；
4. 打开图形按钮，选择箱图框中的 Factor levels together 要求自变量各水平箱图显示在一起（每个因变量独立生成一个箱图）；选中描述框中的茎叶图；选中 Normality plots with tests，要求输出正态概率单位图 (normal Q-Q plot) 并进行正态分布检验 (test of normality) 选中 Spread vs. level with Levene Test 下的幂值估计 (Power estimation)，要求输出散布-水平图 (Spread vs. level plot) 并进行转换幂值估计。

输出结果包括如下几个部分：汇总报告、按性别分组后的描述统计表、M 估计量的四种加权计算结果、极端值分布、正态分布检验、方差齐性检验 (test of homogeneity of variance)、茎叶图、正态概率单位 Q-Q 图、箱图及散布-水

平图。

- 描述统计：主要给出均值，均值的 95% 置信区间 (confidence interval) 及上、下限 5% 调整均值 (trimmed mean)、方差和标准差、最值、全距、四分位差 (四分差的两倍)、偏度和峰度等，其中偏度和峰度值分别为：男性 1.863 和 4.641，女性 1.639 和 2.78，表明数据非正态。

- M 估计量：即均值的最大似然比的稳健估计量，分别用 Huber、Tukey、Hampel、Andrew 四种不同的加权方法得出估计值，与均值进行比较发现都比均值小 (因为数据非正态，均值受到较大影响)。

- 极端值分布：分别列出男、女员工薪水的 5 个最大值和最小值。

- 正态分布检验：给出 Kolmogorov - Smirnov 检验统计量及检验结果，结果中显著性水平 <0.05 ，表明数据分布和正态分布差异极其显著，数据不服从正态分布。

- 方差齐性检验：给出 Levene 检验统计量和检验结果，结果中基于四种方法得到的检验显著性水平 <0.05 ，拒绝方差相等假设。

- 茎叶图：如图 10.21 所示为女性收入茎叶图，茎叶图的主体 (图中) 包括频数、茎 (整数值)、叶 (小数值) 三部分，可用来近似计算各观测值；② 显示共有 11 个极值 (≥ 40800)；表示茎宽为 10 000 (相当于单位)；表示每个叶代表 1 个观测值。如何读取茎叶图中的数据呢？例如 这一行中频数为 5 表示共有 5 个观测值，茎宽都为 3、而叶分别是 66 777，则该 5 个观测值可近似计算为： $(\text{茎} + \text{叶} / 10) \times \text{茎宽}$ ，分别近似为 36 000、36 000、37 000、37 000、37 000。

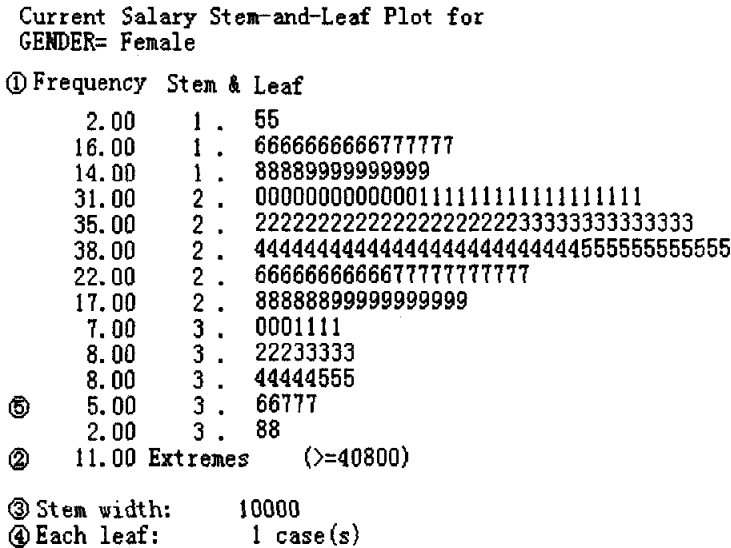


图 10.21 女性收入茎叶图

- 正态概率单位 Q - Q 图 :如图 10.22 和图 10.23 所示为女性正态 Q - Q

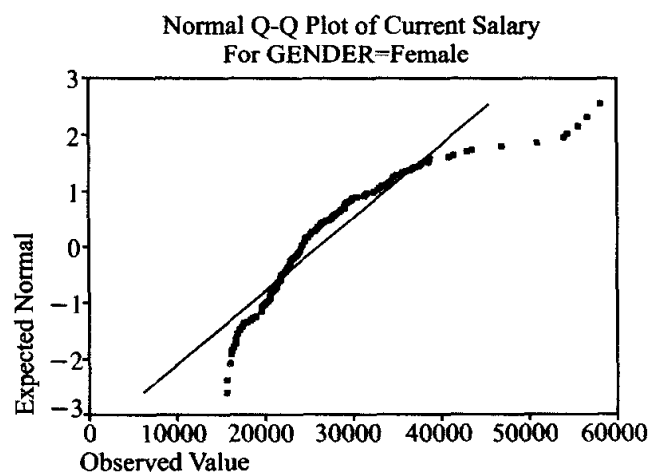


图 10.22 女性收入正态 Q - Q 图

图和离散正态 Q - Q 图(detrended normal Q - Q plot)。图 10.22 中的斜线为正态分布的标准线，正态概率图对此线拟合越好表示数据越接近正态分布，图中大量数据点偏离此线，同样图 10.23 中的大量点偏离了过零点位置的水平线，都表明数据不服从正态分布。

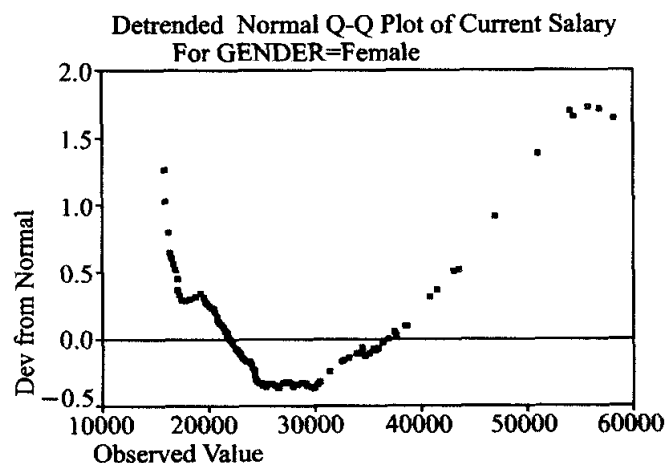


图 10.23 女性收入离散正态 Q - Q 图

- 箱图：如图 10.24 所示，箱图提供了各分组间在因变量上差异的直观显示，同时也提供了数据分布的证据。图中的灰色方箱是箱图的主体，其上、下界为 25% 和 75% 四分位数，其跨度称箱长，变量的 50% 观测值落在这一区域中。中间的粗横线为中数（这也可以看出数据的大致分布形态），方箱上下的纵向直线为触须线（whiskers），它和两端的本体内最大值和最小值共同构成一个“工”字形。图中还可以看到带有数字的 ○ 号和 * 号，分别代表奇异值（方箱外 1.5

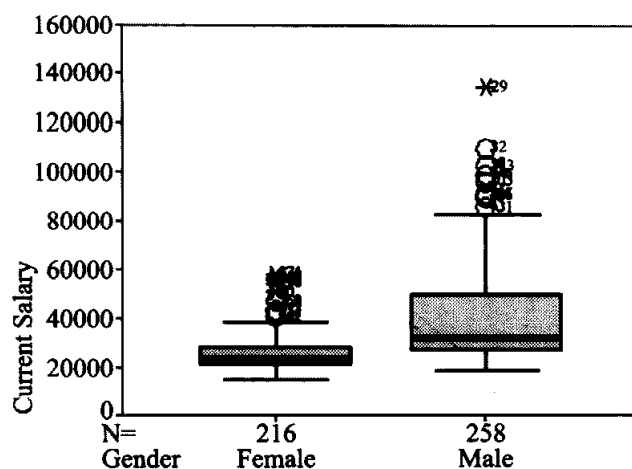
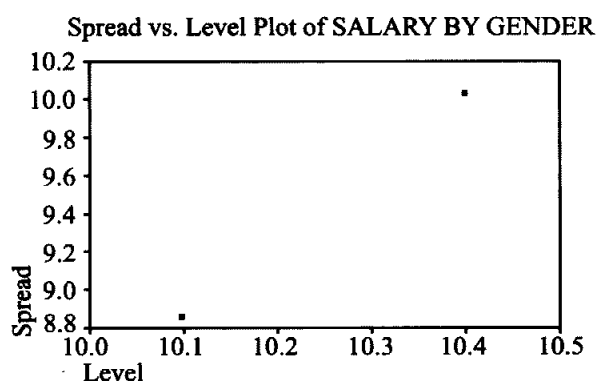


图 10.24 男、女性员工收入箱图

到 3 个箱长范围内) 和极端值(方箱外 3 个箱长范围外), 如第 29 号观测值为极端值 第 32 号观测值为奇异值等。箱图的另一个重要功能是显示出组间差异, 途中明显可以看出, 男性的收入水平普遍比女性高。

● 散布—水平图: 提供了回归线的斜率和方差齐性的 Levene 检验结果。图 10.25 中只有两点表示男、女两组数据的散布—水平分布。下方结果显示要使两分组中数据的方差相等, 建议进行幂转换的幂值为 -2.893 , 两点间回归线的斜率为 3.893 。



★ Plot of LN of Spread vs LN of Level

Slope = 3.893 Power for transformation = -2.893

图 10.25 各性别员工收入散布—水平图

► 本章提要

数据的量化是将原始资料通过类别化转化为计算机或其他机器可识别

的机读格式，即编码的过程，其中最关键的环节是类别化和制定编码框架。

2. SPSS 的数据录入有多种方法，包括直接输入、直接打开原始数据文件、文本导入向导读入、数据库转录等。

3. 数据的初步整理包括对数据进行排序、变换、重编码、重分类、分类汇总、加权，以及数据文件的合并、分割，对缺失值的处理等，主要牵涉到 SPSS 中的数据编辑、数据变换菜单。

4. 描述统计用于描述一组数据分布的全貌，包括数据的分布特征度量、地位度量和相关度量。其牵涉到的 SPSS 过程包括分析菜单下的报表、描述统计、相关等过程。

5. SPSS 描述统计菜单下探索过程用于进行探索性数据分析，即从客观原始数据出发，通过分析数据的耐抗性、残差、重新表达及图示手段来探索数据的内在规律，挖掘数据中的有用信息。

►本章关键术语

编码(coding)	积差相关 (product moment correlation)
编码框架 (coding frame)	斯皮尔曼等级相关 (Spearman ' s rank correlation)
测量尺度(measurement scale)	肯德尔等级相关 (Kendall ' s rank - order correlation)
统计产品和服务解决方案 (Statistical Product & Service Solutions, SPSS)	点二列相关 (point - biserial correlation)
描述统计 (descriptive statistics)	二列相关(biserial correlation)
集中趋势度量 (central tendency measures)	ϕ 系数 Phi coefficient)
离中趋势度量 (dispersion measures)	列联系数(contingency coefficient)
偏度 skewness)	在线数据处理 (online analytical processing, OLAP)
峰度(kurtosis)	次数分布(frequency distribution)
百分位数 (percentile values)	列联表分析(cross tabulation)
百分等级 (percentile rank)	探索性数据分析 (exploratory data analysis, EDA)
标准分数 (standard score)	
相关系数 (correlation coefficient)	

►复习与练习

1. 名词解释

编码，编码框架，描述统计，统计量，参数，次数分布，集中趋势度量，离中趋势度量，分布

趋势度量，地位量数，皮尔逊积差相关，斯皮尔曼等级相关，质量相关，品质相关，探索性数据分析

2. 主要的资料编码方式有哪些？
3. 对于资料中的缺失值应如何处理？
4. 简述描述统计的含义及其内容。
5. 试比较算术平均数、中数和众数的特点及比较三者在不同分布中的位置关系。
6. 简述方差分析的基本原理。
7. 试比较百分位数和百分等级的含义和计算方法。
8. 相关量数包含哪几种？分别简述其基本原理和计算方法。
9. 使用 SPSS 自带的 Employee data. sav 数据对其中各种变量分别尝试进行多种描述统计和探索性数据分析。

► 推荐参考读物

[美] 艾尔·巴比. 社会研究方法. 邱泽奇译. 北京：华夏出版社 2000。

严辰松. 定量型社会科学研究方法. 西安市：西安交通大学出版社，2000

张敏强. 教育与心理统计学. 北京：人民教育出版社，2002 年。

张文彤. 世界优秀统计工具 SPSS 11 统计分析教程（基础篇），北京：希望电子出版社，2002. 第三部分

Hoaglin, D. C. Mosteller, F., & Tukey, J. W 探索性数据分析——Understanding Robust and Exploratory Data Analysis. 陈忠琰，郭德媛译. 北京：中国统计出版社，1998

► 在线资源

统计学基本概念：<http://www.statsoft.com/textbook/esc.html>

有关描述统计、相关分析、卡方检验：<http://www.statsoft.com/textbook/stbasic.html>

统计表：<http://www.statsoft.com/textbook/sttable.html>

统计图：<http://www.statsoft.com/textbook/stgraph.html>

探索性数据分析：<http://www.statsoft.com/textbook/stdatmin.html>

第十一章 推断统计原理和 SPSS 应用概述

小时候玩猜钱币的游戏，猜有字面朝上或国徽面朝上，或许你经常有运气都能够赢，但从长远看会有真正的赢家吗？看电视时经常会碰到关于产品质量报告的节目，质检员如何仅根据少数几个样品的质量就可以推断出整个产品甚至同类产品的质量？每一年高考后，各省都会给出一个分数线，它是根据什么标准制定的？如何正确评价某个学生成绩的高低？假设想知道某种学习方法对于提高学生成绩是否有作用，或者想比较视觉通道和听觉通道对同样的刺激材料的记忆成绩有无差异，应当如何进行统计分析？更具体的，如何利用前面介绍的描述统计量进一步挖掘数据资料中的有用信息，探寻事物或现象的本质？以上问题都涉及推断统计的应用。

学习本章，读者将了解到：

1. 概率和概率分布的概念。
2. 正态分布和二项分布的特点和用途。
3. 抽样分布概念及几种常用抽样分布特点和用途。
4. 统计推断的基本原理和步骤。
5. 常用统计检验法的分类以及使用条件。
6. 如何利用 SPSS 实现常用的统计检验。

第一节 概率分布及其应用

教育和心理学中的现象不同于物理学、天文学等自然科学中的必然现象（inevitable phenomena）或确定性现象（definite phenomena），它往往是不确定性现象（indefinite phenomena），即结果无法预先知道甚至事物本身含义不确定的现象。概率论和统计学是专门用来研究“在多次重复相同条件时单个试验结果呈现不确定性”的随机现象（random phenomena）——例如抛掷骰子、天气预报、心理物理实验中的作业成绩、考试成绩等——的数学工具，其根本目的正是要探讨和发现在大量重复试验条件下随机现象中隐藏的统计规律。作为推断统计（inferential statistics）的准备，本节概要介绍概率及概率分布的知识，更详尽的介绍读者可以进一步参阅专业统计学教材。

一、概率与概率分布

(一) 概率

一定条件下对随机现象所作的一次观测称为一个试验。试验的可能结果即随机事件(random event) A 发生的规律性可由其频率 $f(A)$ 来反映。实际上, 概率的一种定义——统计概率(statistical probability) 或后验概率(posterior probability) 即反映了这种联系:

$$P(A) = \lim_{N \rightarrow \infty} \frac{n}{N} \quad (11-1)$$

据此, 可以将频率视为和概率相对的一个“统计量”, 这样理解有助于解释次数分布和概率分布之间的联系。概率的另一种定义——古典概率(classical probability) 或先验概率(prior probability) 适用于基本事件个数有限、发生可能性相等且互不相容的场合下计算事件的概率, 例如对于掷骰子的试验, 其包含的 6 种基本结果(6 点) 可以组合成更复杂的事件“点大”、“点小”, 若要计算一次抛掷中出现大点的概率, 只需计算该事件(点大) 所包含的基本事件个数在总基本事件个数中占的比例, 即:

$$P(A) = \frac{A \text{ 事件包含基本事件个数}}{\text{总基本事件个数}} = \frac{m}{n} \quad (11-2)$$

(二) 离散型随机变量概率分布和二项分布

若将概率理解为和频率相对, 则概率分布(probability distribution) 可认为和次数分布或频率分布相对, 即当样本容量趋于总体容量时次数分布“理论”上的或理想中的极限形式。概率分布是总体的分布, 而次数分布是样本的分布。概率给出的是单个结果发生的可能性, 概率分布是对随机变量所有可能结果的可能性分布描述, 通常可以写成某个函数式形式。

根据随机变量(简写成 r. v.) 的类型, 概率分布可分为离散型随机变量(discrete r. v.) 概率分布和连续型随机变量(continuous r. v.) 概率分布。常见的离散型概率分布如二项分布(binomial distribution), 其概率分布函数表示的是总数为 n 的大量二项试验(binomial trials) 中, 某事件 A 出现 x 次的概率:

$$P(X=x) = C_n^x p^x q^{n-x} \quad (x=0, 1, 2, \dots, n) \quad (11-3)$$

二项分布主要用于二项试验或贝努利试验(Bernoulli trials)——重复试验间相互独立且每次试验仅包含两个互斥事件(A 和 \bar{A})——中的概率计算。例如投掷硬币、新生儿的性别、教育与心理学中的态度调查、试题中客观题的回答(只有对错两个类别)、题目或任务中的猜测(对或错) 等。

(三) 连续型随机变量概率分布和正态分布

连续型随机变量(如智商、反应时、考试成绩等) 的概率分布必须用变量 X

在某个区间内取值的概率 $P(a < X < b)$ 来表示。对于其概率分布可以这样理解：若将频率直方图的纵坐标转化为频率密度（频率与组距比值），设想样本容量趋于无穷大，分组越来越细（组距趋于 0），则某一范围内的频率将趋近于一个稳定值即概率。这时，直方图各直条最上端中点连线逐渐平滑趋于一条曲线，称概率密度曲线（probability density curve）。记概率密度函数为 $f(x)$ ，则 x 取值于区间 (a, b) 的概率为该区间在曲线下所围区域的面积，用积分来表示：

$$P(a < x < b) = \int_a^b f(x) dx \quad (11-4)$$

此即连续型随机变量概率分布函数表达式，概率密度曲线下总面积为全概率 1。连续型概率分布的典型代表是正态分布（normal distribution）亦称高斯分布（Gaussian distribution），因其发现者高斯（Carl Friedrich, Gauss, 1777—1855）而得名。自然界中凡受许多因素影响且其影响力都很小的现象大多服从正态分布，例如教育与心理学中学生的学业成绩、实验中的误差、人群的智商等。传统推断统计中许多成熟的统计分析方法都是以正态分布为基础，还有不少随机变量概率分布在一定条件下趋于正态分布。因此正态分布在统计学理论和实际应用中占有重要地位。

正态分布是一个典型的单峰、对称（ $X = \mu$ 为对称轴）的钟型分布（如图 11.1），两边低，中间高，且存在拐点 $X = \mu \pm \sigma$ 处）。正态分布实际上是一簇分布，随机变量 X 服从正态分布记为 $X \sim N(\mu, \sigma^2)$ 其中平均数 μ 和方差 σ^2 为其两个参数。 μ 决定对称轴的位置， σ 决定分数的离散程度，越分散曲线跨度越大则必然峰度越低（整个曲线下面积恒定）。

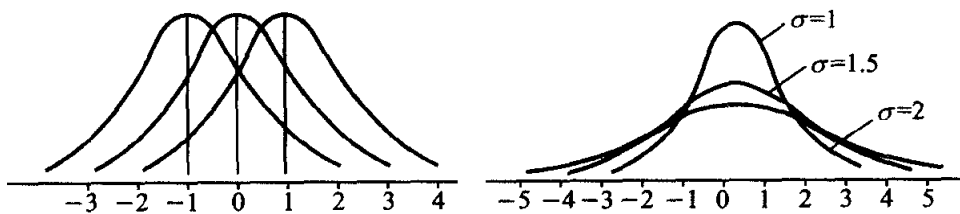


图 11.1 正态分布曲线随参数变化示意

前面已述及 Z 分数，将任意正态分布随机变量 X 值转换成 Z （称标准化过程），则该正态分布成为唯一的标准正态分布（standard normal distribution）其概率密度曲线形状恒定，因此曲线下面积（概率）和横坐标 Z 之间存在着对应关系，据此编制出标准正态分布表，供使用者计算正态分布概率时查阅。如图 11.2 所示，在图形右半边（与左半边完全对称）从 $Z = 0$ 开始逐渐右移，则自该 Z 值引出的竖直线到对称轴间围成的曲边梯形（图中分割线到对称轴围成的阴影部分）面积即对应的中央概率。如 $Z = 0.4, 1.2, 1.96, 2.58$ 时对应的中央概率分

别为 $P = 0.155、0.385、0.475、0.495$ 。求出中央概率，则尾端概率（分割线右侧面积）也不难求出。

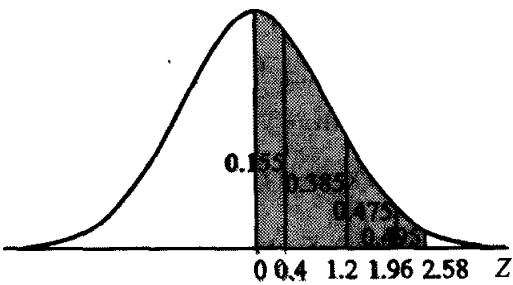


图 11.2 标准正态分布 Z 分数（横坐标）和中央概率对应关系示意

二、概率分布的应用及 SPSS 操作

（一）二项分布应用

二项试验在实际生活中非常普遍，因此二项分布的精确概率计算对于解决在教育与心理学研究中诸如态度调查、客观题的正确率、题目或作业中的猜测等问题很有用。以下结合 SPSS 的操作举例说明。

例 11-1 假定某高校学生使用网络人数比例为 60% 随机调查了 6 名该校学生，求该样本中有 3 名学生使用网络的概率是多少？至少有 3 人使用网络的概率又是多少？

解答：这是一个态度调查，其回答只能区分为是 / 否两个类别，是一个典型的二项试验。设随机变量 $X =$ 样本中使用网络学生人数。抽取 1 名学生相当于一次二项试验，一次试验的成功概率即抽取到使用网络学生的概率 $p = 0.6$ 。试验总次数为 6。则不难算得 X 取值的概率和累计概率分别为：

X	P	ΣP	X	P	ΣP
0	.00410	.00410	4	.31104	.76672
1	.03686	.04096	5	.18662	.95334
2	.13824	.17920	6	.04666	≈ 1
3	.27648	.45568			

即有 3 名学生使用网络的概率和至少有 3 人使用网络的概率分别为 0.276 和 0.456。利用 SPSS 的 compute 过程和相应的函数可以很方便地实现这种二项分布概率计算。

☆ SPSS 操作提示

利用 SPSS 中 Transform => Compute 过程中的二项分布累计函数 CDF. BIN-

OM 可方便地实现二项分布概率的计算。具体方法是：首先构造问题中要计算的各成功次数变量并输入数值（ $X = 0, 1, 2, 3, 4, 5, 6$ ）。接着调用 Compute 过程，设目标变量为 sum_px 表示 6 次二项试验中各成功次数的累加概率，在数值表达式框中选择函数 CDF. BINOM(q, n, p) 要求输入三个参数(?,?,?)分别输入 X 变量、6、0.6，执行即可。如果要得到每个成功次数对应概率，可再利用 Compute 过程中延迟函数 LAG 实现，具体方法是：在 Compute 对话框中令目标变量为 px，数值表达式输入 sum_px - LAG(sum_px)，意为用累计函数值减去其前一个累计函数值即可（LAG(sum_px) 中各记录为变量 sum_px 中该记录前一个记录，由于使用了延迟，因此新的变量中第一个记录为空）。上述过程的 Syntax 如下：

```

COMPUTE sum_px = CDF. BINOM(x,6,0.6).
EXECUTE.
COMPUTE px = sum_px-LAG(sum_px).
EXECUTE.

```

例 11-2 一个包含 10 道四选一选择题的测验，要答对 6 道题才能及格，问：(1) 若某考生完全凭猜测，及格的机会多大？(2) 若某考生经过复习可以排除每道题中的两个错误答案，问及格的机会多大？

解答：二项分布经常被用于检定测验中的“猜测问题”，即由于猜测的存在使得被试得分可能被高估。该题中虽然题目是四选一，但仍属于二项试验，因为反应的结果仍然只能有对 / 错两个类别，只不过猜对的先验概率 P 变成了 $1/4$ 。

设 $X = 10$ 题中猜对题数，问题 1) 实际上要求的是猜对 6 题以上（包括 6 题）的累加概率 $P(X \geq 6)$ 仍采用上面的方法：将函数 CDF. BINOM 的参数设定为 $x, 6, 0.25$ ，执行即可，得到下列结果。

X	0	1	2	3	4	5	6	7	8	9	10
$\sum P$.05631	.24403	.52559	.77588	.92187	.98027	.99649	.99958	.99997	≈ 1	≈ 1

可以看出，完全凭猜测猜对 6 题以上的概率 $= 1 - 0.99649 = 0.00351$ 若以小概率事件原理判断，几乎不可能，实际上猜对 5 题以上的概率就已经不足 0.05。同理，问题 2) 中排除了每题中的两个错误答案，等价于在两个答案中作猜测，显然只需将上述过程中 CDF. BINOM 的第三个参数改为 0.5 即可。算得此时及格的概率为 $1 - 0.82812 = 0.17188$ 。

此外，二项分布的极限分布是正态分布，当 n 较大时($p < q$ 且 $np \geq 5$ 或 $p > q$ 且 $nq \geq 5$) 事件 A 成功次数的平均数 μ 、标准差 σ 有： $\mu = np$ 和 $\sigma = \sqrt{npq}$ 。此时可将其视为正态分布，利用正态分布的良好特性进行概率计算。

（二）正态分布应用

正态分布主要应用于心理与教育测量或测验中，利用标准正态分布中概率和 Z 分数的对应关系划定分数线、确定特定等级或分数界限内人数以及编制测验常模中的各种标准分，其基本原理大同小异。

（1）分数线问题

例 11-3 某公司对新员工工作绩效进行考核。设全体员工的绩效得分呈正态分布，平均得分 80 分，标准差 12 分。现根据考核结果“奖优罚劣”：对得分最高的 10% 员工提升工资和职位，对得分最低的 30% 员工实行换岗再培训。已知甲、乙两名员工的考核得分分别是 90 和 60 分，问他们是否得到提升或换岗？

解答：这是典型的利用标准正态分布和 Z 分数划定“分数线”的问题，要求最高 10% 和最低 30% 的分数线。已知全体员工得分呈正态分布，其总体参数 $\mu = 80, \sigma = 12$ 。则 10% 和 30% 分别为数据分布的右侧尾端概率和左侧尾端概率，查标准正态分布表可以求得它们分别对应的 Z 分数 Z_1 和 Z_2 （如图 11.3），再利用总体参数反求得原始分数线。注意查正态分布表是要查中央概率，即分别查 0.40 和 0.20，同时还要注意 Z 分数的符号，这里 Z_2 应为所查值的负数。不难求得 $Z_1 = 1.28155$ 和 $Z_2 = -0.52440$ ，则求得原始分数线分别为：

$$X_1 = Z_1 \times \sigma + \mu = 1.28155 \times 12 + 80 \approx 95;$$

$$X_2 = Z_2 \times \sigma + \mu = -0.5244 \times 12 + 80 \approx 74。$$

可见甲员工虽然考了 90 分，但仍然进入不到前 10% 行列（因为员工的平均成绩太好）；乙员工考了 60 分，在第二个分数线以下，所以他要被换岗参加再培训。当然，该题也可以先求出甲、乙两员工得分的 Z 分数，查表换算得他们的百分等级（比他们考得更差的员工人数百分比），然后直接和 30% 和 90%（1 - 10%）比较，由此判断他们的去向。

（2）确定特定等级或界限内人数

例 11-4 设某班共 50 名学生，某次数学考试平均分 80，标准差 5 分。甲学生考了 85 分，问（1）其大致名次是多少？（2）如果选拔 90 分以上的同学去参加数学竞赛，问该班共有多少同学被选？假设数学考试成绩服从正态分布。

解答：这是典型的已知分数线求概率或人数的问题，和上例刚好相反。已知考试成绩正态分布的参数为 $\mu = 80, \sigma = 5$ ，则不难求得甲学生得分的 Z 分数为 $Z = (85 - 80) / 5 = 1$ ，查正态分布表换算得对应的右侧尾端概率为 0.1586 即全体学生中只有约 15.86% 的人比他考得好，人数约为 8。所以其大概名次为第九名（从高到低）。同理，求得 90 分在分数分布中对应的右侧尾端概率为 0.0227 人数大约为 1 人。

此外，若测验的得分服从正态分布，将原始分数化为等级评定，在保证各等级（例如 1 - 5 级）等距划分前提下，要确定各等级人数划分比例也利用到标准正态分布概率的求解。例如，通常将正态分布平均数左右三个标准差（即 $-3 \leq Z \leq 3$ ）视为正常数据范围，则可以将总共 6 个标准差等分为 5 分，即各等级 Z 分数界限分别为： $-3 \sim -1.8$ 、 $-1.8 \sim -0.6$ 、 $-0.6 \sim 0.6$ 、 $0.6 \sim 1.8$ 、 $1.8 \sim 3$ ，利用这些界限查表求出各等级对应的概率即是所求的人数划分比例（见图 11.4）。

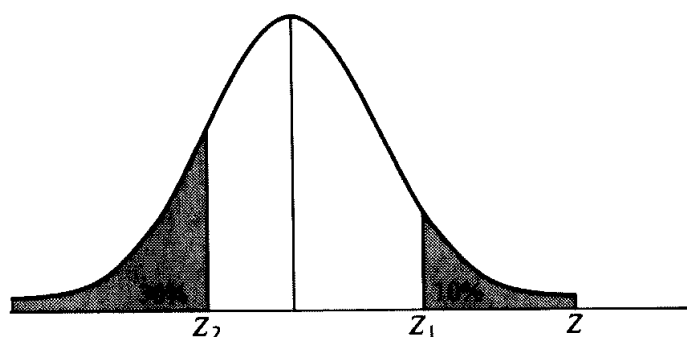


图 11.3 例 11 - 3 中所求分数线示意

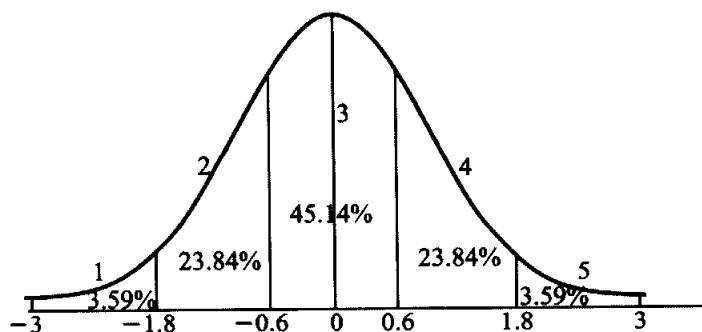


图 11.4 等距等级对应人数示意

☆ SPSS 操作提示

这两种问题的计算同样可利用 SPSS 中 Compute 过程的相应函数求解。

已知比例或概率求解 Z 分数或原始分数可利用逆分布函数 $\text{PROBIT}(\text{prob})$ 或 $\text{IDF.NORMAL}(p, \text{mean}, \text{stddev})$ 求解， PROBIT 函数只要求输入左侧累加概率 prob 即可，求出的是 Z 分数；而 IDF.NORMAL 函数则要求输入左侧累加概率、总体平均数和标准差，因此算出的是原始分数，更加简便。如例 11 - 3 中可设变量 p 分别输入 0.3 和 0.9，将 $\text{IDF.NORMAL}(p, \text{mean}, \text{stddev})$ 参数分别设定为变量 p 、80、12，即可求解出两个分数线。

已知分数求概率可利用累加概率函数 CDF.NORMAL(*q*,*mean*,*stddev*) (10.0 中还有一个 CDF.NORM(*q*) 用于输入 *Z* 分数), 分别输入原始分数、总体均值、标准差, 即可返回左侧累加概率。如例 11-4 中设定变量 *x* 输入 85、90, 输入函数参数为 *x*、80、5, 即可求得左侧累加概率。而 5 等级划分问题中, 设定变量 *x* 输入 -1.8、-0.6、0.6、1.8 函数参数为 *x*、0、1, 即可求得左侧累加概率, 将两个相邻累加概率相减即求得各等级比例。

利用上面两类函数, 也可求得大多数基于抽样分布的统计检验的临界值或尾端概率, 不过大多数情况下, 检验结果会直接在 SPSS 相应的统计检验中给出, 省去了查表的麻烦。

(3) 等级评定和题项难度的等距化

等级评定是心理与教育评价中常见的评定方式, 但不利于不同评定者之间的评分综合, 同时等级评定不具备等距属性, 不利于比较被评价者的心理量差异。此外, 心理与教育测验中题项的难度通常用该题通过率 (或经过猜测校正后的通过率) 表示, 但通过率并非等距尺度, 不利于不同题目间的难度比较。上述问题都牵涉到将等级评定或难度等距化。注意这里所谈的等级评定等距化刚好是上面所谈的等级人数比例划分的反问题, 其实质是将样本数据正态化, 然后利用 *Z* 分数或 *Z* 分数的变式表示测验得分或难度。

等级评定正态等距化的步骤是: ①将等级视为次数分布表的分组, 从小到大求出所有被评价者落入各等级的频率; ②求出各等级频率中点以上 (或以下) 的累加频率作为累加概率; ③将此概率转化为相应 *Z* 分数, 即该等级的等距测量值; ④如果存在多个评价者, 则用各被评价者所得 *Z* 分数均值进行被评价者间的差异比较。看下面的操作例子:

例 11-5 设 3 名教师对 70 名学生进行 1-5 的等级评定, 1 表示最低分, 5 表示最高分, 对评定分数进行正态化。

解答: 先调用 Frequencies 在结果中输出三位教师 (a、b、c) 评价的学生各等级频率百分比 *f*₁ 和累加频率百分比 *f*₂; 三组数据分别再拷贝入数据集, 利用公式 $f_3 = (f_2 - f_1/2)/100$ 求出中点以上累加频率; 调用 PROBIT 函数 输入参数 *f*₃ 求得对应 *Z* 分数即可 (各数据见表 11-1)。

表 11-1 例 11-5 中所求各数据汇总

等级	教师 a 评定				教师 b 评定				教师 c 评定			
	<i>f</i>	频率	中点以上 累加频率	<i>Z</i>	<i>f</i>	频率	中点以上 累加频率	<i>Z</i>	<i>f</i>	频率	中点以上 累加频率	<i>Z</i>
1	9	.129	.0645	-1.52	11	.157	.0785	-1.42	10	.143	.0715	-1.46
2	16	.229	.2425	-.7	18	.257	.2855	-.57	16	.229	.2565	-.65

续表

等级	教师 a 评定				教师 b 评定				教师 c 评定			
	<i>f</i>	频率	中点以上 累加频率	<i>Z</i>	<i>f</i>	频率	中点以上 累加频率	<i>Z</i>	<i>f</i>	频率	中点以上 累加频率	<i>Z</i>
3	16	.229	.4715	-.07	15	.214	.522	.06	19	.271	.5075	.02
4	22	.314	.743	.65	19	.271	.7645	.72	17	.243	.7645	.72
5	7	.1	.95	1.64	7	.1	.95	1.64	8	.114	.943	1.58

根据三位教师的各评定等级的 *Z* 分数，则每个学生的得分可用平均 *Z* 分数表示（将每个学生所得等级 Recode 为相应 *Z* 分数）。注意，下列两个学生的成绩都得到两个 5、一个 4，直接相加或平均无法体现差异，但经过转换后差异就体现出来。

	教师 a	教师 b	教师 c	平均成绩
甲生	4	5	5	$(0.65 + 1.64 + 1.58)/3 = 1.29$
乙生	5	4	5	$(1.64 + 0.72 + 1.58)/3 = 1.31$

对难度的等距化原理和方法是：假设不同难度题目在测验中呈正态分布，然后将各题难度作为概率（*P*）查正态分布表转换为 *Z* 分数（难度在 0.5 以下的 *Z* 为负值 用 $0.5 - P$ 查表；在 0.5 以上的为正值，用 $P - 0.5$ 查表），即为难度指数。考虑到该值存在负值，可用 $Z = Z + 5$ 或 $\Delta = 13 + 4Z$ 进一步转换。

更一般的测验分数正态化常用于编制测验的常模。通常，常模抽取的样本容量比较大，但仍不能保证样本数据（通常关心的是各维度得分）一定服从假想中的正态分布（如智力测验的各分测验），因此需要将原始分数分布正态化，求出 *Z* 分数，然后再换算出各种所谓的标准分。其方法和上述类似：自小数端求出原始分数的次数分布中各分组（通常需要根据该维度得分的全距来具体确定）上限以下的累加频率；用该频率减去各组频率一半作为概率（实际上是求各组中点以下累加频率）；查正态分布表转换为 *Z* 分数 将 *Z* 分数进一步转换为需要的常模标准分。如 *T* 分数就是经过正态化的一种标准分数，其平均数和标准差分别为 50 和 10 即 $T = 50 + 10Z$ 。

第二节 抽样分布和推断统计原理

描述统计关心的是样本信息的抽象和概括。推断统计要求根据样本的局部特性推断总体的情形，主要包括假设检验（hypothesis testing）和参数估计（parametric estimation）两部分。前者对某种关于总体分布情况或其参数的假设进行检验，根据计算结果推断所作假设能否接受；后者则以一次抽样为依据，对总体的某个数字特征（参数）作出估计。本节先介绍抽样分布的有关知识和应用，进而介绍推断统计特别是假设检验的基本原理及主要类型。

一、抽样分布概述

推断统计中要用样本统计量对总体参数作出估计，或者根据样本统计量对总体参数进行假设检验（主要指参数检验（parametric test），必须要了解样本统计量的分布情况。抽样分布（sampling distribution）即样本统计量（或随机变量函数）的理论分布*，是相对于基本随机变量的分布而言的。

（一）样本均值的抽样分布

不同的统计量可以得到不同的抽样分布，例如样本标准差、方差、中数、比例等的抽样分布。每个统计量的抽样分布都可计算出它的均值和标准差等，简称为该统计量的均值和标准差。最常见的是样本均值的抽样分布。设一总体平均数为 μ ，方差为 σ^2 ，称该总体为母总体。现从该总体中随机抽取容量为 n 的样本，样本平均数记为 \bar{x} ，则称该样本平均数的概率分布为均值抽样分布（sampling distribution of means）。通常记该分布的平均数和标准差为 $\mu_{\bar{x}}$ 和 $\sigma_{\bar{x}}$ 。其中，称 $\sigma_{\bar{x}}$ 为抽样分布的标准误（standard error, SE），表示平均数抽样误差的大小。根据中心极限定理（central limit theorem），均值抽样分布总体的两个参数与母总体的两个参数有如下关系：

$$\mu_{\bar{x}} = \mu, \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \quad (11-5)$$

图 11.5 显示了这一关系，随着样本容量的增大，抽样分布的标准误越来越小（因而峰度越来越高），因此，可以通过增大抽样的样本容量来降低样本平均数的抽样误差。若母总体 $X \sim N(\mu, \sigma^2)$ ，则其相应的均值的抽样分布始终是正态分布，即 $\bar{x} \sim N(\mu, \sigma^2/n)$ ；若样本容量足够大（大样本），不论母总体分布如何，只要总体参数有限且总体容量至少是样本容量的 2 倍，则均值抽样分布近似正

作者注：抽样分布一般指理论分布，不同于根据样本次数分布计算而来的经验分布。

态，且其参数仍满足 11-5 式。可见正态分布的应用非常广泛。

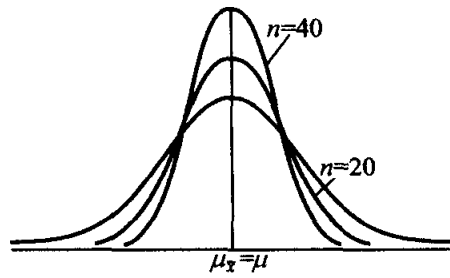


图 11.5 抽样分布 ($n=20, n=40$) 和母总体分布示意

(二) 比例的抽样分布

假设总体无限，某事件 A 成功概率为 p 失败概率为 $q = 1 - p$ ，从中抽取容量为 n 的样本 (即 n 次二项试验) 则对 A 事件发生的比例 P (成功次数 x /样本容量 n) 的抽样分布，有：

$$\mu_p = p, \sigma_p = \sqrt{\frac{pq}{n}} \quad (11-6)$$

注意，事件 A 成功次数 x 的均值 np 和标准差 \sqrt{npq} 同这里抽样分布的均值和标准差仅仅是单位不同。特别地，大样本时 (np 或 nq 较小者 ≥ 5) 该抽样分布为渐近正态分布。同样，当总体比例 (即先验概率 p) 未知时，可用样本比例 (x/n) 估计 P 计算抽样分布均值和标准差。

(三) t 分布

上述大样本统计量的抽样分布为近似正态分布的，常称为大样本理论；与之相对，小样本理论 (small sample theory，或精确抽样理论) 研究的是小样本统计量的抽样分布。比如当母总体标准差 σ 未知时，小样本均值的抽样分布通常不再是正态分布，此时对包含样本均值的统计量 t 有：

$$t = \frac{\bar{x} - \mu}{SE} = \frac{\bar{x} - \mu}{S_{n-1}/\sqrt{n}} \sim t(n-1) \quad (11-7)$$

即统计量 t (其形式上模仿 Z) 服从 t 分布 (t distribution) 或学生分布 (student distribution)，因该分布由高赛特 (William Sealy, Gosset, 1876—1937, 1908) 以 Student 为笔名在一篇文章中提出而得名。其中参数 ($n-1$) 称自由度 (degree of freedom, df)，为样本中变量个数 (样本容量) n 减去估计该统计量时用去的限制条件个数 k ，即统计量中存在的独立、自由变量个数。如样本标准差 S_{n-1} 的计算要用到所有数据均值为 \bar{x} 这个限制条件，因此其自由度为 $n-1$ ；要计算 t 统计量，也需要用到均值条件 (且只用到一个)，其自由度为 $n-1$ 。 t 分布是一个比正态分布低的单峰对称型分布 (如图 11.6 ①)，均值即对称轴 $t=0$ ，标准差为

$\sqrt{df/(df-2)}$ 。 t 分布曲线族随自由度增大而趋于正态分布，即大样本情况下可视为正态分布。 t 分布主要用在小样本条件下正态母总体方差未知时关于均值的推断统计。

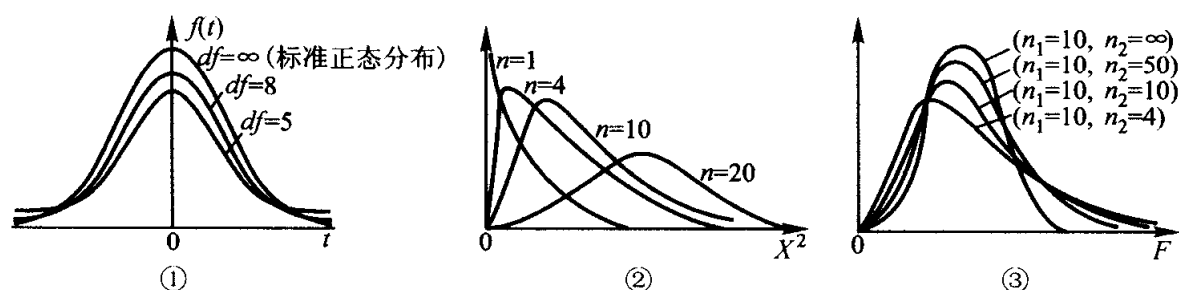


图 11.6 t 分布 (①) χ^2 分布 () 和 F 分布 () 概率密度曲线示意

(四) χ^2 分布和 F 分布

这两个精确抽样分布都主要用于有关样本方差的推断统计。 χ^2 分布 (卡方分布, Chi-square distribution) 由海尔墨特 (Friedrich Robert, Hermert, 1814—1917, 1875) 和皮尔逊 (Karl Pearson, 1857—1936, 1900) 分别提出。设 x_1, x_2, \dots, x_n 是 n 个相互独立的正态变量 (均服从 $N(\mu, \sigma^2)$) 则统计量:

$$\chi^2 = \sum_{i=1}^n \left(\frac{x_i - \mu}{\sigma} \right)^2 = \sum_{i=1}^n Z_i^2 \sim \chi^2(n) \quad (11-8)$$

和 t 分布一样, n 为卡方分布的自由度, 决定卡方分布曲线族的形状和位置。卡方统计量始终为非负数, 因此曲线只出现在直角坐标系的第一象限, 随着自由度增大, 卡方分布由正偏态趋于正态分布 (如图 11.6) ; 其均值和标准差分别为 df 和 $2df$ 。

F 分布 (F distribution) 由费希尔 (Ronald Aylmer, Fisher, 1890—1962) 提出。设两个相互独立变量有 $X \sim \chi^2(n_1)$ 和 $Y \sim \chi^2(n_2)$, 则随机变量:

$$F = \frac{X/n_1}{Y/n_2} \sim F(n_1, n_2) \quad (11-9)$$

F 分布有两个参数; n_1 是分子自由度, n_2 是分母自由度, 随着两个自由度增大, F 分布亦由正偏态趋于正态分布 (如图 11.6 ③) ; 其均值和标准差分别为 $n_2/(n_2-2)$ 和 $2n_2^2(n_1+n_2-2)/n_1(n_2-2)^2(n_2-4)$ 。此外, 从两个正态母总体 X, Y 中抽取出的随机样本 (容量各为 n_1, n_2) 其样本方差之比有:

$$F = \frac{S_{n_1-1}^2}{S_{n_2-1}^2} = \frac{\sum (x - \bar{x})^2 / (n_1 - 1)}{\sum (y - \bar{y})^2 / (n_2 - 1)} \sim F(n_1 - 1, n_2 - 1) \quad (11-10)$$

二、推断统计原理和步骤概述

假设检验亦称统计检验 (statistical test)、显著性检验 (test of significance) 是推断统计的重要内容之一。假设检验方法很多, 常用的如 t 检验、 F 检验和 χ^2 检验等。尽管这些检验方法的用途及使用条件不同, 但其基本原理大同小异。

(一) 假设检验原理

由标准正态分布特性可知, 标准正态曲线下某个 Z 分数范围确定的区域面积等于随机变量 (Z 分数) 相应取值的概率。设若离差智商的两个参数分别为 $\mu = 100, \sigma = 15$, 则不难算得一个 36 人的随机样本其平均智商在 110 以上的概率为 $P(Z > 4) \approx 0.0003$, 这个推断是基于均值的抽样分布 ($\bar{x} \sim N(\mu, \sigma^2/n)$) 意味着如果无限多次地进行有放回抽样, 抽得某个样本均值在 110 以上的可能只有 0.03%。这个小概率通常被认为在一次抽样中不会发生, 那么若某个样本的平均智商确实为 110, 则说明该样本所代表的总体其实和先前的总体存在显著的差别, 抑或这种差异其实是由于随机误差造成的? 这是假设检验要回答的问题, 看下面的例子。

例 11-6 某研究者想了解少年大学生群体的智商是否显著高于一般同龄人, 测查了某大学少年班的 70 名学生的韦氏智商 (假设相应年龄常模 $\mu_0 = 100, \sigma_0 = 15$) 平均智商为 103.3。能否认为少年大学生智力高于一般水平。

解答: 这里的一般智商母总体参数记成了 μ_0, σ_0 , 问题实际还牵涉到另一个总体: 即少年大学生群体 (设其参数为 μ, σ), 是要比较两个总体的均值 (参数) 间有无差异, 即一个显著性检验问题, 更具体要判断 μ 是否显著地大于 μ_0 。但关于 μ (今后就习惯地用 μ 来指代其总体) 的情况则一无所知。其次, 这里的 70 人样本是从 μ 中抽取而来, 既然无法正面回答 μ 是否显著地大于 μ_0 , 不妨采用“反证法”的思路。若两个总体本无“区别”则从 μ 中抽样等价于从 μ_0 中抽样。问题转换为上述的概率判断。在正常人总体中, 抽得 70 人样本均值大于 103.3 的概率是多大呢? 根据抽样分布理论, 70 人样本均值的抽样分布仍为正态分布, 均值 103.3 处在该抽样分布总体中的位置是:

$$Z = \frac{\bar{X} - \mu_0}{\sigma_0 / \sqrt{n}} = \frac{103.3 - 100}{15 / \sqrt{70}} = 1.84$$

不难求得标准正态分布中, $Z = 1.84$ 以上的概率是 0.0329。由此得出, 完全在 μ_0 总体中随机抽样, 抽得 70 人平均智商 103.3 的概率很小, 不足 5%。通常认为一次抽样中不会发生 (参见小概率事件原理)。这样, 上述推导的预期结果和“一次抽样就抽到 103.3 的样本平均数”的事实就有矛盾, 追根溯源, 有理由认为起初关于两个总体无差异的假设并不成立, 即该样本不可能来自于一般学生

智商总体，而是来自于另一个不同的总体（位于 μ_0 右端）。从而得出“少年大学生智力显著高于一般水平”的结论。

以上就是假设检验的主要思想，请读者结合这个例子学习下面的一些重要概念，进一步理解假设检验的原理。

（1）两类假设

研究者关注的研究假设往往是关于“差异”、“有效”、“显著”的假设，比如上述的“少年大学生智力高于一般水平”、“某种抑郁治疗药物有效”、“三种不同的教学方法效果间存在显著差异”、“学业成绩和学习动机强度之间的相关关系显著”等，记为 H_1 ，例如上面例子中的研究假设若用统计术语描述即： $H_1: \mu > \mu_0$ ；例子中已经看到，研究假设通常不能直接加以验证而必须借助于反证法，即取其对立面，首先假设两者无差异，这种包含等式假设即无差假设、零假设或虚无假设（null hypothesis），记为 H_0 ，例如上面例子中的零假设为 $H_0: \mu \leq \mu_0$ ；因研究假设与零假设相对，故也称前者为备择假设（alternative hypothesis）、对立假设。从例子可以发现，零假设是假设检验中推导过程的前提，它起一个“桥梁”作用，通过将已知样本（ x ）和已知总体（ μ ）建立联系而得到 x 的抽样分布，从而能够利用标准正态分布进行概率推断。没有零假设中的等号做桥梁，推断就无法进行；而研究假设则站在零假设的对立面，研究者正是通过“接受”或“拒绝”零假设达到对研究假设正确性的判断。总之，假设检验是根据无差异假设下，推导出的预期抽样结果和抽样事实是否有矛盾来判断是否存在真实的差异，因此习惯也将其称为统计决策。

（2）小概率事件原理和两类错误

例 11-6 中求得在 μ_0 总体中抽得 103.3 以上均值的概率为 0.0329，这是一个小概率，通常认为这种事件在一次抽样中不会发生，称为小概率事件原理，它决定了推导结果是否和抽样事实发生矛盾从而零假设是否有误。确定是否接受零假设的小概率标准称显著性水平（significance level），记作 α ，通常取 $\alpha = 0.05$ 或 0.01。

显著性水平在抽样分布右侧截出一段区域，如图 11.7 所示，在 H_0 为真条件下，图中实线即为 \bar{x} 的抽样分布，其右侧尾端截出 5% 的面积（图中浅灰色）。

若样本均值的 Z 分数落入该区域（计算得到的 Z 统计量值对应的尾端概率小于 0.05）或者 Z 分数超过检验临界值（尾端概率 0.05 对应的 Z 值）则拒绝 H_0 而接受 H_1 ，称该区域为 H_0 的拒绝域（region of rejection）；反之若落在临界值左侧区域则接受 H_0 而拒绝 H_1 ，称 H_0 的接受域（region of acceptance）。若拒绝了 H_0 则说明该样本均值来自于 H_0 拒绝域中的另一个不同的总体 μ （图中虚线）。

重新审查一下上述假设检验过程，应当看到，根据样本均值 \bar{x} 落入接受域或

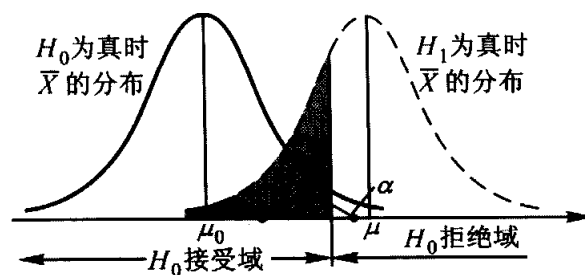


图 11.7 假设检验两类错误示意

拒绝域得出接受或拒绝 H_0 的结论是基于小概率事件原理，但这种结论可能会犯两种错误 第一是 H_0 实际成立却否定了它（ \bar{x} 落入图中浅灰色区域），称弃真错误或 I 型错误（type I error）、 α 型错误——即“错误地丢弃了真实的 H_0 ”统计量 Z 值不论多大其尾端仍有概率，只是推断时将其作为小概率事件忽略不计了，做出正确结论的概率只有 95%，而不是 100%。第二是 H_0 实际不成立却接受了它（ \bar{x} 落入图中深灰色区域），称取伪错误或 II 型错误（type II error）、 β 型错误——即“错误地择取了虚假的 H_0 ”。注意， β 区域是在 H_1 成立（虚线的总体 μ 存在）前提下 x 真实抽样分布尾端被 α 概率临界值截出的区域（图中总体 μ 中 α 临界值左侧部分）； x 原本来自于总体 μ 却因落入了 H_0 的接受域而被误认为来自于 μ_0 。

（3）单侧检验和双侧检验

上述的检验主要关心未知总体均值是否显著大于已知参数，即样本均值比正常水平（100）多出的 3.3 是代表了真实的差异还是因为随机误差所致，这称为单侧检验（one-tailed test），即将拒绝域置于抽样分布一端（例中放在右侧），换句话说，只需判断该样本均值是否大于（超过）此时的右侧临界值即可得出结论。类似的例子还有采用某种新的教学方法时抽样得学生的平均成绩比原来提高 5 分，要判断新方法是否有效（新方法下的总体均值是否显著大于旧方法）。

但有时无法事先断定两个参数的相对位置，只能笼统地判断两个参数是否有显著差异，这时检验的拒绝域（概率为 α ）被分配在抽样分布的两侧（每侧 $\alpha/2$ ），称双侧检验（two-tailed test）。可见，采用单侧或是双侧检验，主要是看研究者的检验目的以及问题是否提供了关于参数相对大小的信息。它们不同的提问方式也决定了两种检验的假设不同，如例子中单侧检验的零假设写为 $H_0: \mu \leq \mu_0$ ；而若为双侧检验则应写为 $H_0: \mu = \mu_0$ ，对应的研究假设为 $H_0: \mu \neq \mu_0$ 。注意，单侧检验零假设虽然是“小于等于”，但检验时只利用了相等的条件，只要统计量 Z 值大于右侧临界值（ Z_α ）即可否定零假设中的所有情形而接受备择假设（想想为什么？）而双侧检验中， Z 值大于右侧大的临界值（ $Z_{\alpha/2}$ ）或小于左侧小的临界值

($-Z_{\alpha/2}$) 都可以拒绝零假设。

(二) 假设检验的一般步骤

上面以最简单条件下的均值显著性检验说明了假设检验的原理,一般地,可以写出关于任何总体参数的假设检验或非参数检验的主要过程:

(1) 根据问题信息和检验目的写出两种假设 H_0 和 H_1 , 通常可以先写出 H_1 再写出 H_0 ; 可以用符号表示, 也可以用语言描述(非参数检验中)。

(2) 选择适当的检验统计量 U (如 Z 、 t 、 χ^2 、 F 等) 参数检验要求 U 必须有精确分布并包含被检验总体参数, 且无未知量; 非参数检验未必有精确分布, 但都有类似的统计检验表。

(3) 根据样本值计算统计量 U 的值 u 。

(4) 据显著性水平 α 和 U 的概率分布, 查表找出拒绝域和接受域的临界值, 注意在双侧检验时要查 $U_{1\alpha/2}$ 和 $U_{2\alpha/2}$ (某些分布中两临界值互为相反数); 单侧检验时只需查一侧 U 即可。

(5) 统计决策, 双侧检验中若 $u < U_{1\alpha/2}$ 或 $u > U_{2\alpha/2}$ 单侧检验中 $u < U_\alpha$ (或 $u > U_\alpha$ 视 H_0 而定), 都拒绝 H_0 接受 H_1 否则接受 H_0 拒绝 H_1 。例如, 上述的正态 u 检验中 $Z_{.05} = 1.64$ 则若 $Z > 1.64$ 或 $Z < -1.64$, 则否定零假设, 认为差异显著; 若改成双侧检验 $Z_{.05/2} = 1.96$ 则若 $|Z| > 1.96$, 否定零假设, 认为差异显著。注意到, 由于单、双侧检验中两端分配概率不同, 双侧检验中的临界值更极端, 同样的差异量, 使用双侧检验更不容易得到显著结论。

(三) 利用参数估计进行统计决策

参数估计和假设检验一样是推断统计的重要内容, 它可以使用样本统计量对总体参数进行推断。区间估计 (interval estimation) 不同于点估计 (point estimation) 只给出基于一次抽样的一个点估计值, 它通过写出一定置信水平 ($1 - \alpha$) 下总体参数的置信区间 (confidence interval) 来进行参数估计, 即总体参数以 $1 - \alpha$ 的概率落在该置信区间内。例如通常关心的总体均值的置信区间, 根据抽样分布知识, 只需知道一个相应的样本统计量和抽样分布标准误即可方便地写出该区间。例如, 根据抽样分布知识, 当总体方差 σ^2 已知时, 有 $P\left(\left|Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}\right| < 1.96\right) = 0.95$, 则 μ 的 95% 置信区间为 $\bar{X} \pm Z_{.05/2} \cdot \frac{\sigma}{\sqrt{n}}$; 总体方差 σ^2 未知时,

$P\left(\left|t = \frac{\bar{X} - \mu}{S_{n-1}/\sqrt{n}}\right| < t_{.05/2}(n-1)\right) = 0.95$, 则 μ 的 95% 置信区间为 $\bar{X} \pm t_{.05/2} \cdot \frac{S_{n-1}}{\sqrt{n}}$ 。

实际上可以将置信区间理解为“可接受的所有假设的集合”(柯惠新, 1992), 可借用区间估计的思想来替代传统假设检验中的“反证法”。例如对于

上例中的问题，同样在零假设条件下，给出所要判断之总体参数（ μ ）的 95% 置信区间为 $103.3 \pm 1.96 \times 1.8 = 99.8 \sim 106.8$ 。已知参数 100 包含在这个置信区间中，因此可认为在 95% 的置信水平（或 5% I 类错误水平）下零假设可以接受，从而完成假设检验。读者或可发现，这种统计决策和传统假设检验并无实质的区别，只是在操作步骤和表述方式上不同，不论是固定置信水平将零假设和置信区间进行比较，还是判定零假设条件下统计量在抽样分布中的位置都可以。

补充讨论 11-1 统计检验的统计功效及评价

实验的目的是获得显著的结果，统计检验的功效（power）指一个统计决策产生显著（差异）结果，即研究假设为真的概率。在前文，读者了解到统计检验中的两类错误，据此，表 11-2 给出了各种不同条件下做出不同统计结论所对应的概率，其中（ $1-\alpha$ ）和（ $1-\beta$ ）分别对应的是正确接受 H_0 和正确拒绝 H_0 的概率。可得其中的 $1-\beta$ 即统计功效，即得到真实差异结果的概率。有关统计功效的其他讨论，请读者参阅第四章补充讨论：一种特殊的内部效度——统计结论效度。

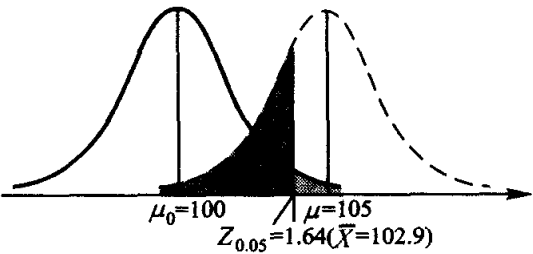
表 11-2 假设检验中四种概率间的关系

决策结果 \ 真实情况	H_0 为真	H_0 为假 (H_1 为真)
	H_0 为真	H_0 为假 (H_1 为真)
拒绝 H_0	错误拒绝 (α)	正确拒绝 ($1-\beta$)
接受 H_0	正确接受 ($1-\alpha$)	错误接受 (β)

由图 11-7 可知，给定条件下 α, β 的大小取决于检验临界值的相对位置，它们呈现一种此消彼长的关系，因此不能同时减小或增大，一般在假设检验中主要关心的是能否有理由拒绝 H_0 从而证实 H_1 ，所以在统计中对第 I 类错误规定较严（这也是为什么 α 通常取 5% 或 1% 的原因），在规定了 α 的情况下尽力避免 β 型错误。

在总体参数未知的情况下（如例 11-6 中不知道少年大学生智商的总体均值）统计功效无法直接确定，但通常可在参数的可能取值范围内确定 β 进而确定 $1-\beta$ 。如例 11-6 中如果少年大学生的平均智商为 105，则原假设检验中单侧 95% 水平临界值 $Z = 1.64$ ，其对应原始智商分数为 102.9。由下页图可知，第 II 类错误概率 β 即 $\mu(105)$ 分布中 102.9 值左侧的面积，不难求得该部分概率大小为 $0.5 - 0.378 = 0.122$ 则可得统计功效为 0.878，即若 $\mu = 105$ ，则该检验正确拒绝错误零假设的概率为 87.8%。对于给定的 α 和样本容量 n, β （参数 θ ）构

成的曲线即接受者操作特征曲线（ Receiver operating characteristic curve, ROC curve）在 SPSS⇒Graphs 中有一个 ROC curve 过程即可用来描绘接受者操作特征曲线。有关操作这里不再详述。



统计决策的功效主要取决于研究假设所预期效应（ effect ）的大小和实验中使用的被试数量。直观上看，研究假设所预期的两个总体分布的重叠部分越小，统计功效越大。称两总体存在差异的程度为效应量（ effect size ），因为它描述了实验处理能够区分两总体的程度。根据 Cohen 的效应量计算公式，大的总体均值差和小的总体标准差意味着较大的效应量和功效；而增大样本容量也会增大功效，因为大样本意味着较小的标准误从而容易检出差异结果。此外，其他一些统计技巧也会影响功效大小，如较宽松的显著性水平（ 如用 5% 代替 1% 或 0.1% 等 ）双侧检验等。

统计功效对于一项实验或研究的评价至关重要，当需要预先保证某研究效应的检出能力和避免犯检验错误以及解释研究结论的可信度时，它显得特别重要。根据 Cohen(1988) 的建议，一项研究功效通常要达到 80% 以上，这个标准视研究的困难程度和成本有所变动。功效过低意味着即便研究假设为真，数据的统计结果也不会产生支持它的结果。当然，研究的评价所包含的内容不仅限于功效。例如读者应当注意统计显著和实际显著的区别，一个差异显著的统计结果在实际应用中未必有多少价值。根据经典统计决策，尾端概率 P 为 5.1% 和 4.9% 会得到迥然相反的统计结论，但这两个结果实质并没有差别。关于研究评价的具体内容，读者可参阅本章有关推荐读物及第十四章元分析。

三、统计检验法的主要类别

经典的假设检验都是在一定的假设条件下（如母总体服从正态分布，数据为等距以上尺度等）对总体参数（如 $\mu \sigma^2$ ）所作的统计推断，即参数检验；伴随实际应用发展起来的非参数检验（nonparametric test）法则放宽了诸如分布形态、测量尺度、样本容量等限定条件，主要用于解决总体分布不明及等级或名义测度数据的假设推断。表 11 - 3 列出各种统计检验条件下的常用统计检验法，其中，单变量的统计检验法多用于分布假设检验，包括前面所谈到的单样本正态 u 检

验、单样本 t 检验，本质上也可以看成是对样本数据是否服从某个分布（样本所在总体均值是否等于已知总体）的假设检验。对于多变量条件，不失一般地将其变量区分为自变量和因变量。读者在后文将陆续看到对这些统计检验法的介绍。其中标星号的属于多元统计法，不在教材讨论之列。

表 11-3 统计检验主要类别简表

单变量		连续等距型		等级型变量		名义型变量			
		单样本正态 u 检验		单样本 $K-S$ 分布拟合检验		χ^2 拟合性检验			
		单样本 t 检验				二项分布拟合检验			
因变量	自变量	名义型变量						部分名义部分连续	连续变量 (一个或多个)
		一个				两个	多个		
		两分类		多分类					
		独立	相关	独立	相关				
单因变量	名义变量	χ^2 拟合性检验/对数线性模型*		χ^2 拟合性检验/对数线性模型		χ^2 独立性检验	对数线性模型		判别分析* /Logistic 回归
	等级变量	秩和检验 中数检验	符号检验 符号等级检验	克-瓦氏 方差分析	弗里德曼 等级方差分析		Logistic 回归*		Logistic 回归
	连续变量	独立样本 t 检验	相关样本 t 检验	单因素完全随机化 方差分析	单因素重复测量方 差分析	两因素方差分析	多因素方差分析	协方差分析*	相关分析/ 回归分析
多因变量	连续变量	多元方差分析*							

第三节 常用统计检验法的 SPSS 应用

统计检验法中最常用的是关于均值和方差的假设检验，这些检验过程在 SPSS 中主要包含在 Analyze 菜单中。本节主要介绍这两种最常用方法所涉及的 SPSS 过程界面及有关的菜单、功能选项。

一、差异显著性检验的 SPSS 应用

SPSS 的统计分析菜单如图 11.8，按照顺序，前两个子菜单包含的统计过程主要完成上一章介绍的描述统计。以下从均值比较开始介绍（Compare Means）。

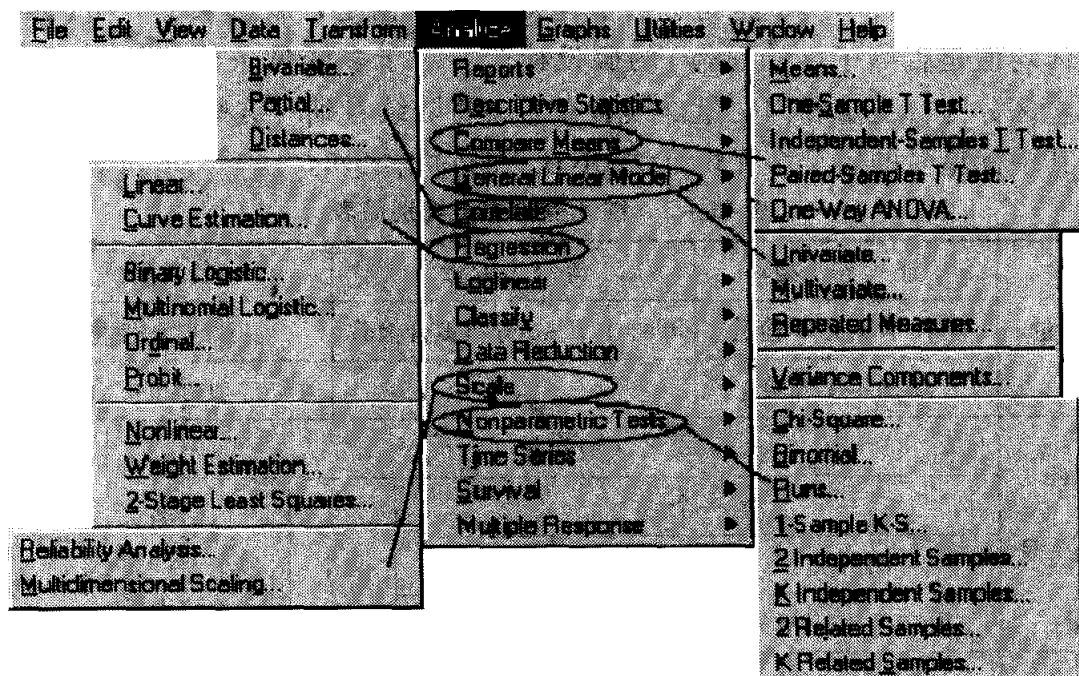


图 11.8 SPSS 统计分析 (Analyze) 菜单各统计分析过程及主要子菜单统计分析功能

这里的“均值比较”实际是关于有某种联系的不同总体间均值参数的差异显著性检验过程，最简单的情况就是两总体均值差异显著性检验。上节已述及，单样本均值的显著性检验可以理解成是对该样本来源总体 (μ) 和一个已知总体 (μ_0) 间的“双总体”差异显著性检验问题。因此，该子菜单中包含的统计过程主要解决两总体均值的差异显著性检验问题，主要指 t 检验。打开 Compare Means 子菜单，其中包括均值 (Means)、单样本 t 检验 (One - Sample T Test)、独立样本 t 检验 (Independent - Sample T Test)、相关样本 t 检验 (Paired - Sample T Test) 和单向方差分析 (One - Way ANOVA) 五个过程，其中单向方差分析可看成 t 检验的推广，但从分析的过程看已经和 t 检验有明显的不同，将和下一个菜单中的方差分析法一起讨论。

(一) Means 过程

该过程更倾向于对按照多个自变量各水平交叉分组的各观测指标样本数据进行描述统计。主对话框和前面所讲 Reports 子菜单的 OLAP Cubes 过程非常类似，可以对单变量或多个因变量在多个分组变量各处理下的各组样本数据计算均值、样本量、标准差 (默认的三个) 以及中数、方差、标准误、总和、极值、全距、峰度、偏度、调和平均数、几何平均数、各样本数据和及频率等绝大多数描述统计量。该过程的主对话框是一个典型的多自变量多因变量统计分析对话框 (如图 11.9)，可从左侧备选变量框中选择多个因变量 (统计指标) 和多个自变量分别

进入 Dependent List 框和 Independent List 框中。如果想得到按自变量各水平交叉组合形成的各处理描述统计量，可以利用自变量的分层设置设定 Layer 1、Layer 2 等层中的自变量。

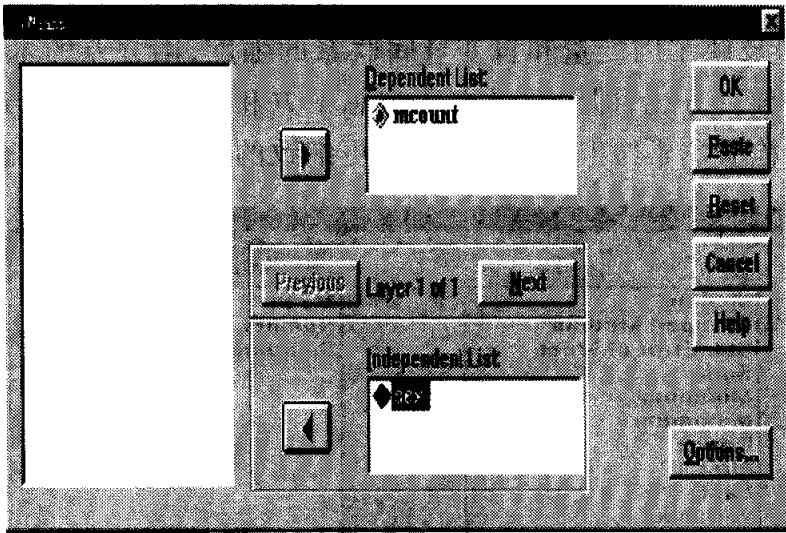


图 11.9 Means 过程主对话框

表 11-4 Means 过程结果示例

		榜样类型			
		男性		女性	
攻击类型		攻击性	非攻击性	攻击性	非攻击性
模仿身	男孩	25.8	1.5	12.4	0.2
体攻击	女孩	7.2	0.0	5.5	2.5
模仿言	男孩	12.7	0.0	4.3	1.1
语攻击	女孩	2.0	0.0	13.7	0.3
.....

*注 据《改变心理学的 40 项研究》P118 表 3.2 “儿童在不同处理条件下攻击反应的平均数”，有删节。

表 11-4 的数据就提供了这样的—个示例：这是班杜拉（Bandura, Ross, & Ross, 1961）对儿童攻击行为的一个著名研究的结果。统计的指标是“不同处理条件下的平均攻击次数”，即因变量；假设不考虑控制组条件，那么研究包括 4 个自变量，分别是给儿童呈现的榜样类型（攻击性 / 非攻击性）、榜样性别以及儿童性别和作出的攻击类型。假设提供每个儿童实验的原始数据，如何得到这样的描述统计结果呢？由于 4 个自变量之间是交叉组合结构，需要分别将四个变量设置入四层（设置的顺序可以随意，统计表的外观和变量的嵌套结构可以改变），否则将无法得到完全交叉后的各处理下描述统计，读者可以自行尝试。

Means 过程的特色是包含多种描述统计量的 Options 功能按钮，打开后（如图 11.10）在备选统计量框下方有一个 Statistics for First Layer 复选组，可分别实施按第一层变量分组的单向方差分析和计算 η (eta) 统计量以及第一层变量和因变量间线性关系检验，主要是提供决定系数 R^2 （参见第十二章）。类似于相关系数， η 统计量反映了因变量和自变量间联系的强度，用方差的术语讲， η^2 是因变量中组间变异百分比，即组间平方和占总平方和的比例。由于这里的方差分析和线性关系检验非常“粗糙”，还需进行更细致的多因素方差分析。

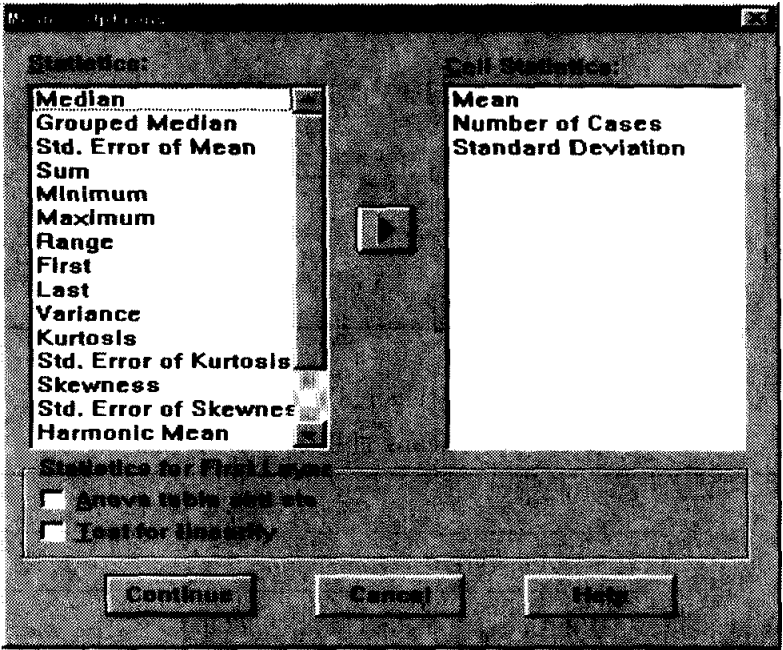


图 11.10 Means 过程 Options

（二）One - Sample T Test 过程

选择何种差异检验方法或公式关键是看样本个数、样本独立性、总体参数是否已知，以及样本容量等统计条件。这里的“单样本 t 检验”主要适用于单组数据（单个连续等距测量型变量）的均值与假设的检验值（给定常数）间的差异检验场合，即上面提到的显著性检验问题。该过程的界面和操作都比较简单，除两个变量框外，需要在下方的 Test Value 框中输入假设的检验值（已知总体均值）；此外打开 Options 按钮，除包含通常的缺失值处理选项外，还需要在 Confidence Interval 框中指定置信度为 $1 - \alpha$ 的置信区间，默认输入为 95%。

由于该统计过程要求给出特定的总体参数检验值，而绝大多数情况下总体参数的情况是很难掌握的，因此该方法在心理学研究中的应用有限。例如，心理测验中往往牵涉到个体能力或测验分数的评价，如果能够找到关于该测验的常模（norm）资料，则常模分数就可以作为一个检验的标准值，即将常模分数输入

Test Value 框中，对某组被试的平均测验得分和常模分数间的差异进行检验。当然，检验变量框可以选择多个变量，意味着可以使用同一个检验值对多组指标进行检验。再看下面的例子：

例 11-7 测得某班 40 名学生某项心理测验成绩，已知测验总体的平均数为 100。问该班学生平均成绩是否与总体平均成绩有差异

解答：根据问题条件，测验总体分布形态和方差均不明，因此适合用单样本 t 检验。该题要求对该班学生的总体平均成绩和测验总体平均成绩 100 间的差异进行比较。在 SPSS 分析菜单中选择 Compare Means => One - Sample T Test 选择变量 score 进入检验变量框，设定检验值为 100，运行分析过程，得到结果分两部分：首先，给出单变量描述统计，即该班学生成绩的均值、标准差以及抽样分布标准误 其次 给出单样本检验结果 如图 11.11。图中分别给出了 t 统计量值 (2.7)、自由度 (39)、双侧显著性水平 ≈ 0.01 (双击后可察看真实值) 样本均值与检验值的差 (6.25) 以及该差值的 95% 置信区间 (1.57 ~ 10.93)。检验结果拒绝无差假说，认为该班平均成绩确实和总体均值存在差异。

请读者思考下面的问题：假设问题条件发生变化，该测验是韦克斯勒智力测验，其总体均值和标准差分别为 100 和 15，且测验总体服从正态分布。问题应如何解决？如何在 SPSS 中实现？考虑到大样本的情况又应如何解决？

One - Sample Test

Test Value = 100						
				Mean Difference	95% Confidence Interval of the Difference	
	t	df	Sig. (2 - tailed)		Lower	Upper
测验成绩	2.700	39	.010	6.25	1.57	10.93

图 11.11 单样本 t 检验结果

(三) Independent - Sample T Test 过程

根据条件，该过程主要适用于两独立样本（数据类型为连续等距测量型且总体服从正态分布）总体均值差异检验场合。打开 Independent - Sample T Test 主对话框 如图 11.12。该对话框与单样本 t 检验的对话框相比，增加了一个分组变量 Grouping Variable。所谓分组变量通常即自变量，用于比较的两个总体一般通过一个外部标志联系在一起，而两组数据就是在该外部变量的两个水平下分别收集而来的，从数据集形式上看，两组数据实际上是一个变量（列），只是按照外部变量的不同水平各自成组。这一点对初学者来说尤为重要。因此，凡是自变量为单个分类变量或性质变量，或可通过取分界点人为划分为两个组别的

连续型变量（通过定义分组来划分），因变量满足上述数据要求的都可以使用该过程（参见第十二章）。注意该过程只能比较两个分组，但自变量水平并不一定是两个，实际上研究者可以任意地选择比较观测指标在自变量某两个水平上的均值。选定分组变量后打开变量框下方的 Define Groups 按钮定义需要比较的类别变量水平（如 Group 1 为 1 男性，Group 2 为 2 女性等）或者输入分界点 Cut point 定义要比较的连续变量的两个人为类别（如考试成绩 60 分以上为及格，以下为不及格等）。同样，和均值比较其他过程一样，待检验的因变量个数可以是多个。

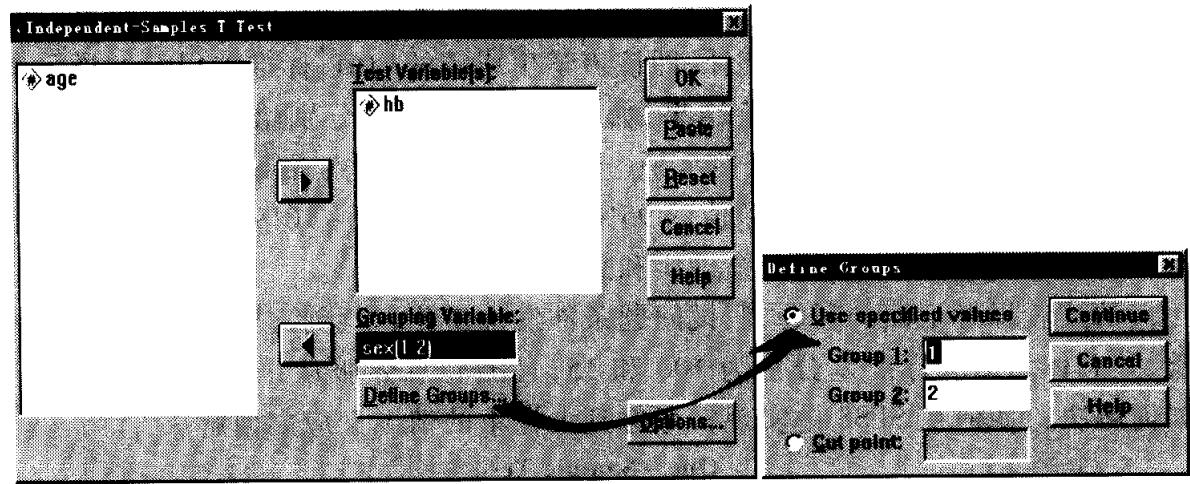


图 11.12 两独立样本 t 检验对话框和功能按钮

（四）Paired – Sample T Test 过程

该过程的适用条件是两相关样本或配对样本（数据类型为连续等距测量型且总体服从正态分布）总体均值差异检验场合。首先要区分什么是“独立样本”和“相关样本”。独立样本指从总体中抽取的、数据间彼此不存在对应关系的样本，通常其样本容量不一定相等。相反，相关样本指观测值间存在一一对应关系的样本，包括两种可能情形：一是对同一组被试进行重复观测得到的两组数据样本；二是将被试按照某种标准进行两两配对后得到的两组数据样本，亦称“关联样本”（related sample），其样本容量必然相等。例如，从同一组被试身上重复观测得到的简单反应时数据和智商分数，调查不同“对”哥哥和弟弟的创造性测验得分，或者不同对母亲的耐心程度得分和其子女的问题行为程度得分（张敏强，2002）等都属于相关样本。

相关样本 t 检验的对话框如图 11.13 所示，界面上除了有常规的 Options 按钮外，主要是一个 Paired Variable 框，用于选择希望进行比较的一对或几对变量：分别选中两个变量，选入配对变量列表，否则无法继续。

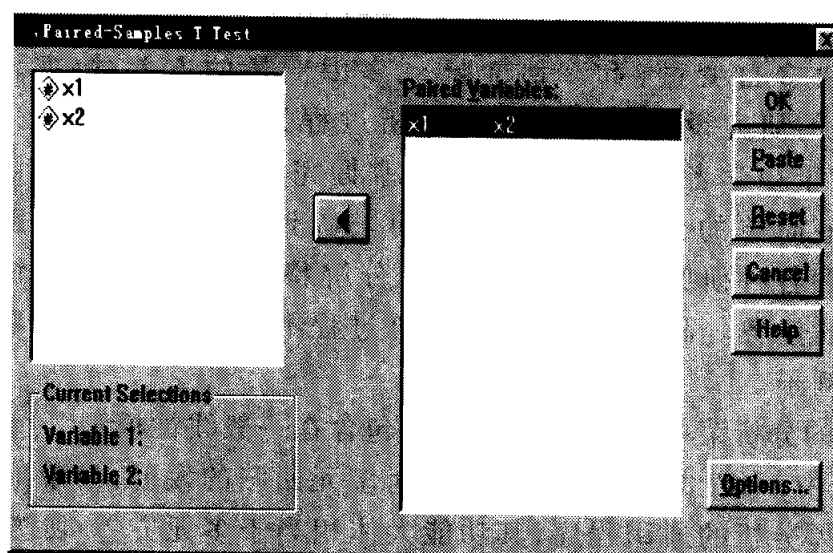


图 11.13 两相关样本 t 检验对话框

二、方差分析和一般线性模型

上述 t 检验分析的自变量只能有一个，从这一点来说，单向方差分析可看成是两组比较 t 检验的推广，它要求自变量有三个以上的水平。不过方差分析的思路已经不再是考虑均值两两之差的抽样分布，而是利用方差的可加性特点，根据不同的研究设计把观测值总变异的平方和及自由度分解为对应不同变异源 (source of variation) 的平方和及自由度，进而获得不同变异源总体方差的估计值；再利用适当的方差比构造 F 统计量，利用 F 检验推断各样本所属总体平均数是否相等。因此，在多组均值比较中，不能用多次 t 检验代替方差分析，其原因之一在于随着两两比较 t 检验次数的增加，会极大地增加第 I 类错误的概率*。

(一) 方差分析和一般线性模型

按照自变量的个数，方差分析有单因素 ANOVA 和多因素 ANOVA，单因素 ANOVA 包括 Compare Means 菜单中的 One-way ANOVA，不过它只适用于单因素完全随机化设计。若考虑无关变量控制，还包括单因素区组设计方差分析和单因素拉丁方设计方差分析等。按照因变量的个数，方差分析包括一元方差分析 (univariate ANOVA) 和多元方差分析 (MANOVA)，此处的“元”指因变量。此外，方差分析还可包括连续性自变量，称这种方差分

* 作者注：假设一次 t 检验犯 I 类错误的概率为 α ，则 k 次连续比较中该概率上升为 $1 - (1 - \alpha)^k$ ，参见 Aron & Aron, 1994 张文彤, 2004。

析为协方差分析 (analysis of covariance, ANCOVA)。而多元方差 / 协方差分析可用来解决多自变量 (和多协变量)、多因变量的方差 / 协方差分析。此外, 区分不同的研究设计其具体方差分析过程又有区别。方差分析的原理比较庞杂, 对于初学者来说, 最主要的是掌握完全随机化设计和随机化区组设计这两种最简单的方差分析过程即可。统计软件省略了手工计算的繁杂, 这时, 掌握各种研究设计对应的方差分解特点, 弄清如何看懂统计结论是读者最大的任务。限于篇幅, 具体计算过程不再赘述, 读者可学习第十二章中的相应例子。

在 SPSS 的分析菜单中, 方差分析功能包含在一般线性模型 * (general linear model, GLM) 中, 一般线性模型可视为是对多元回归的发展, 简单地说, 它是进行方差 / 协方差分析的通用程序, 其功能大体相当于多元方差 / 协方差分析; 在 SPSS 中, 多元方差 / 协方差分析是一个独立的过程, 不过只能在 Syntax 模式下运行, 它和 GLM 的主要区别是二者对分类变量的参数估计方法不同, 前者是将各水平与总均值进行比较, 采用的是偏差对比 (deviation contrast), 而后者是设定某参照水平, 将其他水平与之比较, 即简单对比 (simple contrast), GLM 所提供的功能和灵活性相比 MANOVA 更胜一筹。多元方差 / 协方差分析不是教材讨论的重点, 有兴趣的读者可参阅有关文献。

SPSS 的 GLM 菜单中包括单因变量 (univariate)、多因变量 (multivariate)、重复测量 (repeated measures) 和方差成分 (variance components) 四个过程, 可以解决完全随机设计、随机区组设计、交叉设计、析因设计、拉丁方设计、正交设计、嵌套设计、重复测量设计、裂区设计和均匀设计等类型的方差 / 协方差分析问题。以下重点介绍 Univariate 和 Repeated measures 两个过程的 SPSS 操作特点:

(二) Univariate 过程

Compare Means 过程中的单向方差分析可同时对一个自变量在多个指标 (因变量) 上的效应进行检验, 其界面和功能按钮也和 Univariate 过程类似, 读者可自行练习。

Univariate 过程主要解决的是单因变量的方差分析问题, 从功能来讲相当于多因素方差 / 协方差分析, 即可对连续型自变量进行统计控制。打开主对话框, 可将其界面和单向方差分析进行对比 (如图 11.14), 可以发现 Univariate 过程中备选变量的“去向”更加复杂, 其自变量在这里实际上包括三类, 即固定因素 (fixed factor)、随机因素 (random factor) 和协变量 (covariate) 甚至还需要为自变

* 作者注: 注意一般线性模型和广义线性模型 (generalized linear model) 不是同一个概念, 后者的范畴更大。

量设定权值系数 (WLS weight)。所谓固定因素，指的是某因素所有可能的水平都出现在样本中，这时考察的效应称为固定效应，即样本限定的具体处理间的差异；相反，随机因素指某因素所有可能的取值并不都在样本中出现，这时考察的效应称为随机效应，即随机选取的若干处理间的差异。通过设定固定因素和随机因素，Univariate 过程可以处理固定效应模型、随机效应模型和混合效应模型。要判断某因素是固定因素还是随机因素并非易事，但是固定效应模型的结论不能外推至相似处理，而随机效应模型的结论则可以外推，因此需要审慎确定两类因素。所谓协变量是指自变量中的连续变量，通常它无法在实验中进行随机化控制（如被试的机体变量），需要借助协方差模型将其设定为协变量以考察其效应。

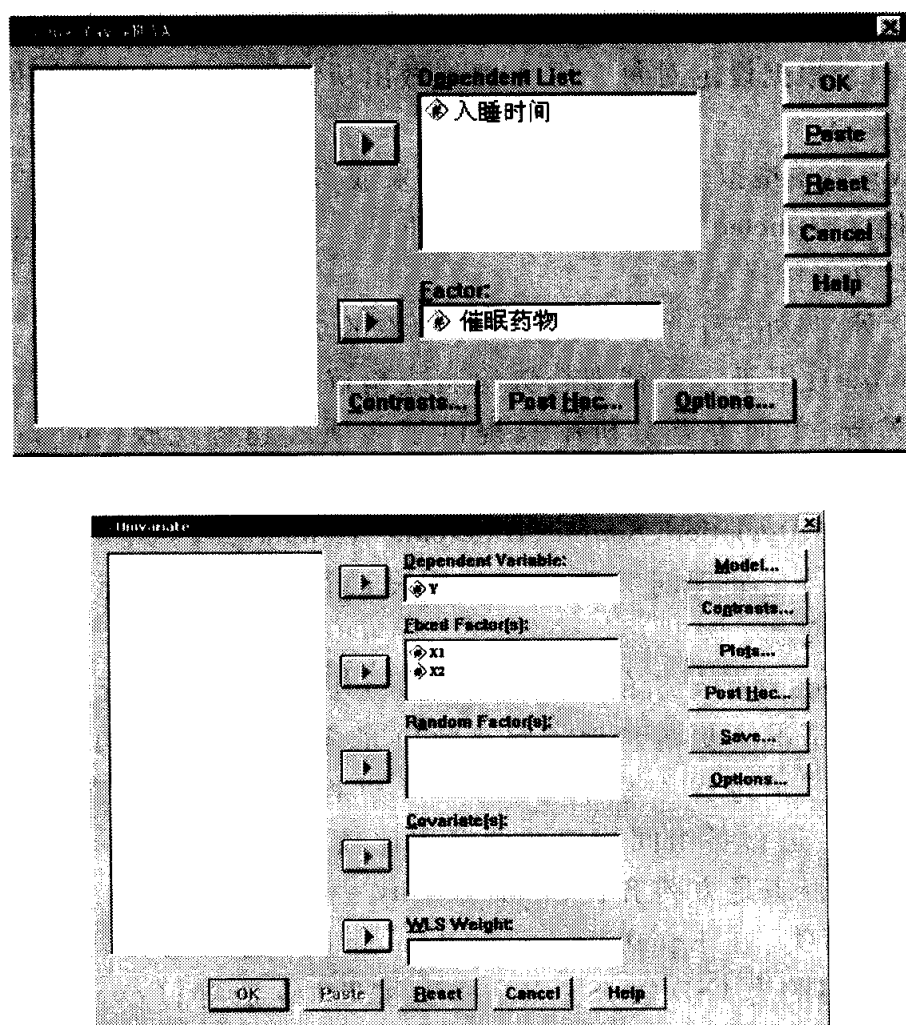


图 11.14 One - Way ANOVA 过程（上）和 Univariate 过程（下）界面比较

Univariate 中的功能按钮有所增加，除了多重比较 Post Hoc 按钮界面和单向

方差分析相同外，新增了 Model、Plots 和 Save 三个按钮，功能分别如下。

- **Model**：设定模型，允许用户在默认的全模型 (full factorial) 之外自行设定模型中所要考察的效应项（包括主效应和各阶交互作用），这可以说是 GLM 最大的人性化特点，提高了统计分析的效率。因为随着模型包含的效应数量增加，GLM 的运算费时增长惊人。经验上，一般三维（三个因素间）以上交互作用可以不用分析。该对话框中还包括选择适当的平方和分解方法和保留模型中截距 (intercept) 项，GLM 中共计四种平方和分解方法，其中默认的第 III 型对于完全（无缺失单元格）非平衡设计（单元格次数不等）具有较好的稳定性，而第 IV 型对于不完全设计效果较好（如图 11.15 ①）。

- **Plots** 画均值图 单向方差分析中 Options 功能按钮中也可以设定画均值图，不过因为只有单因素，画出来的只是一条折线；GLM 中一般都包含多因素，因此其设定要复杂得多。其对话框中“一次”可设定三个因素（协变量不加入作图），SPSS 自动根据设定对每个因变量都做出所设定因素的边际均值图（如图 11.15 ②）。

- **Save**：将模型拟合时产生的中间结果或参数保存为新变量供继续分析用，如预测值 (predicted values)、残差 (residuals)、诊断指标 (diagnostics) 等。有关指标的含义不再详述。

除此之外，另外三个单向方差分析中也有的功能按钮是：

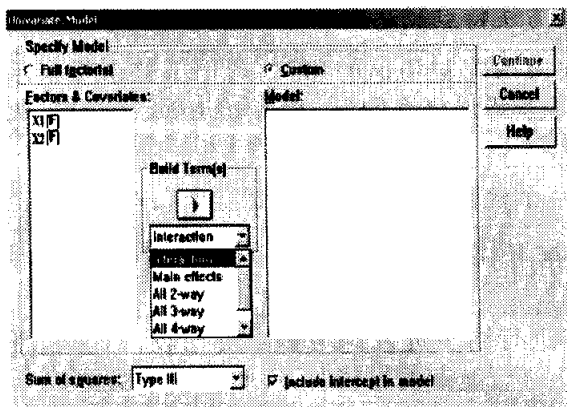
- **Contrast**：用于定义精细趋势检验或精确两两比较。该选项在三个过程中略有差异，单向方差分析中只提供了多项式趋势检验 (polynomial)，可选择从线性趋势一直到最高五次方曲线来进行检验。GLM 过程则提供了 6 种比较方法，如 Deviation、Simple、Difference、Helmert、Repeated 等。该子对话框主要用于对特定处理间关系的事先假设进行检验，其功能和应用请读者自行参考有关文献。

- **Post Hoc**：方差分析的事后 (post hoc) 检验可以使研究者在知道自变量效应显著的情况下更清楚地了解该效应究竟由哪些处理间的显著差异造成。该选项在三个过程中基本相同，GLM 过程中可指定对被选择因素（不包括随机因素）进行各组间均值两两多重比较 (multiple comparisons)。其方法按照各组是否满足方差齐性 (homogeneity - of - variance, or homoscedasticity) 条件分成两类：在 Equal Variances Assumed 复选框组中列出了 14 种方法，其中最常用的为 LSD 法、S - N - K 法和 Tukey 法等；Equal Variances Not Assumed 复选框组中共有 4 种，以 Dunnett's C 法较常用。同时还可以设定两两比较时的显著性水平，默认为 0.05。表 11 - 5 列出了各种方法的特点供参考（如图 11.15 ③）。

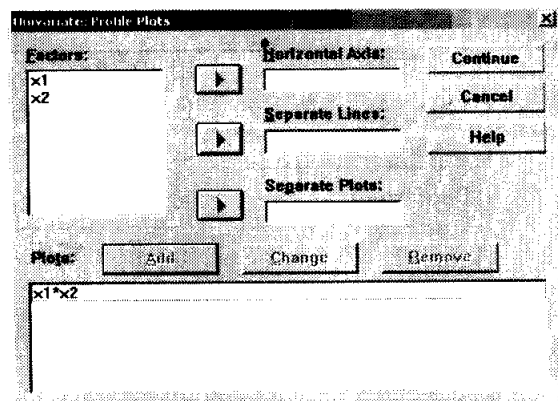
表 11-5 14 种多重比较方法特点及适应条件

方差 整齐 适用	LSD	用 t 检验完成组间两两配对均值的比较,利用了整个样本信息,但仍存在放大 I 型错误概率问题,和 Bonferroni 法一样,侧重减少第 II 类错误
	Bonferroni	同上,但通过设置每个检验的误差率来控制整个误差率
	Sidak	用 t 检验完成多重配对比较,为多重比较调整显著值,但比 2 的界限要小
	Scheffe	对所有可能的组合进行同步进入的均值配对比较,当各水平个案数不相等,或想进行复杂的比较时,用此法较为稳妥
	R - E - G - W F	Ryan - Einot - Gabriel - Welsch F ,用 F 检验进行多重比较
	R - E - G - W Q	Ryan - Einot - Gabriel - Welsch range test,在正态分布范围进行多重比较
	S - N - K	Student - Newman - Keuls,用 Studentized Range 分布进行所有各组均值间的配对比较,对 I 型错误控制较好,是应用最广泛的方法之一,和下法一样较适用于组数相等条件
	Tukey	Tukey's honestly significant difference,用 Studentized Range 统计量进行所有组间均值的配对比较,用所有配对比较集合的误差率作为试验误差率
	Tukey's - b	同上,但其临界值是 Tukey 和 S - N - K 检验相应值的平均值
	Duncan	进行配对比较时,使用的逐步顺序与 S - N - K 检验的顺序一样,但并不是给每个检验设定一个误差率,而是给所有检验的误差率设定一个临界值
	Hochberg's GT2	用正态最大系数进行比较检验和范围检验
	Gabriel	用正态标准系数进行配对比较检验,单元数较大时较自由
	Waller - Duncan	用 t 统计量进行比较检验,使用贝叶斯逼近
	Dunnett	选择最后一组为参照,其他组和其进行配对比较,要激活控制类别框选择对照组
方差 不齐 适用	Tamhane's T2	用 t 检验进行配对比较
	Dunnett's T3	用正态分布范围进行配对比较检验
	Games - Howell	同上,这种方法较灵活
	Dunnett's C	正态分布下的配对比较

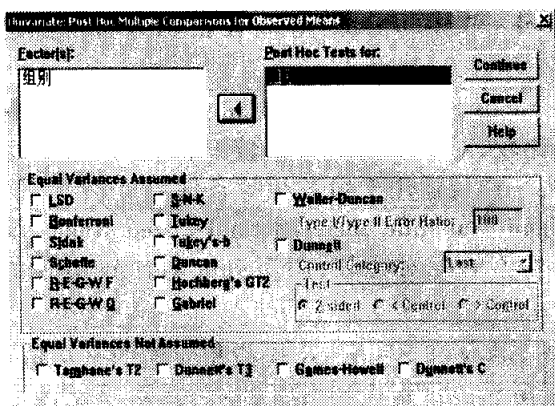
• Options 其他选项,One-way ANOVA 中主要包括描述统计、方差齐性检验、简单均值图复选框;GLM 中,增加了各处理估计边际均值(estimated marginal means)、主效应比较、效应量估计、参数估计、精细比较系数矩阵、散布-水平图、残差图、拟合劣度检验、一般估计函数表等。对于实际应用者来说,上述大多数选项并不一定常用,读者只需了解主要操作的实现方法并能够阅读相应分析结果即可(如图 11.15④)。



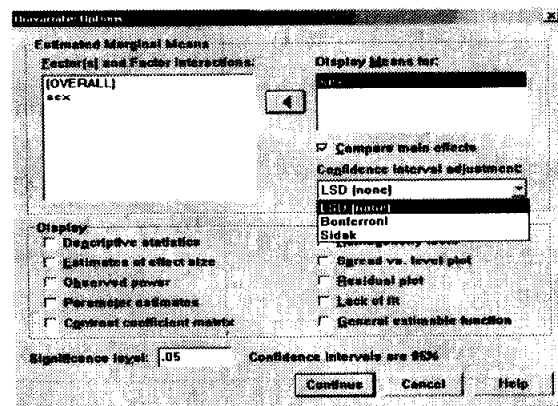
①



②



③



④

图 11.15 Univariate 过程主对话框和主要功能按钮

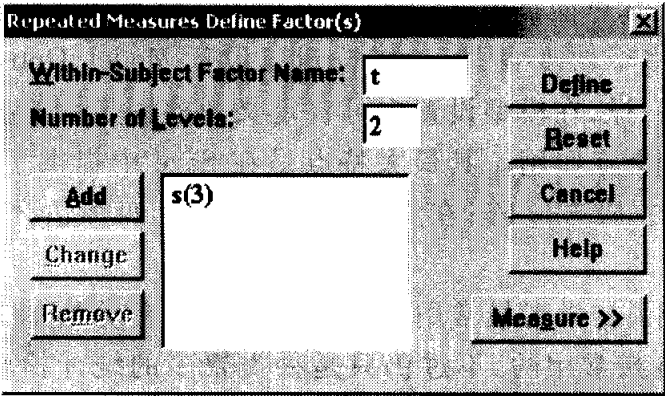
(三) Repeated Measures 过程

CLM 中对于被试内设计 (within - subject design) (包括牵涉到被试内变量的混合设计) 数据的方差分析是通过重复测量 (Repeated Measures) 过程实现的。被试内实验设计中每个被试接受多个甚至全部处理。某因素的所有水平均实施了重复观测, 则称该因素为被试内因素, 如测量的时间点; 而在重复测量中对于被试个体保持恒定的因素称被试间因素, 如性别。重复测量设计的重复观测间存在相关, 违反了一般方差分析的独立性假设, 因此其方差分析有独特的方法: 它是将总变异分解为被试间变异、被试内 (重复观测) 变异, 并从这两种变异中各自分离出被试间因素效应、被试内因素效应及对应的误差项, 其 F 检验的误差项要比一般的方差分析复杂得多。这种设计要求处理对被试没有长期的影响, 如学习、记忆和疲劳等效应。实验时每个被试接受处理的顺序应该进行随机化。

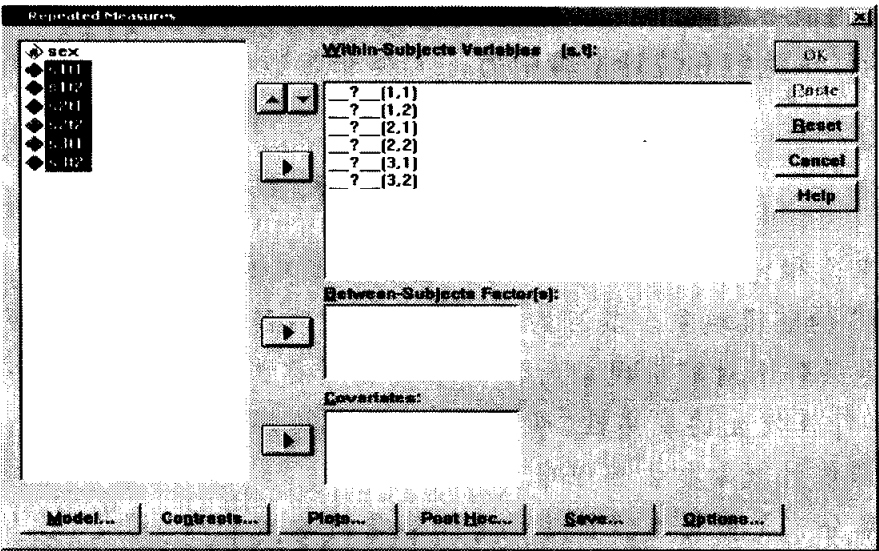
调用 Analyze => Repeated Measures 后, 首先进行被试内因素定义 (如图 11.16) , 因为数据集中的不同列代表了不同实验处理下的重复观测, 因素“隐藏”在变量列中。输入变量名, 再输入该变量水平数, 按 Add 添加。例如先后定义了两个因素 s 和 t 水平数分别为 3 和 2 ; 按 Define 键后进入主对话框, 左侧的备选变量框中除了有一个被试间变量性别 (sex) 外, 其他六个变量 (列) 均为同

一群被试的重复测量结果。由于两个被试内因素交叉后形成 6 个处理，不论每个被试接受这些处理的顺序如何，假设输入数据集后按顺序分别记为 s1t1、s1t2、s2t1、s2t2、s3t1 和 s3t2（假设序号分别表示该处理由两因素的哪两个水平组合而成）。选中六个变量到右边，可以看到各变量的记号刚好对应标准的处理序号格式即 s1t1(1,1)、s1t2(1,2)、……。同时可以看到，该过程同样允许选择被试间变量甚至协变量，即处理被试间—被试内混合设计的方差 / 协方差模型（图 11.16）。若是重复测量收集了多个指标（多因变量），则可打开 Measures，通过 Measure Name 来设定不同的观测指标。

该过程主对话框中的各功能按钮设置和结构同 Univariate 过程，需要注意的是在 Model 中，被试间变量和被试内变量是分别设置的。Post Hoc 中，只能设定对被试间变量进行多重比较，要对被试内因素进行事后比较，需要在 Options 中选择检验被试内因素主效应（compare main effects）。有关其分析过程和结果解释请参阅第十二章有关内容。



①



②

图 11.16 Repeated Measures 过程主界面示意

（四）GLM 中的其他过程

GLM 中的 Multivariate 过程可以解决多元方差 / 协方差分析问题，单向方差分析或多因素方差分析都只涉及一个因变量，而多元方差分析则包含多个相互关联的因变量，因此，其效果不等同于多次“一元”（单因变量）方差分析结果的复合。因为其考察各处理在多个指标上的差异是同时完成的，牵涉到因变量的多元联合分布。多元方差 / 协方差分析对于样本规模和因变量等有一些假设条件，大都可以利用相应的功能选项进行检验，读者可参阅推荐读物中的文献。从功能看，Multivariate 过程能够实现除重复测量方差分析之外的所有方差 / 协方差分析，前面所讲的单向方差分析和多因素方差 / 协方差分析都可以视为 Multivariate 过程的特例。其 SPSS 界面和操作特点类似于 Univariate 过程。多元方差分析不是教材讨论的重点，这里不再赘述。

GLM 中还包括一个方差成分（Variance Components）分析过程，也属于高级统计过程，主要用来进行随机效应方差分析模型的统计分析，也可以同时包含对固定效应的检验。其界面类似于 Univariate 过程，即可设置单因变量、多个固定因素或随机因素、多个协变量；但功能按钮只包括 Model、Options 和 Save 三个。和 Univariate 过程一样，在模型设置中可自定义考察固定因素间、随机因素间或固定 \times 随机因素间交互作用。其最大特点是在 Save 中提供了方差成分估计量以及各成分间相关 / 协方差阵；在 Options 中提供了四种估计随机效应方差的方法。

上面主要介绍了初学者最常用的均值检验和方差分析的方法，其牵涉到的 SPSS 统计过程也是最常用的，不过仍只是种类繁多的统计检验法的一小部分。此外，属于初等统计内容的相关分析、列联表分析过程在第十章中已有所涉及，其详细的应用请参阅第十二章“关联推断”。限于篇幅，本节所述两大类 SPSS 统计过程的详细例解和统计检验中另一大类非参数检验的 SPSS 应用将在第十二章介绍。

►本章提要

1. 概率论和统计学专门用于研究“在多次重复相同条件时单个试验结果呈现不确定性”的随机现象。其目的是要探讨和发现随机现象的统计规律。
2. 一定条件下对随机现象所作的一次观测称为一个试验。试验的可能结果为随机事件 A 发生的规律性（即概率）。可由其频率 $f(A)$ 来反映。概率的定义有两种，即统计概率和古典概率。
3. 概率分布是当样本容量趋于总体容量时次数分布“理论”上的或理想中的极限形式。根据随机变量的类型，概率分布可分为离散型概率分布和连续型概率分布。

4. 二项分布可用于解决在教育与心理学研究中诸如态度调查、客观题的正确率、题目或作业中的猜测等问题。正态分布主要用于心理与教育测量或测验中利用标准正态分布中概率和 Z 分数的对应关系划定分数线、确定特定等级或分数界限内人数以及编制测验常模中的各种标准分。

5. 抽样分布是样本统计量的理论分布，是推断统计的基础。常见的抽样分布有样本均值的抽样分布、比例的抽样分布、 t 分布、 χ^2 分布和 F 分布等。

6. 假设检验是推断统计的重要内容之一，其目的是对关于总体分布情况或其参数的假设作检验。常用的假设检验法有 t 检验、 F 检验和 χ^2 检验等。

7. 参数估计也是推断统计的重要内容，是以一次抽样为依据，使用样本统计量对总体参数进行推断的方法，主要有点估计和区间估计两种。

8. SPSS 中进行最常用的差异显著性检验和方差分析的过程主要牵涉到分析菜单中的均值比较、一般线性模型子菜单。它们是统计分析中最常用的统计过程。

►本章关键术语

必然现象 (inevitable phenomena)	χ^2 分布 (Chi - square distribution)
随机现象 (random phenomena)	F 分布 (F distribution)
推断统计 (inferential statistics)	零假设 / 虚无假设 (null hypothesis)
统计概率 (statistical probability)	备择假设 (alternative hypothesis)
古典概率 (classical probability)	显著性水平 (significance level)
概率分布 (probability distribution)	拒绝域 (region of rejection)
二项分布 (binomial distribution)	接受域 (region of acceptance)
正态分布 (normal distribution)	I 型错误 (type I error)
标准正态分布 (standard normal distribution)	II 型错误 (type II error)
假设检验 (hypothesis testing)	单侧检验 (one - tailed test)
参数估计 (parametric estimation)	双侧检验 (two - tailed test)
参数检验 (parametric test)	区间估计 (interval estimation)
抽样分布 (sampling distribution)	点估计 (point estimation)
均值抽样分布 (sampling distribution of means)	非参数检验 (nonparametric test)
标准误 (standard error, SE)	单样本 T 检验 (one - sample t test)
中心极限定理 (central limit theorem)	独立样本 T 检验 (independent - sample t test)
小样本理论 (small sample theory)	相关样本 T 检验 (paired - sample t test)
t 分布 (t distribution)	单向方差分析 (one - way ANOVA)
自由度 (degree of freedom, df)	

多元方差分析 (multivariate analysis of variance, MANOVA)

协方差分析 (analysis of covariance, ANCOVA)

一般线性模型 (general linear model, GLM)

固定因素 (fixed factor)

随机因素 (random factor)

协变量 (covariate)

多重比较 (multiple comparisons)

方差齐性 (homogeneity - of - variance or homoscedasticity)

被试内设计 (within - subject design)

►复习与练习

1. 名词解释

必然现象, 随机现象, 推断统计, 概率分布, 抽样分布, 参数检验, 非参数检验, 假设检验, 离散型随机变量, 连续型随机变量, 正态分布, 二项分布, t 分布, χ^2 分布, F 分布, I 类错误, II 类错误, 点估计, 区间估计, 固定因素, 随机因素, 协变量, 方差分析, 多重比较, 一般线性模型

2. 如何理解概率分布及其种类间的差异?

3. 例 11 - 7 中, 假设问题条件发生变化, 该测验是韦克斯勒智力测验, 其总体均值和标准差分别为 100 和 15, 且测验总体服从正态分布。问题应如何解决? 如何在 SPSS 中实现? 考虑到大样本的情况又应如何解决?

4. 简述二项分布、正态分布及其特点。

5. 试简述各种抽样分布的概念和特点。

6. 什么是标准误? 标准误与标准差有何联系与区别?

7. 如何理解假设检验的原理? 试简述其一般步骤。

8. 如何理解两类错误? 试简要说明什么是小概率事件原理。

9. 单侧检验和双侧检验的依据是什么?

10. 简述点估计和区间估计的基本原理和步骤。

►推荐参考读物

张敏强. 教育与心理统计学. 北京: 人民教育出版社, 2002

柯惠新. 调查研究中的统计分析法. 北京: 北京广播学院出版社, 1992

张文彤. 世界优秀统计工具 SPSS 11 统计分析教程 (基础篇). 北京: 希望电子出版社, 2002

张厚粲. 心理与教育统计学. 北京: 北师大出版社, 1998

Aron, A. & Aron, E. N. Statistical methods in psychology. New Jersey: Prentice - Hall, 1994

Heiman, G. W. Understanding research methods and statistics: An integrated introduction for psychology (2nd ed). Boston: Houghton Mifflin Company, 1998

➤ 在线资源

统计学基本概念 : <http://www.statsoft.com/textbook/esc.html>

有关 t 检验 : <http://www.statsoft.com/textbook/stbasic.html>

有关方差分析 / 多元方差分析 <http://www.statsoft.com/textbook/stanman.html>

第十二章 研究设计的 SPSS 统计分析

设想某个人力资源经理试图考察员工的薪酬和工作满意度之间的可能关系，以帮助其确定薪酬框架，数据是使用问卷测得的员工工作满意度和月收入，如何用统计方法实现这种关系的考察呢？又设想他试图考察部门负责人和普通员工的工作满意度间可能存在的差异，如何可靠地推断这两种不同职责间确实存在显著的差异呢？假设问题进一步深入，设想不同的性别、多种职责类别、不同工作部门、不同工作年限等因素都可能影响员工的工作满意度，又如何考察这些不同因素对工作满意度的可能效应呢？本章内容照应教材第二编，对于定量研究设计特别是真实实验设计、准实验设计和非实验设计中可能用到的统计分析方法均有涉及。统计方法有其自身的体系，从数据分析的角度看，统计方法可以举一反三——不同的研究设计自然可能使用相同的统计方法，同一个研究也可能使用不同的统计方法来挖掘不同类型的信息。本章重点在于阐明不同的研究设计和研究目的和可能的统计方法间的联系，并结合 SPSS 统计软件的分析菜单提供操作样例。

学习本章，读者将了解到：

1. 探索和度量变量间线性关系的统计方法——相关分析。
2. SPSS 相关分析操作及相关分析应用。
3. 线性回归分析的基本内容和步骤。
4. SPSS 线性回归分析操作及其应用。
5. 卡方检验的主要内容及用途。
6. SPSS 进行卡方独立性检验的操作。
7. 不同条件下 SPSS 进行数据组间比较的操作。
8. 因素设计方差 / 协方差分析的 SPSS 操作。

第一节 关联推断

真实实验设计的目的主要在于探查关于变量间的因果关系假设，而诸如调查法等非实验设计则主要立足于探查一些描述性假设（descriptive hypotheses）如关于变量间相关关系的假设。为了不和“相关分析”统计法相混淆，将探究变量间相关或关联程度的方法统称为“关联推断”法。其在 SPSS 中主要对应 Corre-

late Regression 子菜单以及 Descriptive Statistics 和 Nonparametric Tests 子菜单中的部分命令。本节介绍关联推断中常用的统计法及其 SPSS 操作特点。

一、相关分析及其应用



在研究的起始阶段，研究者最先想到的可能是两个行为或心理过程之间是否存在关联，例如高尔顿最先引入相关概念源于其对父代和子代间的身高存在关联的假设；另一个研究者根据经验观察假设人的攻击行为或许和环境温度存在关联等。借助于相关分析（correlation analysis）不仅可以计算相关系数以度量这种关联的程度，还可对这种关联的真实程度进行显著性检验。

（一）相关分析的 SPSS 操作

第十章中接触到相关系数作为相关程度的度量，涉及到不同测量尺度的相关指标，例如用于等距尺度的积差相关系数、用于等级尺度的斯皮尔曼相关系数和用于名义尺度的列联系数等。SPSS 中 Analyze 菜单下的相关过程（Correlate）主要只包含等距和等级两种尺度数据的相关分析，而更广泛意义上的相关分析特别是名义尺度数据的关联分析则包含在 Descriptive Statistics \Rightarrow Crosstabs 过程中。

1. Bivariate 过程

Bivariate 意即“双变量”，主要是进行多变量间两两线性相关分析的过程，也是最常用的相关分析过程。第十章已述及，Bivariate 过程提供了三种相关系数计算方法（Pearson、Kendall's tau-b 和 Spearman）Pearson 相关系数用于双列等距正态数据，能较好地反映变量间的线性关系（即简单相关）*；而 Spearman 和 Kendall's tau-b 等级相关系数是适用于等级型数据或非正态等距型数据的非参数相关分析法，前者主要利用两变量的秩（rank）——即数据在样本中按从小到大所占等级的大小作线性相关分析，后者则主要用于反映有序分类变量间的相关性。需要提醒的是，由于三种计算方法对数据测量尺度要求逐渐降低，通常满足更高要求的数据也可用更“低级”的相关指标进行分析，但统计功效会下降，表现在其计算的相关系数绝对值越来越小。

在进行相关分析时，利用散点图（scatter plot）可直观地帮助确定两变量间是否存在相关趋势，该趋势是否为直线趋势以及相关的程度等，即以两变量分别为横、纵坐标，将其成对观测值在平面直角坐标系描点，如图 12.1 ①、所示：散点分布的形状预示着相关的程度，①中显示形状接近椭圆表示两变量具有直线型关系，形状越扁平，相关程度越大；显然，表示两变量间几乎没有相关。

作者注：皮尔逊相关还可用于计算质量相关中的点二列相关系数，后者常用于评价测验中题项的区分度（discrimination）。

通过椭圆的朝向还可以判断相关的方向，比如 中椭圆长轴朝向第一、三象限，表示正相关，反之表示负相关。需要注意的是，有时两变量间的相关关系为 中所示的曲线型，这时用积差相关系数等适用于直线相关的相关系数往往无法准确地表达这种相关关系和相关程度。SPSS 中提供多种散点图的画法，其中最简单的是调用 Graphs = > Scatter，选择其中的 Simple，然后选择好横、纵坐标变量即可。

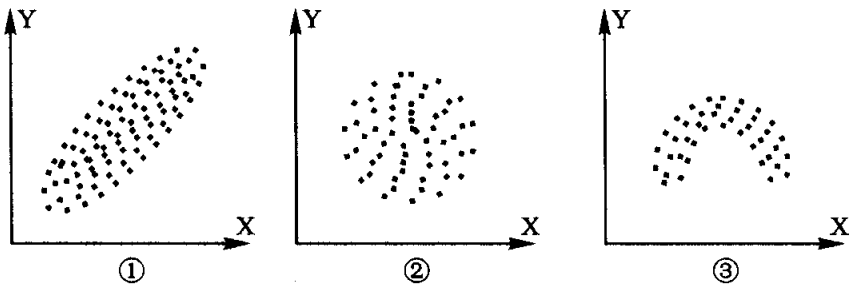


图 12.1 相关散点图示意

2. Partial 过程

偏相关(partial correlation)系数是对在控制其他额外变量情况下两变量间线性关系的描述，通常直接计算出来的相关系数反映了变量值数量变化的联系，这种联系并不一定完全反映了两变量间真正的联系，因为其间可能还夹杂了其他变量的影响。例如，有人要考察婚姻满意度 (x) 和结婚时间 (y) 之间的联系，简单相关系数可能会产生误导，例如可能会得出结婚时间越长，婚姻满意度越低的结论；但若控制住生活水平 (z) 的影响，则它们之间的联系就会显著地下将。可以将偏相关理解成是对两变量间“真正的、纯粹的”关联的反映，因此也称其为“净相关”。和积差相关系数一样，计算偏相关系数通常也要求数据服从二元正态分布。一阶偏相关系数（控制一个变量）计算公式如公式（12-1）：

$$r_{xy.z} = \frac{r_{xy} - r_{xz}r_{yz}}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \quad (12 - 1)$$

假设婚姻满意度的例子中三个变量的简单相关系数分别为： $r_{xy} = -0.5$ ， $r_{xz} = -0.5$ ， $r_{yz} = 0.3$ 为考察变量间的净相关 计算得 $r_{xy.z} = -0.42$ 即若控制了生活水平的影响，实际上婚姻满意度和结婚时间的相关程度要低一些。看一个实际研究的例子来具体分析 SPSS 的偏相关分析操作。

例 12-1 某研究者欲考察员工的工作满意度和收入满意度的关系，但研究者考虑到这种关联可能会受到员工对其上司满意度的影响，假设三种满意度均采用 5 点自陈式评分，1 表示“非常不满意”，5 表示“非常满意”。试进行分析推断。

解答：首先不妨回顾一下该研究属于第七章中谈到的何种类型设计？研究者并没有对被试施加系统的实验处理，而是基于事后观测结果探讨变量间的关系，所以这是一种典型的相关研究设计（见第七章第四节）。调用 Analyze⇒Correlate⇒Partial 过程，指定“工作满意度”和“收入满意度”两变量进入 Variable 分析变量框，指定“上司满意度”变量进入 Controlling for 控制变量框，试图分析在控制了对上司满意度影响情况下员工工作和收入满意度间的关系。为了和前面的 Bivariate 过程比较，打开 Options 按钮，勾中零阶相关（Zero-order correlation）复选框。运行结果如图 12.2 所示，首先给出零阶偏相关系数，相当于前面 Bivariate 过程求三变量间两两相关系数，相关系数均在 0.01 水平上显著（**）；控制了上司满意度（supersa）后，求得工作满意度和收入满意度间净相关为 0.4818，比原先的 0.6610 要小。

- - - PARTIAL CORRELATION COEFFICIENTS - - -			
Zero Order Partial			
	JOBSAT	SALSAT	SUPERSA
JOBSAT	1.0000	.6610**	.5630**
SALSAT	.6610**	1.0000	.6179**
SUPERSA	.5630**	.6179**	1.0000
* - Signif. LE .05 ** - Signif. LE .01 (2-tailed)			
- - - PARTIAL CORRELATION COEFFICIENTS - - -			
Controlling for.. SUPERSA			
	JOBSAT	SALSAT	
JOBSAT	1.0000	.4818**	
SALSAT	.4818**	1.0000	
* - Signif. LE .05 ** - Signif. LE .01 (2-tailed)			

图 12.2 例 12-1 偏相关分析计算结果示意

3. Distances 过程

即所谓的“距离分析”，该过程主要用于对不同记录或不同变量间的不相似性（dissimilarity）即距离（distance）或相似性（similarity）进行分析，并通过一个接近矩阵（proximity matrix）显示出来。它通常不是一个独立的过程，而是给出关联量度或距离量度以便进行进一步的高级统计分析，如因素分析、聚类分析、多维量表分析等，这里不再赘述。

（二）相关分析应用

作为一种统计分析技术，相关系数不仅可直接用于研究变量间的关系，而且在测量领域特别是量表以及题项的评价也非常有用。第四章中提到测量的信度

和效度评定中常会用到相关系数。 SPSS 的 Analyze⇒Scale 菜单中有一个专门的信度分析 (reliability analysis) 过程，举一例说明信度分析的操作和结果解释。

例 12 -2 分析中学生应对方式量表 (黄希庭等， 2000) “ 问题解决 ” 和 “ 求助 ” 两个分量表的信度。

解答：数据中只列出了两个分量表所对应的各题项的序号、含义，容量 300 的一个虚拟样本。调用 Analyze ⇒Scale⇒ reliability analysis 先选择“ 问题解决 ” 分量表对应的 8 个题项 (t2、t4、t13、t15、t20、t26、t29、t35) 进入分析题项 (item) 框中 选择分析模型 Alpha(cronbach's α 信度系数) 并打开 Statistics 统计按钮 (其界面如图 12.3)，其中提供了题项和分量表的描述及汇总统计以及真分数理论假设考察的方法，可为分析量别信度和题项筛选提供依据。在 Descriptives for 描述统计量复选组中勾中最常用的是 “ 删除项目后的量表统计 ” Scale if item deleted 在“ 量表内题项间 ”Inter-item 复选组中勾中相关项。

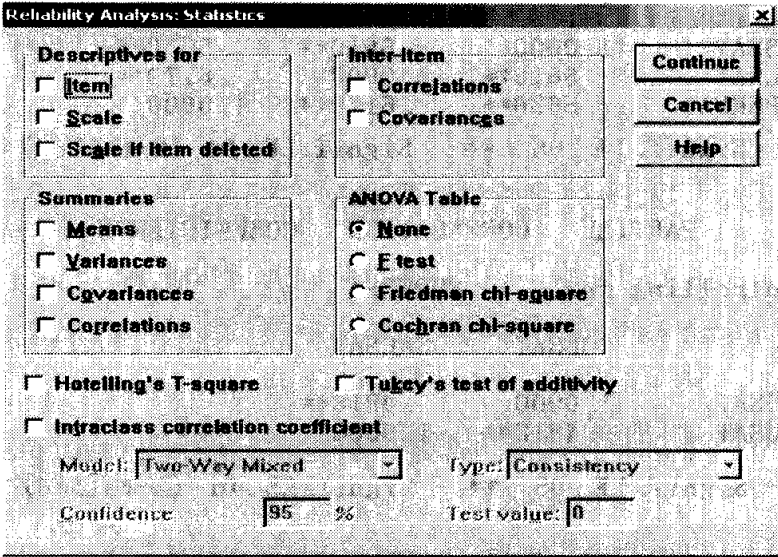


图 12.3 信度分析统计按钮界面

运行结果首先给出各题项间相关矩阵见表 12 - 1 按照信度理论 量表内题项间不应当存在高相关，表中某些题项间相关系数超过 0.5，可对这些题项仔细考察；接着对 “ 问题解决 ” 量表中各题项与量表总分进行了描述统计。① ~ 分别给出了删除该题项后量表的均值、方差、与量表总分间相关系数及 Alpha 系数，其中与量表总分间相关低表明该题项未达到分量表测量目的，可考虑删除；Alpha 系数若突然变大表明该题项区分性差，删除之可增大信度。最后对 300 个记录进行信度分析的结果，该量表的 Alpha 信度系数为 0.8102。

表 12-1 例 12-2 数据信度分析结果

	Correlation Matrix							
	T02	T04	T13	T15	T20	T26	T29	T35
T02	1.0000							
T04	.1999	1.0000						
T13	.2729	.2921	1.0000					
T15	.2490	.3011	.4309	1.0000				
T20	.3336	.2623	.4547	.3722	1.0000			
T26	.2432	.2405	.3277	.3780	.2564	1.0000		
T29	.3035	.3482	.5187	.4672	.4938	.3894	1.0000	
T35	.3151	.3777	.3301	.3967	.3590	.4295	.4421	1.0000

RELIABILITY ANALYSIS - SCALE (ALPHA)				
Item-total Statistics				
	Scale Mean ① if Item Deleted	Scale Variance ② if Item Deleted	Corrected Item-Total Correlation ③	Alpha ④ if Item Deleted
T02	23.6800	25.6899	.4031	.8065
T04	22.7033	26.0288	.4236	.8024
T13	23.0100	24.5584	.5710	.7823
T15	23.0233	24.4777	.5609	.7835
T20	23.6500	23.8737	.5451	.7857
T26	23.2633	24.6094	.4799	.7959
T29	22.9800	23.4043	.6521	.7694
T35	22.8867	24.6426	.5751	.7819
Reliability Coefficients				
N of Cases = 300.0			N of Items = 8	
Alpha = .8102				

除了 Cronbach's α 系数外, Scale 过程还提供了分半 (split-half) 信度、Guttman 信度、平行 (parallel) 信度和严格平行 (strict parallel) 信度等模型, 其中 Guttman 信度用于计算信度系数的区间估计下界, 后两种模型则采用极大似然法直接估计信度系数, 是对真分数模型的直接应用。

除了 Scale 中提供的这些信度指标外, 常用的其他信度指标还有评分者信度, Scale 的 Statistics 子对话框的 ANOVA Table 单选组实际上提供了不同被试 (记录行) 评分间的评分者信度, 根据评分的量尺, 可以选择适用于连续等距量尺的 F 检验 或非参数的 Friedman χ^2 检验、Cochran χ^2 检验 (后两种参见第三节多组比较)。若各次度量检验显著性水平不超过 0.05, 表明各评分者间评价不存在差异, 评分者信度可靠 (此例中评分者过多, 结果不具实际意义)。此外, 对于等级量尺评分, 读者比较熟悉的是肯德尔 W 系数 (和谐系数, Kendall's coefficient of concordance W), 这是一种计算多列数据间相关性的相关系数, 在选择上述 Friedman χ^2 检验时, 结果会同时计算该系数, 非参数检验的 K Related Samples 过程中 (参见第三节) 也提供了该系数的计算选项。该子对话框最下方提供的组内相关系数 (Intraclass correlation coefficient, ICC) 则提供了利用概化理论进行信度评价的三种模型, 已经不属于本书讨论的范畴, 有兴趣的读者可参阅有关

文献。

此外，另两个重要的信度指标重测信度和复本信度可利用相关分析的 Bivariate 过程计算两次重复测量或两个复本间的相关系数。直接利用 Correlate 过程还可以用来计算量表或测量的一些效度指标，例如第四章提到的多特质—多方法矩阵，就是利用某量表和其他量表或测量的相关系数来评价区分效度和聚合效度。再如准则效度的计算，也可通过计算某测量和效标的相关来评价。有关内容请读者参考第四章。

二、回归分析及应用

相关分析通常给出一个相关系数来度量变量间的关系程度，其中的变量不分主次，处于同等地位；与之不同，回归分析（regression analysis）更进一步给出变量间关系的数学模型，它通常会区分出自变量和因变量，即用一个数学关系式反映一个或多个自变量对“一个”因变量的预测关系；另一方面，也可利用该数学模型限定自变量的取值从而实现对因变量变异的控制。尽管如此，仍不能简单地将回归分析视为是对变量间因果关系的探讨，因为视研究者的目的和对变量间关系的定义，完全可以将回归分析中自变量和因变量的位置对调。

（一）简单回归分析原理

最简单的回归分析即一个自变量与一个因变量的一元回归分析；若是多个自变量，则称多元回归（multiple regression）分析。根据回归方程函数关系的形式，回归分析又分为直线回归和曲线回归。回归分析中建立回归方程和回归方程的检验两个步骤是最基本的内容，这里只介绍最简单的一元线性回归即简单回归分析的原理。

1. 建立回归方程

在一元线性回归中，自变量 X 和因变量 Y 间有：

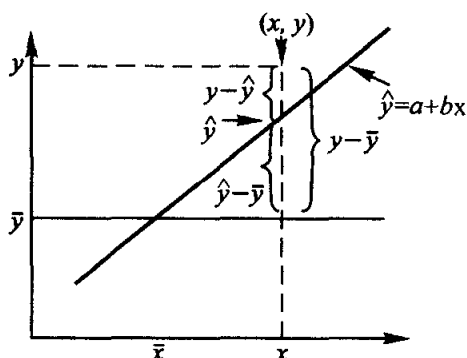
$$y_i = \alpha + \beta x_i + \varepsilon_i \quad (i = 1, 2, \dots, n) \quad (12-2)$$

$$\hat{y} = a + bx \quad (12-3)$$

(12-2) 式的线性回归模型实际上包括 (12-3) 式的回归方程 (regression equation)， Y 的观测值分为估计值 \hat{y} 和 ε_i 两部分， ε_i 为相互独立且都服从 $N(0, \sigma^2)$ 的随机误差称残差。 α 和 β 分别表示回归方程中的常量和回归系数 (coefficient of regression)，分别代表回归直线的截距和斜率； a 是 α 的估计值， b 是 β 的估计值。

通常，建立回归方程就是利用最小二乘 (least square) 原理估计出回归方程中的各项参数，其思路是假设用回归直线去尽可能逼近所有的散点，则最优拟合直线应当使得所有各点距该直线的离差平方和 (想想离差平方和的意义) 最小。如下图所示，因变量的总离差平方和可分解为回归平方和和残差平方和两部

分即：



$$\begin{aligned}\sum (y - \bar{y})^2 &= \sum [(\hat{y} - \bar{y}) + (y - \hat{y})]^2 \\ &= \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2 + 2 \sum (\hat{y} - \bar{y})(y - \hat{y}) \\ &= \sum (\hat{y} - \bar{y})^2 + \sum (y - \hat{y})^2\end{aligned}\quad (12-4)$$

显然 和点 (x, y) 对应的最优拟合直线上的点即为 (x, \hat{y}) 各点距最优拟合直线离差平方和最小即残差平方和 $\sum (y - \hat{y})^2$ 达到最小，则重新记 (12-4) 为：

$$SS_y = SS_R + SS_{resi} \quad (12-5)$$

数学上通过求偏导解得系数 a 和 b ：

$$\begin{cases} b = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2} = \frac{\sum xy - (\sum x)(\sum y)/n}{\sum x^2 - (\sum x)^2/n} = \frac{S_{xy}}{S_x^2} \end{cases} \quad (12-6)$$

$$\begin{cases} a = \bar{y} - b\bar{x} \end{cases} \quad (12-7)$$

2. 回归方程和回归系数的显著性检验

根据上述的最小二乘法总可以由 n 对观测值得到一个回归方程 $\hat{y} = a + bx$ ，不过这时它反映的 X 和 Y 间的线性关系可能为假，需要首先对该方程的显著性进行检验。由 (12-4) 式知道，因变量的总变异分解为回归平方和和残差平方和两部分，若残差平方和达到最小，则很自然地会期望回归平方和占总变异的比例（即回归方程所解释的方差变异）达到最大。称这个比例为 x 对 y 的决定系数（coefficient of determination）记为 R^2 ，决定系数反映了回归模型拟合度的高低，其平方根称复相关（multiple correlation）系数，即：

$$R^2 = \frac{SS_R}{SS_y} = \frac{\sum (\hat{y} - \bar{y})^2}{\sum (y - \bar{y})^2} \quad (12-8)$$

回归方程显著性检验的零假设是回归方程无效，实质是构造回归方差和残差方差的方差比统计量进行 F 检验。为了构造 F 统计量，两部分平方和还需要比上各自的自由度，即：

$$F = \frac{MS_R}{MS_{resi}} = \frac{SS_R/df_R}{SS_{resi}/df_{resi}} \sim F(df_R, df_{resi}) \quad (12-9)$$

上式中称 MS 为均方 (mean square), 即方差的无偏估计量, 两部分自由度间有:

$$df_y = df_R + df_{resi} \quad (12-10)$$

在一元线性回归分析中, 回归自由度等于自变量的个数, 即 $df_R = 1$; y 的总自由度 $df_y = n - 1$; 残差自由度 $df_{resi} = n - 2$, 检验的结果若分子的回归均方显著地大于残差均方, 则认为回归方程显著或有效; 反之, 则不显著, 表明所建立的回归方程并无实际意义。该结果以方差分析表的形式给出, 计算过程不再详述。

(二) 线性回归分析的 SPSS 操作

SPSS 中分析菜单下的回归 (regression) 实际上包含了很多复杂的过程, 包括多元线性回归模型、岭回归、非线性回归、Logistic 回归、概率单位回归、路径分析模型等。这里只介绍线性回归 (Linear) 分析并说明 Linear 过程的操作特点。对高级回归模型感兴趣的读者可参阅推荐读物。

调用 Analyze = > Regression = > Linear 其界面如图 12.4 所示, 也包括自变量框和因变量框, 其中的自变量允许多个, 且可通过 Block 设定多种自变量组合分别进行分析。

Method 下拉列表提供了 5 种回归方法, 这里选择默认的强迫纳入法 (Enter), 所选自变量全部进入方程。其余四种方法分别为: Stepwise——逐步回归法, 逐个加入显著的变量并剔除不显著的变量, 实际上是向前引入法和向后剔除法的综合; Remove——强迫剔除法, 按设定条件剔除自变量; Backward——向后剔除法, 即自变量由多到少逐个从回归方程中剔除; Forward——向前引入法, 自变量由少到多逐个引入方程。下方的 Selection variable 相当于 Select cases 可以设定规则选择符合条件的记录进行分析, 额外的标签变量和加权变量不再赘述。注意若选择 Enter 之外的其他方法, 应当配合 Options 子对话框设定模型拟合判据 (Stepping Method Criteria), 选择使用 F 概率作为引入变量或剔除变量的判据, 如设定 Entry = 0.05, 表示当某变量的显著性水平 < 0.05 时引入方程; 设定 Removal = 0.10, 表示当某变量显著性水平 > 0.10 时, 剔除出方程。若选择使用 F 值作为判据则同样需要设定引入或剔除变量的 F 检验临界值。默认地, 回归分析在这里设置模型中包括常数项。

该界面中其他三个功能按钮提供了各种复杂的参数估计和模型检验、评价功能, 以下据这些子对话框中选项的功能分别说明。

1. 回归分析的假设条件检验

线性回归分析的基本要求包括:

(1) 自变量和因变量间呈线性趋势; 可通过作变量间的散点图加以检验, 例如例子中牵涉到三个变量, 可调用 Graphs = > Scatter 选择矩阵散点图 Matrix 或 3-D 散点图, 分别考察两自变量和因变量的线性趋势。

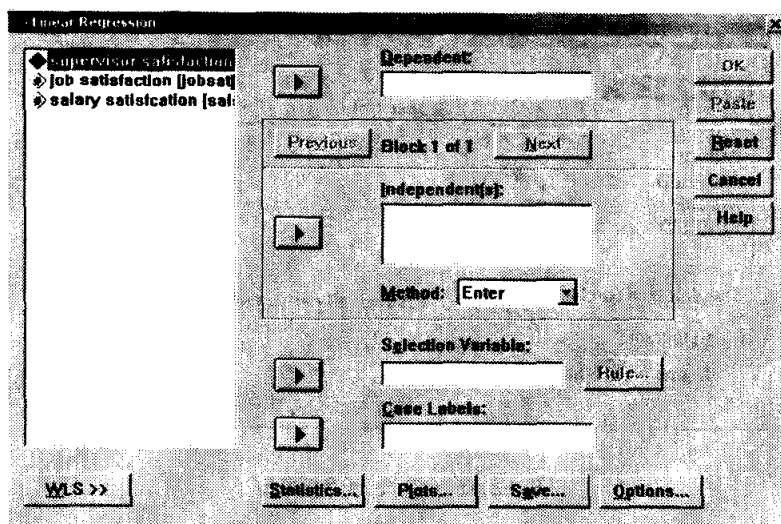


图 12.4 Linear Regression 主对话框

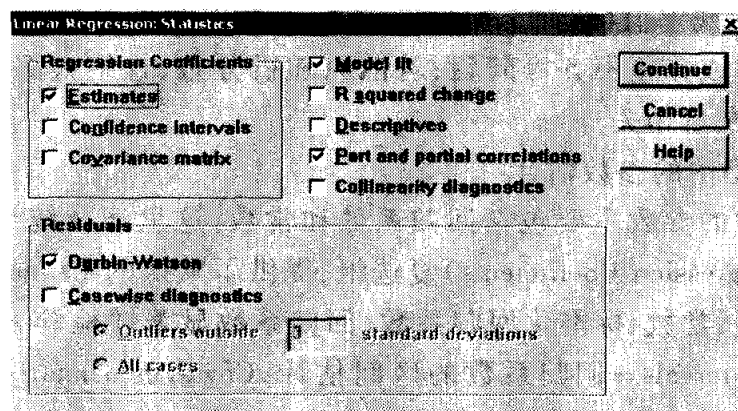


图 12.5 统计子对话框

(2) 残差项 ε_i 服从 $N(0, \sigma^2)$; 该条件的检验可通过 Plots 子对话框中的作图命令实现。如绘制标准化残差图, 可选择作带有正态曲线的残差直方图 (Histogram)、残差累计概率图 P-P 图)、选择输出偏残差图 (partial plots) 等。

(3) 残差相互独立不存在自相关 (autocorrelation); 可采用的检验方法很多, 如统计 (Statistics) 子对话框中 (图 12.5) 的残差分析 (Residuals) 复选组 提供了 Durbin - Watson 残差独立性检验和个案奇异值诊断 (输出指定 n 个标准差外的奇异值), 输出的 D 统计量若在 2 附近表明残差相互独立, $D < 2$ 表示相邻残差正相关; 反之, 表示负相关。

(4) 残差的方差齐性, 指残差的分布是常数, 与自变量和因变量无关。可通过绘制散点图加以检验, 如 Plots 中的 Scatter 组, 可设定按照不同变量组合画各种散点图 (图 12.6), 除因变量预测值外, 其他参数还有: ZPRED—标准化预测

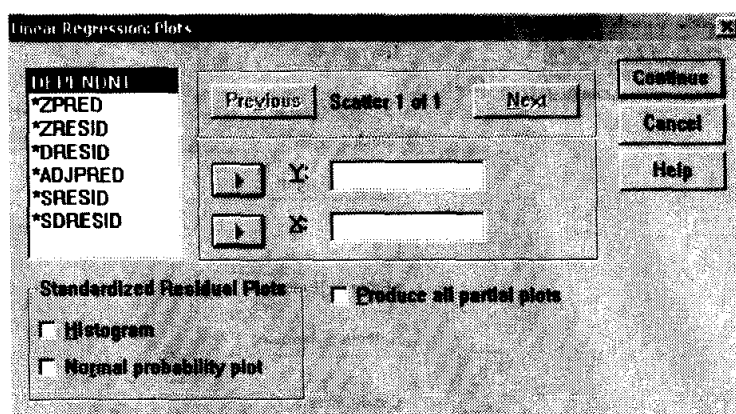


图 12.6 绘图子对话框

值 ;ZRESID—标准化残差； DRESID—剔除残差； ADJPRED—校正后预测值； SRESID—学生化残差； SDRESID—学生化剔除残差。如可绘制因变量预测值（横轴）与学生化残差值（纵轴）散点图，若散点呈现明显的规律性，认为存在自相关性或非线性或方差不齐等问题；若散点随机分布，且绝大部分落在 ± 26 范围内，认为不存在自相关。

2. 回归模型中的参数估计

回归模型中的参数主要包括回归系数和常数，这里涉及统计（Statistics）中的回归系数（Regression Coefficients）复选组 提供了默认的 Estimates—估计回归系数及其标准误，常数、标准化回归系数、回归系数显著性检验 t 值及显著性水平；Confidence intervals—回归系数的区间估计；Covariance matrix—变量相关阵和回归系数协方差阵。

3. 回归模型拟合性检验

模型的拟合性检验是一个复杂的过程，在 Statistics 子对话框中提供了一些检验方法，如模型拟合（Model fit）检验输出进入和剔除出模型的变量，提供复相关系数 R 、决定系数 R^2 、调整 R^2 值等；模型拟合过程中 R^2 、 F 值等的变化（ R squared change）情况（用于决定是否剔除或保留某自变量）；输出符合设定条件记录的描述统计；部分相关（某自变量进入方程后的 R^2 增加量平方根）和各阶偏相关（part and partial correlations）矩阵；共线性诊断（Collinearity diagnostics）表等。此外，上述的残差图示也可以用来判断模型拟合效果，如散点随机分布可认为模型拟合效果好。

实际上回归分析所包含的过程远不止这些，例如有时需要利用回归方程进行关于确定自变量值因变量总体均值 Y 的估计和单个因变量实测值 Y 的估计，以实现预测。不过在心理学研究中这种预测并无多少意义，限于篇幅不再

介绍。

(三) 例题求解和结果分析

1. 简单回归分析结果分析

对上述例 12 - 1 中的数据采用两种方法进行回归分析。首先采用默认的 Enter 法对“薪水满意度”和“工作满意度”两变量进行回归，不改变其他选项。结果首先给出引入或剔除的变量表，这里只有一个自变量。其次给出模型摘要表 (model summary) 即上述 Model fit 所进行的模型拟合性检验 (如表 12 - 2)，可看出复相关系数 (这里等于简单相关系数，等于偏相关系数)、决定系数和调整后决定系数分别为 0.661、0.437、0.433，两变量间存在相关，回归方程可以解释因变量总变异中的 43.7%。估计标准误 (即公式 (12 - 9) 中的残差均方 MS_{resi} 的平方根) 1.009。

表 12 - 2 模型拟合性摘要表

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.661 ^a	.437	.433	1.009

a. Predictors: (Constant), salary satisfaction.

表 12 - 3 给出了回归方程显著性检验的方差分析表，可看出“薪水满意度”对“工作满意度”的回归分析中，回归自由度和总自由度分别为 1、149，残差均方 1.017，F 统计量为 114.81， $p < 0.001$ ，表明回归方程有效。

表 12 - 3 回归方程方差分析表

ANOVA ^b						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	116.786	1	116.786	114.810	.000 ^a
	Residual	150.547	148	1.017		
	Total	267.333	149			

a. Predictors: (Constant), salary satisfaction.

b. Dependent Variable: job satisfaction.

表 12-4 回归分析参数估计和显著性检验

Coefficients ^a					
Model		Unstandardized Coefficients		Standardized Coefficients	t
		B	Std. Error	Beta	
1	(Constant)	.942	.232		4.062
	salary satisfaction	.716	.067	.661	10.715

a. Dependent Variable: job satisfaction.

表 12-4 给出的是非标准化及标准化回归参数及其标准误的估计结果，可以发现，在标准化回归方程中常数项消失，此时的回归系数称标准化回归系数 (standardized coefficient of regression)。据此可以写出回归方程为： $\hat{y} = 0.932 + 0.716x$ 或 $Z_y = 0.661 \cdot Z_x$ 。联系前面的结果和 (12-6) 式可以发现回归系数和相关系数存在这样的关系：

$$b_{y.x} = r \cdot \frac{S_y}{S_x} \tag{12-11}$$

则当原始数据标准化后，一元回归分析中的标准化回归系数正好等于相关系数 r (例子中为 0.661)。表中右半边为各回归参数的显著性检验结果，通常采用 t 检验， t 统计量为：

$$t = \frac{b - \beta}{SE_b} \sim t(n-2), \text{ 其中回归系数标准误 } SE_b = \sqrt{\frac{MS_{resi}}{SS_x}} \tag{12-12}$$

注意到前面的回归方程显著只表明回归方程总体上有意义，但不排除在多元回归中，个别自变量和因变量间其实并不存在显著的线性联系。因此，在多元回归中，除了对回归方程的总体有效性进行检验，还需要对各自变量的显著性进行检验，即回归系数的显著性检验。回归系数显著性检验要考察的零假设为 $H_0: \beta_i = 0$ 。不过对于一元线性回归，可以证明该 t 检验的结果和前面的方差分析结果一致。

2. 多元回归分析结果分析

接下来采用 Stepwise 法对三个满意度变量进行多元回归，仍将“工作满意度”作为因变量，打开 Statistics 按钮，添加选择 Confidence intervals、R squared change、Part and partial correlations、Durbin - Watson 四项输出；打开 Plots 按钮 选择所有复选项并要求作标准化因变量预测值 (X 轴) 与学生化残差值 (Y 轴) 残差散点图。结果如下：

引入或删除变量表 (如表 12-5) 中显示两个变量依次按照设定判据被引入模型。

表 12-5 变量引入 / 剔除表
Variables Entered/Removed^a

Model	Variables Entered	Variables Removed	Method
1	salary satisfcation	.	Stepwise(Criteria: Probability - of - F - to - enter <= .050 , Probability - of - F - to - remove >= .100).
2	supervisor satisfaction	.	Stepwise(Criteria: Probability - of - F - to - enter <= .050 , Probability - of - F - to - remove >= .100).

a. Dependent Variable: job satisfaction.

模型摘要表的内容相比更加丰富（如表 12-6），除复相关系数、决定系数和调整后决定系数外，由于逐步回归法依次引入两个自变量，由单变量变化到多变量模型（见表下的备注）引起的模型拟合指标变化也很有意义。从结果可看出，引入第二个变量对模型整体的变异贡献影响不大（ R^2 由 0.437 只增加了 0.039）；两变量逐步引入的 F 检验结果也无大的变化。注意，Durbin - Watson 检验 D 值靠近 2 显示残差独立性可保证。这里，多变量模型的复相关系数为 0.69，复相关系数或多重相关系数不同于偏相关系数，它度量的是多个变量（此处为两个自变量）的联合效应和某一个变量（此处为因变量）间的相关，反映了变量集和某个变量间的线性关联程度。其计算公式为：

$$r^2_{x,yz} = \frac{r^2_{xy} + r^2_{xz} - 2r_{xy}r_{xz}r_{yz}}{1 - r^2_{yz}} \quad (12-13)$$

这里可将其理解为因变量 y 和其估计值 \hat{y} 间的简单相关，其显著性检验等价于回归方程的显著性检验。

表 12-6 模型拟合性摘要表
Model Summary^c

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Change Statistics					Durbin – Watson
					R Square					
					Change	F Change	df1	df2	Sig. F Change	
1	.661 ^a	.437	.433	1.009	.437	114.810	1	148	.000	
2	.690 ^b	.475	.468	.977	.039	10.830	1	147	.001	1.992

a. Predictors: (Constant), salary satisfcation.

b. Predictors: (Constant), salary satisfcation, supervisor satisfaction.

c. Dependent Variable: job satisfaction.

方差分析表（图略）仍显示多元模型中回归方程总体上显著， F 值为 66.633, $p < 0.001$ 。表 12-7 左半部分的回归系数显著性检验表明两回归系数均显著（请读者尝试写出回归方程），中间给出回归系数的置信区间。

表 12-7 回归分析参数估计、显著性检验及相关

Coefficients ^a												
		Unstandardized	Standardized	95% Confidence Interval for B							Correlations	
		Coefficients	Coefficients									
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound	Zero-order	Partial	Part	
1	(Constant)	.942	.232		4.062	.000	.484	1.401				
	salary satisfaction	.716	.067	.661	10.71	.000	.584	.848	.661	.661	.661	
2	(Constant)	.547	.255		2.146	.033	.043	1.050				
	salary satisfaction	.549	.082	.506	6.666	.000	.386	.711	.661	.482	.398	
	supervisor satisfaction	.289	.088	.250	3.291	.001	.116	.463	.563	.262	.197	

a. Dependent Variable: job satisfaction.

表 12-7 的右半部分还提供了各阶偏相关系数和部分相关 (part correlation) 系数, 从中可以看出各自变量和因变量的线性趋势, 如“薪水满意度”、“上司满意度”与因变量的偏相关系数分别为 0.482、0.262; 部分相关系数分别为 0.398、0.197, 可以看出相比而言薪水满意度对因变量变异的解释力更大。

图 12.7 提供了几种回归分析假设条件的检验图, (a) 是标准化残差直方图, 如前所述可帮助判断残差是否服从正态分布 (注意是样本数据), 这里注意在 0 附近并没有形成峰点; (b) 是标准化残差的正态 P-P 图, 从中可看出散点分布大致吻合正态假设条件下的期望分布 (对角线), 并且在横坐标 0.5 右侧的散点更为密集 (这与直方图照应); (c) 是要求所作的残差散点图, 用于检验方差齐性假设, 可看出残差散点在横向上分布均匀, 满足齐性条件, 此外还可看出并不存在三个标准差之外的奇异值; 最后给出的是偏误差散点图, 可看出各自变量对因变量的边际影响, 可以发现“薪水满意度”和因变量间仍存在明显的线性关系, 但“上司满意度”和因变量间则看不出明显的线性关系 (见 d))。

三、卡方检验和列联表分析

上述相关分析和线性回归主要用来处理等距或等级尺度变量间的关联问题*。在第十章品质相关中, 读者了解到对于诸如性别、职业、态度等名义变量或类别变量, 通常只能得到计数数据, 即变量各类别取值的频数, 此时若要考察该两变量间频数分布间的关联性, 可利用卡方检验 (χ^2 test) 对这种关联的显著性进行检验, 而利用品质相关系数可以度量关联程度的大小。

(一) 卡方检验概述

χ^2 检验的一般问题是利用皮尔逊 χ^2 分布对多个类别中的实际观测频数与理论分布频数的差异进行检验, 它主要用于解决样本次数分布对某个总体分布

作者注: 相关分析和线性回归也可以用来处理涉及名义变量的问题, 例如线性回归中虚拟变量的使用。

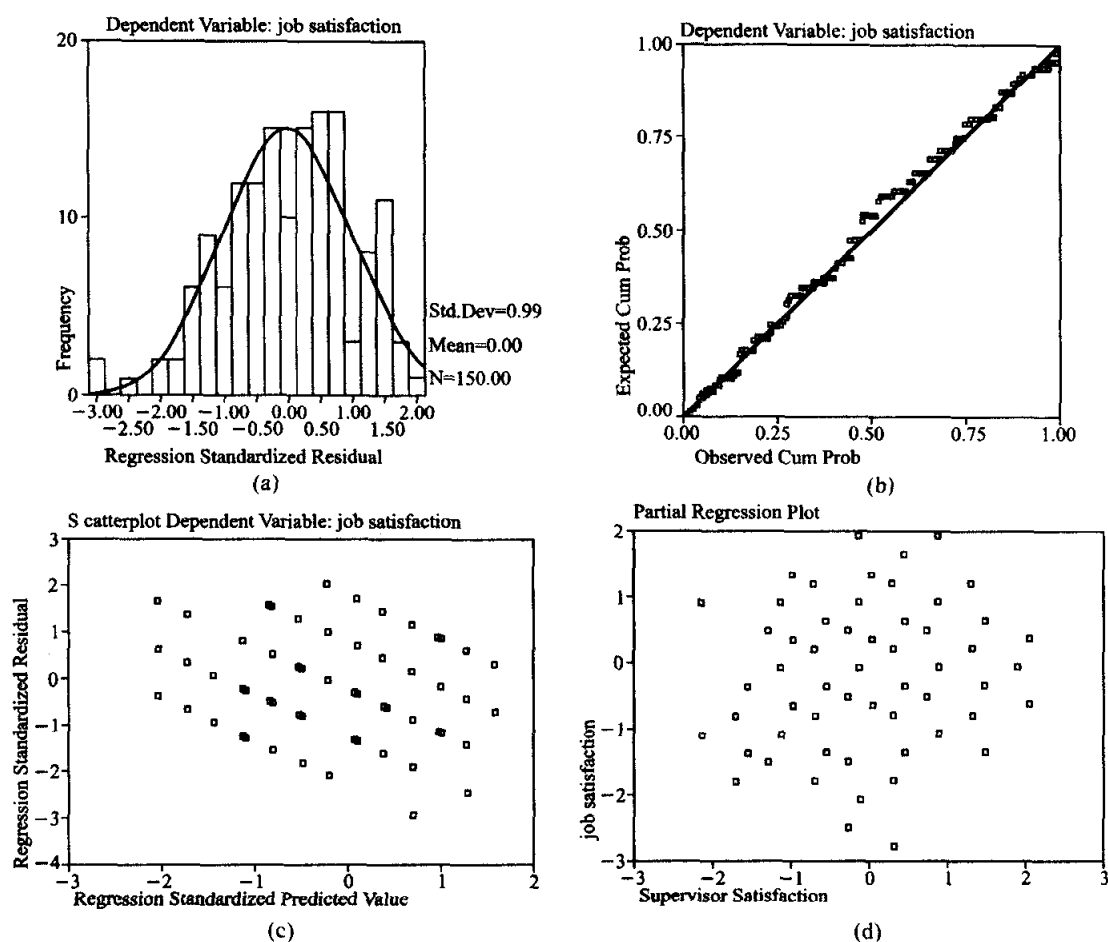


图 12.7 回归残差分析图

或特定分布的拟合性检验 (test of goodness of fit) 和两个或多个分类变量次数分布间的独立性检验 (test of independence) 两类问题。但本质上都是基于皮尔逊 χ^2 检验公式和定理, 即在样本容量足够大时, 所有类别实际观察次数 (f_o) 与理论或期望次数 (f_e) 之差的平方除以理论次数之和所得之统计量近似服从卡方分布 (12-14 式), 然后利用 χ^2 分布检验实计数和理论次数间的差异是否超过一定置信水平下的临界值推断出该差异可归为随机误差或真实差异。

$$\chi^2 = \sum_{i=1}^k \frac{(f_o - f_e)^2}{f_e} \sim \chi^2(k-1) \quad (12-14)$$

χ^2 检验应用于计数数据的分析, 对于总体的分布不作任何假设, 因此也被认为是非参数检验法的一种。由于其自由分布的特性, χ^2 检验是统计学中应用最广泛的推断方法之一。请看下面的研究实例:

- 在拟合性检验中, 若将 f_o 定义为一个分类变量的两项或多项分类的实计数, 将 f_e 的分布定义为均匀分布——即各项分类间次数分布相等, 称这种 χ^2 检验为“无差假设”检验。例如随机调查某样本学生对是否赞成文理分科的态度, 分别得到两类回答的人数, 要比较两类态度间的人数有无显著差异。

• 第十一章提及的二项分布问题和比例的显著性检验问题实际上是上述问题中的特例，即分类变量只有两个类别的情况。例如要比较今年 160/300 (300 人中升学 160 人) 的升学比例和往年 45% 的升学率间有无显著差异。唯一的区别是前两种所使用的指标是概率或比例，而 χ^2 使用的是次数。

• 更一般的假设分布拟合性检验中，根据假设分布（可以是经验性的，也可以是某理论分布）计算出 f_e ，利用公式 12 - 14 比较其与实计数分布间的差异显著性。例如，检验样本中某种疾病的发生率和该疾病发病率间的差异检验；再如，次数分布的正态性检验等。

• 在独立性检验中，检验独立假设条件下推算出的理论次数和实计数之间有无显著差异，推断两个分类变量间的次数分布是否存在关联或一般地说两分类变量间是否存在关联。这种检验通常基于一个 $R \times C$ 的列联表 (contingency table)——即两个变量的交叉次数分布表进行。例如第十章中提到的“性别与考试成绩”、“父母吸烟和子女吸烟关系”两个四格表问题都属于这种情况。

• 从更广泛的意义上说， χ^2 检验还可以考察在控制若干分类变量情况下两个特定分类变量间的独立性，即多维列联表问题，例如在控制专业影响情况下考察性别与毕业是否选择考研间的独立性。

值得注意的是， χ^2 检验作为一种非参数检验法，虽然能够考虑多个分类变量间的关联，但它无法考虑多变量间的交互作用，这时借助对数线性模型 (Log - linear model) 或 Logistic 模型等高等统计方法能够较好地解决这一问题。此外，不论是拟合性检验还是独立性检验，都对样本量和单元格期望次数（通常应 > 5）有所要求。这里，作为推断关联性的方法，只介绍涉及两个变量的 χ^2 独立性检验，即列联表分析及其 SPSS 操作。

（二）SPSS 中的 χ^2 独立性检验

在 SPSS 的 Analyze \Rightarrow Nonparametric tests 中的 Chi - square 过程其实主要用于解决拟合性问题，而 Descriptive Statistics \Rightarrow Crosstabs 才能进行独立性检验。通过第十章品质相关和上面的介绍，读者对独立性检验的原理应当有所了解，下面通过例子说明 Crosstabs 列联表卡方检验的 SPSS 基本操作。

例 12 - 3 对第十章品质相关中“父母吸烟与否与子女是否吸烟”的四格表数据进行 χ^2 独立性检验。

解答：对于这种没有原始数据的既成四格表或 $R \times C$ 表数据，需要经过转换才能调用列联表分析，在 SPSS 数据集中设立两个变量 parent、childs 构造出两变量各水平交叉组合形成的 4 个单元格，并用另一变量 count 列出各单元格对应实计数 数据保存为 data12 - 3。接着很重要的一步是利用记录加权将前两个变量的频数表达出来，方法是：调用 Data \Rightarrow Weight Cases，选中 count 变量进入 Weight cases by 的次数变量框（此过程在例中等价于在数据集中前两个变量下

按照次数依次输入 65 行“ 1—1”、25 行“ 1—2 ”等(1—吸烟 ;2—不吸烟)。

接着调用 Crosstabs 过程，在主对话框（图略，其中选项和子对话框功能请参阅推荐读物）中设定 parent、childs 分别为行、列变量 打开统计量按钮 其中提供了多种“相关量数”或检验法（参见第十章表 10 - 7）。根据各种量数的使用条件，判断例中使用的是独立样本，因此选用适用于独立样本名义尺度的相关度量，这里选择最主要的皮尔逊 χ^2 、 ϕ & Cramér's V 和 Lambda 检验法。打开单元格 Cells 按钮 勾选 Counts⇒Expected 计算期望次数。运行后主要结果如表 12 - 8。表中(1)中斜体加粗部分是四格表的实计数和边缘次数和数据集中一致，直观可看出两变量存在关联；但根据独立性假设推导出的期望次数，任何一行或一列两单元格的次数比都等于对应的边缘次数之比（如 $45/35 = 45/35 = 90/70$ ， $45/45 = 35/35 = 80/80$ ），反映了极端的独立情形。但实计数的这种差异是否达到显著水平呢？表 12 - 8(2)、(3)、(4) 分别提供了 χ^2 、Lambda 和 Phi & Cramér's V 检验结果：(2)给出了 χ^2 及其他一些默认统计量和显著性检验结果，其结论大致相同，都表明两变量存在极显著关联（注意若单元格期望次数过小，应当参考 Yates 连续性校正 χ^2 统计量及其检验结果）；(3)从变量预测的角度也验证了两变量间的关联，并提供了大致的关联程度，(4)也有类似结论。

表 12 - 8 例 3 χ^2 独立性检验输出结果

PARENT * CHILDS Crosstabulation					Chi-Square Tests				
		CHILDS			Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
		吸烟	不吸烟	Total					
PARENT 吸烟	Count	65	25	90	Pearson Chi-Square	40.63 ^b	1	.000	
	Expected Count	45.0	45.0	90.0	Continuity Correction ^a	38.63	1	.000	
不吸烟	Count	15	55	70	Likelihood Ratio	42.71	1	.000	
	Expected Count	35.0	35.0	70.0	Fisher's Exact Test			.000	.000
Total	Count	80	80	160	Linear-by-Linear Association	40.38	1	.000	
	Expected Count	80.0	80.0	160.0	N of Valid Cases	160			

a. Computed only for a 2x2 table.

b. 0 cells(.0%) have expected count less than 5. The minimum expected count is 35.00.

① 列联表分析数据构造示意

② χ^2 检验结果

Directional Measures						Symmetric Measures			
			Asymp.	Approx.	Approx.				Approx.
		Value	Std. Error ^a	T ^b	Sig.			Value	Sig.
Nominal by Lambda	Symmetric	.467	.081	4.622	.000	Nominal by	Phi	.504	.000
Nominal	PARENT Dependent	.429	.097	3.479	.001				
	CHILDS Dependent	.500	.074	5.164	.000	Nominal	Cramer's V	.504	.000
Goodman and	PARENT Dependent	.254	.068		.000 ^c				
Kruskal tau	CHILDS Dependent	.254	.068		.000 ^c	N of Valid Cases		160	

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

c. Based on chi-square approximation.

a. Not assuming the null hypothesis.

b. Using the asymptotic standard error assuming the null hypothesis.

③ Lambda 检验结果

④ ϕ & V 检验结果

第二节 组间比较的统计分析

组间比较的对象是研究者收集的两组 / 多组观测指标，即因变量，比较的目的实际上是要判断某个分组变量（自变量）对观测指标是否有显著的“效应”或作用，即同一个因变量在自变量的不同分组上是否有显著差异。这种统计思路不同于第一节中的关系推断，不过通常这仍然不能视为是在对自变量和因变量间的“因果关系”进行推断，除非是在非常严格的真实验设计中。本节结合各种研究设计实例介绍单自变量两组比较和多组比较的统计分析，其对应的 SPSS 功能界面参见前面有关章节。

一、两组比较

严格来说，“组间比较”只是一个习惯性的称呼，它和十一章中提到的均值比较有很多类似，不过这里谈的组间比较范围更宽泛一些，它不仅指对“均值”参数的假设检验，也包括对组间的其他指标或度量的比较，比如非参数检验中对中数、秩等指标甚至计数数据的差异显著性检验。组间比较中的所谓“组”所指非常宽泛，既可能是研究设计中的实验组、控制组，也可能是某个自变量的多个分组（即水平），或者前测组和后测组等。两组比较涉及到的研究设计类别非常多，如真实验设计中的单因素两水平也是最常见的。以下结合表 11-3 的统计方法分类表，从数据或变量特性角度分别举例介绍。

（一）等距型数据的两组比较法

像 u 检验和 t 检验这样的参数检验法对于数据的分布形态和量尺有较高的要求（如数据总体服从正态，总体方差已知或未知；数据为连续等距尺度以上等），等距测度是最基本的要求。SPSS 中对等距变量进行两组比较的方法主要包括在 Analyze \Rightarrow Compare Means 过程中：

1. 独立样本 t 检验

两组均值的差异显著性检验中最特殊的情形是两母总体的方差都已知时采用正态 u 检验，只需直接计算 Z 值查表即可，不再赘述。通常情况下，总体方差总是未知 应采用 t 检验。看下面的例子：

例 12-4 在探寻观看暴力电视对儿童攻击行为影响的实验中，将 50 名攻击行为“同质”（前测无差异）的儿童随机分成两组，实验组 25 名儿童观看一段时间暴力动画片，同时，控制组儿童观看非暴力动画片，事后记录在某时间段内观察到的两组儿童攻击行为发生的次数。完成数据分析并得出结论。

解答：这是一个完全随机化设计（参见第五章），实验只有实验组和控制组

两组，因此，采用独立样本的 t 检验完成统计。调用 Analyze \Rightarrow Compare Means \Rightarrow Independent Samples T Test 将变量“攻击行为次数”和“组别”分别拖入检验变量框和分组变量框（并设置两组为 0 和 1，即控制组和暴力影片组）。执行检验后得到结果，先给出分组描述统计，接着给出独立样本 t 检验表如表 12-9 所示。表中内容按列分成三部分，第一部分给出 Levene's 方差齐性检验（亦可用第十章中提及的有关检验方法加以检验）结果，可看出显著性水平 sig. 远大于 0.05，表明两组方差整齐。第二部分给出 t 检验结果，注意表中按照方差齐性检验的两种结果分行给出检验结果，这里应当看上面一行的“方差相等”的结果。包括 t 统计量、自由度、显著性水平（这里默认用双侧检验）以及 t 统计量的均值、标准误等中间结果；显著性水平表明检验极其显著，实验组、控制组攻击行为差异明显。最后一部分给出的是均值差的置信区间（通常为 95% 置信水平）的上下限。

注意，该题可能更适合用单侧检验，即假设实验组的攻击行为比控制组更多。不过由于 SPSS 直接给出了统计量对应的尾端概率（sig.），因此只需要看这个概率是否小于既定的显著性水平即可。若沿用这里的双尾概率，则在单侧检验中，该概率只需小于 0.10（双侧小于 0.05）即可得到显著差异结论。

2. 相关样本 t 检验

表12-9 独立样本 t 检验 SPSS 输出结果

Independent Samples Test									
Levene's Test for Equality of Variances					t - test for Equality of Means				
					95% confidence Interval of the Difference				
					Sig. (2 - tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal variances									
攻击 行为 次 数	assumed	0.030	.863	-4.99	48	.000	-4.04	.810	-5.669 -2.411
	Equal variances								
	not assumed			-4.9947	937	.000	-4.04	.810	-5.669 -2.411

对相关样本的统计分析实际上是通过构造配对数据的“差”成为一个新的变量，从而转换成对差值均数和总体均数 0 之间的显著性检验问题。不过这两种过程的操作过程和对话框结构仍有非常大的差异，在实践中初学者往往难以

区分独立样本和相关样本的条件造成分析错误，甚至在数据录入的最初阶段就弄错了数据排列方式。所以，这里需要强调相关样本和独立样本 t 检验统计过程的区别。看下面的例子：

例 12-5 某研究者欲研究镜画作业成绩是否受被试视野中方位线索的影响，令 10 名被试分别先后（顺序平衡）完成两种实验条件下的若干相同难度镜画作业，记录下各自的平均错误次数（设只考虑正确率这一指标）。试完成数据分析。

	催眠药物	入睡时间	var	var	var
1	1	1.90			
2	1	.80			
3	1	1.10			
4	1	1.60			
5	1	4.60			
6	2	.70			
7	2	-1.60			
8	2	-.20			
9	2	-1.20			

	before	after	var	var	var
1	70.00	48.00			
2	76.00	54.00			
3	56.00	60.00			
4	63.00	64.00			
5	63.00	48.00			
6	56.00	55.00			
7	58.00	54.00			
8	60.00	45.00			
9	65.00	51.00			

图 12.8 两独立样本（上）和两相关样本（下）数据集格式

解答：这是一个典型的单因素被试内设计，或者也可理解为最特殊的一种随

机区组设计（即一个被试为一个区组），应采用相关样本 t 检验。分析之前，先看一下这种被试内设计在数据集格式上和被试间设计的差异。第十一章提到相关样本 t 检验要求往 Paired Variable 框中输入要进行比较的“成对”变量（如图 11.13、图 12.8）。这就要求，配对的两个变量必须是两个“不同”的变量（列），这时要比较的两组数据在数据集形式上才是真正的“两个”变量。这和独立样本 t 检验中两组数据本质上为一列而必须通过一个外部变量的不同水平加以区分是不同的。这种特性是由被试内设计的重复测量特性所决定的，可以这样认为，独立样本 t 检验中要区分分组变量（自变量）和指标，而相关样本 t 检验中要区分不同处理下的指标，自变量“隐藏”在各个列中。请读者比较图 12.8 中的两种情况加以体会。总之，在 SPSS 中一个变量的数据可以为一系列、也可以为多列（视变量的含义），而同一个被试的数据则一定要录入为同一个记录（行）。

调用 Compare Means \Rightarrow Paired Samples T Test 将“有 / 无方位线索”两个变量选中后拖入检验变量框，执行检验。结果除了分组描述统计外还进行了两变量间的相关分析（显著相关），相关样本 t 检验如表 12-10 所示。表中内容和独立样本类似，对配对数据差（paired differences）的检验表明，显著性水平 sig. 约为 $0.057 > 0.05$ ，所以检验结果差异不显著，即有无方位线索对被试作业成绩无显著的影响。不过从结果看，检验的显著性水平非常接近显著，对这种边缘显著的结果下结论时应当小心。

表 12-10 相关样本 t 检验输出结果
Paired Samples Test

		Paired Differences							
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference		<i>t</i>	df	Sig. (2 – tailed)
					Lower	Upper			
Pair	有方位线索—								
1	无方位线索	1.10	1.59513	.50442	– .0411	2.2411	2.18	9	.057

☆ SPSS 操作提示

选择独立样本 t 检验或者相关样本 t 检验关键是看研究设计中是否将观测数据一一配对。若是，则是后者，数据录入 SPSS 中时没有单独成列的自变量，自变量的两个水平隐含在两列变量中，这个规则对于多个配对组或多次重复测量也适用（例如在重复测量设计中）。这时在数据分析时很重要的工作就是要正确“配对”要比较的两个处理组。相反，若是前者，则因变量和自变量必须各自

成列，这时很重要的工作是正确“定义”要比较的两个处理组。

（二）等级型数据的两组比较法

等级型数据的均值或方差并无实际意义，因此该类数据的组间比较主要牵涉到非参数检验中关于中数、秩这样的指标的推断统计；对于数据总体分布不明（非正态）以及小样本的情况也可以使用非参数检验方法。就这类方法对应的研究设计而言，和上面所讲的参数检验的组间比较方法有很大联系；但从非参数检验相比参数检验的特点看，其对信息的利用不够，精细和准确程度也不如参数检验法，因此，这类方法更多地见于非实验设计或一些预备性的探索研究中。该类检验在 SPSS 中主要包含在 Analyze 菜单的 Nonparametric Tests 菜单中。

1. 两独立样本非参数检验

例 12-6 某研究者假设师范专业学生对课堂教学的评价会和非师范专业学生存在差异。选用两个专业学生共同观摩一次课堂教学并进行评价，满分 30 分。试进行统计分析。

解答：该研究设计可视为一种非实验设计，因为两组被试的选择并没有经过严格的随机化控制。该问题可采用 Nonparametric Tests 中的两独立样本 (2 independent samples) 检验或多独立样本 (k independent samples) 中的克一瓦氏 H 检验及中数检验进行统计检验。以前者为例，打开其主对话框后和上述独立样本 t 检验一样，设置分组变量和检验变量分别为“组别”和“评价”，其中，组别定义为数据集中出现的 1、2。下方的方法复选框提供了四种检验法：曼-惠特尼 U (Mann-Whitney U) 检验、克尔莫哥洛夫-斯米尔洛夫 Z (Kolmogorov-Smirnov Z) 检验、Moses 极值反应 (Moses extreme reactions test) 检验、沃德-沃尔福威茨游程 (Wald-Wolfowitz runs) 检验，选择前两种方法，分析后分别给出各方法的描述统计和推断检验（如表 12-11）。其中，最常用的曼-惠特尼 U 检验主要是给出分组秩统计量，两组的平均秩分别为 5.79 和 10.61，秩和 (sum of ranks) 分别为 40.5 和 95.5。同时还给出等价的 Wilcoxon 秩和检验、常用的 u 检验结果，显著性水平表明边缘显著 ()；这里的 Kolmogorov-Smirnov Z 检验 () 是单样本方法的推广，它将两个样本累计分布曲线进行比较寻找整体的差异，检验结果不能拒绝无差异假设 (sig. ≈ 0.084)。注意，从两种方法利用的信息看，Mann-Whitney U 的结果在该例中相对更有效。

2. 两相关样本非参数检验

例 12-7 某心理学家欲考察在办公室工作环境中播放流行音乐能否提高员工的工作满意度，他选择了一组员工，令其在两种（有无音乐）工作环境下分别工作一周，接着收集了他们的工作满意度评价（满分 100），试进行统计分析。

表 12 - 11 独立样本非参数检验输出结果

Test Statistics ^b		Test Statistics ^a		
	评价			评价
Mann-Whitney U	12.500	Most Extreme	Absolute	.635
Wilcoxon W	40.500	Differences	Positive	.000
Z	-2.013		Negative	-.635
Asymp. Sig. (2-tailed)	.044	Kolmogorov-Smirnov Z		1.260
Exact Sig. [2 * (1 - tailed Sig.)]	.042 ^a	Asymp. Sig. (2-tailed)		.084

a. Not corrected for ties.

a. Grouping Variable 组别.

b. Grouping Variable 组别.

① Mann - Whitney U 检验结果

② Kolmogorov - Smirnov Z 检验结果

解答：这也可看成是一个被试内设计。由于被试的评价分布不明，考虑采用非参数检验法 可调用 Nonparametric Tests 中的两相关样本(2 related samples) 检验或多相关样本(k related samples) 中的弗里德曼检验及肯德尔 W 检验。以前者为例，其主对话框类似于相关样本 t 检验，将两个变量配对选入检验变量框，选中下方方法复选框中的威尔考克森符号等级检验 (Wilcoxon signed ranks test)、符号检验 (Sign test)，分析仍先给出两列数据的描述统计。Wilcoxon 检验同时考虑了两组数据的相对大小和差数的等级并给出检验结果 (如表 12 - 12①、②)，检验结果近似 Z 值为 -0.939, P 近似为 0.348，表明差异不显著。符号检验仅利用了差数符号信息，从 中可看出正、负差数几乎相等，检验结果 (如表 12 - 12 精确概率近似 1，没有差异。方法复选组中的麦克纳马 (McNemar) 检验就是上节提到的 χ^2 独立性检验中的相关样本独立性检验法，但它只适用于两个二分类变量，重点考察两配对组间分类的差异；另一种 Marginal Homogeneity 检验则是 McNemar 法的扩展，适用于等级型多项分类资料。

表 12 - 12 相关样本非参数检验输出结果

		Ranks		
		N	Mean rank	Sum of Ranks
有音乐 - 无音乐	Negative Ranks	8 ^a	9.56	76.50
	Positive Ranks	7 ^b	6.21	43.50
	Ties	0 ^c		
	Total	15		

a. 有音乐 < 无音乐 .

b. 有音乐 > 无音乐 .

.. 无音乐 = 有音乐 .

Test Statistics	
	有音乐 - 无音乐
Z	-.939 ^a
Asymp. Sig. (2 - tailed)	.348

a. Based on positive ranks.

b. Wilcoxon Signed Ranks Test.

②

二、多组比较

前已述及,多组均值的差异比较不能用多次 t 检验而应使用方差分析,按照数据的测量尺度和分部形态,对于等距型数据应使用单因素方差分析,对于等级或名义型数据应使用非参数检验法。两组比较法通常不能用于多组比较,但反过来多组比较法几乎都能用于两组比较,此外,多组比较也应区分独立样本和相关样本,方便考虑,对方差分析只介绍独立样本多组比较法,而相关样本(重复测量)的多组比较则在第三节重复测量方差分析中集中介绍。

(一) 等距型数据的单因素方差分析(独立样本)

在参数检验中,独立样本多组比较的单因素方差分析可通过 Compare Means \Rightarrow One - way ANOVA 或 GLM \Rightarrow Univariate 过程实现,其对话框和操作界面和有关功能的一般介绍参见第十一章。注意, One - way ANOVA 过程只处理单因素完全随机化设计数据,对于区组设计或拉丁方设计则应使用 Univariate 过程。看这样两个例子。

1. 单因素完全随机化设计方差分析

例 12 - 8 若将上述例 12 - 4 中的研究加以拓展,成为一个单因素实验组控制多组后测设计(参见第五章),如设立三个实验组,分别给予暴力动画片刺激、中性动画片刺激和助人动画片刺激,又应如何分析?

解答:这是一个典型的单因素完全随机化设计,数据中分组变量增加了两个水平,虽然这里已经没有设立控制组的必要,但一般仍然将其作为一个水平。调用 One - Way ANOVA 过程,仍将因变量和分组变量分别纳入相应位置。单向方差分析提供了三个功能选项。 Contrast 子对话框暂且不用设置,打开多重比较子对话框,选择事后检验(post hoc)方法,这里在方差整齐条件下选择最常用的 LSD 法和 S - N - K 法;方差不齐条件下选择 Dunnetts' s C 法。选择 Options 子对话框中的描述统计、方差齐性检验及作均值图(means plot)三个复选项。

Test Statistics	
	有音乐 - 无音乐
Exact Sig. (2-tai	1.000 ^a

a. Binomial distribution us.

b. Sign Test.

③

输出结果首先给出各组在因变量上的主要描述统计，方差齐性检验（图略），Levene 统计量为 7.088, $P < 0.001$ ，表明方差不齐。接着给出了一个 ANOVA 表（如表 12-13），这是方差分析的主要结果，从中可看出单因素完全随机化设计的变异分解形式：即将因变量的总变异分解为由处理差异造成的组间（between groups）变异和误差（包括未加控制的无关变量造成的系统误差）对应的组内（within groups）变异两部分。这里由于只对单因素的主效应进行检验，因此只有一个 F 统计量（ $F = 34.948$ ），对应的显著性水平 $\text{sig.} < 0.001$ （双击可察看具体值），即该主效应显著，表明各组在攻击行为上差异显著。表中的均方（mean square）是各变异的方差估计量，由对应行的离差平方和（sum of squares）比上自由度 df 得来（其中组间自由度即为组数 - 1，总自由度为样本量 - 1），总平方和和总自由度均分解成对应的两部分。组间均方比上组内均方形成 F 统计量，然后借助 F 分布进行方差差异的显著性检验，其基本思路是若分子的组间变异显著地大于组内（误差）变异，则认为组间效应显著，各组存在差异。

表12-13 单向方差分析表

ANOVA					
攻击行为次数					
	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	572.880	3	190.960	34.948	.000
Within Groups	524.560	96	5.464		
Total	1097.440	99			

上表只表示实验因素的主效应显著，欲找出各实验组的具体差异，还需要进行多重比较。在不考虑方差齐性条件下，表 12-14 和表 12-15 分别显示了两种均值比较方法的结果：S - N - K 法列出了各组均值的“同质子组”（homogeneous subsets），即在一定的显著性水平下，4 个均值被按照大小分成三个子组（即表中 subset 下的三列），各子组间均值差异显著而组内均值差异不显著。可以看到：其中助人影片组均值最小为第一子组，中性影片组和控制组其次，为第二子组，而且两组均值差异不显著；最后是暴力影片组为第三子组。有时可能看到一个均值同时位于两个相邻子组上（例如中性影片组横跨 1、2 两组）则表示该均值和其所在两子组的其他均值差异不显著。表中最下方的显著性水平是子组内均值的差异比较显著性。LSD 法和 Dennett's C 法都采用两两比较的方式，将各组依次组合进行比较，可看出除中性组和控制组无差异外，其余均差异显著，结

论和 S - N - K 法类似。

表 12 - 14 S - N - K 法多重比较结果
攻击行为次数

组别		N	Subset for alpha = .05		
			1	2	3
Student-Newman-Ke	助人影片组	25	5.04		
	中性影片组	25		7.08	
	控制组	25		7.60	
	暴力影片组	25			11.64
Sig.			1.000	.434	1.000

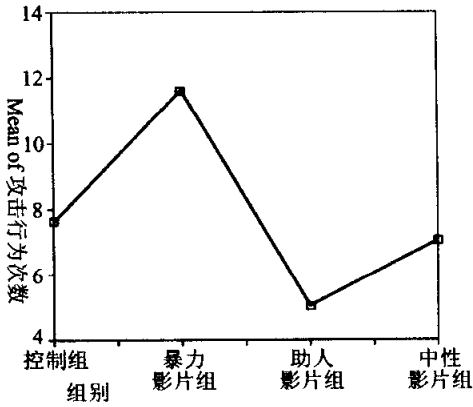


图 12.9 例 12 - 8 均值图

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 25.000.

表 12 - 15 LSD 法和 Dunnett's C 法多重比较结果
Multiple Comparisons

Dependent Variable 攻击行为次数

		Mean		Sig.	95% Confidence Interval		
(I)组别	(J)组别	Difference (I - J)	Std. Error		Lower Bound	Upper Bound	
LSD	控制组	暴力影片组	-4.04 *	.661	.000	-5.35	-2.73
		助人影片组	2.56 *	.661	.000	1.25	3.87
		中性影片组	.52	.661	.434	-.79	1.83
	暴力影片组	控制组	4.04 *	.661	.000	2.73	5.35
		助人影片组	6.60 *	.661	.000	5.29	7.91
		中性影片组	4.56 *	.661	.000	3.25	5.87
	助人影片组	控制组	-2.56 *	.661	.000	-3.87	-1.25
		暴力影片组	-6.60 *	.661	.000	-7.91	-5.29
		中性影片组	-2.04 *	.661	.003	-3.35	-.73
中性影片组	控制组	-.52	.661	.434	-1.83	.79	
	暴力影片组	-4.56 *	.661	.000	-5.87	-3.25	
	助人影片组	2.04 *	.661	.003	.73	3.35	

续表

		Mean		Sig.	95% Confidence Interval	
(I) 组别	(J) 组别	Difference (I - J)	Std. Error		Lower Bound	Upper Bound
Dunnett C	控制组	暴力影片组	-4.04 *	.810	-6.27	-1.81
		助人影片组	2.56 *	.618	.86	4.26
		中性影片组	.52	.719	-1.46	2.50
	暴力影片组	控制组	4.04 *	.810	1.81	6.27
		助人影片组	6.60 *	.598	4.95	8.25
		中性影片组	4.56 *	.702	2.62	6.50
	助人影片组	控制组	-2.56 *	.618	-4.26	-.86
		暴力影片组	-6.60 *	.598	-8.25	-4.95
		中性影片组	-2.04 *	.467	-3.33	-.75
	中性影片组	控制组	-.52	.719	-2.50	1.46
		暴力影片组	-4.56 *	.702	-6.50	-2.62
		助人影片组	2.04 *	.467	.75	3.33

* The mean difference is significant at the .05 level.

当然，该例中方差不齐，理论上应当参照方差不齐的多重比较结果，不过在应用中，若数据分布不是明显偏态，方差齐性对结果的影响通常不大。最后作均值图(如图 12.9)可使得各组间的差异情况一目了然，结合数值分析结果可以清楚地判断各组均值孰大孰小。

2. 单因素拉丁方设计方差分析

例 12-9 研究者探讨文章的生字密度对学生阅读理解的影响，实验的自变量为生字密度，设立了 5:1、10:1、15:1、20:1 四个水平。研究者假设阅读理解成绩会随着生字密度的增加而下降，考虑到来自不同班级的学生成绩可能会有差异，同时测试时间也可能影响成绩，因此将班级和时间两个无关变量都设置成 4 水平，选取 4 个班各 8 名学生，每 2 名学生为一个单元格随机分配到标准拉丁方格中去，测得其阅读理解成绩(数据据舒华，

1994, P52 例)。

解答：拉丁方设计相比区组设计多控制了一个无关变量，从两个方向控制了实验误差，这种方差分析中的变异分解是将完全随机化设计中（不考虑两个无关因素）的组内变异分解成区组变异和残差变异，从而减小了误差变异，提高了 F 检验的功效。因此，从统计控制上比完全随机化设计更有效。

实际上， K 因素拉丁方设计的方差分析可以看成是不考虑两个无关变量和其他因素间交互作用的 $K + 2$ （若是区组设计，则为 $K + 1$ ）因素的方差分析，因此，这里调用 Univariate 过程，设定阅读理解成绩、生字密度、班级和时间分别为因变量和自变量（固定因素）；打开 Model 按钮，选择自定义（Custom），选择只考察三个因素的主效应。运行结果方差分析表见表 12-16，可以发现该表的结构和单向方差分析不同，第一行的 Corrected Model 是对整个方差分析线性模型的检验，其显著性水平 < 0.001 表明模型有意义，有效应显著，第二行的截距是对总均值为 0 的检验，无特别意义。第三行以下是对各效应的检验，其中生字密度和班级主效应显著，时间主效应不显著。可发现该结果的实质等价于一个不考虑交互作用的三因素方差分析，总平方和和总自由度分别分解成三个因素的主效应和误差效应所对应平方和、自由度（例中三个因素的自由度均为 3，总自由度为 31）；此外，三个主效应的 F 统计量中误差均方均为 Error 项，实际上包含了单元格内（被试内）误差和真正的残差，当二者没有差异时这种合并可得到对实验误差的更好估计。最下方还给出了该三因素主效应模型的决定系数，可解释总变异的 92%。

由于主要关注生字密度因素的效应，因此可对该变量进行多重比较，表 12-17 的 S-N-K 法结果表明，除 5:1 和 10:1 两水平无显著差异外均差异显著，基本上验证了研究的预期假设；在 Univariate 主界面的作图子对话框中可设定多因素模型中的均值图作法（图 12.10），由于二维平面均值图纵坐标总是因变量，因此，一个图中最多包含两个因素，其中一个用横坐标（horizontal axis）表示，另一个用不同图例（separate lines）表示；如果再增加一个因素，则只能借助不同的图形（separate plots）表示。这里，可将主效应显著的生字密度、班级分别设置为横坐标、不同图例和不同图形，按 Add 提交（在多因素场合，允许对因素进行组合作图），图形见图 12.11，从图中可直观看各种文章的阅读理解成绩变化趋势，还可看出班级变量各水平的差异情况（也可作多重比较），图中还可看出若考虑生字密度和班级的交互作用也是不显著的（折线平行），虽然该设计已决定了要分析该交互作用必然要“牺牲”一个主效应。

表 12-16 例 12-9 方差分析表

Tests of Between-Subjects Effects

Dependent Variable 阅读理解分数

Source	Type III Sum of Squares	df	Mean Square	F	sig.
Corrected Model	247.625 ^a	9	27.514	28.48	.000
Intercept	1275.125	1	1275.12	1320	.000
生字密度	190.125	3	63.375	65.61	.000
班级	56.125	3	18.708	19.37	.000
时间	1.375	3	.458	.475	.703
Error	21.250	22	.966		
Total	1544.000	32			
Corrected Total	268.875	31			

a. R Squared = .921(Adjusted R Squared = .889)

表 12-17 生字密度多重比较结果

阅读理解分数

Student-Newman-Keuls^{a,b}

生字密度	N	Subset		
		1	2	3
10:1	8	3.88		
5:1	8	4.38		
15:1	8		7.00	
20:1	8			10.0
Sig.		.320	1.00	1.00

Means for groups in homogeneous subsets are displayed. Based on Type III Sum of Squares

The error term is Mean Square(Error) = .966.

a. Uses Harmonic Mean Sample Size = 8.000.

b. Alpha = .05.

(二) 非参数多组比较

若多组数据不满足参数检验的要求（如等级或名义型数据），应当采用非参数的多组比较方法。

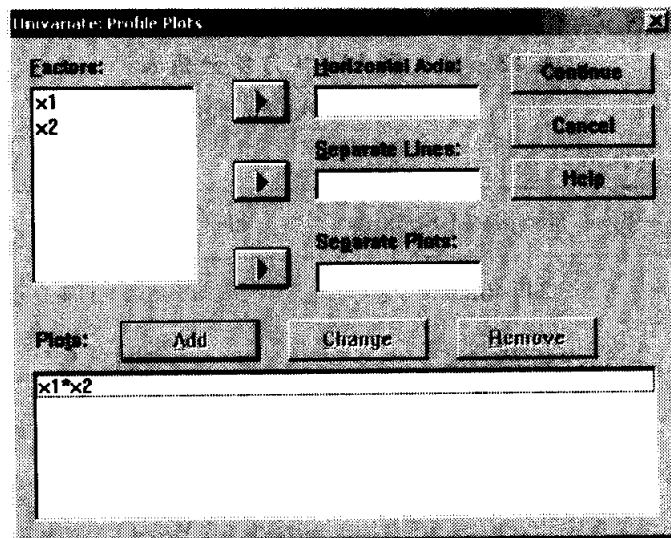


图 12.10 Plots 作图界面示意

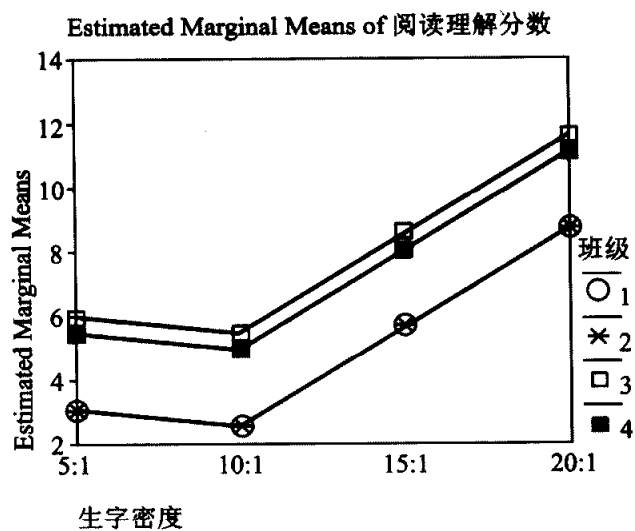


图 12.11 生字密度 × 班级均值图

1. 等级型数据的非参数方差分析

顾名思义，非参数方差分析法在研究设计上其实和单因素方差分析并无区别，只是其主要用于等级型数据或不满足参数检验条件的场合，故亦称等级方差分析 (ANOVA by ranks)。

例 12-10 上述例 12-8 中的数据若不满足参数检验的条件，使用非参数方差分析如何进行？

解答：例子属于单因素完全随机化设计，其对应的非参数检验法为 Nonparametric tests => k independent samples 中的克-瓦氏 Kruskal - Wallis 单向方差

分析和中数 Median)检验 主界面类似于 One - way ANOVA , 选择攻击行为次数和组别分别进入检验变量框和分组变量框 , 并定义该分组变量值的全距最小为 0, 最大为 3 选中两种方法 运行结果分别见表 12 - 18 和表 12 - 19。

表 12 - 18 克—瓦氏单向方差分析结果

Ranks			
	组别	N	Mean Rank
攻击行为次数	控制组	25	50.44
	暴力影片组	25	80.98
	助人影片组	25	23.34
	中性影片组	25	47.24
	Total	100	
Test Statistics ^{a, b}			
攻击行为次数			
Chi-Square		50.413	
df		3	
Asymp. Sig		.000	

a. Kruskal Wallis Test.

b. Grouping Variable 组别 .

表 12 - 19 中数检验结果

Frequencies					
		组别			
		控制组	暴力影片组	助人影片组	中性影片组
攻击行	> Median	12	23	0	12
为次数	<= Median	13	2	25	13
Test Statistics ^b					
攻击行为次数					
N		100			
Median		7.00			
Chi-Square		42.513 ^a			
df		3			
Asymp. Sig.		.000			

a. 0 cells(.0%) have expected frequencies less than 5. The minimum expected cell frequency is 11.8.

b. Grouping Variable 组别 .

克—瓦氏 H 检验首先将 k 个样本的所有数据按大小排秩 , 然后分别求出各

样本的秩和，在平衡设计中若各组秩和相当，表明各样本来自同一总体；结果上方的表显示了各组秩和，下方的 H （克-瓦氏检验的统计量，服从自由度为 $k - 1$ 的 χ^2 分布）检验表明 $P < 0.001$ ，各组差异显著。而中数检验实际上是将各样本中数据按照总数据序列中数分割成 $2 \times k$ 的列联表（上方表），然后直接进行 χ^2 独立性检验，显著表示各样本来自中数不同的总体，检验结果同上。非参数检验的缺陷是无法进行具体差异比较。

例 12 - 11 考察三位选修课教师的教学效果是否存在差异，选取了 30 名学生根据其成绩、性别等特征配对成 10 个区组，每组 3 人，分别评价一位教师的教学效果，满分 25 分。试进行统计分析。

解答：该研究设计是一个典型的区组设计，若无法满足参数检验条件可考虑使用 Nonparametric tests => k related samples 中的有关过程。该主界面类似于相关样本 t 检验，选择三个教师的评价得分进入检验变量框（其中 block 变量只作为记号，不参加分析），选中弗里德曼 Friedman）双向方差分析和 Kendall's W 检验两种方法（其中 Cochran's Q 只适用于二分类数据），运行结果见表 12 - 20、
、
，两种检验法都先给出各样本的平均秩，采用相同的卡方检验，结果并不显著；其中 Kendall's W 还会给出相应的和谐系数值。

表 12 - 20 例 11Friedman 检验和 Kenndall's W 检验结果

Ranks		Test Statistics ^a		Test Statistics	
	Mean Rank	N	10	N	10
TEACHER1	1.70	Chi-Square	5.400	Kendall's W	.270
TEACHER2	2.60	df	2	Chi-Square	5.400
TEACHER3	1.70	Asymp. Sig.	.067	df	2
				Asymp. Sig.	.067
a. Friedman Test.					
①		②		③	

2. 名义型变量的多组比较

对于名义型指标变量，只能得到其各分类的计数数据。因此，比较多组的频数分布间有无显著差异实际上是单样本的 χ^2 拟合性检验。其对应的 SPSS 过程包括 Nonparametric tests 菜单中的 χ^2 检验（Chi - square）、二项分布检验（Binomial）和游程（Run）检验等，这里不再赘述。

第三节 多因素设计方差分析简介

析因设计或多因素设计、因素（型）设计相比单因素设计的最大优势是，它

不仅可以分别考察每个因素的效应，而且可以考察多个因素之间的交互作用。这种交互作用在统计分析中必须作专门的分析，因此，多个单因素方差分析不能替代析因设计方差分析。本节通过实例简单介绍 GLM 的 Univariate 和 Repeated Measures 两个过程的操作，了解析因设计方差 / 协方差分析的基本内容。

一、多因素设计的方差 / 协方差分析

（一）多因素设计方差分析

多因素设计方差分析的一般过程类似于单向方差分析，不过其重点不在于对因素主效应的考察，因为作为每个单因素主效应度量的平方和和均方实际上是在不考虑其他因素时从总变异中分解出来的，并不总能完整地反映因素的效应，只有分析因素间的交互作用（称简单效应检验）才可以分辨出各因素对于因变量变化的真实效应。看下面的例子。

例 12 - 12 某零售商人想了解市场中货架样式对于产品销量的影响，他设计了一个实验：将商场中的货架设置成三种高度，即齐膝高、齐腰高和齐眼高；又设计了两种货架宽度，即半宽和整宽，控制了其他条件。6 种货架样式各收集了三个产品销售记录数据，试帮助他完成数据分析（据 J M Lattin et al., 2003, P422）。

解答 这是一个 3×2 完全随机化设计，其数据格式也非常典型：两个因素和一个因变量分别构成数据列（三个变量：col1—货架高度、col2—货架宽度和 col3—销售量）。共 6 个处理 18 个记录，每个处理下有三个重复数据。分析操作过程如下：调用 General Linear Model \Rightarrow Univariate, 选择 col1、col2 进入固定因素框，col3 进入因变量框。打开 Plots 定制均值图，这里只有两个因素，将 col1 设置为横轴、col2 设置为不同图例（separate lines）。按 Add，下面提交框中显示“col1 * col2”。接着打开 Post Hoc 框，设定对 col1 进行多重比较（因 col2 只有两个水平，不用多重比较），同样选择 S - N - K 法和 Dunnetts' s C 法。

1. 基本结果解释

方差分析见表 12 - 21，由于是完全随机化设计，所有只有组间效应，模型中的总变异被分解成两因素主效应和一个交互作用效应，从平方和及自由度的构成看出，这三部分即第一行 Corrected Model，这里相当于组间效应，检验表明组间效应显著，其中还给出了整个模型的决定系数和校正决定系数；第二行截距表示总均值不为 0，并无实际意义；从第三行开始是各因素效应检验，即都将各效应的均方和统一的误差均方（即组内均方）相比构造 F 统计量，结果三个效应检验都显著，在这种情况下，应当直接分析交互作用，注意此时对于因素 col1 的多重比较并无意义。此外，从绘制的均值图

(图 12.12)可直观验证交互作用显著(折线不平行),进入简单效应检验过程。

表 12 - 21 例 12 - 12 方差分析表 (tests of between - subjects effects)

变异源	平方和	自由度	均 方	<i>F</i>	<i>P</i>
Corrected Model	696.000 ^a	5	139.200	19.886	.000
Intercept	138338.000	1	138338.000	19762.571	.000
货架高度	152.333	2	76.167	10.881	.002
货架宽度	470.222	1	470.222	67.175	.000
货架高度 * 货架宽度	73.444	2	36.722	5.246	.023
Error	84.000	12	7.000		
Total	139118.000	18			
Corrected Total	780.000	17			

a. R Squared = .892 (Adjusted R Squared = .847)

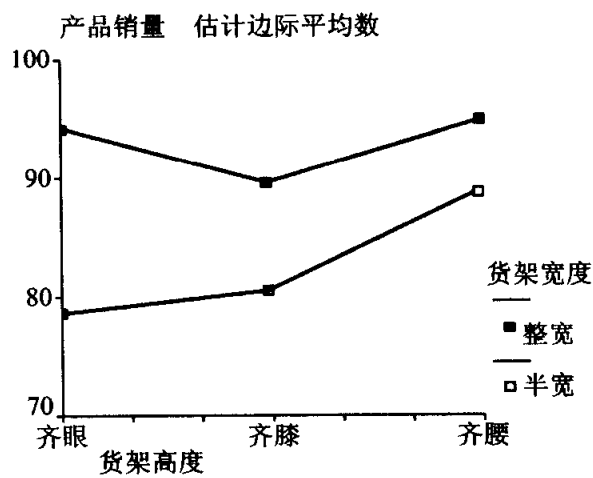


图 12.12 货架高度和货架宽度边际均值图

2. 简单效应检验

进行简单效应检验就是要检验在某个因素的不同水平上另一因素的各水平间差异不一致。即要分别检验在“货架宽度”的两个水平上，三种“货架高度”间有无差异（效应显著仍需多重比较）或者在“货架高度”的三个水平上两种“货架宽度”有无差异。这里，有多种方法可供选择。结合上面的例子分述如下：

- 使用条件筛选法固定某个因素的各水平进行单因素方差分析，即在固定某因素的一个水平时不考虑其他水平上的数据。固定水平的因素有几个水平即做几次少一个因素的方差分析以分别检验在其各水平上各效应情况。其方法是选择 Data => Select cases 分别选择 col1(或 col2) 的各水平进行另一因素的单向方差分析；更简便的办法是选择 Data => Split files 选择 Compare groups (各次检验合成一个表) 或 Organize output by groups (各次检验结果分别呈现) ，再选择根据 col1(或 col2) 分组。例子中用该法固定 col1 的方差分析表见表 12 - 22。

表 12 - 22 条件筛选法简单效应检验结果

ANOVA					
COL3					
COL1		Sum of Squares	df	Mean Square	F Sig.
齐膝	Between Groups	121.500	1	121.500	36.45 .004
	Within Groups	13.333	4	3.333	
	Total	134.833	5		
齐腰	Between Groups	54.000	1	54.000	5.226 .084
	Within Groups	41.333	4	10.333	
	Total	95.333	5		
齐眼	Between Groups	368.167	1	368.167	50.20 .002
	Within Groups	29.333	4	7.333	
	Total	397.500	5		

- 使用 MANOVA 的 DESIGN 语句 (舒华 ,1994)。这种方法和上述方法的唯一不同是它使用了所有数据的组内均方作为误差项变异的估计，每次检验的误差项相同。即两种方法对各效应的均方计算相同，但前一种方法实际上没有考虑全部数据，总的组内均方被拆分了（平方和和自由度分别拆分），因此每次检验都要分别计算误差项均方。显然，若各分组误差项不等，两者的检验结果可能会有较大差异。例子中用该法同样固定 col1 的检验结果和 MANOVA 语句见

表 12 - 23 , 由于各分组误差并无差异 (方差齐性) , 所以两者的结果相差不大。

表 12 - 23 MANOVA 语句简单效应检验语句及结果

manova col3 by col1(1,3) col2(1,2) /error=within /design= col2 within col1(1) col2 within col1(2) col2 within col1(3).	MANOVA					
	***** Analysis of Variance *****					
	18 cases accepted.					
	0 cases rejected because of out-of-range factor values.					
	0 cases rejected because of missing data.					
***** Analysis of Variance ***** -- design 1 Tests of Significance for COL3 using UNIQUE sums of squares Source of Variation SS DF MS F Sig.	6 non-empty cells.					
	1 design will be processed.					

	WITHIN CELLS 84.00 12 7.00					
	COL2 WITHIN COL1(1) 121.50 1 121.50 17.36 .001					
COL2 WITHIN COL1(2)	54.00 1 54.00 7.71 .017					
	COL2 WITHIN COL1(3) 368.17 1 368.17 52.60 .000					

• 使用 Univariate => Options 中的精细比较系数矩阵 (Contrast coefficient matrix) 或 LMATRIX 语句设定具体比较 (张敏强 , 2004) 这里不再赘述 , 有兴趣的读者请参阅有关文献。

3. 多因素方差分析中的模型设定

当多因素方差分析中要考虑的变异源较多时 , 应当善用模型设定优化检验结果 ; 有时模型设置不正确 , 可能无法得到希望的结果 , 最常见的问题如在各处理下无重复数据时纳入了交互作用 (例如前面提到的拉丁方设计的例子) 或者纳入了设计中不存在的因素。通常 , 多因素方差分析需要经过多次反复尝试 , 有时需要将高阶 (如三阶以上) 不显著的交互作用项从模型中去除或者合并到误差项中以增大检验的功效。此外 , 在 Model 子对话框中提供了四种平方和分解方法 其中 默认的第 III 型对于完全 (无缺失单元格) 非平衡设计 (单元格次数不等) 具有较好的稳定性 , 而第 IV 型对于不完全设计效果较好。在 Options 子对话框中也提供了许多精细分析选项 , 可用来细化分析和帮助自定义模型的分析。这里不再赘述。

(二) 协方差分析简介

协方差分析是针对研究设计中所出现的连续型自变量或无关变量 , 将线性回归和方差分析结合起来 , 运用回归方法校正各组平均数和 F 检验误差项 , 再用方差分析方法检验校正后的主效应 , 以实现协变量的统计控制和分析各变量效应的方法 , 可看作方差分析的延伸和扩展 (黄希庭 , 2004) 。除方差分析的一般限定条件外 , 协方差分析通常还需假定协变量和因变量呈线性关系以及各组回归斜率相等。协方差分析可按照因素多少分为单因素协方差分析和多因素协方差分析 , 也可按照协变量多少分为一元协方差分析和多元协方差分析。在 GLM 的所有过程中 , 协方差分析都被整合到相应的三个过程中 , 这里通过一个多因素设计例子介绍利用 Univariate 进行协方差分析的基本过程。

例 12-13 某研究者欲检验学生成绩在三种教学方法和性别两个因素上的差异情况，选取男、女各半共 24 名被试，随机分成三个组分别接受三种教学方法，在教学一段时间后测得学生成绩。考虑到学生先前学习成绩可能会影响后测成绩，事先测得他们在实验前的成绩。试完成数据分析（据吴明隆，2003，P202）。

解答：从研究设计上讲，由于被试的选取和分配遵循了随机化原则，因此可视为一个多因素前测后测设计（第五章实验组控制组前测后测设计的推广）。由于要排除前测成绩这个连续型无关变量的影响，因此采用 Univariate 进行协方差分析 选择 after 作为因变量 ,sex、method 作为固定因素 ,before 作为协变量框。一旦设定协变量，多重比较功能将失效，需借助 Options 中的比较主效应（Compare main effects）复选项实现事后比较。按照默认设置，该模型中只控制前测成绩而不对它与两个因素的交互作用进行考察。在绘图子对话框设置 method 为横轴、sex 为不同图例。

方差分析表和均值图分别见表 12-24 和图 12.13，可见协变量主效应显著，性别和教学方法的交互作用显著，可结合均值图进一步进行简单效应检验和事后比较得出结论。

表 12-24 例 12-13 协方差分析结果

Tests of Between-Subjects Effects

Dependent Variable 后测成绩

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	1012.334 ^a	6	168.722	30.68	.000
Intercept	20.636	1	20.636	3.752	.070
BEFORE	189.500	1	189.500	34.45	.000
SEX	16.648	1	16.648	3.027	.100
METHOD	322.439	2	161.220	29.31	.000
SEX * METHOD	98.473	2	49.237	8.952	.002
Error	93.500	17	5.500		
Total	10310.000	24			
Corrected Total	1105.833	23			

a. R Squared = .915 (Adjusted R Squared = .886)

表 12-25 列出了校正前测成绩前（左）、后（右）各处理下后测成绩的均值，可以发现，两个结果是不同的，校正后的均值估计值是根据模型中协变量的估计值 29.33 计算而来的。注意：这里只按照默认模型进行分析，实际上若要检验各

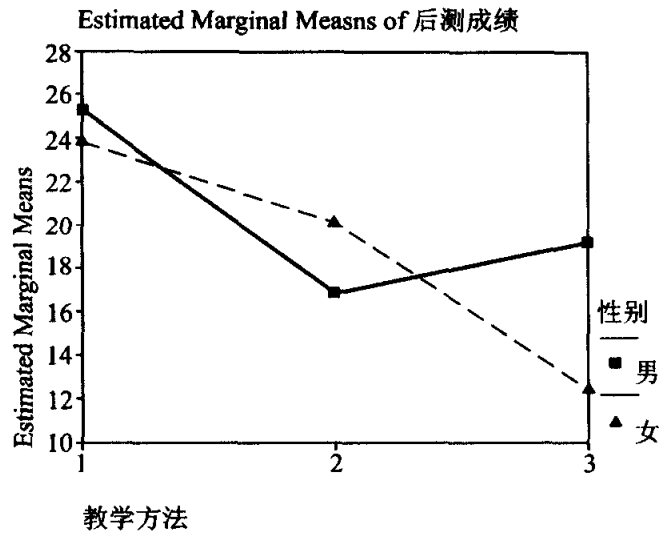


图 12.13 例 12 - 13 边际均值图

组回归线平行假定或者需要考虑协变量和因素间的交互作用，还需要自定义模型将其他两个二阶交互作用添加进来。请读者自行尝试分析并比较两种模型结果的差异。

表 12 - 25 校正协变量作用前、后各处理后测成绩均值

Descriptive Statistics					1. 性别 * 教学方法				
Dependent Variable 后测成绩					Dependent Variable 后测成绩				
性别	教学方法	Mean	Std. Deviation	N	性教学方法别方法	Mean	Std. Error	95% Confidence Interval	
								Lower Bound	Upper Bound
男	1	25.50	4.79583	4	男 1	25.334 ^a	1.173	22.850	27.808
	2	10.75	2.87228	4	2	16.825 ^a	1.564	13.525	20.125
	3	24.75	5.67891	4	3	19.091 ^a	1.518	15.888	22.294
女	1	25.25	2.62996	4	女 1	23.835 ^a	1.197	21.309	26.361
	2	14.50	4.04145	4	2	20.076 ^a	1.509	16.892	23.260
	3	16.75	2.75379	4	3	12.339 ^a	1.393	9.401	15.278

a. Evaluated at covariates appeared in the model 前测成绩 = 29.3333.

二、重复测量方差分析示例

(一) 重复测量设计及其数据格式

被试内设计或重复测量设计是令被试接受重复测量因素上的所有处理，由

于在所有重复实验处理中被试的各方面特点都假设为保持恒定，其对由于个体差异导致的无关变异的控制更优于随机区组设计，因此受到越来越多研究者的青睐，特别是应用于时间序列设计中。重复测量设计也分为单因素和多因素两种，若研究设计中还包含有被试间因素，则成为混合设计，不过因素是否设置为重复测量还取决于对于单个被试试验量和被试数量的综合权衡，因为重复测量必然会节省被试，但也增大了被试作业量。SPSS 的 Repeated Measures 就是用来处理混合设计的方差 / 协方差分析的过程。在第十一章中已经介绍了其操作特点。值得重点强调的是：重复测量中的每次观测都作为 SPSS 数据集中的一列（一个变量），但这并不意味重复了几次就有几个“因变量”——实际上，这几列数据表示的是同一个因变量在不同处理下的观测值。前已提及相关样本两组比较法是最简单的重复测量设计；同样，重复测量设计的数据格式也符合“行表示一个被试记录，列表示一次观测”这一 SPSS 数据集基本规则。此外，相关样本两组比较法中的两列数据实际上隐含了自变量的两个水平；同样，重复测量型数据也隐含了自变量的信息。不同的是，由于可能牵涉到多个被试内变量，因此，这里的列不单纯对应一个自变量的不同水平，而是不同自变量各水平交叉组合后的处理。在进行重复测量方差分析之前，很重要的工作即定义自变量或因素，即对各因素水平交叉组合和数据集中的各重复观测列进行匹配，否则就无法检验因素效应或发生错误。以下结合一个例子简单介绍重复测量方差分析操作过程。

（二）例题求解

例 12-14 例 12-9 中若研究者将生字密度因素设置为被试内因素，设立 5:1 10:1、20:1 三个水平，此外考虑到文章主题熟悉性也会影响阅读理解成绩，因此设立了两种主题（1—熟悉和 2—不熟悉）的实验材料，选取 8 名被试随机分成两组，分别阅读两种主题材料并测得其阅读理解成绩（据舒华，1994, P93 例）。

解答：这是一个典型的 2×3 二因素混合设计，调用 Analyze \Rightarrow Repeated Measures 后，首先进行被试内因素定义（如图 11.11）。先输入变量名 dense 表示生字密度，输入水平数 3 添加。按 Define 键，将三列重复观测数据拖入定义变量 dense 的三个对应水平；并选择 topic 为被试间变量；设定按 dense * topic 作均值图。结果说明如下：

表 12-26 输出对整个模型的多元方差分析检验结果，提供了 Pillai's Trace、Wilks' Lambda、Hotelling's Trace、Roy's Largest Root 四种统计量，第一种统计量最保守也最稳健，结果显示，被试内因素及其与被试间因素的交互作用都极显著。表 12-27 给出了球度检验（test of sphericity），它是一个风向标，若检验结果不显著（服从球形假设），则看后面的一元分析结果，其效能较上面的多元检验高，尤其在样本容量较小时；否则，应该看上面的多元检验或一元分析的

校正结果。SPSS 提供了 Greenhouse - Geisser、Huynh - Feldt 和 Lower - bound 三种校正系数，两者结论不一致时主要看多元检验结果。这里数据不服从球形假设(sig. = 0.026)，多元检验和一元分析结论一致。接着 SPSS 默认给出重复测量间的多项式对比结果，其提供的线性趋势和二次曲线趋势检验在这里不太有意义，可通过设定 Contrast 对话框改变。若要一般地对因素进行事后检验，可在 Options 对话框中的 Display Means for 框中设定比较 dense 因素的主效应（即事后比较）。表 12 - 28 给出的是利用 LSD 法进行生字密度事后比较结果。接下来给出被试间效应检验结果，结果显示主题熟悉性效应显著（略）。最后由图 12. 14 对照上述检验结果可更精细地对各效应情况得出结论。

表 12 - 26 例 12 - 14 多元检验结果

Multivariate Tests						
Effect		Value	F	Hypothesis df	Error df	Sig.
DENSE	Pillai's Trace	.976	101.429 ^a	2.000	5.000	.000
	Wilks' Lambda	.024	101.429 ^a	2.000	5.000	.000
	Hotelling's Trace	40.571	101.429 ^a	2.000	5.000	.000
	Roy's Largest Roo	40.571	101.429 ^a	2.000	5.000	.000
* TOPIC	Pillai's Trace	.962	62.857 ^a	2.000	5.000	.000
	Wilks' Lambda	.038	62.857 ^a	2.000	5.000	.000
	Hotelling's Trace	25.143	62.857 ^a	2.000	5.000	.000
	Roy's Largest Roo	25.143	62.857 ^a	2.000	5.000	.000

表 12 - 27 例 12 - 14 球度检验结果

Mauchly's Test of Sphericity							
Measure: MEASURE_1							
Within Subjects				Epsilon ^a			
Mauchly's				Greenhous	Huynh	Lower-	
Effect	W	Chi-Square	df	e-Geisser	-Feldt Sig.	bound	
DENSE	.233	7.290	2	.026	.566	.724	.500

表 12-28 例 14 多元检验结果
Pairwise Comparisons

Measure: MESURE_1

(I) DENSE	(J) DENSE	Mean		Sig. ^a	95% Confidence Interval for Difference ^a	
		Difference (I - J)	Std. Error		Lower Bound	Upper Bound
1	2	-2.125 *	.239	.000	-2.711	-1.539
	3	-4.375 *	.346	.000	-5.222	-3.528
2	1	2.125 *	.239	.000	1.539	2.711
	3	-2.250 *	.144	.000	-2.603	-1.897
3	1	4.375 *	.346	.000	3.528	5.222
	2	2.250 *	.144	.000	1.897	2.603

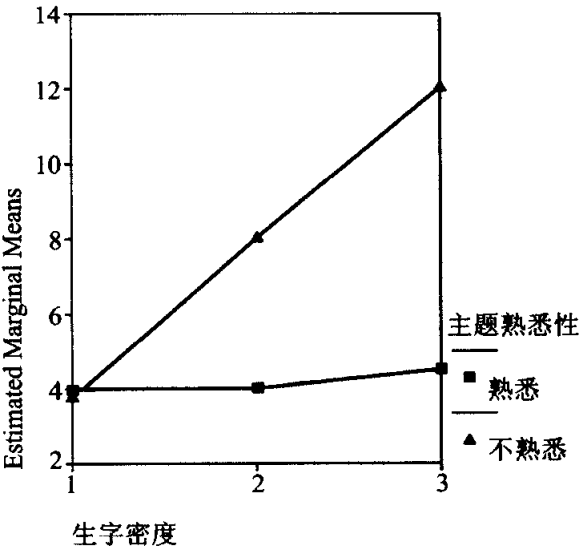


图 12.14 例 12-14 边际均值图

补充讨论 12-1 多元统计分析简介

心理学的研究目的是探索人类行为和心理过程中各种变量间的关系，多元统计分析(multivariate statistical analysis)可以满足研究多变量间关系的需要。虽然关于多元统计分析法的范畴存在争论，但普遍认同的是：多元统计分析是一组用来分析两组或多组观测间联系的统计程序的统称，因此亦称复变量分析、多

变量分析（黄希庭,2004），如多元方差 / 协方差分析、多元回归等都是典型的多元统计法之一。通常，可根据研究的模型类型、因变量的数目、因变量及自变量的测量等级来建立多元统计法的分类框架（郭志刚，1999）。这里根据模型类型、方法的目的和变量的量尺对常见的多元统计法进行分类（见表 12 - 29）并概括介绍各种方法的特点，仅供参考。

多元统计法按照其探讨的变量间关系类型可分成两类：一类是用于解释变量间相依关系或进行变量分类的相依模型（interdependence model），另一类是用于探讨自变量和因变量间因果关系的因果模型（dependence model）。按照多元统计法的目的，可分为探索性和验证性两类，前者旨在探索多组观测数据的内部结构，例如进行数据简化或变量分类，探讨变量间关系等；后者旨在验证某个研究假设或理论构念。而变量的量尺则大致可分为测量型（metric）和非测量型（nonmetric），前者主要包括比例和等距尺度，后者主要指定性的等级和名义尺度。

表 12 - 29 多元统计分析法分类及特点

多元统计法	相依关系 或 因果关系	探索性 或 验证性	测量型或 非测量型	分析目的
主成分分析	相依关系	探索性	测量型	降维,简化数据
探索性因素分析	相依关系	探索性	测量型	解释相关模式,发现隐含特质
验证性因素分析	相依关系	验证性	测量型	验证测量的模型
多维量尺分析	相依关系	探索性	皆可	根据事物相似性建立变量的空间表达
对应分析	相依关系	探索性	非测量型	建立非测量型数据间联系的空间表达
聚类分析	相依关系	探索性	皆可	根据事物相似性建立变量分组
多元回归分析	因果关系	探索性	测量型	建立多个自变量对因变量的预测模型
典型相关分析	因果关系	探索性	测量型	解释两组变量间的共变关系
结构方程建模	因果关系	验证性	测量型	包含潜变量的因果模型
多元方差分析	因果关系	验证性	皆可	具有离散型自变量的典型相关分析特例
判别分析	因果关系	皆可	皆可	具有离散型因变量的典型相关分析特例
对数线性模型	相依关系	皆可	皆可	建立多个分类变量间的关系模型

在 SPSS 的分析菜单中除包含本教材中介绍的一些基础统计分析法外，还提供了迄今大多数成熟的多元统计分析法，整合了许多用于多元分析的统计程序。以下对这些多元统计技术作一概括，有兴趣的读者请参阅有关文献。

主成分分析(principal components analysis)将一组相关变量转换成另一组不相关变量的数据转换方法，是探索性因素分析(exploratory factor analysis)中求初始解的主要方法。通过对数据的重新表达，变量所包含的信息按照大小依次分配到各主成分上，则可适度去除信息量较少的主成分使得总的信息量不致损失过多，从而达到对多元数据降维和化简的目的。具体请读者学习第十三章对主成分分析和探索性因素分析的介绍。

探索性因素分析和验证性因素分析(confirmatory factor analysis)统称因素分析，这类方法旨在确定变量间的潜在公共变异源，观测变量的变异可分解为潜在公因子和独特因子。前者旨在从一组相关变量中提取出少数几个潜在变量(公因子)，寻找数据的基本结构和简化数据。而后者用于检验关于因素结构和因素间关系的假设，即检验根据事先假设推导出的因素结构和基于观测数据导出的因素结构是否拟合，因此也可看成是结构方程建模(structural equation modeling)的特例。

多维量表分析(multidimensional scaling)是基于研究对象间的相似性或距离获得关于对象的低维空间图示表征的一组方法的统称，其目的是对研究对象进行分类或分组，揭示数据的潜在结构。

对应分析(correspondence analysis)是将多个分类变量间的频数列联表转换成以散点图示以直观表达变量间联系的方法。

聚类分析(cluster analysis)旨在分析各研究对象在多个测量型变量值上的相似性或距离进行对象分组，对个案的分类称为 Q 型聚类，对变量的分类称为 R 型聚类。

多元回归旨在建立多个自变量和一个因变量的关系模型(称回归方程)，从而可通过自变量的变化来预测因变量的变化。其因变量为测量型变量，自变量为测量型或虚拟变量，回归系数的求解通常采用最小二乘估计。

典型相关分析(canonical correlation analysis)旨在分析两组变量之间的整体相关关系，可视为简单相关和多重相关的扩展。

结构方程建模旨在研究潜在变量间的复杂因果关系，它由多个有内在联系的多元回归方程组成一套路径分析(path analysis)联立方程组，每个方程都通过多元回归求解参数，并用路径分析分解变量间的直接作用和间接作用。主要采用极大似然估计求解回归系数，常用的分析软件如 LISREL、AMOS 等。

多元方差分析用于检验两个以上分类均值在多个因变量上是否存在显著差异，分为单因素和多因素多元方差分析，此处的“元”指因变量。

判别分析(discriminant analysis)旨在研究两组或多组之间在一套鉴别变量上的差别,它就一些个案的分类和其他多元变量的已知信息,确定分类与多元变量间的数量关系,建立判别函数,并据之对其他已知多元变量信息的个案进行判别分类。可视为因变量为分类变量,自变量为测量型或虚拟变量的因果预测。

对数线性模型是对列联表分析的拓展,旨在建立多个分类变量影响下的列联表单元格观察频数的数量模型,分析多个变量的效应及交互作用,探讨多变量的独立性和相依性。

►本章提要

1. 探讨变量间的相关或关联程度是心理学研究的常见问题,这类统计检验法包括相关分析、回归分析和卡方检验等常用统计检验法。其对应的 SPSS 统计过程也有很多。

2. SPSS 相关过程可以进行等距型、等级型数据的简单相关分析和偏相关分析以及用于多元统计方法的距离分析。

3. 相关分析常应用于量表或问卷的信、效度分析,虽然其分析往往有具体的统计过程或程序,但大都运用了相关的计算;使用相关过程可以直接计算诸如评分者信度、重测信度、复本信度以及准则效度、区分效度和聚合效度等指标。

4. 回归分析旨在建立变量间关系的数学模型,线性回归是假设变量间呈直线性关系,利用最小二乘法,通过寻找最优拟合直线所满足的条件对回归方程中的参数进行估计,并通过方差分析检验回归方程的有效性。

5. 卡方检验是度量变量间频数分布间关联性的统计检验法,它既可以用来解决单个多项分类变量的次数分布无差假设检验问题,也可用来解决连续型变量次数分布对某假设分布的拟合性问题;还可用来对变量间次数分布的独立性进行检验。SPSS 中的列联表提供了更一般的整套关联推断检验法。

6. 通过组间比较探讨某自变量对观测指标是否存在显著效应是心理学研究的另一类常见问题,当然这类问题并非都是因果关系推断。该类检验法可按照组数多少、研究设计类型和测量尺度加以分类,通常所讲的 t 检验、单向方差分析等参数检验法以及非参数差异检验法。

7. 因素型设计方差分析的主要特点在于考察因素间的交互作用,利用一般线性模型中的一元过程和重复测量过程可以解决大多数类型的单因变量因素型设计方差/协方差分析问题。

8. 多元统计分析是一组用来分析两组或多组观测间联系的统计程序的统称,可根据研究的模型类型、因变量的数目、因变量及自变量的测量等级等标准对多元统计法进行分类 SPSS 中提供了大多数成熟的多元统计分析法。

►本章关键术语

等级方差分析 (ANOVA by ranks)	验 (Kolmogorov - Smirnov Z test)
对数线性模型 (log - linear model)	克 - 瓦氏单向方差分析 (Kruskal - Wallis one - way ANOVA)
多元回归 (multiple regression)	离差平方和 (sum of squares)
弗里德曼双向方差分析 (Friedman two - ways ANOVA)	曼 - 惠特尼 U 检验 (Mann - Whitney U test)
符号检验 (Sign test)	偏相关 (partial correlation)
复相关 (multiple correlation)	威尔考克森符号等级检验 (Wilcoxon signed ranks test)
回归方程 (regression equation)	相关分析 (correlation analysis)
回归分析 (regression analysis)	协方差分析 (analysis of covariance , ANCOVA)
回归系数 (coefficient of regression)	信度分析 (reliability analysis)
决定系数 (coefficient of determination)	秩和 (sum of ranks)
均方 (mean square)	中数检验 (median test)
卡方独立性检验 (Chi - square test of independence)	重复测量设计 (repeated measure design)
卡方拟合性检验 (Chi - square test of goodness of fit)	
克尔莫哥洛夫 - 斯米尔洛夫 Z 检	

►复习与练习

1. 关联推断可以通过哪些 SPSS 过程加以实现? 各自适用于何种场合? 简单相关和偏相关、部分相关有何联系和区别?
2. 举例说明相关分析在量表信、效度评价中的应用。
3. 阐述线性回归分析的一般原理和步骤。
4. 举例说明卡方检验的原理和用途。
5. 在例 12 - 12 中若要考察因素 col2 的主效应, 该因素只有两个水平, 为什么不需要进行多重比较?
6. 在例 12 - 12 中二因素交互作用显著, 为什么对因素 col1 的多重比较结果无意义?
7. 在例 12 - 13 中设定考察所有交互作用进行协方差分析, 并根据分析结果判断哪种模型更加合理, 应当采信哪种分析结果?

►推荐参考读物

伯恩斯坦, 统计学原理 (下) ——推断性统计学. 史道济译. 北京市: 科学出

版社,2002

夏普. 社会科学统计学. 王崇德译. 北京市: 科学技术文献出版社, 1990

张厚粲. 心理与教育统计学. 北京市: 北京师范大学出版社, 1998

张敏强. 教育与心理统计学. 北京市: 人民教育出版社, 2002

张文彤. 世界优秀统计工具 SPSS 11 统计分析教程(基础篇). 北京市: 希望电子出版社, 2002

弗里德曼等. 统计学(第二版). 魏宗舒等译. 北京市: 中国统计出版社, 1997

Goodwin C. J. Research in psychology: Method and design(3rd ed). New York: John Wiley & Sons, 2002

Heiman G. W. Understanding research methods and statistics: An integrated introduction for psychology(2nd ed). Boston: Houghton Mifflin Company, 1998

► 在线资源

相关分析、 t 检验、 χ^2 检验等: <http://www.statsoft.com/textbook/stbasic.html>

方差分析 / 多元方差分析: <http://www.statsoft.com/textbook/stanman.html>

线性回归: <http://www.statsoft.com/textbook/stmulreg.html>

非参数检验: <http://www.statsoft.com/textbook/stnonpar.html>

信度和项目分析: <http://www.statsoft.com/textbook/streliab.html>

实验设计: <http://www.statsoft.com/textbook/stexdes.html>

第十三章 数据化简技术

——因素分析法

心理学家是如何考察像自我、人格、智力这类所谓的“心理构念”的？因为这些构念似乎都是内潜的、不可观测或难以捉摸的。可否通过一些可观测的指标或变量来度量它们呢？例如黄希庭等编制的青少年自我价值感量表包含个人取向自我价值感和社会取向自我价值感两个亚因素，贝克抑郁问卷包含消极态度或自杀、躯体症状和操作困难三个因素，而王登峰等编制的中国人人格量表则包含七个因素，这些所谓的“因素”是如何从大量的自陈式测题中获得的呢？又假定我们编制了一个包含四个因素共 60 个题项所构成的大学生责任心量表，怎样才知道这个量表中的哪些题项符合量表编制要求呢？一个成熟的、经得起考验的量表是如何问世的？对于这些问题的解答涉及到因素分析技术，本章将讨论以下问题：

1. 因素分析法的主要概念和基本原理。
2. 进行因素分析的基本条件。
3. 利用 SPSS 进行探索性因素分析的操作过程。
4. 因素分析法在解决实际问题中的应用。

第一节 因素分析原理概述

因素分析（**factor analysis**）法的主要目的在于：通过寻找或确定几个较少的假想“因子”（**factor**）来反映多个观测变量中蕴含的大部分信息，从而浓缩或化简观测数据。同时，浓缩后的因子代表了数据间的基本结构，通过得到的因子估计值使研究者更方便地掌握数据的本质特质以及因子和观测变量之间的关系。这两种目的表现在心理学研究应用中，前者主要是编制心理测量问卷、量表；后者主要是探查因素或变量之间的关系、结构。因此，因素分析法既是一种数据处理技术，同时也是一种重要的研究方法。本节先利用图解了解因素分析法的基本原理，并通过因素分析模型介绍其中的一些重要概念和使用条件，接着对 SPSS 软件中关于因素分析的菜单选项及其功能进行汇总，并用一个案例教会你如何看懂因素分析的结果。

一、因素分析基本原理

(一) 因素分析和多元统计分析

因素分析又称因子分析，是多元统计分析 (multivariate statistical analysis) 技术的一个分支，也是心理学研究中常用的一种数据化简技术 (data reduction technique)。诸如自我、人格、智力等这样的心理概念都是一些潜在变量 (latent variable)，只能通过对一些可观测的外显变量 (manifest variable) 的测量间接反映。但正如前面问题中提到的那样，这样做可能会产生大量观测变量，使得对数据的分析和描述陷入混乱；更主要的是大量变量间可能存在的高度相关还会极大地削弱某些多元统计方法的效果，如在多元回归分析中由于自变量间的高度线性相关导致的多重共线性 (multiple collinearity) 现象会产生极大的危害 (郭志刚, 1999)。

因此，包括因素分析在内的许多多元统计分析技术的首要目的在于化简数据，如主成分分析、对应分析、典型相关分析、多维量表法等。这些方法在化简数据的同时又依然能够保留原始数据中的大部分信息，所不同的是它们适用的条件和场合各不相同，因素分析法是心理学研究中常用的一种针对等距及以上尺度类型变量的数据化简技术。根据研究者对因子数量和因子结构是否有事先设定，因素分析又分为探索性因素分析 (exploratory factor analysis) 和验证性因素分析 (confirmatory factor analysis) 两种，通常说的因素分析主要是指前者，本章也采用这种简称。因素分析的原理相当复杂，但因为它在问卷及量表编制中非常有用，如果立足于问题解决和统计软件操作，则学习这种非常有用的多元统计技术又是可能而且十分必要的。

(二) 因子分析原理图解

在方差分析法中，考察自变量的效应是通过分析因变量分数的变异源来实现的，可能的变异源包括研究者关心的心理特质和其他无关变量以及误差，可用方差来表达这种差异的变异量。因素分析是假设大量观测变量背后潜藏着少数几个维度 (dimension)，称为公共因素或“公因子” (common factors) 每个观测变量总变异中的绝大部分都能够被这几个公因子所解释，不能为公因子解释的部分称为该变量的“特殊因子” (unique factor)。这样，所有观测变量一般都可以表示为公因子和特殊因子 (包括误差) 的一个线性组合，称为因素分析的“公因子模型” (common factor model) (Tucker & MacCallum, 1997)。事实上，研究者主要关心的是能够代表较多信息的公因子，因此，一般所谓的“因子”主要是指公因子。

以因素分析中最常用到的主成分分析法为例，它将一组相关的观测变量转换成少数几个相互独立的主成分，即因子，在这个变换过程中不改变变量的总方差，转换后第一个因子具有最大方差，即能够解释最多的变异，称为“第一主成分”随后抽取的因子依次称为“第二主成分”，……第 k 主成分，其解释的数据

变异量逐渐递减。借助最简单的二维空间中的图解看因子和观测变量间的关系 如图 13.1 左所示，两个观测变量 x_1 、 x_2 构成的样本空间中，成对样本点的分布近似一个椭圆（ x_1 和 x_2 相关），按照上述原则，第一主成分应当选择这个椭圆的长轴（ f_1 ），因为样本点距离它平均最近，在这个方向上数据分布总体上离散程度最大，将样本点全部投影到这条线上，投影点可以表示最多的变异（方差），包含的信息最多；如果再抽取一个主成分，显然应当是和长轴垂直的椭圆短轴（ f_2 ），但它对样本点的代表程度显然比第一主成分差，包含的信息较少。两个主成分相互正交（orthogonal）保证了它们所代表的变异不会重合。

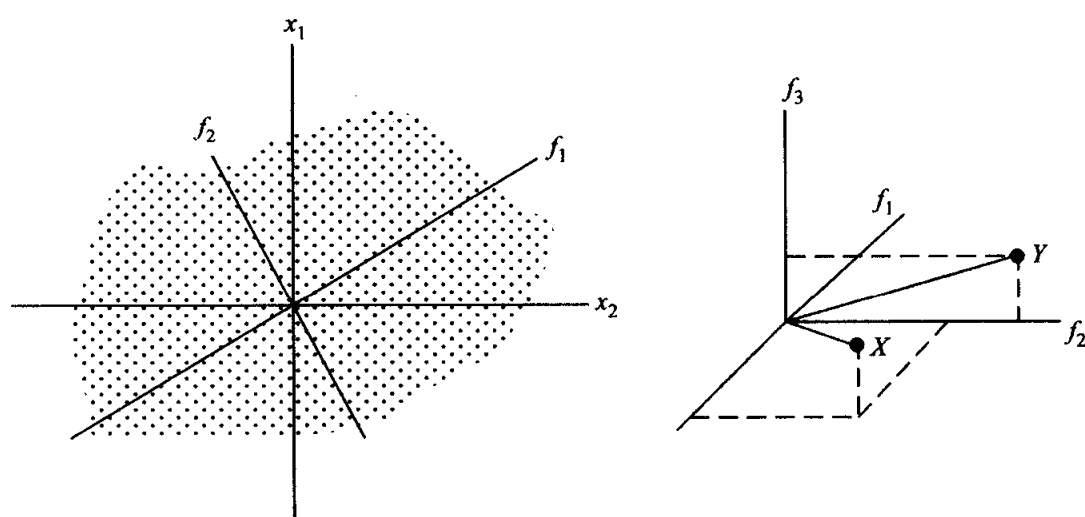


图 13.1 主成分原理示意图

主成分分析求解直接用到的数据是观测变量间的相关系数矩阵或协方差矩阵。通过求出少数几个主成分，变量间原来的关系就可以用因素和变量间的相关矩阵来反映，称为“因素结构”(factor structure)。由于两变量间的积差相关系数等于两变量分数之标准分数两两相乘的平均值用几何图解法来说明公因子和观测变量间的这种关系，如图 13.1 右所示，假设有三个公因子 f_1 f_2 f_3 构成一个立体直角坐标系 $f_1f_2f_3$ ，如果将所有观测变量数据都经过标准化而统一单位，则观测变量 X 、 Y 得分的标准分数均可表示为该坐标系中的向量 (vector)，其在三个坐标轴上的投影值分别为 $(-2, 2, 1)$ 和 $(0, 3, 1)$ ，分别为变量在因子上的“因素负荷”(factor loading)，而向量的长度即“公因素方差”，或变量的“共同度”(communality)。此外，数学上可以推知 X 、 Y 两变量间的相关系数等于两变量共同度与它们所夹角度之余弦的乘积。因此可以说，各变量的因素负荷、夹角由公因素方差、相关系数决定，而不随因素解的变化而变化，所谓的“求因素解”无非是为说明向量的空间位置而在空间中设置适当的坐标轴而已。

按照这种原则抽取的主成分数量等于变量数目，数据由彼此相关的观测变量转换成相互独立（表现在图形上相互正交）的主成分（ f_1, f_2 彼此不相关），总的变异量不会有任何损失，而如果将第二主成分省略且不至于使总的信息量损失过多，就可以达到数据化简的目的，即仅用一个主成分表示两个观测变量。同理，在三维空间中样本点的分布近似于一个椭球，第一主成分应选择第一长轴。如果椭球很长很细，则第一主成分就足以反映原来变量的信息；而如果椭球像一个铁饼，则还需要选择第二长轴作为第二主成分。

提取出的主成分（公因子）对数据总变异的解释能力称为该主成分（公因子）的“贡献”（contributions），主成分分析的原则就是求解因素负荷矩阵，使得抽取出的第一主成分能够解释最多的变量方差（变异量）。然后从去除了第一主成分的变异量的因素负荷矩阵中求出第二主成分，它能够解释剩余的最大方差并且保证和第一主成分相互正交（相关为 0 无重合信息）以此类推直到求出所有的主成分。

二、因素分析模型及条件

（一）因子分析一般模型

看一个具体例子来了解一下因素分析模型。

例 13-1 某研究者收集了 100 名被试对六种政策的回答数据，其中六个变量 col1—col6 分别表示“政府应当投入更多的教育经费”、“政府应当投入更多的经费减少失业”、“政府应当控制大企业”、“政府应当通过用校车送孩子上学加速废止种族隔离”、“政府应当增加少数民族的工作配额”和“政府应当扩展领先计划*”（J M Lattin et al. 2003, P124）。数据间相关系数如表 13-1 所示。

表 13-1 各观测变量间相关系数矩阵（ $P < 0.01$ ）

变量	col1	col2	col3	col4	col5	col6
col1	1.0000	.6008	.4984	.1920	.1959	.3466
col2	.6008	1.0000	.4749	.2196	.1912	.2979
col3	.4984	.4749	1.0000	.2079	.2010	.2445
col4	.1920	.2196	.2079	1.0000	.4334	.3197
col5	.1959	.1912	.2010	.4334	1.0000	.4207
col6	.3466	.2979	.2445	.3197	.4207	1.0000

作者注：Head Start Programme，领先计划或先起计划，美国政府实施的旨在扶持少数民族和贫困人口的教育工程。

$$\begin{cases} x_1 = l_{11}f_1 + l_{12}f_2 + \cdots + l_{1m}f_m + \varepsilon_1 \\ \dots\dots\dots i = 1, \cdots, k \quad j = 1, \cdots, m \\ x_i = l_{i1}f_1 + \cdots + l_{ij}f_j + \cdots + l_{im}f_m + \varepsilon_i \\ \dots\dots\dots \\ x_k = l_{k1}f_1 + l_{k2}f_2 + \cdots + l_{km}f_m + \varepsilon_k \end{cases} \quad (13-1)$$

a) f_j 代表 m 个假想公因子 ($m \leq k$)，其值称为因素分或因子值 (factor score)，表示各观测变量共有的因子，解释了变量之间的相关关系；

b) ε_i 代表特殊因子，一般作为残差项，表示各观测变量中不能为公因子解释的变异部分；

c) 公因子前面的系数 l_{ij} 称为因子负荷，表示第 i 个变量在第 j 个因子上的相对权数，相当于标准化回归系数。

（二）模型中的主要概念

因子负荷是因素分析模型中最重要的统计量，它反映了在第 i 个变量中第 j 个因子被反映的程度，相当于给因子加了一个权数。 $|l_{ij}|$ 绝对值越大表示因子和变量间的关系越密切，因子反映变量的信息就越多。因素负荷构成的 $k \times m$ 矩阵称为因素矩阵 (factor matrix) 或因素模式 (factor pattern)，因为它反映了因素与变量关系中因素的性质。因素结构是指因素与变量间的相关关系，可以证明 (芝祐顺, 1999)，在正交模型中，因素模式等于因素结构，因素负荷就是变量与因素间的相关系数。因此在正交模型中不加区别地统称为“因素负荷”。

$$r'_{ij} = l_{i1}l_{j1} + l_{i2}l_{j2} + \dots + l_{im}l_{jm} \quad (13-2)$$

这样构造出的一个新的相关系数矩阵称为导出相关矩阵或衍生相关阵 (repro-

duced correlation matrix)。显然，如果由因素负荷计算出的变量间相关系数和从观测数据计算出的变量间原始相关系数差别很小，则说明模型很好地拟合了观测数据，因素解是合适的。

(2) 公因子方差

公因子方差又称为共同度、公共方差，它表示的是观测变量 x_i 方差中被公因子所解释的部分比例（标准化后变量的总方差为 1），等于所有公因子负荷的平方和，即：

$$h_i^2 = \sum_{j=1}^m l_{ij}^2 \quad (13-3)$$

显然，公因子方差越大，说明变量能被所有公因子解释的程度越高，原变量信息被保留的程度也更大（因此常被解释为某变量的信度）。

(3) 因子贡献和贡献率

公因子方差反映的是单个变量被所有公因子解释的部分比例，因子贡献反映的则是单个因子解释的数据总方差。显然，所有公因子的累计贡献等于所有变量的共同度之和；如果公因子数等于变量数（主成分分析）则也等于原观测变量的总方差。公因子 j 的贡献记为 V_j ，它等于和因子 j 有关的所有因子负荷的平方和：

$$V_j = \sum_{i=1}^k l_{ij}^2 \quad (13-4)$$

实践中更常用“贡献率”这个相对指标来表示各因子在解释方差上的相对重要性。实际上由于数据标准化后总方差就等于变量个数 k ，因此 V_j/k 恰好表示因子 j 的贡献率。将因子的贡献率从大到小排列，计算前 m 个主成分的累计贡献率可以帮助判断应当选取多少个因子，即将后面贡献率很小的主成分省略而不至于损失过多的解释力。

(4) 特征值和特征向量

根据矩阵代数知识，求解主成分主要通过求解变量相关矩阵的“特征方程”得到若干个“特征值”(eigenvalue)和对应的单位特征向量(eigenvector)来实现，有多少个主成分，就有多少个特征值和对应的特征向量，它们分别记为 λ_j 和 $V_j = (v_{1j}, v_{2j}, \dots, v_{kj})$ 。可以证明（芝祐顺，1999），各主成分特征值等于其因子贡献。因此，特征值也可以用于判断公因子的数量，一般认为要达到数据化简的目的，每个公因子应至少能解释一个变量的方差，即保留特征值不小于 1 的主成分作为公因子。

(三) 因素分析的假设条件

因素分析结果受到许多因素的影响，但有一些基本条件必须满足方适合进行因素分析。从宏观层面的研究构想说，因素分析着重在于分析变量间的共同变异，因此变量背后必须存在“潜在”的维度，否则得到的因素分析结果就难以理解。变量除包含共同变异外还有区别于其他变量的特殊变异，其中既包含来自自身的特殊变异，也有随机误差或测量误差的变异，因此完整的因素分析模型中，残差项还应分解成特殊因子和误差两部分。关于模型中这些参数的不同假设可能对应着不同的分析过程。此外，因素分析通常首先要求观测变量分数以及公共因素、独特因素都用标准分数表示，即变量和因素的均值均为 0，方差均为 1；其次，公共因素和独特因素间不相关，各独特因素之间也不相关（尤其在公因子分析法中）；另外，由于因素分析的目的是找出观测变量中潜在的因素结构，因此要求观测变量间应有较强的相关关系，一般要求绝大部分相关系数应不低于 0.3；此外，正交模型要求公因子之间应当相互独立，尤其是在主成分分析中。

同任何统计方法一样，因素分析对数据类型以及数据的分布也有一定的要求，等距型数据是一个基本要求，一些特殊情况下等级型数据也可以接受（Kim & Mueller, 1978）。由于基于样本观测数据的因素分析结果往往要推论到总体甚至参加进一步的假设检验（参见第三节），数据最好是多元正态数据，而其中有些参数估计方法（如极大似然法）更是必须要求多元正态数据。此外，因素分析求解特别是主成分分析是基于变量间的皮尔逊积差相关系数，因此要求变量间是线性关系，否则应予以剔除或进行适当的转换。样本量越小，样本数据的分布和线性关系越需要加以检验。

实际应用者往往还要考虑关于样本量适当大小。通常可以立足于“被试一变量比”（subjects-to-variables ratio, STV ratio），如至少每个变量有 10 次观测值，STV 比至少为 5 等（Bryant and Yarnold, 1995）。新的观点认为因素的理论结构越清晰，必需的样本量越小，但起码也得是 50 或 100。要得到可信的因素又需要多大样本呢？研究发现：具有 4 个以上负荷高于 0.6 或 3 个以上负荷高于 0.8 的因素或主成分是可信的，而不论样本量多少；若样本量大于 150 则具有 10 个以上负荷较低（ <0.4 ）的因素是可信的；而当因素只有很少的低负荷时不足取，除非样本量大于 300（Guadagnoli & Velicer, 1988）。事实上，因素分析对于样本量的要求远不如对变量数的要求高，这一点不同于回归分析。有关的讨论读者可参阅其他参考文献。

补充讨论 13 - 1 因素分析导致心理学的重大进展

——斯皮尔曼、卡特尔、瑟斯顿与因素分析技术

斯皮尔曼 (Charles Edward Spearman, 1863—1945) , 英国心理学家、统计学家, 曾任英国心理学会主席。早年从事哲学研究, 后因不感兴趣投笔从戎并在英国军队服役长达 15 年。斯皮尔曼在德国师从冯特 (Wilhelm Wundt) 学习心理学, 1904 年在他 45 岁时获得博士学位, 后回国在伦敦大学学院 (University College) 任教至 1931 年退休, 之后先后应邀在哥伦比亚大学等地进行教学和科研。其代表作有《人的能力》 (Abilities of Man, 1927 ; 目前国内出版的“ 20 世纪心理学通览 ” 丛书中有收录)、《智力本质和认知原理》 (The Nature of Intelligence and the Principles of Cognition 1923) 等。



受统计学家高尔顿 (Francis Galton) 的影响, 经过严格统计学训练的斯皮尔曼十分关注智力问题。他在测量农村在校儿童的智力水平的过程中发现, 不同测验的得分存在高度相关。他相信由于测量的不准确和信度问题, 实际观测到的这种相关通常会低估两个被试测验得分的真正相关。因此, 如果知道准确的信度值, 就可以对这种“衰减”问题进行校正, 利用一种校正公式, 他发现了所有类型的能力间存在很大程度的相关, 导致了所谓“一般智力因素”(g 因素) 概念的诞生。1904 年, 他在攻读博士学位其间发表了堪称心理学历史上伟大里程碑的著名论文《一般智力的客观测定和量度》 (General Intelligence Objectivity Determined and Measured), 提出了其著名的智力“二因素说”: 即个体的所有认知作业背后潜藏着一个“一般智力”因素; 同时存在许多具体能力, 即所谓的“特殊智力”因素 (s 因素) 部分地决定着在这些任务上的作业。他强调一般因素, 认为一个具有高 g 得分的个体必然能在另一个 g 因素测验中得高分。

在建立该学说的过程中他首先使用并创立了今天称为“因素分析”的统计技术, 并用该技术验证了其智力“二因素说”。虽然有人认为他的“g 因素”过于简单化, 所用的统计程序也存在稳定性问题。但无论如何, 他的“g 因素”仍被许多人接受为一个合理的心理构念, 而他创立的因素分析技术今天也成为被广泛应用的成熟技术。

卡特尔 (Raymond Bernard Cattell, 1905—1998), 杰出的英裔美国心理学家、

人格理论家，美国《普通心理学评论》杂志评出的“二十世纪一百位最著名的心理学家”之一。1929年在伦敦大学获哲学博士学位，先后在英国埃克塞特大学（Exeter University）、美国克拉克大学及哈佛大学任教，1932至1937年创办并主持英格兰莱斯特市心理诊所（City of Leicester Child Guidance Clinic），1945年后一直任伊利诺大学教授。代表作有：《人格：一种系统、理论和实在的研究》（Personality: A systematic, theoretical, and factual study, 1950）、《多变量实验心理学手册》（Handbook of multivariate experimental psychology, 1966）、《人格和能力的遗传性：研究方法和发现》（The inheritance of personality and ability: Research methods and findings 1982）等。



卡特尔早年师从斯皮尔曼学习因素分析法，他坚信该方法具有足够的客观性和信度。在实施自己的测验和因素分析后，他扩展了斯皮尔曼的智力“二因素说”，牵涉到智力发展中的更多因素，包括动机和心理状态。卡特尔对于因素分析技术的主要贡献在于，他发展了一些实用的统计技术和方法，如用于确定因子数目的碎石图检验和比较因素解的相似性系数等。卡特尔相信所有的理论都应当从实际研究中产生，正是借助于经验观察和客观测量，他不仅对其老师的智力理论做出发展，提出“晶体智力和流体智力”理论，更率先将因素分析技术用于人格研究，通过对人格的因素分析来寻找基本的人格要素以探索有关行为的详尽理论，产生了大量理论模型和心理测量工具（典型的如卡特尔16PF）。

他认为，人格心理学家的任务是找出人格的那些最终的、不可分解的根源特质而非表面特质，以更好地理解人类行为。他收集了大量的素材（如系统分析了英语语言中所有可能的人格特质描述词，借用了阿尔波特等人的语料库），使用了大量方法（如同伴评价、自陈式测验、客观行为测量）和广泛的样本群（包括大学生、军职人员和工人等），最终将大量描述词简化为16个潜在的维度，即16种根源特质，并于1949年出版了含有187个题项的“16种人格特质问卷”，至今仍得到广泛应用，也为今天被广泛接受的“大五”人格理论奠定了基础。

瑟斯顿（Louis Leon Thurstone, 1887—1955），美国心理学家、心理测量学家。他最初学习电气工程，曾受教于大发明家爱迪生，后来转学心理学。1917年在芝加哥大学获得博士学位，先后在卡耐基技术学院（Carnegie Institute of Technol-

ogy)、芝加哥大学任教,曾担任美国心理学会主席,心理测量学会首任主席。其代表作有:《智力的本质》(The Nature of Intelligence, 1924)、《心智的向量》(Vectors of the Mind, 1935)、《基本心理能力》(Primary Mental Abilities, 1938)、《多因素分析》(Multiple Factor Analysis, 1947)。



瑟斯顿对因素分析技术的贡献主要是发展了多因素分析技术,并最早将因素分析法用于编制智力测验和知觉测验。他的“基本心理能力”智力理论主要是依据他在芝加哥大学实施的两项大规模研究的结果,即人的智力由大约 6 到 7 种基本心理能力决定而非斯皮尔曼所认为的只依赖于一般能力和某种特殊能力。这些基本心理能力相互联系但又彼此独立,个体在不同能力上各有强弱差异,因此不存在斯皮尔曼所谓的中心或核心的一般智力。根据其智力理论编制出的智力测验不再只给出个体的单一分数(如 IQ),而是会给出一个个体在各种测验上的得分剖面图,从而全面地反映其智力状况。

1934 年,瑟斯顿和夫人基于其智力理论发展了一套包含 56 个测验的成套测验,并在大学施测,通过采用一种叫做重心法的因素分析方法将数据区分成 13 种因素,后来删减为 7 种,即言语理解、字词流畅性、数字、空间、机械记忆、推理和知觉速度,在之后的深入研究中瑟斯顿又发现某些因素之间存在相关,可以抽取出一个“二阶一般因素”。他开创的这种群因素智力观随后被广泛接受,对后来的智力理论发展产生深远影响。

启示:心理学理论上的重大突破往往源于方法学上的进步。

心理学家们往往是在其解答实际生活问题的努力中做出重要贡献的。

没有对前人研究成果和资料的运用、积累、总结和批判是不可能有什么重大发现的。

三、SPSS 因素分析功能选项

在 SPSS10.0 中,探索性因素分析包含在 Analyze⇒Data Reduction 菜单中,在数据窗口下单击“Factor”选项即进入因素分析的主对话框。和通常的 SPSS 主对话框一样,该对话框除了变量框外还提供了多个功能子对话框,每个子对话框中又包含了许多选项,目前探索性因素分析已发展得相当成熟,研究者可以根据实际情况灵活地选择不同的方法来得到最终的结果。要完全理解和掌握这些不同的选项和方法的特点和功用对初学者来说并非易事,需要较深的数学知识。为便于大家尝试和进一步学习,将其名称和功能简列如表 13-2 所示,其具体用

途会陆续介绍。

表 13 - 2 SPSS 因素分析各对话框及选项功能

子对话框	功能	选项	选项功能
I : Descriptives (描述统计)	提供观测变量、初始解及相关矩阵有关统计量	①Statistics	是否计算单变量描述统计和初始解
		②Correlation Matrix	是否计算相关系数矩阵、显著性水平、决定系数、逆矩阵 (Inverse)、导出矩阵 (Reproduced) 及残差、反映像矩阵 (Anti-image) 以及 KMO 测度和 Bartlett 球度检验
II : Extraction (因子抽取)	提供因子抽取有关选项	①Method	七种解法依次是:主成分分析、最小二乘、广义最小二乘、极大似然、主轴因子法、 α 因子法、映像因子法
		②Analyze	采用相关阵或协方差阵求解,默认前者
		③Extract	规定抽取准则为最小特征值 (默认 > 1) 或指定因子数
		④Display	是否显示未旋转前因子解或碎石图
		⑤Maximum iterations for Convergence	指定因子解收敛的最大迭代次数,默认 25
III : Rotation (因子旋转)	提供因子旋转有关选项	①Method	旋转方法选择依次是:不旋转、方差极大法、直接斜交极小法 (输入 $\delta \leq 0.8$)、四次方最大正交法、平均正交法、快速斜交旋转 promax 法 (输入 $k > 1$)
		②Display	是否显示旋转后因子解 (需选择旋转) 或因子负荷图
		③Maximum iterations for Convergence	指定旋转收敛的最大迭代次数,默认 25

续表

子对话框	功能	选项	选项功能
IV: Scores (因子值)	提供因子 值计算及 处理选项	①Save as variables	是否将因子值作为新变量保存在数据文件中
		②Method	三种因子值计算法依次是:回归法、Bartlett法和Anderson-Rubin法
		③Display factor score coefficient matrix	是否显示因子值系数矩阵
V: Options (杂项)	提供缺失 值处理及 显示格式 选项	①Missing values	缺失值处理方式
		②Coefficient Display Format	因子负荷显示方式按绝对值大小排列或隐藏绝对值小于指定值因子负荷(默认0.1)

四、例题求解和结果阅读

实际上 对于初学者来说 完全可以不理睬这些子对话框 只按照 SPSS 中的默认选项直接选定分析变量进行分析即可。下面给出的便是根据例 13 - 1 中的相关矩阵使用默认选项得出的因素分析结果。表 13 - 3 给出的是用默认的主成分分析法抽取前后的公因子方差（共同度），从中可以看出各变量初始共同度（initial communalities）是对每个变量中由所有成分或因素解释的方差的估计，对主成分抽取过程而言，它始终等于 1（标准化的结果），抽取共同度（extraction communalities）是抽出的公因子所解释的变量方差估计，其取值较大表示抽取出的主成分能很好地代表原变量；反之，某个共同度取值较小则需要重新抽取，从表中看绝大部分都在 60% 以上。

表 13 - 3 共同度 (Communalities) 估计结果

	Initial	Extraction
COL1	1.000	.732
COL2	1.000	.704
COL3	1.000	.601
COL4	1.000	.595
COL5	1.000	.697
COL6	1.000	.530

表 13 - 4 给出的是根据特征值大于 1 准则抽取出的各因子特征值、贡献率和累积贡献率，分成两块，分别显示的是初始解、抽取主成分解释的方差。第一块区域为“初始特征值”(initial eigenvalues)，从中看出主成分分析的初始解并不减少变量维度，仍是 6 个主成分，且若基于相关矩阵求解，特征值之和等于主成分数目。第一列“Total”表示每个主成分的特征值或其所解释的原始变量变异量。依次，第二列表示各主成分所解释变量总方差的百分比；第三列表示前 n 个主成分解释总方差的累计百分比。第二块区域为抽出解，各公因子方差贡献可以用因素负荷平方和 (sums of squared loadings) 表示。一般抽取特征值大于 1 的主成分，所以该例中，前两个主成分被抽取作为公因子。2 因子解释了观测变量总变异的 64.33%，可以接受。因此可以用这两个公因子代替原来的 6 个变量。

表 13 - 4 初始解抽出解及各因子贡献率

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.633	43.889	43.889	2.633	43.889	43.889
2	1.227	20.444	64.333	1.227	20.444	64.333
3	0.684	11.393	75.726			
4	0.560	9.341	85.067			
5	0.505	8.416	93.483			
6	0.391	6.517	100.000			

表 13 - 5 给出的是各观测变量在两个主成分上的因素负荷矩阵或成分矩阵 (component matrix，即因素模式)，也即因素结构矩阵。结合这三个结果，计算一下有关的量：

抽取后变量共同度等于 2 因子负荷的平方和： $0.751^2 + (-0.410)^2 = 0.732$, $0.735^2 + (-0.404)^2 = 0.704$, $0.682^2 + (-0.369)^2 = 0.601$ 。自己试一下计算出其他的共同度。

因子 1 的特征值和解释的方差为： $0.751^2 + 0.735^2 + 0.682^2 + 0.652^2 + 0.554^2 + 0.575^2 = 2.633$

因子 2 的特征值和解释的方差为： $(-0.410)^2 + (-0.404)^2 + (-0.369)^2 + 0.324^2 + 0.537^2 + 0.605^2 = 1.227$

表 13 - 5 因素负荷矩阵 (Component Matrix)

	Component	
	1	2
COL1	0.751	-0.410
COL2	0.735	-0.404
COL3	0.682	-0.369
COL4	0.554	0.537
COL5	0.575	0.605
COL6	0.652	0.324

当然也可以根据因素负荷矩阵重新计算变量间相关系数，如变量 1 和 2 间的相关系数计算为： $r'_{12} = 0.751 \times 0.735 + (-0.410) \times (-0.404) = 0.718$ 显然，所有这些相关系数如果和表 13 - 1 中的数据相差不大表示因素分析模型拟合良好。实际上这个导出相关阵可以由 Descriptives 子对话框中的 Reproduced 复选项直接给出（见表 13 - 6），读者可以尝试和原相关阵作比较。同时还给出它和原相关阵的差值，称剩余相关阵（residual correlation matrix）。从表 13 - 6 中还可以看出，导出相关阵中的主对角线元素就是各变量的估计共同度。

表 13 - 6 导出相关阵 (Reproduced Correlations)

变量	col1	col2	col3	col4	col5	col6
col1	0.732					
col2	0.718	0.704				
col3	0.664	0.651	0.601			
col4	0.196	0.191	0.180	0.595		
col5	0.184	0.179	0.169	0.643	0.697	
col6	0.357	0.349	0.325	0.535	0.571	0.530

☆ SPSS 操作提示

通常情况下，因素分析可以基于原始变量数据，程序可以自动计算出变量间相关矩阵进行分析。某些特殊情况下，研究者也可以基于变量间的相关矩阵、协方差阵甚至已经做出的因素负荷矩阵来进行因素分析。但这时要借助于 SPSS 的 Syntax 语句来实现。这已超出了读者的接受能力，这里给出利用表 13 - 1 的相关矩阵进行因素分析的 Syntax 语句以供参考。分析之前，先要将原相关矩阵

输入或读入 SPSS 数据窗口，然后在相关阵前面增加两列标示变量 ROWTYPE_ 和 VARNAME_ 取值分别为 CORR 和 6 个变量名，再增加第一行数据 ROWTYPE_ 列取值 N, VARNAME_ 留空，其他 6 列变量取值 100 表明样本量。语句如下：

```
FACTOR
  /MATRIX = IN (COR = * )
  /MISSING LISTWISE /ANALYSIS col1 col2 col3 col4 col5 col6
  /PRINT INITIAL EXTRACTION
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /ROTATION NOROTATE
  /METHOD = CORRELATION .
```

第二节 因素分析的步骤

一个完整的因素分析过程要解决以下几个主要问题：第一，因素分析适合度的检验；第二，在给定的 k 个观测变量下确定合适的因素数 m ；第三，求解因素负荷矩阵 $[l_{ij}]$ ；第四，对因素作旋转，以便于进行合理的解释。实际上这四个相互关联的问题是因素分析争论最多、研究成果最丰富的问题，也是对于初学者来说最困难的部分，必须反复体会、反复尝试才能很好理解。本节的目的不在于讨论艰深的理论原理，而主要讲解如何运用 SPSS 软件完成一个完整的因素分析过程，并进一步熟悉 SPSS 因素分析命令中各子对话框的设定和功能。

SPSS 因素分析主对话框的结构不再赘述，逐一打开其 5 个子功能对话框（各子对话框选项的功能见上节表 13-2），可以对分析过程进行精细的设置。仍采用上节中的例 13-1 说明操作过程。

一、数据适当性考察和因素数确定

因素分析模型有一些假设条件，其对观测变量也有一些要求。筛选观测变量是一个复杂的问题，它涉及到研究理论构想、编制量表前开放式问卷项目的归纳整理和项目分析等。模型中包含了无关的变量或者剔除了相关的变量可能极大地影响分析的结果。这里给出一些常用的标准，可通过对数据的描述统计或者预分析加以考察，下面列出一些经验准则，符合这些条件的变量最好予以剔除或加以变形。

- 标准差过低，通常表现为观测变量中被试的反应趋同；
- 重测信度低（建议 0.2）；

- 最大负荷值 l_{ij} 小 (建议 0.4) ;
- 共同度 h_i^2 小 (建议 0.16) ;
- 最大负荷值 l_{ij} 与共同度 h_i^2 之比小 (建议 0.5) ;
- 最大两负荷值 l_{ij} 与共同度 h_i^2 之比小 (建议 0.25) ;
- 取样适当性系数 (MSA) 过小 ;
- 多极变量 , 即一个变量在几个因子上的负荷都较大。

(一) 样本数据适当性考察

样本数据适当性考察实际上牵涉到因素分析模型的拟合度检验问题 , 有兴趣的朋友请参考文献。

考察数据适当性 , 首先可以看样本相关矩阵 , 如果变量之间相关程度普遍较低则不大可能找到便于解释的公因子或者达不到简化数据的目的 , 一般大部分相关系数应当不低于 0.3 (郭志刚, 1999) , 但过高的相关可能表明存在 “多重共线性” , 变量间的信息存在冗余。此外 , SPSS 还提供了其他一些度量来进行检验 :

(1) Bartlett 球度检验 (Bartlett's test of sphericity)

通过构造一个近似 χ^2 统计量从整体上检验相关矩阵 , 即 H_0 : “相关矩阵是一个单位阵” , 显然 , 其显著性水平要至少小于 0.05 才能拒绝 H_0 , 说明各个变量间存在相关 , 适宜进行因素分析。例子中 , 近似 χ^2 统计量 = 135.516, $P < 0.001$ 拒绝 H_0 。

(2) 反映像相关矩阵 (Anti-image correlation matrix)

其元素等于偏相关系数的负数。偏相关系数反映了在排除其他变量影响下变量间的相关系数 , 公因子存在时 , 偏相关系数实际上是特殊因子 (一般被视为误差) 间的相关系数估计 , 应当接近于零。所以一个好的因素模型中反映像相关阵里远离对角线的元素 (绝对值) 应该比较小。读者可自行操作 , 例中绝大部分远离对角线的元素都小于 0.1 , 可以接受。

(3) KMO 取样适当性度量 (Kaiser-Meyer-Olkin measure of sampling adequacy)

用于比较观测变量间的相关系数与偏相关系数的相对大小 , 它是变量间相关系数平方和占这两种系数平方和的比率。显然 , KMO 值越接近 1 越好。一般规定 : 0.9 以上 , 极好 ; 0.8 以上 , 较好 ; 0.7 以上 , 一般 ; 0.6 以上 , 较差 ; 0.5 以上 , 差 ; 0.5 以下 , 不可接受。同时 , 每个变量的 KMO 值恰好为反映像相关矩阵的对角线元素 , 记为 MSA (Measures of Sampling Adequacy)。KMO 度量可以考察 “多重共线性” 问题。例子中 , KMO 值为 0.749 , 可认为比较好。

☆ SPSS 操作提示

以上度量或检验方法在 SPSS 因素分析功能中都是 Descriptives 子对话框下

的复选项，如图 13.2。该对话框分为 Statistics 和 Correlation Matrix 两部分，选中 Statistics 部分的两个复选框分别完成单变量的基本描述统计（均值、标准差）和给出初始解。

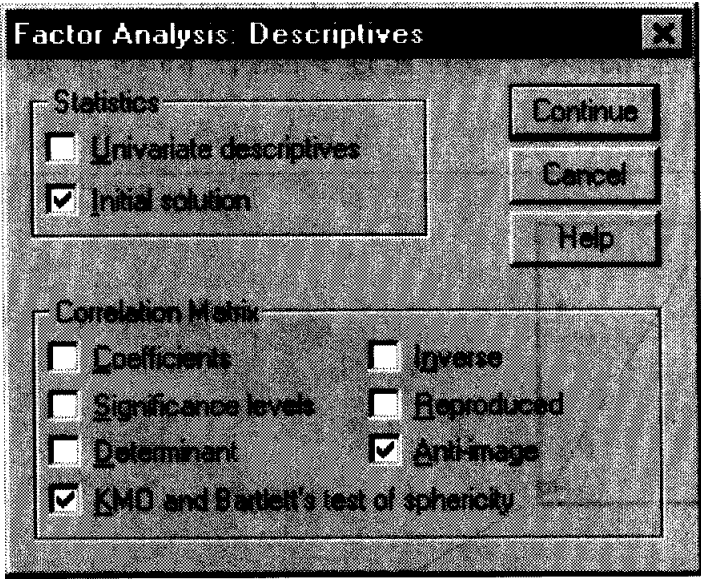


图 13.2 Descriptives 菜单子对话框

（二）确定公因子数

大多数因素分析问题中，公因子数目是未知的。在主成分分析中，有多少个变量就有多少个主成分，但因素分析是要在不遗失太多信息的前提下尽量减少公因子数。但是，因子数选择多少才恰当呢？从因素结构拟合原相关矩阵的角度，研究者提出了变量数与因素数间的制约关系和因素数的若干边界（芝祐顺，1999）如在公因子方差未知情况下 因子数 m 和变量数 n 必须满足：

$$m \leq \frac{(2n + 1) - \sqrt{8n + 1}}{2} \quad (13 - 5)$$

实际应用中，常根据一些具体情况下求出的公因子特征值的大小作为甄选因子的标准，比较公认的方法是 Kaiser 准则(Kaiser criterion, Kaiser, H, F, 1960)：即选取特征值大于 1 的主成分作为初始因子。因为原始变量标准化后的方差为 1，代替它的公因子方差应当至少能解释原来一个变量的方差，否则代替就没有意义了。Kaiser 准则常常作为求解初始因子时使用，但当变量数少于 20 时 抽取出的公因子往往偏少，而变量数大于 50 时，抽取出的公因子又往往偏多。类似的还有将临界特征值定为 0.7 的 Jolliffe 准则等，其选出的因子数可能更多。例 13 - 1 中根据 Kaiser 准则前两个主成分被抽取为公因子。有鉴于此，卡特尔 (1966) 提出了一种利用特征值变化趋势确定因子数的陡阶检验准则 (scree test criterion)，即按照主成分的提取顺序（因子贡献递减）将其特征值画成直观的碎

石图(scree plot) 如图 13.3 所示，图中纵坐标表示初始解中各主成分的特征值，通常可以选取陡坡上的主成分。例 13 - 1 中前两个因素大的陡急坡度与后面因子的缓慢坡度之间有明显的转折点，表明第 3 因素后的主成分贡献甚小，可以舍弃，所以选取前两个主成分。这个结论和上面的 Kaiser 准则相同。一般使用“陡阶检验”准则选出的因子数可能过少，结合 Kaiser 准则可以收到较好的效果。

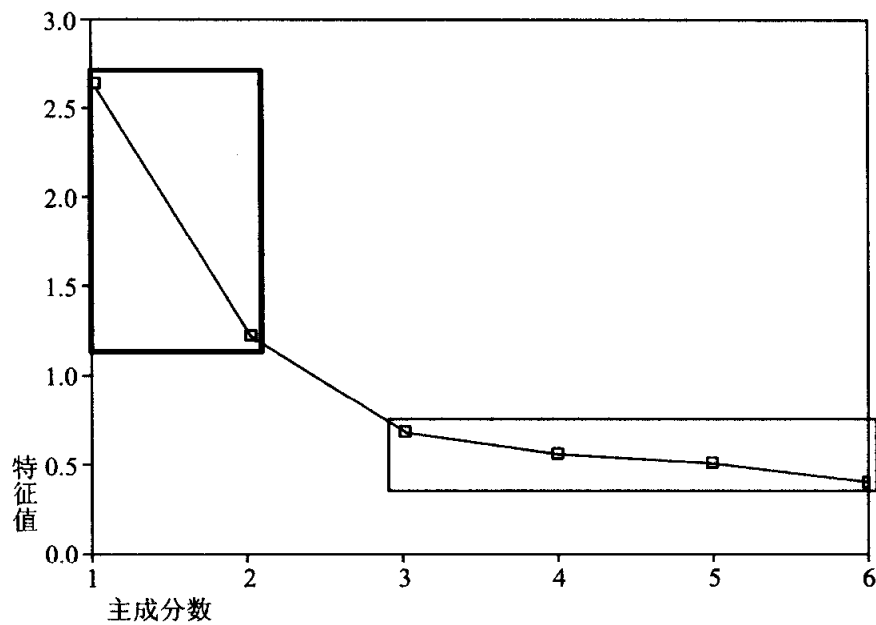


图 13.3 陡阶检验碎石图 (scree plot)

此外，一些特殊的参数估计方法也有不同的确定因素数目的准则，如在极大似然估计法中根据模型拟合度的变化决定因素数（即增加因子不会使拟合度显著性水平显著减小）等（芝祐顺，1999）。不论采用哪种准则，关键是看保留的因子解释了原变量的方差多少，由于被连续抽取出的因子所解释的变量方差越来越少，因此，在初始分析结果中考察前两个因子累计贡献率（见表 13 - 4）也是一个重要参考，一般这个比例应达到 80% ~ 90% 以上。当然，在实际应用中往往很难保证这个比例，有些研究者甚至也接受 50% 以上的比例（例 13 - 1 中将近 65%），因此上述准则都不是客观标准，往往还要根据具体情况具体分析。同一组原始数据其因素数目并不是一成不变的，需根据研究的构想、继往研究的结论，特别是根据因子的可解释性等来衡量最恰当的因素数目。

☆ SPSS 操作提示

SPSS 在 Extraction 子对话框中提供了 Kaiser 准则和碎石图检验的选项，其中 Kaiser 准则是系统默认准则，如前所述，用户也可以实际情况选择改变临界特征值（如变成 0）或者指定抽取因子数目，后者通常用于既定因素模型的验证或

者已确定因子数情况下求解因素负荷矩阵的情况下。

二、求解因素模式

所谓求因素分析的解就是计算出能够说明原始变量相关关系的因素模式，即因素负荷矩阵 $[l_{ij}]$ 。求因素解有许多不同的方法，其依据的原理也各有不同。前面所讲的主成分分析法（PCA）实际上是一种独立于因素分析的数据化简方法，也是 SPSS 默认的分析方法。

与主成分分析不同，另一类方法是以主轴因子法（principal axis factoring）为代表的主因素分析（principal factor analysis）法（PFA）或公因子分析（common factor analysis）法。此类方法着眼于解释变量间的相关关系，即假设公因子可以完全解释变量间的相关关系，而不一定能完全解释变量的变异（即不考虑特殊因子）。因此，此类解法首先面临的问题是估计公因子方差，即初始共同度估计。

（一）共同度估计

同样是基于相关矩阵，主成分分析法和公因子模型的分析方法对于相关矩阵对角线上元素的处理是不同的：前者都是 1，不存在估计；而后者则是以某种估计量代入对角线。常用的处理方法有：

（1）反复迭代法：用上一次因素分析的结果（如主成分分析的抽取共同度）作为下一次分析的初始共同度代入主对角线，反复迭代直至结果差异很小时。此法受选定的因素数目影响较大，但对因素负荷量影响很小。

（2）最大相关系数法：用每个变量和其余变量相关系数中（即相关矩阵中非对角元素）绝对值最大者作为估计共同度，此法当变量数较少时效果可能不理想。

（3）SMC 法：用每个变量和其余变量的复相关系数平方（squared multiple correlation, SMC） R^2 ，此法是对最大相关系数法的改进，对于因素数显著地少于变量数时十分有效（Tucker et al. 1969），也是许多因素模式解法中默认的方法。

因素分析中，因素数目和共同度估计关系密切，往往需要研究者根据研究的目的进行权衡，越需要准确地探求因素的场合，对共同度估计的要求越高。共同度大小说明了变量能被公因子解释的程度，但还要视因子的可解释程度——很高的共同度并不一定能很好地解释因子，而较低的共同度却可能有很好的因子解释力。

（二）因素模式解法

主成分法和公因子法是因素分析的两类基本方法，主成分法的其核心是在方差不变的前提下求解相互独立的主成分。在抽取因子时，其初始共同度总是 1。主成分法对总体的分布没有什么限制，因此适用范围广，因素分析常将其结

果作为初始解。如果说主成分法适用于用少量因子代替原变量从而简化数据的场合，那么公因子法则更适用于检测变量间的数据结构或者对变量方差不了解的场合，由于它不考虑变量方差中不能被因子解释的部分，因此在实际应用中更常被使用。

SPSS 中提供的常用公因子法有（如图 13.4）：

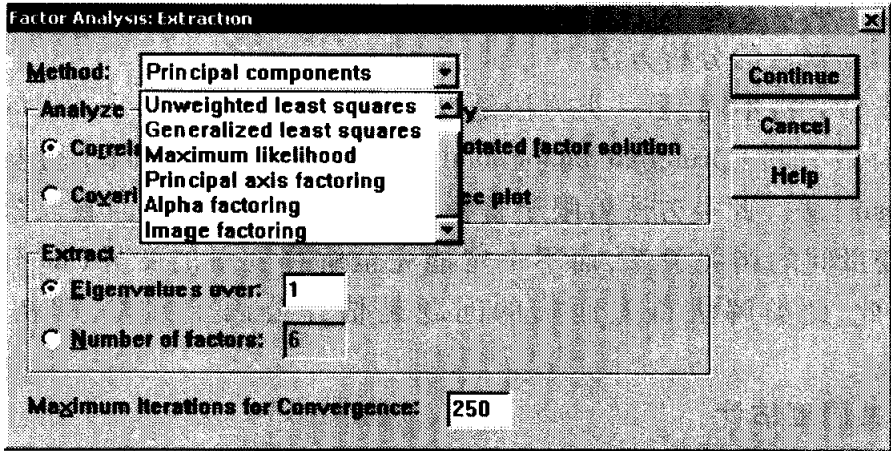


图 13.4 Extraction 选项子对话框

（1）主轴因子法

前已述及，其程序类似主成分法，不同的是用估计共同度（SMC）代替相关矩阵主对角线上的元素，形成的新矩阵称为调整相关矩阵（Adjusted correlation matrix），然后解该矩阵的特征方程得到因子解。

（2） α 因子法（alpha factoring）

此法借鉴心理测量学中同质性信度的概念，认为其他抽取方法都未考虑变量的抽样代表性问题而假设其固定，它认为因子分析中包括的变量是来自潜在变量空间中的一个样本，抽取出的因子解应当使抽取的公因子和假设的公因子有最大相关，因此只有 α 系数为正数（特征值必然大于 1）的公因子才被抽取（Kaiser & Caffrey, 1965）。此法一般适用于决定公因子。

（3）映像分析法（image factoring）

变量中可以由其他观测变量预测的部分称为该变量的映像（image），即 R ；另一部分为该变量的特有部分，称为反像（anti-image）。该法利用映像理论进行分析，其对共同度的处理也使用 SMC 估计。映像分析法假设数据为从被试总体所测得的变量总体，因此结果受抽样影响较大，往往难以做出合理解释，抽出的公因子累计方差贡献往往低于其他方法。

上述公因子法主要是立足于使求出的因子模式尽可能“再现”变量间的相关关系，由于变量的特殊变异只影响原相关矩阵主对角线元素，因此只要求其因

子解能拟合非主对角线元素，而公因子方差只是因子解的副产物。而下列的两种公因子法则立足于因子解应当使得导出相关阵和观测相关阵之差（即剩余相关阵）列向量元素的平方和达到最小，称为最小残差平方和（minimum squared residual, MINRES; Harman & Jones, 1966）。类似于验证性因素分析（CFA），这两种方法借助于卡方统计量度量导出阵对原相关阵的拟合优度（goodness of fit）。这两种方法的求解过程比较复杂，计算机运算耗时也较长。但其对公因子只能解释部分变量方差的认识本质上仍和其他公因子法一样。

（4）最小二乘法 least Squares）

该法回避了对共同度的估计，包括一般最小二乘法（unweighted least squares）和广义最小二乘法（或加权最小二乘法 generalized least squares），二者的区别是后者在迭代过程中每次用特殊因子方差的倒数调整相关矩阵，给特殊因子方差更大的变量的相关系数赋予更大的权值。

（5）极大似然法 maximum likelihood）

该法假定观测变量来源于多维正态总体，通过构造样本的似然函数（likelihood function）（Tucker & MacCallum, 1997）使其达到最大求得因子解。类似于广义最小二乘法，也用特殊因子方差的倒数加权来调整相关矩阵。

在实际应用中选用何种抽取方法不仅要看有关模型的一些假设（如是否服从多维正态分布、数据是否正态化等），更要结合因子分析的目的和对变量方差的了解程度。变量数较少时，主成分法可以最大限度地保证因子的累计方差贡献；随着变量数的增加，相关矩阵主对角线元素的重要性降低，两种方法的差异越来越小，这时使用公因子法应当更关注公因子方差的估计方法而非因子解，可以考虑多种方法相互参照。而在样本量很大时，后两类基于模型拟合的方法精度更高。牵涉到因子数目的决定，通常可以使用几类方法，结合特征值大小及变化趋势以及对模型的拟合优度来综合权衡。

☆ SPSS 操作提示

图 13.4 显示的 Extraction 对话框中除可以选择抽取方法外，还可设定分析矩阵是基于相关阵还是协方差阵。注意到上述的所有分析过程都是基于标准化数据，此时相关矩阵就是协方差阵。由于单位的不同，基于相关阵和基于协方差阵计算出的因素解可能差异很大，通常为统一各变量的单位，都使用相关矩阵，只有在原始数据集丢失而只有协方差阵这样的特殊情况下才使用该选项（分析过程要借助于 Syntax 才能实现）。

此外，大多数因素抽取方法计算共同度时都需要对相关阵进行迭代（iteration）计算，特殊情况下，SPSS 默认的迭代次数（25）可能不够，则需要手工将其改大一些，一般可参照结果中显示收敛（converge）需要的迭代次数。同样的操作也见于旋转对话框中，如图 13.5。

三、因素的解释和命名——因子旋转

因素解求出后，因素分析过程并没有结束，这时的因子解并没有什么实际意义。求初始解是按照因子解释变量方差的多少顺序提取因子的，因此，绝大多数变量在第一个因子上肯定都有显著的负荷，其后的因子解释方差比例逐渐递减。这时的因素负荷矩阵中大多数因子都和许多变量相关（见表 13-5）尤其是靠前的第一因素，负载了过多的变量，很不便于解释。这时就需要改变因子轴的位置，重新分配各因子所解释的方差比例，从而得到易于解释的“简单结构”，这个过程常称为因子旋转（rotation）。根据矩阵代数的知识，因子旋转实际上是对因素解进行的一个变换（正交旋转是标准正交变换），得到的模型对数据的拟合程度不发生改变，同时也不改变每个变量的共同度。

（一）简单结构规则

简单结构规则（rule of simple structure）是瑟斯顿 L L Thurstone, 1947 提出的五个使因素与变量间具有简单的单纯关系以便于解释的因素变换规则（Tucker & MacCallum, 1997）。进一步简化，这个规则一般可表述为：

- （1）在各因子上只有少数变量有较高的负荷，其他变量上的负荷（绝对值）很低；
- （2）每个变量只在少数因子上有很高的负荷；
- （3）任取两因子，负荷都低的变量尽量多一些；
- （4）任取两因子，每个变量只能在一个因子上有较高负荷。

笼统地说，因子旋转的目的就是调整因素模式中的行、列值向 0 和 ± 1 极化，使某些变量的负荷尽可能往某个因子上集中，而另一些变量的负荷尽可能往另一个因子上集中，以此类推，使得每个因子上仅“负载”几个变量。根据因素旋转时对因子解所作的变换的性质（如是否是一个标准正交变换），因子旋转可分为正交旋转（orthogonal rotation）和斜交旋转（oblique rotation）两种方式：正交旋转是使因子轴之间保持 90 度角（因子不相关），而斜交旋转中因子轴间不一定正交。这里主要介绍几种常用的正交旋转法，而斜交旋转留待下一节中具体介绍。

（二）正交旋转法

SPSS 中提供的正交旋转包括三种最常用的方法，其差别仅仅在于因素负荷矩阵中对行、列负荷处理的具体准则不同（如图 13.5），它们也都同属于所谓的“正交极大准则”（orthomax criterion）（范畴 芝祐顺, 1999）。

（1）方差最大法（varimax）

旋转因子轴使各因子（列）上与该因子有关的负荷平方的方差最大，即拉开列上各变量的负荷差异，这样某些小的负荷可以忽略，从而减少该因子

负载的变量数。这时的因素结构是最容易解释的，因此是最常用的旋转方法。

(2) 四次方最大法 (quartimax)

从因素负荷矩阵的行出发，旋转因子轴使各变量（行）上因子负荷平方的方差达到最大，即拉开行上的负荷差异，使每个变量只在因子（列）上一个有较高负荷而在其他因子上的负荷尽可能低。但该方法的一个缺陷是容易产生综合因子，大部分变量在该因子上都有较高负荷，实际上是通过牺牲对因子解释的简洁性来换取对变量解释的简洁性。

(3) 平均正交法 (equamax)

或平均最大法，是将前两个规则进行加权平均后的准则。该法综合考虑了对因子和对变量的可解释性。注意在因子数为 2 时，其结果等同方差最大旋转。

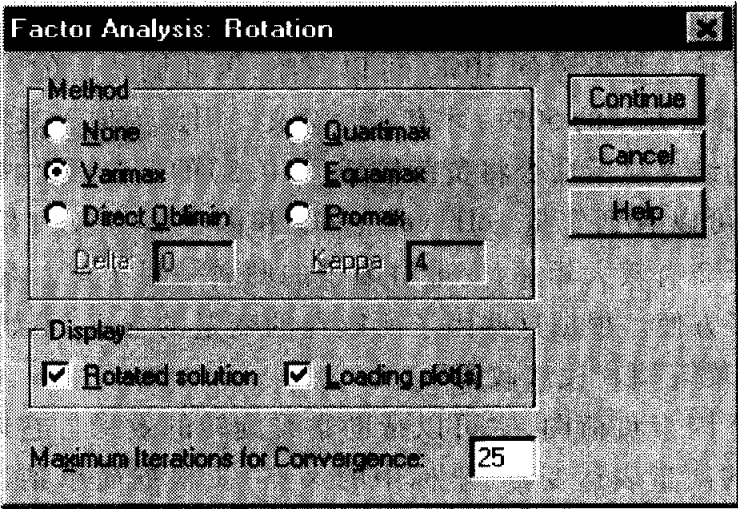


图 13.5 Rotation 选项子对话框

(三) SPSS 因素旋转操作及结果阅读

例 13 - 1 中，如果不按照 SPSS 默认的方式进行分析，希望进行因素旋转，在 Rotation 子对话框中选择 Varimax 旋转，自动增加显示旋转解，选中复选框因素负荷图(loading plot)，其他选项不变再进行分析。分析结果除给出主成分分析法的初始解、抽出解外，还给出了旋转解。表 13 - 7 显示的是旋转前后因子负荷平方和（特征值和贡献率）。可以看出它的变化，在抽出解的基础上，旋转解保留了原来的 2 个因素，累计方差贡献不变，但各因素特征值大小平均了，相应地，所解释的方差比例相比抽出解也更加平均了，2 因子各解释了约 35% 和 30% 的方差。这正符合各因素对原变量信息的解释能力相当的期望，不至于一些因素

负载过大，而另一些因素过小。

表 13 - 7 抽出解和旋转解

Component	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.633	43.889	43.889	2.082	34.702	34.702
2	1.227	20.444	64.333	1.778	29.632	64.333

再来看经 Varimax 法旋转后的因素负荷矩阵（见表 13 - 8），这就是一个比较典型的“简单结构”。经过对列上负荷的调整，除变量 col6 外其余变量只在 1 个因子上有较大的负荷。根据因素负荷的意义，变量在某因素上的负荷量为 0.3 表示因素解释了该变量方差的 10%，可以忽略，因此调整后的因素负荷最好超过 0.7(解释原变量方差的 50%) 才有实际意义。因此为便于观看，可通过 SPSS 选项对话框中的结果显示命令将负荷较低（<0.3）的值隐藏并按负荷大小排列（如图 13.6）。可以看到变量 col1、col2、col3 和因子 1 有较大相关，变量 col4、col5、col6 和因子 2 有较大相关，变量对因子的归属一目了然。客观上为这些因子的命名提供了方便。根据这些因子所解释的变量的含义，不妨将因子 1、2 分别命名为“经济政策”和“民族政策”。

最后看一下因素负荷图，它可以帮助察看旋转的效果。它以因子为轴，因子负荷为坐标的向量表达变量（坐标中的 x 点）。根据简单结构规则，良好的旋转应当使得这些向量尽可能落在坐标轴附近，且各坐标轴附近积聚的变量数或者各因素的累计贡献应大致平均。在二维坐标系中，对于正交模型的正交旋转很容易理解（对比图 13.7 左右，左图是在不旋转情况下做出的因素负荷图），只需把坐标轴旋转到尽可能接近变量处即可。多维坐标系中的旋转方法就是将其分解成多个二维平面，分别进行简单结构的旋转，直到得出稳定、一致的结果。因素负荷图的另一个功用是可以帮助判断极端变量的存在，即个别无法通过旋转使其负荷集中到某因素上的变量可以考虑删除并重新进行因素分析。结果中还提供了旋转对应的因素变换矩阵（component/factor transformation matrix）。读者可以试一下，将该矩阵右乘旋转前因素负荷矩阵就得到旋转后因素模式。

表 13 -8 旋转后因素模式 (Rotated Component Matrix)

	Component	
	1	2
COL1	0.842	
COL2	0.826	
COL3	0.763	
COL5		0.832
COL4		0.765
COL6	0.306	0.661

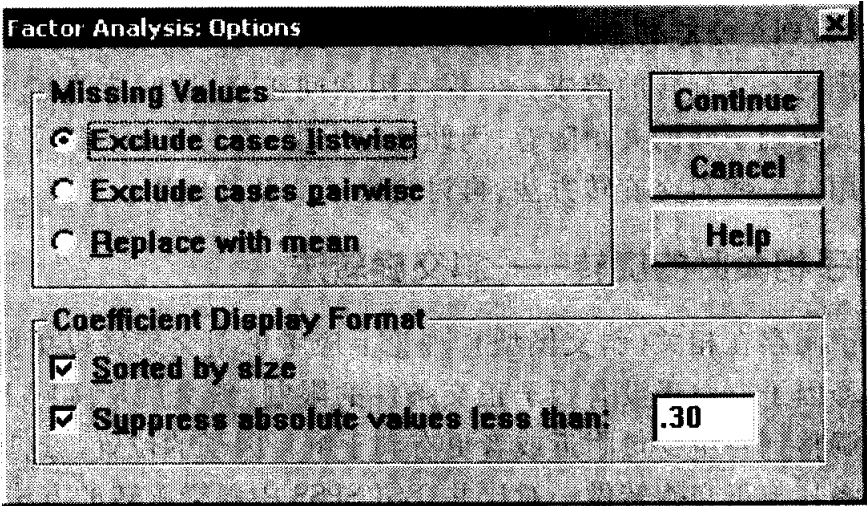


图 13.6 Options 选项子对话框

综上所述，该例中经方差最大法旋转后的因素解是可以接受的，利用主成分分析法得到了两个相互独立的因素用来代替原来的 6 个变量并对其进行了命名，信息损失了大约 35%。本节没有列举用公因子法进行因素分析的例子，有兴趣的读者可以选用其他因素模式解法进行尝试。到此为止，一个完整的因素分析过程即告结束。但大多数因素分析研究的目的还不仅在于探索公因子，还要进一步利用分析结果进行评价、分类，这种情况常见于使用量表或问卷对被试进行评判。这些内容将在下一节中介绍。

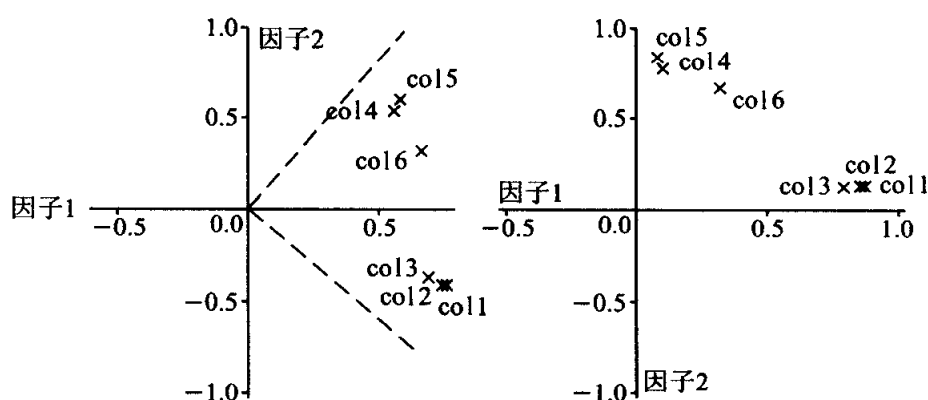


图 13.7 旋转前后的因素负荷图 (Varimax)

第三节 斜交旋转和因子值应用

如果因素分析的目的只是要简化数据，而不是要得到理论上有意义的几个因子，则通常做到正交旋转整个因素分析即可结束。但实际研究中，因素间的关系往往很难满足正交旋转的要求——保持因素间正交，这时应当考虑使用斜交旋转。斜交旋转因子间的夹角随意，因此理论上说，它对于解释因子更有利。本节先介绍常用的一些斜交旋转方法，再讨论因子值及其应用。

一、因子相关时的旋转——斜交旋转法

如果因子间存在相关（斜交模型），理论上应考虑斜交旋转以更好地形成简单结构和解释因子。由于没有因子正交条件的限制，斜交旋转实际上就是尽可能将因素轴调整到各组变量附近或者更有利于解释因子的位置。这里结合旋转的图形意义将其与正交旋转做一对比并介绍 SPSS 中斜交旋转的方法。

（一）正交旋转和斜交旋转的对比

（1）斜交模型的因素模式和因素结构

要了解二者的差异，须重新提及“因素模式”和“因素结构”两个概念。在旋转因素空间中可以用因素负荷向量表达变量作出因素负荷图，同样，在斜交模型中也是如此（如图 13.8）。设 2 因素空间中变量 x_i^* (x_i 去除残差项) 在因子 f_p 上的负荷为 l_{ip} 因子 f_p 和因子 f_q 之间的相关系数为 a_{pq} ($p=q$ 时为 1) 则由第一节中的因素分析模型和矩阵知识可以推知，变量 x_i 和因子 f_p 间的相关系数为：

$$r_{ip} = |x_i^*| \cdot |f_p| \cdot \cos \theta = |x_i^*| \cdot \cos \theta \quad (13-6)$$

$$r_{ip} = l_{ip} a_{pp} + l_{iq} a_{pq} = l_{ip} + l_{iq} \cos \phi \quad (13-7)$$

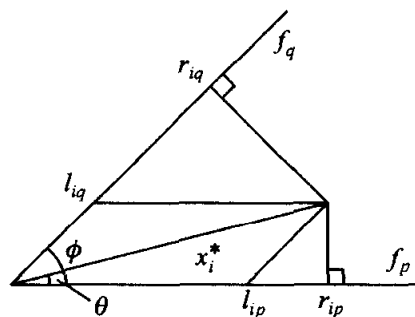


图 13.8 斜交模型中的模式负荷和结构负荷

变量 x_i^* 的向量长度乘以因子的向量长度和变量 x_i^* 和因子 f_p 间的夹角 (θ) 余弦, 相当于向坐标轴的垂直投影 ((13-6) 式, 因因素轴向量长度为 1), 也等于所有与该变量有关的因素负荷与因子间相关系数乘积之和 ((13-7) 式)。特别地在正交模型中, 因素间相关阵为单位阵, 所以因素模式等于因素结构 ((13-7) 式)。但在斜交模型中, 真正的因素负荷 (因素模型中因素前的系数) 不是向坐标轴垂直投影而是平行投影的结果。所以在斜交模型中, 因素模式和因素结构是不同的, 有些资料中也将其元素分别称为模式负荷 (pattern loading) 和结构负荷 (structure loading) 图 13.8 显示模式负荷要小于结构负荷, 变量 x_i^* 在 f_p 轴上的负荷大于在 f_q 轴上的负荷, 所以应当归为因子 f_p 。

(2) 斜交模型的简单结构

斜交模型区分出因素模式和因素结构, 相应地二者都存在简单结构的问题。表 13-9 显示的是用例 1 的数据进行的斜交旋转结果, 从因素模式看, 更接近于一个标准的简单结构, 特别是变量 col6 在因子 1 上的负荷更小了, 但从因素结构看则离简单结构较远。反之, 具有简单因素结构的斜交解也不一定符合简单模式结构的条件。通常大多数研究者更愿意报告模式矩阵, 因为结构矩阵中包含的因素间相关关系往往不便产生简单结构, 当因素间相关系数较大时就很难区分某变量是否单独负载于某个因素了。而模式矩阵元素是斜交因素线性组合的系数, 因此类似于标准化回归系数, 它考虑了其他因素对变异的解释。和正交解的一个区别是, 斜交解的因素负荷可能会超过 1。也有一些方法是从寻求简单因素结构入手, 例如下面要提到的 Procrustes 旋转。具体要求何种简单结构则要视研究者的侧重点了。

(二) SPSS 斜交旋转方法

由于斜交旋转结果的争议性, SPSS 提供的斜交旋转法并不多, 常用的有两种, 都需要进行参数设置:

表 13-9 例 1 斜交旋转后模式矩阵和结构矩阵

	Direct Oblimin Rotation($\delta = 0$)				Promax Rotation($k = 8$)			
	模式矩阵		结构矩阵		模式矩阵		结构矩阵	
	因子 1	因子 2	因子 1	因子 2	因子 1	因子 2	因子 1	因子 2
col1	0.855	0.002	0.856	0.294	0.863	-0.020	0.855	0.305
col2	0.839	-0.001	0.839	0.286	0.847	-0.022	0.839	0.297
col3	0.774	0.004	0.776	0.269	0.781	-0.016	0.775	0.279
col4	-0.040	0.784	0.228	0.771	-0.052	0.790	0.245	0.770
col5	-0.080	0.859	0.213	0.831	-0.093	0.865	0.233	0.830
col6	0.199	0.636	0.416	0.704	0.191	0.634	0.430	0.706

(1) 直接斜交极小法 (direct oblimin)

最常用的斜交旋转法是在广义斜交极小法的基础上提出的一种利用因素模式简单结构准则进行的斜交旋转法 (Jennrich & Sampson, 1966), 由参数 δ (Delta) 控制倾斜程度, 该参数控制因子轴的倾斜程度, 其中 $\delta \leq 0.8$ 。一般 δ 取大负值时表示因子间倾斜程度越低, 越不相关。

(2) 快速斜交旋转法 Promax)

是一种强调在因素结构外部寻找旋转准则的斜交旋转法 (Hendrickson & White, 1964), 即在积累了大量关于变量内容和多次因素分析资料基础上构造一个理想的因素结构——假设矩阵, 然后寻找能使得实际因素结构最大程度拟合假设矩阵的旋转变换。它可以看成是普罗克拉斯提斯*变换 (Procrustes transform, Hurley & Cattell, 1962) 的一种特例, 通过扩大初始简单结构中元素的大小差距 (2 次方或 4 次方) 来取得简单结构。其参数 k (Kappa) 即乘方数, $k > 1$ 。此法比直接斜交旋转法的计算速度快, 因此常用于大数据集。

斜交旋转结果往往受分析者对斜交参数的定义影响, 很大程度上取决于分析者的主观经验, 导致斜交旋转的“高风险性”, 同时也不利于研究结果的交流。因此碰到因子间高度相关的情况, 往往代之以减少因子数目或者做高阶因素分析, 导致斜交旋转在实际应用中的功用被削弱。

前面的例 1 中采用正交方差极大法旋转结果, 变量 col6 在第一因子上的负荷高于 0.3 (见表 13-8), 分别采用直接斜交极小法 ($\delta = 0$) 和 Promax 斜交旋转法 ($k = 8$) 求得结果如表 13-9, 负荷阵和结构阵都十分类似。由于斜交解的求得实际上是针对抽取解, 所以用主成分分析法抽取出的公因子仍然可以进行斜交旋转, 上面的旋转就是根据例 1 中的主成分解而来的。斜交旋转结果不再显

作者注: 普罗克拉斯提斯 (Procrustes), 古希腊传说中的强盗, 拦路截人并将其放在铁床上, 比床长的人, 砍去长出的部分; 比床短的人, 强行将其拉长。可见该方法的特点。

示变换矩阵而代之以因素间相关矩阵，两种方法结果分别为 0.341、0.377 表明两因子间存在着中等程度的相关。比较两种方法得到的因素负荷图和图 13.7 的差异（斜交解因素负荷图也用直角坐标系表示）。读者也可以自己改变旋转参数进行尝试。需要强调的是斜交解中变量的共同度不能通过因素负荷矩阵中对应行的负荷平方和求得；同样，因子解释的方差也不能用对应列的负荷平方和求得。所以 SPSS 结果中无法列出斜交解的因子贡献。

至此，得到了比较满意的结果，因素分析可以告一段落了。但因素分析要解决的问题原不仅仅如此。如当斜交旋转后因素间的相关关系很难解释时，许多研究者会借助所谓“阶层因素分析”（hierarchical factor analysis）的程序将“一阶”斜交分析中得到的因素间相关矩阵作为“二阶”分析的出发点抽取出二阶因子，则变量的变异又分解成二阶因子解释的“公共变异”和一阶因子单独解释的“独特变异”。此处就不再详细介绍。上述因素分析的步骤如图 13.9 所示。

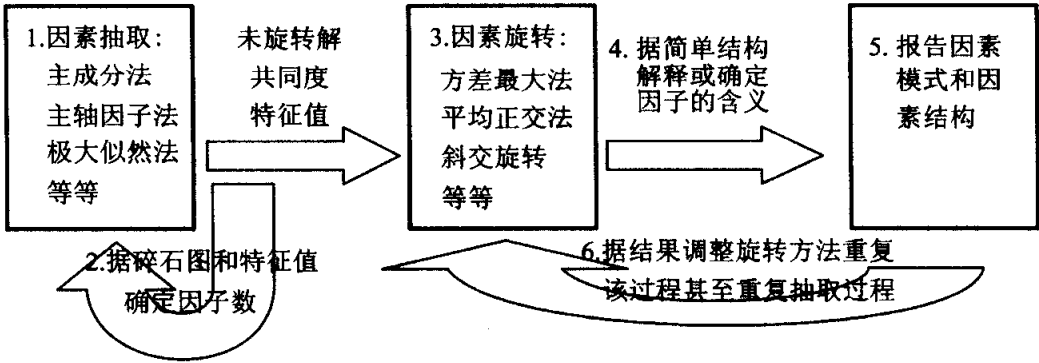


图 13.9 因素分析流程图

注意上述步骤并非一次即能达到理想效果，而可能要经过多次反复。解释因子时要最终赋予因子理想的含义，往往需结合专家意见或者借鉴以往研究。一个简单的规则是至少 80% 的变量都能够分配给适当的因子。

二、因子值的意义及应用

根据第一节中提到的因素分析数学模型，因子解使得能够用少量因子线性地表达一组观测变量，其中因素负荷即线性组合的系数。某些情况下，研究者还希望利用因素分析的结果进行进一步的研究，如根据各因子得分对某个自变量或样本进行分类、评价。这时客观上还要获得因子的度量，给出因子在每个案例上的值。当然，可以选用因素中负荷特别高的变量求和，或者如总加量表（sum-mated scale）那样求因素中各变量的总和或平均数来表示，但注意到各变量在因子上的负荷不同，所以应当赋予变量不同的权值，即因子值。实际上，求因子值的过程就是求因素分析模型的逆过程，目的是用观测变量的线性组合来表达因

子。计算案例 i 在因子 p 上的因子值是用该案例每个变量的标准化分数 x_{ji} 乘以相应的因子值系数 (component's score coefficients) w_{pj} 之和。

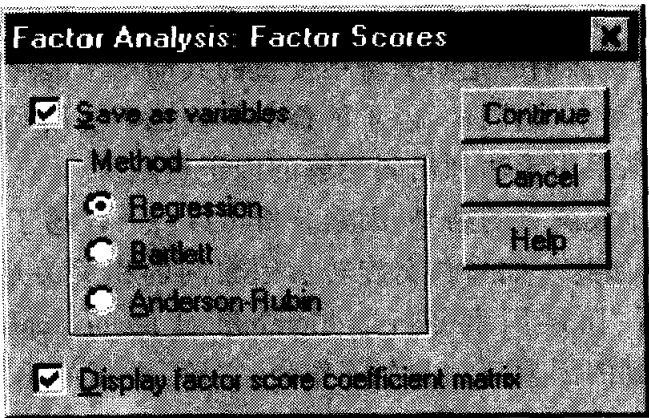


图 13.10 Factor Scores选项子对话框

因子值系数通常需要估计，但对于主成分法未经旋转求得的因子解可以直接得到因子值系数，通常是相应的因素负荷比上该因素的特征值。因此若不比较因素值单位，此时因素负荷就是因素值的估计。因子值系数的估计方法很多，SPSS 提供的三种计算因子得分的方法（见图 13.10）实际上都是基于最小二乘原理估计因子值系数的方法，分别为：回归法、Bartlett 法和 Anderson-Rubin 法，只是定义误差的方式不同。其中回归法是求解使真因子得分和因子得分估计值的误差平方和达到最小的因子值系数，这样得出的因子得分可能相关，它是 SPSS 中默认的方法；Bartlett 法的误差是独特因素得分估计值，而 Anderson-Rubin 法在其基础上增加因素间相互正交的条件。SPSS 可以通过 Scores 子对话框中的 Save as variables 复选框将标准化因子值作为新变量保存在当前数据文件中，计算出的因子值均值为 0，默认的变量名为 FAC1_1、FAC2_1、FAC3_1（分别对应因子 1、2、3）等，其中第二个数字表示第一次分析过程。例 1 中由于没有原始数据，所以无法计算因子值。

要全面地理解和掌握因素分析法还需要学习许多知识，要经过许多练习和反复尝试。但通过本章的学习，相信对于大多数实际应用者来说已经足够。为了完整地呈现利用统计软件实施因素分析法的全过程，读者可结合下面的实例进行模仿学习，尝试用不同的方法多做几次。进一步体会如何运用因素分析思想解决科学研究和实际生活中的问题。

实例学习 13 - 1

例 13 - 2 问题取自 *Analyzing Multivariate Data*（多元数据分析）一书英文

版(Lattin et al. ,2003)。一项消费者调查研究旨在了解消费者对谷类速食产品的态度和属性之间的关系。研究者选取了 12 种品牌速食产品对消费者进行调查，最终得到 116 名有效被试的 235 份评价（被试不一定评价完所有产品。思考一下：这种情况下，输入 SPSS 中的原始数据是怎样排列的？）。研究者根据事先调查选取了 25 个反映不同产品属性的指标，要求每位被试从这 25 个方面评价几种喜爱的食物品牌，并在 5 点量表上标明每个品牌具有该属性的程度。25 种属性如下 分别对应 25 个变量，数据中的前两个变量为被试编号和品牌编号。

充实	天然	纤维质	芳香	方便	咸味	令人满意	能量高	有趣
儿童口味	乏味	经济	健康	老少皆宜	高热量	平常	脆性	规范
甜味	加水果	经过加工	质量	试尝	倒胃口	有营养		

要达到了解产品属性和消费者评价甚至购买行为间的关系首先必须将数据简化 ,25 个变量对于产品评价来说太多了，因此可以使用因素分析来寻找这些属性背后的潜在因素。这是研究的主要目的。

首先，考虑一下选用什么样的公因子抽取方法？由于研究者对变量的方差并不了解，也不要求解释的总变异很大，所以考虑使用公因子分析法抽取因素。但可先利用主成分法大致掌握因素数目。

其次，进行因素分析适合度考察。检验表明数据适合因素分析，主成分分析抽取出 5 个特征值大于 1 的公因子，但从碎石图看（如图 13.11）似乎特征值变化的转折点在 4、5 主成分之间。用主轴因子法指定抽取 4 个因子，并选择 Varimax 正交旋转。

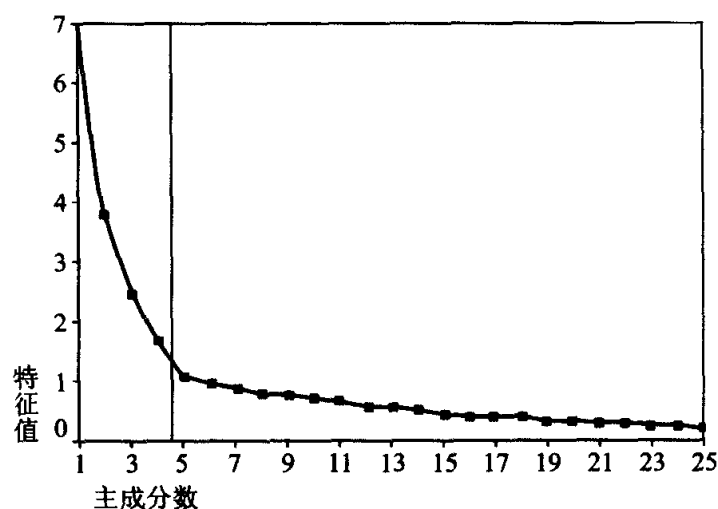


图 13.11 例 13-2 陡阶检验碎石图

从结果看 ,4 个因子旋转前特征值分别为： 6.086,3.364,2.055 和 1.116 ,共解释了变量总方差的 50.5% ，旋转后特征值分别为： 5.202、2.661、2.473、

2.285，第一因子方差贡献较大，负载变量数最多。考察共同度（见表 13 - 10）发现方便、经过加工和倒胃口三个变量共同度都较低（用下划线标出），表明这些变量中大多数信息是该属性特有的。其原因很多，比如可能抽取因子数较少、缺乏其他的相似变量、测量误差过大（如可能方便属性在所有速食产品中相当而不易区分）等，因此这三个变量的真正作用可能无法确定。

表 13 - 10 抽取前后变量共同度（PAF）

	初始	抽出		初始	抽出		初始	抽出
充实	0.633	0.569	儿童口味	0.636	0.725	甜味	0.664	0.731
天然	0.636	0.615	乏味	0.314	0.241	加水果	0.488	0.444
纤维质	0.703	0.703	经济	0.372	0.308	经过加工	<u>0.297</u>	<u>0.212</u>
芳香	0.612	0.623	健康	0.750	0.774	质量	0.614	0.549
方便	<u>0.230</u>	<u>0.171</u>	老少皆宜	0.599	0.594	试尝	0.561	0.590
咸味	0.448	0.486	高热量	0.417	0.420	倒胃口	<u>0.334</u>	<u>0.337</u>
令人满意	0.615	0.606	平常	0.408	0.461	有营养	0.711	0.728
能量高	0.571	0.522	脆性	0.410	0.356			
有趣	0.483	0.461	规范	0.521	0.395			

旋转后因素模式见表 13 - 11，变量因素负荷最大的用粗体表示，少数变量有两个大小相当的最大负荷（单下划线标出较大者）或者最高负荷都较低（虚线标出）。因子 1 上负荷的变量有：充实、天然、纤维质、令人满意、能量高、健康、规范、质量和有营养，负荷都较高，将其命名为“健康”或“有益性”较好。

因子 2 上负荷的变量有：芳香、咸味、高热量、甜味和经过加工，“经过加工”变量的负荷只有 0.374，因子命名为“人工性”，其含义和天然相对，这从因子 1 负荷的大多数变量都和因子 2 呈负相关也可看出。但注意因子 1 和因子 2 并不相关，所以因子 2 不宜命名为“不健康”。

因子 3 上负荷的变量有：方便、儿童口味、经济和老少皆宜，“方便”的负荷不足 0.4，而令人满意、有趣、脆性三变量的负荷也在 0.4 附近，但它们在其他因子上有更高的负荷；该因子似乎可以命名为“大众性”或“普及性”。

最后，因子 4 上负荷的变量有：有趣（乏味）、平常、脆性、加水果、试尝和（倒胃口），括号中的变量具有负的负荷（负相关），比如乏味得分越高，因子 4 上的得分越低，因此可以命名为“趣味性”。

表 13 - 11 旋转后因素模式矩阵 (Varimax Rotation)

	因子			
	1	2	3	4
充实	0.706	0.088	0.199	0.151
天然	0.753	-0.209	0.055	0.037
纤维质	0.821	-0.116	-0.121	0.021
芳香	0.069	0.702	0.074	0.347
方便	0.238	0.063	<u>0.325</u>	0.064
咸味	-0.092	0.686	0.016	-0.083
令人满意	0.626	0.077	0.424	0.171
能量高	0.660	0.078	0.192	0.210
有趣	0.163	0.176	0.417	<u>0.478</u>
儿童口味	-0.025	0.034	0.850	0.010
乏味	0.033	0.014	0.094	-0.481
经济	0.069	-0.281	0.415	-0.229
健康	0.829	-0.288	0.051	0.046
老少皆宜	0.062	-0.055	0.761	0.090
高热量	-0.114	0.627	-0.007	0.120
平常	-0.147	-0.062	0.068	-0.657
脆性	0.074	0.146	0.373	<u>0.436</u>
规范	0.613	-0.100	-0.027	0.089
甜味	-0.184	0.817	-0.053	0.165
加水果	0.376	0.187	-0.267	<u>0.443</u>
经过加工	-0.236	<u>0.374</u>	0.026	-0.126
质量	0.647	-0.245	0.204	0.170
试尝	0.245	0.233	0.336	0.602
倒胃口	-0.165	0.067	-0.225	-0.505
有营养	0.831	-0.177	0.051	0.056

总的说来，这个结果还是可以接受的，读者可以自行尝试，使用斜交旋转基

本上很难产生更好的结果。我们用回归法计算因子值并保存在原数据集中。由于这里我们实际上是关心各品牌的分类和评价，所以这里列出了 12 种品牌评价的 4 个因子的平均分（以品牌为自变量使用 Analyze⇒Compare means⇒Means 过程）如表 13 - 12 ；特定情况下也可以直接对样本进行分类。

此外根据均值可以画出散点图，以表明各种品牌在四个因子上的相对位置，这样就可以对品牌进行综合评价和比较。方法是将表中数据作为分析数据集，调用 Graphs⇒Scatter 选择 Simple 点击 Define，选择各因子作为坐标系的横、纵坐标，品牌变量作为标签变量，重复执行后得到的是 12 种品牌在各坐标系中的散点图（图 13. 12 ）。

表 13 - 12 12 个品牌 4 因子均值

品牌	N	因子 1	因子 2	因子 3	因子 4	品牌	N	因子 1	因子 2	因子 3	因子 4
		有益性	人工性	大众性	趣味性			有益性	人工性	大众性	趣味性
1	15	.3482	-.3190	-.8861	-.3756	7	18	.6227	.6831	-.4104	.5755
2	13	.5175	.4989	-.2340	.6583	8	21	-1.1649	-.4284	.6051	.0626
3	16	-.0135	.2275	-.4198	.4752	9	23	-.3058	-.2139	.1545	-.0550
4	27	-.5536	.1171	.5741	.0242	10	12	.6767	-.3028	-.3239	.8566
5	14	.5458	.2328	-1.009	.6023	11	25	.3648	-.5920	.2183	-.9316
6	24	-.4247	.8091	.5480	.2855	12	27	.3263	-.4121	-.0712	-.8837

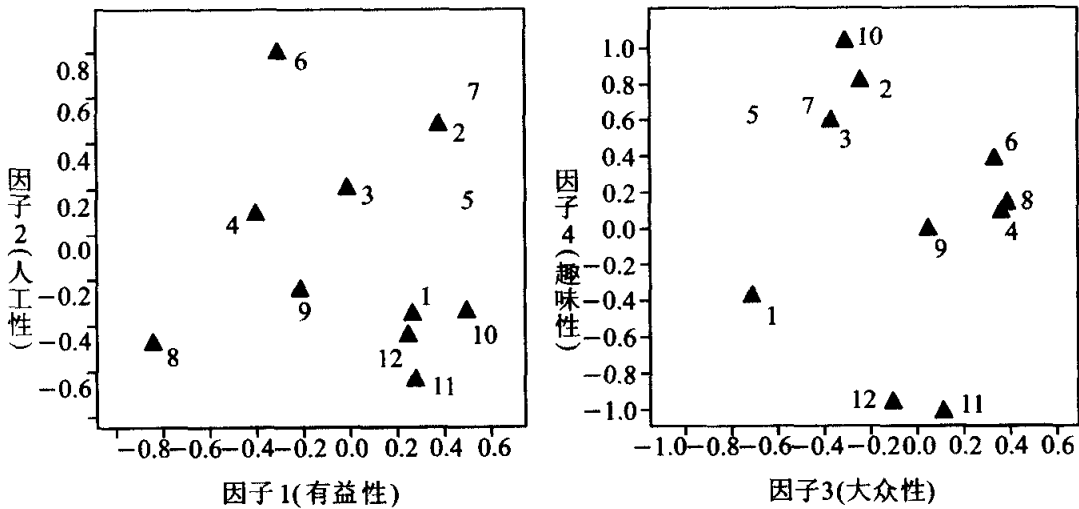


图 13. 12 12 个速食品牌 4 因子均值散点图（部分）

从散点图的分布看，品牌 11、12 很类似，都是有益、天然和无趣；品牌 2、5、7 也属同类，都是有益、人工和有趣的，等等。

►本章提要

1. 因素分析法是一种用于简化数据并探求数据间结构的多元统计方法，它用几个较少的公因子来反映众多观测变量的信息。因子分析有一系列的假设条件，利用 SPSS 进行因素分析首先要检验这些条件，对数据进行仔细考察。

2. 因素分析大致包括主成分分析法和公因子分析法两类，主成分法立足于将原变量的变异完全转移到相互正交的主成分中，然后抽取方差贡献靠前的若干主成分作为公因子。

3. 公因子法包括一系列方法，但都立足于解释原变量间的相关关系来求得因素解。它首先要估计变量的共同度，更适用于对变量方差不了解和重在探求数据间结构的场合。

4. 因子旋转分为正交旋转和斜交旋转两类，但目的都是通过适当的坐标变换使因子和变量间的关系和因子的意义更容易解释，它不改变模型对数据的拟合程度，也不改变变量的共同度。

5. 因子值作为因素分析的产物，可用来代替原变量以便于更清楚地认识因素、变量间的关系，实际应用中常用来进行个体的评价、分类。

►本章关键术语

陡阶检验准则 (scree test criterion)

公因子分析 (common factor analysis)

公因子模型 (common factor model)

共同度 (communality)

贡献 (contributions)

简单结构规则 (rule of simple structure)

潜在变量 (latent variable)

数据化简技术 (data reduction technique)

探索性因素分析 (exploratory factor analysis)

特殊因子 (unique factor)

特征值 (eigenvalue)

外显变量 (manifest variable)
斜交模型 (oblique factor model)
斜交旋转 (oblique rotation)
因素分 / 因子值 (factor score)
因素分析 (factor analysis)
因素负荷 (factor loading)
因素结构 (factor structure)
因子旋转 (rotation)
正交旋转 (orthogonal rotation)
主成分分析 (principal components analysis)
主因素分析 (principal factor analysis)

► 复习与练习

1. 主成分法的基本原理是什么？

2. 辨析下列几组概念：

公因子、主成分、特殊因子、正交模型、斜交模型、因子负荷、共同度、因子方差贡献、特征值；

相关矩阵、导出相关矩阵、剩余相关矩阵、因素模式、因素结构、反映像相关矩阵、因素变换矩阵、因素值系数矩阵；

初始解、抽出解、旋转解、主成分法、公因子法、简单结构、正交旋转、斜交旋转

3. 因素分析的假设条件有哪些？ SPSS 中如何检验？

4. 因素分析的主要步骤是什么？主要解决哪些问题？

5. 主成分法和公因子法的主要区别是什么？

6. 正交旋转和斜交旋转的区别是什么？

7. 如何根据 SPSS 提供的分析结果重新进行因素分析？

► 推荐参考读物

郭志刚. 社会统计分析方法——SPSS 软件应用. 北京：中国人民大学出版社, 1999

吴明隆. SPSS 统计应用实务——问卷分析与应用统计. 北京：科学出版社, 2003

张敏强. 教育与心理统计学第十二、十三章. 北京：人民教育出版社, 2002

张文彤. 世界优秀统计工具 SPSS 11 统计分析教程（高级篇）. 北京：希望电子出版社, 2002

► 在 线 资 源

因素分析综合介绍的网上资源：<http://www.fa100.info>

主成分分析和因素分析网上资源：<http://www.statsoft.com/textbook/stfa-can.html>

第四编 研究间的比较与交流

本编共两章,将主要讨论两个问题:元分析的研究方法、研究报告的撰写与交流。学习完本编之后,可以了解一种科学的文献综合方法,如何撰写研究报告以及如何进行学术交流。

总结评价已有的研究成果既是每项心理学研究必须进行的工作,也是一类独立的研究。元分析是完成这类综述研究的一种科学方法。面对迅速累积的定量研究资料,传统综述方法的科学性和有效性越来越受到质疑。元分析通过引入了数量化手段等科学方法,试图改进传统的文献综述,使综述研究更科学、规范,研究结果更可信。第十四章不仅讨论了元分析的一般问题,还着重探讨了元分析的过程,并配以实例学习,以便掌握。

写作研究报告是研究的最后环节,担负着表达研究、深化研究的使命。作为学术表达手段,研究报告要求遵循客观、准确、简约的写作风格和公认的写作方法;作为研究的组成部分,它是对研究的创造性综合与提升。研究报告各部分的写作规范是十五章的重点内容,包括如何确定报告题目、选定关键词、撰写前言、表达研究方法、研究发现,以及进行讨论、得出研究结论和著录参考文献等。本章另一个重点内容是关于研究报告的交流,包括评估报告的标准,以及如何向期刊投稿、如何准备会议交流等。

第十四章 研究的综合—— 元分析

任何科学研究都必须建立在前人研究的基础之上。所以，查阅文献、进行文献综述就是每个科研工作者在进行科研之前必须完成的工作。但是，目前全世界的科技期刊已超过 4 万种，我国的中文期刊也已超过 6 千种。因此，现在仍使用传统的文献综述方法已经远不能达到文献综述的目的。此外，随着实证科学的发展，常常会出现这样的情况：在许多科学领域有众多研究者就同一问题对不同的对象或对同一对象在不同的条件下进行实证研究，如胎教对婴儿智力影响的研究就是这样。因此会在很短时间内出现大量的关于同一问题的各自独立的研究，但是，这些研究的结果又很可能是不一致的，面对这些结果，我们应该相信哪一个呢，又如何进行文献综述呢？显然，面对这样的问题传统的文献综述是无能为力的，因此从上个世纪起就有学者探索新的对大量独立的实证研究进行综合的方法，这也就是我们下面要讨论的元分析。在本章中，我们将讨论以下问题：

1. 元分析的一般问题
2. 元分析的实施方案

第一节 元分析的一般问题

元分析的英文是 meta - analysis，由美国教育心理学家 Glass 在 1976 年首次命名。Meta 是英文中的前缀，有“更加全面或超常规的综合”的意思。国内将其译为元分析、荟萃分析、整合分析、汇总分析、二次分析、集成分析、再分析、汇后分析等，还有许多学者直接使用“Meta 分析”。本节讨论元分析的涵义、目的与意义、局限性与质量控制等基本问题。

一、元分析的涵义

格拉斯 (Glass, 1976) 对元分析的定义是：以综合已有的发现为目的，对单个研究结果进行综合的统计学分析方法。之后许多统计学家也做了类似的界定，即把元分析仅看作是对以往研究结果进行定量合并的统计分析方法。因此，有人 (Milos, 1989) 把这类元分析称作典型或定量元分析 (classic meta-analysis or quantitative meta-analysis)。

随着元分析方法的应用范围的扩大和自身的发展，学者们认识到元分析中应该既有定量分析也有定性分析，不再将元分析简单看作一种统计分析，而是作为综合多个同类研究的结果，对研究效应进行定量合并的分析研究过程和系统方法，如埃伦伯格 (Ellenberg, 1988) 将其定义为：对具有共同研究目的相互独立的多个研究结果给予定量分析、合并分析，剖析研究间差异特征，综合评价研究结果。萨克斯 (Sacks et al, 1987) 对其的定义是：对以往的研究结果进行统计学的合并和严谨的系统综述方法。

简单地说，元分析就是应用特定的设计和统计学方法对以往的研究结果进行整体的和系统的定性与定量分析。它是回顾性的，是对传统综述的一种改进，是概括以往研究结果的一种方法，包括大量的方法和技术，具有全面、系统和定量的特点，可用于对以前的具有不同研究设计和不同时期收集到的资料进行整合。它最初应用于随机对照试验，现在已扩大到非实验研究。

二、元分析的目的和意义

科学研究都必须建立在前人研究的基础之上，而实践工作者也非常关心一些基本问题的研究结论，如某种心理疗法的效果究竟如何？因此，总结前人的研究是一项很重要的工作，这就是所谓的文献综述 (literature review)。但传统的文献综述常常难以让人满意，因为它容易出现以下几个问题：

以定性分析或描述为主，难以给出一个定量的结论。同时，当所涉及的实验数量不断增加时，得出错误结论的概率也随之增加。

如果有关研究的数量众多如有 500 个，那么要从这么多的研究结果中得出一个一般性的结论已超出了个人的能力。

综述者究竟使用了哪些文献常常不得而知。因为综述者一般不会详细交待文献检索的范围和方式，比如是否同时进行计算机检索和手工检索，计算机检索中使用了哪些数据库，哪些关键词，是否检索了硕士论文和博士论文，检索了哪些语种的外文文献等。综述者一般也不会交待进行综述时排除了哪些检索到的文献，其依据是什么。

传统综述中没有足够重视研究质量、样本大小等因素对研究结论带来的影响。

使用同样的文献，不同的综述者可以得出不同的结论。综述者常常没有或难以说清其是依据什么最后得出一个结论，特别是对于那些有争议的问题，是如何看待和判断不同甚至是完全相反的研究结果的。因此综述者很可能基于自己的见解，采取简单化的方式处理一些研究结果，从而得出与自己观点类似的结论。

总之，传统的文献综述难以使用系统方法来对所综述内容的原始数据进行收集、综合，也未进行定量综合，因此常常只是罗列以往的研究结果，同时研究结论不可避免带有主观性。

元分析作为一种定量的综合文献的方法和传统的文献综述有很大的不同，其可以在很大程度上克服传统的文献综述中的这些问题。因为元分析是较高一级逻辑形式上的文献综述，其以原始研究结果为单位，设计严密，强调对有关研究进行全面的文献检索，有明确的文献纳入和排除标准，系统地考虑了研究的对象、方法、测量指标等对分析结果的影响，对纳入文献进行严格评价，并在此基础上对结果进行定量的合并。所以，与传统的文献综述相比，元分析能最大限度地减少各种偏向，确保结论的科学性、客观性和真实性。

具体地说，元分析具有以下作用：

解决研究结果的矛盾，定量估计研究效应的平均水平，为进一步的研究和做出决策提供全面的文献总结。应用元分析可以得到同类研究的平均效应水平，使有争议甚至相互矛盾的研究结果得出一个定量的结论，同时使效应估计的有效范围更精确。

提高统计分析的功效。有时由于样本量较小等原因，使研究结果不能得出统计学上的显著差异，但这并不一定就是处理无效应。应用元分析把许多同类研究结果进行合并分析，增大了样本量，就可以使一些相对较弱的效应也显现出来，提高对结论的论证强度和对效应的分析、评估力度，起到改进和提高统计学检验功效的目的。

揭示和分析多个同类研究的分歧。由于研究水平、条件、抽样等原因，可能使同类研究的质量有较大差异，多个研究结果也可能存在较大分歧。元分析可以揭示出单个研究中的不确定性，并通过异质性（齐性）检验（tests for heterogeneity）等方法考察研究间异质性的原因，揭示文献异质性的来源，估计可能存在的各种偏向（bias）。

为确定新的研究问题和对新实验的设计提供帮助。通过元分析能发现以往研究的不足之处，回答单个研究中尚未提及或不能回答的问题，揭示单个研究中存在的不确定性，并据此提出新的研究假说、课题和方向。

具有处理大量文献的能力，不受研究数目的限制。

⑥ 节省研究费用。由于研究对象数量的限制、各种干扰因素的影响以及研究本身的偶然性等原因，许多研究结果可能不一致甚至相反，要解决这个问题除了使用元分析外还可以通过严格设计大规模随机试验来进行验证。但这种办法费时、费力，使用元分析则可以节约费用。

⑦ 研究发表偏向（publication bias）等问题。

总之，元分析是获取和评价大量文献的科学方法。在当今知识爆炸的时代，

当需要系统总结、分析以往的工作，为科学决策提供依据时，元分析是合并现有信息的最好方法之一。其有助于研究者和实践工作者对文献进行分析与评价，从而在较少的时间、人力和物力的投入下获取更多的信息。目前，元分析已经广泛用来探讨一些有争议的问题，如按摩治疗（*Massage Therapy*）是否有效（*Moyer, Rounds, & Hannum, 2004*），是否存在普遍性的积极归因偏向（*Mezulis, et al, 2004*）。

三、元分析的局限性与质量控制

虽然与传统的文献综述相比元分析具有许多优点，但也存在着不少局限与问题，关于元分析的争论也从来没有停息过，如有人（*Feins, 1995*）认为元分析方法本身就存在一些缺陷，其结果的精确性较差，不适合外推，甚至有时会造成错误的结论。

（一）各种偏向及其控制

可以说，在元分析的各个步骤中均有可能产生偏向。偏向（*bias*）的存在会对元分析的结果产生较大影响，甚至会使元分析产生错误的结论。常见的偏向可以划分为汇集偏向、选择偏向和研究内偏向三类。

选择偏向是指根据文献纳入标准选择符合要求的文献时产生的偏向，主要是由于元分析者有时会根据自己想要得到的结果和已有的知识，有目的地选择或放弃某些文献造成的。

研究内偏向是在资料提取时产生的偏向，主要包括从纳入研究中提取的数据信息不准确和对研究质量的评价不恰当。

最后，也是最难控制的是汇集偏向问题。元分析是对以往研究的再次分析，一个好的元分析应包括所有与课题有关的资料。由于多种原因，不可能收集到同一研究问题的全部资料，这就是汇集偏向。出现汇集偏向有多种原因：一是因为研究者检索用词不当或检索策略失误导致漏检或误检文献；二是因为检索文献时限定为某种语言文献而导致检索不全面；三是一项研究结果以系列研究形式发表，或重复发表而导致被多次重复检索；四是发表偏向（*publication bias*）即具有统计学显著意义的研究结果（阳性结果）较无显著性意义的结果（阴性结果）或无效的结果被报告和发表的可能性更大的倾向。

要减少汇集偏向对元分析结果的影响，首先，要全面收集发表和未发表的研究，不过搜索未发表的研究是相当困难的。其次，要识别和控制发表偏向。最后，还可以应用有关统计方法评估各种偏向对研究结果的影响。

总的来说，元分析目前还不能很好解决潜在的各种偏向问题，这也是进行元分析必须注意的问题。

（二）收集到的资料中缺少元分析所需的数据

有许多收集到的文献常由于以下原因而不能被利用：①只收集到摘要，文献对最初的研究结果进行有选择的报道，文献对原始数据描述不完整。这就使进入元分析的文献大大减少，从而降低了元分析的综合能力。

（三）关于各独立研究的质量评价问题

各个研究的质量常常是参差不齐的，如果把高质量和低质量的研究合并在一起，并给予相同的权值（weight），就会带来结果的偏向甚至是错误。低质量的研究在结果估计时就可能已经存在错误和无法纠正的偏向，包含了低质量的研究会增加结果变量的变异度和偏向，增大犯第一或第二类错误的概率。因此，很多学者都推荐使用各种质量评价（quality assessment）办法，并建议将质量评分作为权值以调整结果。但质量评分又增加了主观标准，不同的评价者可能会得出不同的结果，同时可能严重地混淆异质性来源。究竟该如何解决这个问题，目前还没有定论。

（四）“桔子与苹果问题”及其控制

对元分析的最尖锐的批评之一就是认为其不应该将研究对象、结果测量指标、实验设计以及测量方式等不同的各项研究所得的结果结合在一起，因为这就好比将桔子与苹果拿来比较，是很难得出正确结论的。目前，主要是通过进行异质性检验来缓解这个问题。

（五）异质性的处理问题

关于如何处理异质性（heterogeneity）的资料问题还存在争议。有学者推荐用质量评价法评价每个异质成分，如没有研究质量方面的严重缺陷，则可以按相同变量进行分层合并分析或是利用随机效应模型进行合并分析。也有学者提出不应武断地用某种权重公式来强求结果的合并，应首先尽量使各个研究达到一致，使研究的某些特点不成为元分析结果混杂的来源。目前常用的解决方法是通过进一步核实资料的可靠性与处理方式，找出异质性的来源，但不轻易剔除异质性文献。

（六）元分析本身的质量控制问题

目前，元分析的质量评价还缺乏一个公认的和科学的标准，这在一定程度上也导致了元分析方法的误用和滥用。虽然没有可靠的标准，但是要保证元分析的质量必须注意以下几个问题（Sacks等,1987;Thacker,1988）：

要有合理的研究设计和文献检索策略，保证文献的查全率和查准率，进行严格的文献筛选和质量评价。

选择恰当的统计学方法进行分析；

注意控制各种偏向；

进行异质性检验，以保证合并分析的可比性和分析结果的可靠性；

对结果进行敏感性分析，并予以合理的解释和应用。

实际上，上述的许多问题是综述研究的各种方法的通病，只是传统的文献综述忽略了这些问题而元分析则把它们暴露无遗。尽管存在这些问题，元分析还是受到很多研究者的青睐，因为元分析本身就是在解决这些问题的过程中提出来的新方法。应该说，元分析方法提出的时间还不长，还需要在逐步克服这些问题的过程中不断深入、完善，不少学者（Field, 2001; Hall & Brannick, 2004; Cheung & Chan, 2004）也正在研究改进该方法的各种措施。

第二节 元分析的基本过程

同任何科学研究一样，元分析也包括提出问题、收集和分析数据、报告结果等基础研究过程（Petitti, 1994）。下面就详细探讨元分析的基本操作过程。

一、提出问题 制定计划

与其他任何科学研究一样，元分析首先应提出问题，进行科研设计并制定研究方案。提出一个好问题是保证质量最基本的一步。

问题的范围取决于研究意义、可行性、基础理论是否支持等多种因素。范围太窄，得出结论的推广性差，范围太宽则针对性差，且耗时耗力，难以完成。元分析课题一般来自心理学研究中不确定或有争议的问题，如心理咨询上某些干预措施的有效性难以确定，工作满意度与工作绩效的关系尚未得出明确的结论。

提出的研究问题应包含四个要素：研究对象、研究设计、处理因素（干预措施）、研究效应（因变量、结果变量）。

确定元分析课题后，应制订详细的研究计划书，计划书应包括研究目的、研究意义等背景材料，文献检索的途径和方法，文献纳入标准（inclusion criteria），数据收集的方法及统计分析步骤，结果的解释等。

为了尽可能减少选择偏向，应根据研究目的、专业知识等来制定文献的纳入标准，通常会考虑以下几个方面的因素：

研究对象。应对纳入元分析的研究对象的特征如年龄、性别、职业等作出规定。

研究设计。如果把各种设计类型的资料都纳入，研究间可能有较大异质性，因此要明确规定哪些设计类型的研究可以纳入。理想的设计方法应该是随机对照试验（randomized controlled trial, RCT）、和盲法观察（blinded observation）的实验。

结果变量。应规定报告了哪些可以量化的、具有可比性的变量如平均数（Mean）、 P 值、 t 值等的研究才纳入元分析。

研究开展的时间或文献发表的年限和语种等。如果某一课题涉及的研究时间很早，应考虑选取哪个时间段的文献。此外，还要考虑仅纳入某一种语言的文献是否会引起语种偏向的问题。

样本量。为防止小样本研究高估处理效应所产生的偏向，应对纳入文献的样本量做适当规定。

需要注意的是，将纳入标准制定得过严或过宽都会带来一定的问题。首先，如果标准太严，虽然保证了进入元分析的研究间的同质性，但可能使选出来的文献很少，这就达不到通过元分析增加统计学功效、定量估计研究效应平均水平的目的。反之，如果标准太宽，又可能出现所谓的“桔子与苹果问题”，使研究间的异质性太大，大大降低元分析结果的可靠性和有效性。

二、检索相关文献

系统、全面的收集有关文献是元分析不同于传统文献综述的重要特征，因此，进行元分析时应尽可能充分利用信息来源，尽可能多地收集与主题有关的研究资料，包括发表和未发表的文献。在制订元分析的研究计划书时，就应确定检索策略 (search strategy)。在实际操作中一般是先进行预检索，大概确定出检索范围，再根据预检索的结果修改检索策略。检索时可根据需要进行必要的限定，如研究对象、语种、出版年限、出版类型等。

在检索文献时应综合考虑检索结果的敏感性和特异性。首先要保证较高的查全率（敏感性），因为漏检了重要文献就可能直接影响元分析结论的可靠性和真实性。但也要同时考虑到查准率（特异性），常常会因为过于追求查全率使检索范围太宽泛，这就会给后面的文献鉴别、筛选工作带来麻烦。因此，在制定检索策略时还可以咨询专业图书馆管理员或信息检索人员，避免漏检和误检。一般来说，全面的文献检索都会使用到以下一些方法。

1. 计算机检索

可以利用的计算机检索工具和数据库很多，具体请参见第二章的有关内容。计算机检索一般是通过关键词进行，因此需要注意：关键词可能会随着时代而发生一定变化，有时变化还很大，选用正确的关键词进行检索对文献查全十分重要。

2. 手工检索

通过人工逐篇翻阅有关文献，主要是对计算机未能检索到的文献再次进行检索，因此检索范围应尽可能的广，应包括专著、书、单行本、期刊、正式或非正式的论文、学术会议论文集、学位论文、政府出版物以及各种私人交换资料等“逃逸文献 (fugitive literature)”。

3 从有关研究的参考文献中追踪查询

仔细查阅已经检索到的文献的参考文献，并根据这个线索继续查询。

4. 专家询问法

向该有关领域的专家、同事收集相关信息，特别是对一些基本内容符合要求，但报道不详细的文献如摘要等，应通过与作者联系获得元分析所必需的信息。

三、对文献进行编码，并根据纳入标准，选出符合要求的纳入文献

由于不是所有检索出的文献都符合元分析的要求，因此需要对文献进行挑选。为了防止选择性偏向，并在一定程度上控制研究间的异质性，应严格根据制定出的纳入标准来筛选检索到的文献。整个过程一般包括三个步骤：①收集反映研究特征的描述性统计量，包括反映研究的外在特征，如起始时间、地点等；反映研究设计质量的因素，如样本量、是否随机化、是否采用盲法等；反映研究内容的变量，如性别、年龄、职业，结果变量的均数、标准差等。对研究特征进行分类、编码，一般有两种编码方法：一种是方法编码，即按照发表日期、研究设计、研究效度、测量指标等进行编码；另一种是内容编码，即根据实验者职业、被试特征、实验处理类型和测量类型等进行编码。对编码后的研究特征进行初步分析，严格根据纳入标准选取研究。

选择文献时一般先初选，即通过浏览标题、文摘等剔除那些明显不合格的文献。然后通读全文，根据文献纳入标准进行仔细的鉴别筛选，对存有疑问的文献，可以先纳入，待联系原文作者获取相关信息或分析评价后再作取舍。遇到同一个研究分多阶段发表、重复发表或既有摘要又有全文发表的情况，应特别注意对其进行鉴别和区分。最后，对于排除的研究要有详细的记录并逐一列出排除的原因，使读者能够了解该元分析有无选择性偏向。

四、纳入研究的质量评价

研究的质量是指一个研究在设计、实施和分析过程中防止和减少系统误差（偏向）及随机误差的程度。每个研究的质量将直接影响元分析的质量，应该对纳入研究的质量进行评价并详细报告评价结果。

在研究的质量评价方面还没有适用于各种研究的统一的“金标准”，不同学者提出的各种评分标准也有待进一步的验证、完善。通常使用的标准会涉及到以下几个方面：

选择偏向的情况，即是否真正做到随机分配。

实施偏向的情况，即除要研究的处理外，其他因素在两组是否一致。

测量偏向的情况，指结果测量时是否存在系统误差，是否采用盲法观察或判断以及是否有选择性地报告研究结果。

是否存在排除偏向，即被试退出试验的情况在实验组与对照组之间是否有系统的差异。

样本量的大小，大样本的研究比小样本的研究更可靠。

⑥ 统计方法是否正确。

此外，如何根据有关标准完成质量评价工作，如何应用质量评价的结果（得分），目前也没有统一的或标准的做法，在实际操作中常见的办法有：两个或两个以上的评价者独立评价纳入文献，如有不一致就分析其原因，使得最后的质量得分尽可能符合统一的标准。确定一个质量得分，选择得分在分数线以上的研究进行元分析。即对所有研究进行质量评分，用中位数、平均数或是其他某个界限值为抉择分数线。将质量得分进行某种变换后，对结果变量进行加权。

五、正确进行数据提取

确定了纳入分析的文献后就需从文献中提取各种所需的数据。为了便于操作，可以根据不同的研究设计不同的专用表格来记录，一般包括基本信息、研究特征、结果变量等内容。由于数据提取工作量大，容易出错，所以为保证数据收集的质量，最好由两人以上的研究者独立进行数据提取工作，然后进行交叉核对，并在文中详细叙述观察者的一致性。此外，在数据提取中应采用盲法，即隐去那些对资料提取者可能产生影响的因素，如期刊名、作者、作者单位、基金资助情况等，以减少选择性偏向。提取数据信息后，对不一致的文献应复核并请专家评议。如果文献中缺乏所需要的数据，应与原作者联系补充。

六、资料的统计学处理

统计处理是元分析最重要的步骤之一，整个统计学处理过程可由专门的元分析软件完成。可用于元分析的统计方法很多，不同的方法适合于不同的情况，因此应根据研究资料的性质选择正确的统计分析方法。这个过程主要包括以下几个步骤：

1. 明确资料类型，选择恰当的效应指标。

元分析需要将收集到的统计量（多个独立研究的结果）合并（或汇总）成某个单一的效应量（effect size, ES）或效应尺度（effect magnitude）即用某个合并统计量反映多个独立研究的综合效应。因此在元分析前必须根据提取到的数据资料的类型确定使用哪些反映研究效应的统计指标。通常提取到的数据资料包括计量资料（连续性变量）、计数资料（即离散型的二分类变量）以及假设检验统计量。计量资料常用的效应指标包括相关系数 r 、均数差（mean difference）、加权均数差（weight mean difference, WMD）和标准化均数差（standardized mean differ-

ences, SMD 等假设检验统计量包括 t 值、 P 值、 χ^2 值以及 F 值等。

由于在心理学的研究中结果数据多以平均值、样本数和标准差的形式报告，因此这里主要介绍标准化均数差（SMD）这种效应量。标准化均数差（SMD）的计算方法有多种，如有格拉斯最早提出的估计值 Δ ，赫奇斯提出的估计值 g 和 d 等。这里介绍的是常用的 d 值。

$$d = (M_e - M_c) / S \quad (14-1)$$

其中 M_e 是实验组的平均数， M_c 是控制组的平均数， S 是标准差。

格拉斯（Glass, 1976; 1981）最初建议使用控制组的标准差，后来赫奇斯等（Hedges, Olkin, 1985）提出使用实验组和控制组的联合标准差（pooled standard deviation）更能减少偏向和变异。联合标准差的计算公式如下，其中 N_e 是实验组的样本量， N_c 是控制组的样本量。

$$S = \sqrt{\frac{(N_e - 1)S_e^2 + (N_c - 1)S_c^2}{N_e + N_c - 2}} \quad (14-2)$$

效应值 d 的方差为

$$S_d^2 = \frac{N_e + N_c}{N_e N_c} + \frac{d^2}{2(N_e + N_c)} \quad (14-3)$$

2. 检验纳入研究的异质性

元分析需要对纳入研究的结果进行异质性检验（tests for heterogeneity）也称同质性检验（tests for homogeneity），也就是齐性检验，以判断多个研究是否具有同质性，之后才能根据其结果选用正确的统计分析模型。异质性检验的方法主要包括图表法，如森林图、率分布图和各种卡方检验（Chi-square test）法。目前广泛使用的是 Q 检验，计算公式如下，其零假设是 $H_0: Y_1 = Y_2 = \dots = Y_i$ 即所有的效果量都来自相同的总体，也就是说研究是同质的。

$$Q = \sum_{i=1}^k W_i (Y_i - \bar{Y})^2 \quad (14-4)$$

其中 k 是参与元分析的研究个数； W_i 是权值（weight），可以用纳入分析的研究的联合方差的倒数、效应值方差的倒数及样本量等来估计； Y_i 是第 i 个研究的效应值； \bar{Y} 是 k 个研究的效应值的平均数，计算公式如下。

$$\bar{Y} = \frac{\sum W_i Y_i}{\sum W_i} \quad (14-5)$$

Q 值服从自由度（ df ）= $K - 1$ 的 χ^2 分布，因此可以用卡方检验来确定纳入研究异质性情况。

3. 根据异质性检验结果选择统计模型，进行统计分析，得到效应合并值的点估计和区间估计值。

当异质性检验的结果为不显著，即 $p > 0.05$ 时，可认为多个独立研究具有同质性，应选择固定效应模型（fixed effect model）来计算合并统计量。若同质性检验的结果为显著，即 $p < 0.05$ 时，可认为多个研究有异质性，此时，可使用敏感性分析，亚组分析，剔除特大、特小的或是方向相反的统计量等异质性处理方法，使之达到同质后，再使用固定效应模型。若经异质性分析和处理后，多个独立研究的结果仍然不具有同质性则可选择随机效应模型（random effect model）。若异质性过大，则应考虑放弃元分析，只作一般的统计描述。

需要注意的是， Q 检验的功效不高，不拒绝 H_0 只能说明研究间的变异较小。因此，如果 Q 检验的统计量在临界值附近，应同时采用两种模型，比较参数估计是否有差异，以使元分析的结论更可靠。

固定效应模型下有不少统计方法适用于不同的情况，如果效果量是标准化均数差，那么计算效应合并值的点估计即平均效果量的方法同公式 14-5。

除了要得到效应合并值的点估计，还应得到其区间估计值，即算出效应值的置信区间（confidence interval, CI）。标准化均数差的置信区间的计算公式为：

$$\bar{Y} \pm z_{\alpha/2} S_{\bar{Y}} \quad (14-6)$$

$$S_{\bar{Y}} = \sqrt{1 / \sum_{i=1}^k W_i} \quad (14-7)$$

如果选用随机效应模型可选择 Fisher 法、Winner 法或 Stouffer 法进行统计，这里不再详细介绍，如果效果量是标准化均数差（SMD）则平均效果量和区间估计的公式类同公式 14-5 和 14-6，只是要将两个公式中的 Y 改为 d ， $S_{\bar{Y}}$ 改为 S_d 即可。

$$S_d = \begin{cases} S_d^2 - S_e^2 \cdots S_d^2 > S_e^2 \\ 0 \cdots \cdots \cdots S_d^2 \leq S_e^2 \end{cases} \quad (14-8)$$

$$S_d^2 = \sum W_i (d_i - \bar{d})^2 / \sum W_i \quad (14-9)$$

$$S_e^2 = \frac{4k}{\sum W_i} \left(1 + \frac{d^2}{8} \right) \quad (14-10)$$

由于一些文献中并不报告平均值、标准差和样本量等信息而只报告了假设统计量如 t 值、 χ^2 值，于是 Glass(1981) 等学者提出了以假设统计量 t 值、 F 值或确切的 P 值来转换效应值的办法，如假设检验统计量如 t 值、 F 值、 χ^2 值都可以被转换为相关系数，然后再用 Fisher 的 z 转换法来计算效应值。

关于如何看待效应合并值的大小的问题，有学者（Cohen, 1977）提出了一个对效应大小的经验解释，即 0.2 是“小”效应 0.5 是“中”效应 0.8 是“大”效应。

4. 合并假设检验统计量

得到了效应合并值，通常还不能满足我们的需要，因为效应合并值不能直接

做统计推断，也就无法知道效应是否显著。通过合并文献中的假设检验统计量则可以解决这一问题。常用的合并假设检验统计量的方法有 Fisher(p 值)法、Winer(累积 t 值)法和 Stouffer(逆正态)法，其中较为简便实用的是 Fisher 法。

Fisher 法可分为两步：如果文献中没有给出确切的 p 值的话，将各检验统计量如 t 值、 χ^2 值转为 P 值，注意此处是指具体的确切的 P 值，不是 $P > 0.05$ 或 ≤ 0.05 的定性数字。按以下公式将 P 值转为 χ^2 值，公式中的 P_i 为各独立研究的单侧概率。在判断时，自由度为参加元分析的文献数 (k) 的两倍，即 $df = 2k$ ，具体的查表等检验方法与一般的 χ^2 检验相同。

$$\chi^2 = -2 \sum \ln p_i$$

5. 根据情况进行亚组分析

亚组分析 (subgroup analysis) 就是根据不同的研究特征如被试的年龄、职业，将各独立研究分为不同组，然后再分别进行合并分析，比较各组及其与总的合并效应间有无显著性差异。当一些因素如性别、年龄、研究质量对纳入研究的结果有较大影响时则应考虑做亚组分析，以使研究结果更具针对性。如两种不同设计类型的研究结果不适合直接合并，应该通过亚组分析分别揭示效应的合并值。

需要强调的是，元分析的统计过程比较复杂，不同资料所选用的统计方法不同，不同分析家所提出的效应值、权值等的计算方法也不同，深入的学习需要参考元分析的专著或教材。具体的数据计算则可借助元分析的专用软件或 Excel、SAS、SPSS 等完成。

七、结果的敏感性分析

为了保证元分析结论的稳健性，还应对元分析的结果进行敏感性分析。所谓敏感性分析 (sensitivity analysis) 就是分析当条件 (变量) 发生变化时，结果是否也发生变化，即研究结论的稳定性如何。通常的敏感性分析会从以下几个方面来进行：选用不同统计模型或方法重新分析资料，考察结论有无变化。按研究质量评价标准从纳入文献中纳入或排除那些质量有问题、有争议的文献后重新进行元分析，考察结论有无变化。③根据样本量的大小对纳入的文献作分层元分析，考察结论有无变化。改变纳入标准，考察结论有无变化。排除异常值后，重新进行元分析与未排除异常值的结果进行对比。

八、元分析结果的分析和讨论

得到元分析结果后还应对其进行分析和讨论。有关的分析与讨论一般会涉及以下一些方面：①分析变异来源。如果纳入元分析的研究间有异质性，应检验异常值，讨论异质性的来源及其对效应合并值的影响。异质性的来源主要有研

究纳入标准不一致，各个研究的基线水平、处理、结果变量不同等。讨论各种偏向的识别和控制。如前所述，元分析过程中出现各种偏向在所难免，因此应该通过计算出失效安全系数或画漏斗图等方式对偏向情况及其可能造成的影响进行说明和评估，并在文中交待为了避免偏向做出了哪些努力。讨论元分析结果的意义。对元分析结果的解释必须十分小心，特别是心理学研究中的有些结果是似是而非或相互矛盾的。在讨论元分析的结果时，应结合研究背景和实际意义进行讨论，必要时也可以比较大样本的单独研究和元分析结果的一致性。

九、根据统计结果作出正确、全面的结论

元分析结论的内容主要包括：①说明根据所纳入研究的综合分析，能否最后得出某一效应如实验处理、心理疗法有效或无效的结论。②如果现有资料尚不足以以下结论，那么有什么趋势？提出是否应该进行进一步的研究如更大样本的实验的建议。

需要注意的是：由于元分析是多个独立的同类研究的综合分析，在增大样本量和检验效能的同时，也可能出现假阳性，因此对有统计学意义的合并结果还要结合专业知识判断其有无意义。此外，还应该说明研究的资助来源，让读者确定研究的可信度。最后，元分析的结论也不是绝对的，作者还可以不断地收集新的资料，完善或更新结论。

实例学习 14-1 应用元分析做研究的一个实例

国外心理学界对元分析的应用已经有较长的历史，也比较成熟，国内才刚刚起步。也正因为如此，国外元分析的应用研究比较深入和复杂，对于一些基本情况的交待如公式使用情况等不是十分清楚，这对初学者是不利的；国内研究对元分析方法的应用有时虽然显得粗糙但比较基本和简单，而且中文文献更便于阅读，对于初学者了解和掌握元分析的基本方法和操作程序是有益的。以下就根据在第二节中介绍的元分析的基本方法，对“音乐疗法辅助治疗慢性精神分裂症的 Meta 分析”（由智勇、王洁贞 2002）一文进行示例性评讲。

一、提出问题，制定计划

该文的问题来源是探讨 1980 年代音乐疗法传入我国以来，音乐疗法辅助治疗慢性精神分裂症的疗效情况。其研究对象是 1989 年 1 月—2001 年 12 月底符合入选标准的应用音乐疗法辅助治疗慢性精神分裂症的国内临床试验资料。其文献纳入标准（涉及到了研究问题中应包含的研究对象、研究设计、处理因素和研究效应四个要素）是：符合 CCMD-Ⅱ-R 或 DSM-Ⅲ-R 中精神分裂症诊断标准②研究对象应是急性期已控制的精神分裂症患者，主要临床症状为阴

性 ;③必须是临床对照试验（包括随机对照试验和自身对照试验）。随机对照试验应是包括试验组（音乐疗法与抗精神病药物合用）和对照组（仅用抗精神病药物）的试验；自身对照试验为进入试验后应用音乐治疗； 检测指标需运用阴性症状评定量表 (SANS)和 / 或简明精神病量表 (BPRS)进行测量，并以最后一次测定作为试验结束的最终评分。

可以看出，该文的问题提出有根据，也制定了比较详细的计划。

二、检索相关文献

文献收集方法是以精神分裂症 (schizophrenia)、音乐疗法 (music therapy)、临床试验 (clinical trials) 为主题词及相关截词、自由词进行检索，查找中国生物医学光盘数据库 (CBMdisc,1989—2001 年) MEDLINE 光盘数据库 (1989—2001 年)、维普科技期刊数据库 (1989—2001 年)。语种仅限于汉语。共检索到文献 24 篇。

可以看出，作者的文献收集还是存在一定问题，比如没有查找心理学类的数据库，没有进行手工检索等。

三、对文献进行编码，并根据纳入标准，选出符合要求的纳入文献

根据纳入标准该文删除了 13 篇文献，其给出的原因如下：①非临床对照试验； SANS 或 BPRS 评分数据不完整 ;③不是以 SANS 或 BPRS 评分作为结果测量的标准 ;④研究重复，或侧重点不同。

该文还给出了纳入文献的基本情况（其中前 6 个为随机对照实验，后 5 个为自身对照实验），见表 14 - 1，从中可以看出该文的文献编码情况。

表 14 - 1 纳入研究的基本情况

研究	实验组/控制组的样本量	得病时间(年)	实验进行的时间(月)	实验随机化情况	测量量表
1	26/28	7 - 36	2	是	SANS, BPRS
2	23/23	> 5	6	不完全随机化	SANS, BPRS
3	38/40	-	3	是	SANS, BPRS
4	16/16	> 5	3	是	SANS, BPRS
5	38/38	1 - 25	1	是	SANS
6	30/30	> 2	2	是	SANS
7	32	≤ 3	2	不是	SANS
8	64	≥ 2	1	不是	SANS, BPRS
9	26	2 - 21	3	不是	SANS, BPRS
10	124	0.5 - 20	3	不是	SANS, BPRS
11	11	5 - 22	3	不是	SANS

可以看出，作者对文献进行了编码，并根据纳入标准进行选择，还给出了排除文献的理由。

四、纳入研究的质量评价

该文报告采用 Jadad 质量评分法对纳入文献进行质量评价，但并未详细报告具体的质量评价情况，仅是在讨论部分有所提及。

五、正确进行数据提取

该文未报告数据提取（data extraction）情况，应该说这也是该文不够严谨的一个地方。

六、资料的统计学处理

该文没有明确报告使用哪种指标作为效应量，但根据文中的有关论述可以推测该文使用了标准化均数差（SMD）和 t 值作为效应指标。该文进行了异质性检验（齐性检验），并规定齐性检验的检验水准为 $\alpha = 0.20$ 。若 $\alpha > 0.20$ 选用固定效应模型；若 $\alpha < 0.20$ ，选用随机效应模型。这里需要说明的是该文选用 $\alpha = 0.20$ 为检验标准与通常使用的 $\alpha = 0.05$ 的显著性水平的方法不同，但该文并未交待其理由，因此，建议读者以后还是使用 $\alpha = 0.05$ 的检验标准。此外，当前有许多学者都提出当固定效应模型不适宜时不加分析的转向随机效应模型是不恰当的，因此建议读者可以先考虑采用亚组分析、选用多因素模型等方法对变异来源进行分析后再做进一步的模型选择的决定。

经过统计，该文得到了效应合并值的点估计和区间估计值。该文采用 Winer（累积 t 值）法对假设检验统计量进行了合并，这也就完成了效应合并值的假设检验。此外，该文还分别对随机对照试验和自身对照试验的 SANS 总分和 BPRS 总分进行了元分析，这可以看作是进行了亚组分析。为了便于理解该文的统计过程，以下列出部分数据和统计结果，在表 14-2 中列出了 11 个临床试验（其中前 6 个为随机对照实验，后 5 个为自身对照实验）的 SANS 总分的资料，在表 14-3 中列出了 11 个临床试验的元分析结果，在表 14-4 中列出了 6 个随机对照试验 SANS 总分的元分析结果，在表 14-5 中列出了 5 个自身对照试验的 SANS 总分和 BPRS 总分的元分析结果。

表 14-2 11 个临床试验的 SANS 总分的资料

研究	接受处理前 （试验组）		接受处理前 （对照组）		接受处理后 （试验组）		接受处理后 （对照组）	
	M	S	M	S	M	S	M	S
1	18.54	11.45	16.39	12.30	12.35	10.21	15.50	12.15
2	43.17	12.24	43.91	14.29	30.09	3.98	39.30	14.17

续表

研究	接受处理前 (试验组)		接受处理前 (对照组)		接受处理后 (试验组)		接受处理后 (对照组)	
	<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>	<i>M</i>	<i>S</i>
3	55.40	19.00	55.68	18.64	45.18	16.26	52.73	15.26
4	51.68	9.48	47.81	12.65	34.62	10.21	42.81	15.15
5	53.10	20.90	45.30	17.40	27.50	14.50	46.50	20.30
6	60.42	20.83	61.47	29.45	43.68	20.11	60.37	29.43
7	37.20	4.30	—	—	28.10	3.20	—	—
8	31.53	20.14	—	—	24.03	19.95	—	—
9	83.08	24.17	—	—	49.92	28.65	—	—
10	30.65	12.14	—	—	19.87	5.14	—	—
11	72.09	20.70	—	—	31.36	16.38	—	—

表 14-3 11 个临床试验 SANS 总分及 BPRS 总分的元分析

	<i>u</i>	<i>P</i>	χ^2	<i>P</i>		
SANS	16.00	<0.01	58.29	<0.01	1.11	0.12 - 2.10
BPRS	17.02	<0.01	122.18	<0.01	1.66	-0.28 - 3.60

表 14-4 6 个随机对照试验 SANS 总分的元分析

Winer 法		异质性检验		\bar{d}	95% CI
<i>u</i>	<i>p</i>	χ^2	<i>P</i>		
6.08	<0.01	6.10	>0.25	0.68	0.46 - 0.90

表 14-5 5 个自身对照试验的 SANS 总分和 BPRS 总分的元分析

量表	Winer 法		异质性检验		\bar{d}	95% CI
	<i>u</i>	<i>p</i>	χ^2	<i>P</i>		
SANS	13.34	<0.01	58.29	<0.01	1.17	0.02 - 2.32
BPRS	18.76	<0.01	122.18	<0.01	2.05	0.28 - 3.82

总的来说，该文的统计过程还是比较清楚的，但作者没有明确指出使用了哪些效应指标，而且仅给出了使用公式的出处，没有给出具体的公式，不太便于读者理解。此外，该文将随机对照试验中的试验组在试验前后的 SANS 总分和 BPRS 总分进行元分析是不恰当的，因为试验组接受的试验处理是音乐疗法与抗精神病药物合用，这样做的话就无法将抗精神病药物的试验效应剥离开。最后，该文其实还可以根据试验时间、患病时间等研究特征进行更细致的亚组分析。

七、结果的敏感性分析

该文是采用“改变纳入标准，考察结论有无变化”的方法进行了敏感性分析，并在文中做了如下交待：“无论合并随机对照试验、自身对照试验及所有临床试验都得出音乐疗法辅助治疗慢性精神分裂症有一定疗效的结论，尽管其平均效应值略有差别。”

八、元分析结果的分析和讨论

在讨论部分，该文对元分析结果的意义，纳入研究的质量等问题进行了简单讨论，给出了一些建议。但基本没有涉及到各种偏向的识别和控制的问题以及对变异来源的分析。

九、根据统计结果作出正确、全面的结论

在摘要中该文给出了结论，即：“音乐疗法辅助治疗慢性精神分裂症近期疗效较好，其远期疗效有待进一步探讨。”

►本章提要

1. 元分析就是应用特定的设计和统计学方法对以往的研究结果进行整体的、系统的定性与定量分析。它是回顾性的，是对传统综述的一种改进，是概括以往研究结果的一种方法，包括大量的方法和技术，具有全面、系统和定量的特点，可以用以对以前的具有不同研究设计的和不同时期收集到的资料进行整合。

2. 发表偏向是指具有统计学显著意义的研究结果（阳性结果）较无显著性意义的结果（阴性结果）或无效的结果被报告和发表的可能性更大的倾向。

3. 要保证元分析的质量必须注意以下几个问题：要有合理的研究设计和文献检索策略，保证文献的查全率和查准率，进行严格的文献筛选和质量评价。

选择恰当的统计学方法进行分析；注意控制各种偏向；④进行异质性检验，以保证合并分析的可比性和分析结果的可靠性；⑤对结果进行敏感性分析，并予以合理的解释和应用。

4. 制定文献的纳入标准时通常会考虑研究对象、研究设计、结果变量、研究开展的时间或文献发表的年限和语种以及样本量等因素。

5. 元分析中提出的研究问题应包含四个要素：研究对象、研究设计、处理因素（干预措施）、研究效应（因变量、结果变量）。

6. 全面的文献检索会使用计算机检索、手工检索、从有关研究的参考文献中追踪查询和专家询问法等方法。

7. 研究的质量评价方面会涉及到选择偏向的情况、实施偏向的情况、测量偏向的情况、是否存在排除偏向、样本量的大小以及统计方法是否正确等方面。

8. 为保证数据收集的质量，最好由两人以上的研究者独立进行数据提取工作，然后进行交叉核对，并在文中详细叙述观察者的一致性。此外，在数据提取中应采用盲法，即隐去那些对资料提取者可能产生影响的因素，如期刊名、作者、作者单位、基金资助情况等，以减少选择性偏向。

9. 元分析需要将收集到的统计量（多个独立研究的结果）合并（或汇总）成某个单一的效应量或效应尺度，即用某个合并统计量反映多个独立研究的综合效应。

10. 元分析需要对纳入研究的结果进行异质性检验，目前广泛使用的是 Q 检验。

11. 亚组分析就是根据不同的研究特征如被试的年龄、职业，将各独立研究分为不同组，然后再分别进行合并分析，比较各组及其与总的合并效应间有无显著性差异。

12. 敏感性分析就是分析当条件（变量）发生变化时，结果是否也发生变化，即研究结论的稳定性如何。

13. 元分析的结果讨论一般会涉及分析变异来源、讨论各种偏向的识别和控制、讨论元分析结果的意义等方面。

► 本章关键术语

元分析(meta-analysis)	联合标准差(pooled standard deviation)
文献综述(literature review)	
异质性(heterogeneity)	异质性检验(tests for heterogeneity)
偏向(bias)	
发表偏向(publication bias)	同质性检验(tests for homogeneity)
选择偏向(selection bias)	权值(weight)
纳入标准(inclusion criteria)	固定效应模型(fixed effect model)
效应量(effect size, ES)	随机效应模型(random effect model)
标准化均数差(standardized mean differences, SMD)	亚组分析(subgroup analysis)
	敏感性分析(sensitivity analysis)

►复习与练习

1. 名词解释

元分析、发表偏向、Q 检验、亚组分析、敏感性分析、研究的质量、标准化均数差

2. 简答与论述

- (1) 什么是桔子与苹果问题，该如何控制？
- (2) 简述 Fisher(p 值) 法的基本过程与公式。
- (3) 简述传统的文献综述的问题与元分析的价值。
- (4) 简述元分析中存在的各种偏向及其控制办法。
- (5) 如何进行元分析的质量控制？
- (6) 如何进行全面的文献检索？
- (7) 对文献进行编码，并根据纳入标准，选出符合要求的纳入文献的基本步骤是什么？
- (8) 研究的质量评价从哪些方面进行？
- (9) 如何应用质量评价的结果(得分)？
- (10) 如何正确提取数据？
- (11) 如何进行敏感性分析？
- (12) 如何讨论元分析的结果？
- (13) 论述元分析的基本操作过程？

►推荐参考读物

Sutton A. J. , Abrams K. R. , Jones D. R et al. Methods for meta-analysis in medical research. Chichester: John Wiley & Sons Ltd, 2000

Rosenthal R. . Meta-analytic procedures for social research. Newbury Park CA : Sage 1991

Glass G. V. , McGaw B & Smith M. L. . Meta-analysis in social research. Beverly Hills, CA : Sage 1981

Hedges L. V & Olkin I. . Statistical methods for meta-analysis. New York : Academic Press 1985

Hunter, J. E & Schmidt F. L. . Methods of meta-analysis: Correcting error and bias in research findings. Newbury Park, CA: Sage 1990

►在线资源

想要了解元分析以及有关软件的更多知识和信息可以访问：

<http://www.meta-analysis.com/>和 <http://statistics.com/>

第十五章 研究报告的撰写与交流

到目前为止，你已经学习到了很多心理学研究的概念和程序，知道了如何针对研究的问题查阅文献、进行研究设计、实施实验、搜集和分析资料，也许对所要做研究进行了细致的思考。为了进一步完善研究，还必须考虑如何表达研究成果。我们之所以知道不同研究者的贡献，就是因为他们将自己的研究报告出来了。假定我们进行了一项研究，如何将自己的研究写成他人读得懂的报告？如何确定自己的报告是否规范？又有哪些渠道和方式让别人知道自己的研究发现呢？对于这些问题的解答，将牵涉到如何撰写与交流研究报告。研究报告的撰写和交流是心理学研究中不可缺少的环节。在本章中，将讨论以下问题：

1. 研究报告的概念与类别
2. 研究报告的写作风格
3. 如何撰写研究报告的各个部分
4. 评估研究报告的标准及发表与交流的注意事项

第一节 研究报告概述

一、什么是研究报告

把研究的历程与结果用文字或口头形式加以正式表达称为研究报告（research report）。

撰写研究报告的基本目的是信息交流。通过及时规范的报告，可使研究为他人所知，展现研究的价值与功能，有助于研究成果的推广和交流。报告中提供的研究方法信息，可以让其他研究者评价研究的质量，也可以使其他研究者重复研究和发展该研究，促进心理学的发展。研究报告还反映研究的水平、价值以及研究者的严谨态度。

心理学研究报告的基本特点是理论性、创造性和规范性。

理论性又称学术性。心理学研究报告是一种学术性文章，它要求运用心理学的原理和方法，对所研究领域的问题进行分析、论证和抽象概括。虽然研究报告是基于实验、观察、访谈等具体的方法，获得的是具体的资料，但报告绝不止于客观描述，而是要提炼、加工，遵循逻辑与实证的法则从理论上作出说明。

创造性是研究报告的重要特征。一方面，研究报告的创造性来自于研究的特征，研究的质量决定了报告的水平。心理学研究的目的就在于发现，为人类创造新知识，如果研究只有继承，没有新见解、新发现，没有创新，就失去了报告撰写的意义和价值。另一方面，创造性也是研究报告写作过程的特点，优秀的写作是创造性的（Solso, 2002）。写作和思考是联系在一起的，研究报告有时并不是思路清楚才写的，写作中经常有思想斗争，有时甚至中断写作。

研究报告要遵循公认的表达规范。为了达到交流的目的，避免给读者带来阅读困难，研究报告要遵循公认的表达方式和结构形式。不同研究报告因研究主题、方法、目的而表现出的特色，也应该在公认的表达规范内体现。一般会依次表达题目、作者姓名和单位、摘要、关键词、前言、方法、结果、讨论、结论、致谢、参考文献和附录等内容基本要素，有的内容如致谢、附录等可以根据研究的具体情况取舍。

二、研究报告的类型

各类报告涉及的内容纷繁复杂，表现的形式也不尽相同，按照不同的标准可以将其分为若干类型。

（一）应用型和学术型报告

按照读者对象的不同可以把研究报告分为应用型报告和学术型报告。

应用型研究报告的主要阅读对象是实践工作者，是为解决实践问题而作，其结构较简单，表达较通俗易懂。学术型研究报告的主要阅读对象是专业研究者，通常为了解决理论问题而作，要求表达的学术规范性较强，大多在学术期刊上发表。

（二）质的研究和量的研究报告

按照研究的取向可以划分为质的研究报告和量的研究报告。

量的研究报告主要包括实验研究报告和相关研究报告。实验类研究报告应该重点描述实验中如何操纵自变量、排除无关因素的影响、测查因变量的过程，以及如何揭示出研究现象的因果联系。相关类研究报告重点叙述在调查等研究中，研究工具、研究过程的特征，以及如何识别和检验变量及其相互关系，为进一步提出变量间可能的因果关系奠定基础。

质的研究报告是非实验研究的报告。与量的研究报告一样，质的研究报告以清楚地言明研究过程为目的，但写作涉及的规则较少、对结构要求也不多。质的研究报告倾向于长篇叙事，因为在收集资料、分析类属、组织证据时，较少使用单一的技术，报告的资料比较难以浓缩。除了呈现事实证据与诠释之外，可使用多种写作手法以及文学式的写作风格，让读者产生共情、了悟等主观感受。

（三）研究报告和研究结果报告

按照研究报告与研究进行的先后关系，可以把研究报告划分为研究报告和研究结果报告，通常所说的研究报告即研究结果报告。两种报告在文献综述、假设陈述、方法等方面的结构基本相同。

研究报告陈述的是将要做的事，只有对研究结果的设想和讨论思路的设想。研究报告通常在小范围内交流，如课题评审委员会或口试委员会等。表达的重点是要做什么，使他人了解研究者所要进行的研究，以确定研究的价值或提出改进建议。

结果报告陈述的是已经做过的事，具有实际的研究结果和对这些结果的讨论。结果报告表达的重点是已经做了些什么，使他人了解已经进行的研究，以对其进行评估。

（四）学位论文、学术期刊报告和会议报告

按照研究报告的写作目的可以划分为学位论文、学术期刊报告和会议报告。

学位论文是高等院校毕业生用以申请授予相应学位而提出作为考核和评审的研究报告。学位论文分为学士、硕士和博士三个级别，用以申请相应的学位。学位论文报告行文最为详细，是对整个研究工作全过程的详尽描述，要求既要充分表达研究的成果，又要表明已掌握的有关知识、方法或技术，具备了申请该学位的学术水平。

学术期刊报告是在专业刊物上发表的报告。写作期刊报告要求作者不仅具有高水平的科研能力，而且具有良好的文字功夫，能够简明扼要说明问题。这类报告虽然追求简洁，但需要列出较为详尽的参考文献，引导读者去寻找更详细的论据和材料。

会议报告是各种会议上发布的研究报告。会议报告需要突出重点，多以研究项目中的某个小题目为报告内容，以便说深说透。对听众感兴趣的方法、发现或结论，应作为重点详细阐述。如果研究还没有明确、肯定的发现，可分析造成当前状态的原因，可能的解决途径，也可提出具有挑战性的问题。

三、研究报告的写作风格

研究报告是陈述性和说明性相结合的应用文文体。作为一种科技文体，在表达方式上应以说明为主，将研究对象、存在的问题、研究方法、结果等内容解释清楚，使读者了解和信服。在语言运用上要求客观、准确、简洁。

客观(objective)就是使用中性的语言词汇把事实告诉读者，避免使用主观或带感情成分的文字，不企图去说服读者。在文献回顾时，应报告正反面的资料，不能只引用自己喜好的资料，或是断章取义；在报告研究过程时，不能故意隐瞒研究中所出现的问题，例如实验控制变量控制不当，也应客观陈述；在结果报

告时，不管研究假设是否得到支持，应客观报导所得结果，不能隐藏真正的研究结果，也不能根据结果而删改资料分析前所提出的研究假设；讨论部分是运用逻辑和实证的力量吸引读者阅读，启迪思考，而不是限制读者。

准确 (precise) 是学术报告最基本的要求，要求使用公认的学术语言表达，用词恰当、搭配合理，以合乎逻辑和语言习惯的方式行文。在行文中陈述要真实可靠，避免模糊的词汇，仔细区分近义词、同义词在含义上、用法上的细微差别。变量和术语在行文的过程中不要转义，以免造成混淆，对于第一次使用的术语则要给出明确的界定。对于复杂的观念或数据，如果无法用一种方式解释清楚，就需要变换角度表达。需要具体介绍的就不要含糊其辞，例如“大部分”、“很少”就让读者无法知道究竟是多少，“使用了某个测验”也没有说清楚究竟是哪个测验。

简洁 (parsimonious) 是对学术报告的重要要求，即用最少的文字将研究表达清楚。既是节约有限的出版空间，也是对读者时间的尊重。要求实施过程不作过多描绘，观点的陈述不做繁琐论证，行文表达直截了当，不拐弯抹角，坚决删除不必要的文字。文献探讨部分切忌为增加篇幅而找一些无关的资料充数；方法部分除非是新异的方法，不需要详尽无遗的交代，特别是专业学术报告，其读者都理解研究中所使用的常用方法，更要避免不必要的啰嗦；结果部分则可以合理地运用图表配合文字表达。

将过长的表达或者在文中多次出现的词用简略语代替，或用外文符号代替，会使行文简洁。但在一篇报告中大量使用缩略会给读者的阅读带来不必要的负担。因此，追求简洁的同时，也要避免不必要的省略。

另外，一篇好的研究报告还应该兼具朴实与生动的特点。朴实是避免华丽的词藻，不能使用随意的夸张和奇特的比喻，避免口语化以及晦涩难懂的文字；而生动则是在准确前提下提高文章的可读性。

补充讨论 15-1 研究报告的写作过程

研究报告与一般的“作文”类似，但又具有自己的规定性和特点，所有的这些都可能令初学者感到无法胜任。总结初学者们的体会，提出以下几点建议：

1. 开始的时候不要犹豫。写作不是研究结束时的总结，而是个过程，有的研究甚至将写作贯穿始终。拖延不但危害研究的进程，也容易使研究者感到不安，如在质的研究中如果不及时处理资料，研究者最后面对的可能是堆积如山的资料。

2. 不要怕修改。无论是准确、简洁的文风，还是学术性与创造性的要求，都使得研究报告的写作充满了挑战。随着写作技巧的提高和对问题认识的推进，

修改对于每一篇报告几乎都是必需的。

3. 遵循出版指南。各种写作指南看起来很复杂，但它将复杂的报告过程简化为一些具体需要做的工作和注意事项，事实上简化了研究报告的写作。

第二节 研究报告各要素的写作

研究报告的各个组成部分都有一定的写作规范，遵循这些规范可以降低研究报告的写作难度，有助于读者阅读评价、研究，便于出版、促进交流。

一、题目

题目(title)即研究报告的名称，是研究主题思想、内容焦点的表达。

研究报告的题目十分重要，它是读者判断和决定是否阅读该文章的重要依据，好的标题往往可以争取较大的读者群，进而使论文体现本身应有的价值；题目也是对论文进行检索、收录和引用的重要依据，是检索本论文的主要标识。所谓的“题好文一半”就是强调题目的重要性，必须用心斟酌选定。

一个好的研究报告题目应该具备准确、概括和简洁的特征。准确是确定报告题目最基本的要求，即题目准确地表达研究的中心内容，使读者能通过题目了解报告的主题。概括是指题目做到涵盖全篇内容，避免题目与报告实际内容的不匹配情形，做到范围界定恰当，题文相符。简洁是指用简短明确的文字反映报告的主题，一般题目不超过 20 个字。要尽量避免使用多余的词语，如“有关……的研究”、“一项……的研究”完全可以省略。

研究报告的题目通常有变量式、主题式和结论直陈式。

形式 1：变量式。这类题目由研究的主要变量组成，如

输出方式和输出时间对数字广度的影响（宣宾、张达人，2004）

加工速度和工作记忆在认知年老化过程中的作用（李德明、刘昌、陈天勇等 2003）

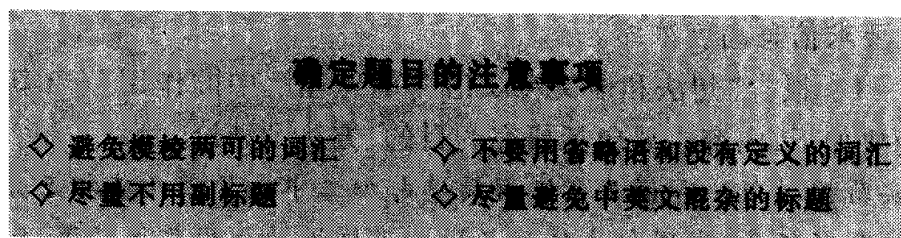
形式 2：主题式。这类题目表达研究的主题，其涉及的变量往往不易通过标题表达出来，如

7 岁超常和常态儿童的信息加工速度（邹枝玲、施建农、恽梅等，2003）

中文句子中双音节兼类词句法分析历程初探（张亚旭、刘友谊、舒华等，2003）

发展性协调障碍与书写困难个案研究（孟祥芝、周晓林、吴佳音，2003）

形式 3：结论直陈式。这类题目直接表达研究的结论，如



二、作者姓名和单位

作者姓名和单位 (names and affiliation) 是研究报告的署名问题。署名一是为了表明文责自负，二是记录劳动成果，三是便于读者与作者的通讯联系及文献检索 (作者、所属机构索引)。

作者的署名是不加任何称谓的，例如“教授”、“博士”等头衔均不需要。如果需要对作者信息进行说明，则可以加作者注 (author notes)。作者注通常给出本文通讯作者的 E-mail、电话等信息，供读者联系；也可以对本研究报告作出说明，例如公开发表的研究报告是学位论文的一部分，可以在此说明；另外，还可以在作者注中致谢研究基金或对研究作出帮助的人。

多作者的研究报告按署名顺序列为第一作者、第二作者……。对研究工作与报告撰写实际贡献最大的列为第一作者，贡献次之的列为第二作者，依次类推。如果作者属于不同的单位，要分别列出。

三、摘要

摘要 (abstract) 是报告内容不加注释和评论的简短概述。应用型研究报告中篇幅较长的摘要称为执行提要 (executive summary)。学术期刊上发表的研究报告、学位论文都要求有摘要和外文摘要 (通常为英文)。

摘要是与报告主要信息量等同的完整短文，是研究是否有价值的最简单表现形式。摘要的功能表现在两个方面，首先，摘要补充题目的不足，担负着吸引读者和将报告主要内容准确地介绍给读者的任务；其次，摘要为情报文献检索数据库的建设和维护提供方便，直接利用规范的摘要，可以避免在加入文摘杂志或数据库时，由他人编写摘要可能产生的误解、欠缺甚至错误。

从内容上看，摘要是全文的高度浓缩，提供的信息包括研究的目的、对象、方法、结果、结论和应用范围等。实际写作中并不要求每篇摘要都要具备以上六个方面，但是研究对象和研究结果是必不可少的。

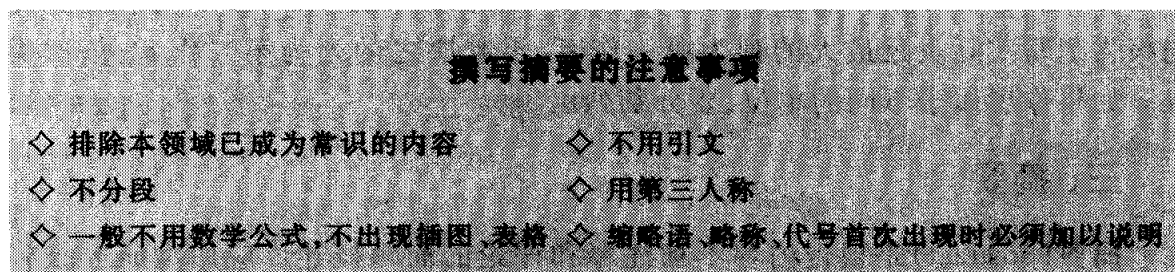
从形式上看，摘要撰写要求正确、精炼、具体、完整。正确，就是忠实于原文，使用规范化的名词术语；精炼，就是简明扼要；具体，就是要把关键的步骤、方法、数据、结论交待清楚；完整，就是要语意连贯，能独立成文。摘要长短各有不同，

从几十个字到一整页，通常在 300 字以内。摘要行文简短又要表达研究的主要内容，因此要求信息高度浓缩。

例 15-1 报告“句法和语义在汉语简单句理解中的作用”（江新、荆其诚，1999）的摘要如下，内容主要包括研究的目的、方法及结论。

两个实验探讨句法和语义在汉语简单句加工中的作用及其相互关系。实验一采用施事判断任务，发现句法作用与语义可逆性的作用都很显著；实验二采用快速句子判断任务，发现句法作用是主要的，语义可逆性无显著作用；两个实验都发现句法与语义可逆性之间的交互作用不显著。这些结果表明：（1）在汉语简单句理解的过程中句法因素的作用是主要的，语义因素的作用是次要的，只有当实验任务更多地强调语义分析时，语义因素才有显著的作用。（2）句法因素、语义因素分别作用于汉语简单句理解过程的不同阶段，句法分析是一个可以独立于语义的自主加工过程，支持句法自主理论。

研究报告附外文摘要是为了国际交流，通常用英文表达。英文摘要通常是题名、摘要和关键词的英译。在撰写英文摘要时，其内容除了与中文对应之外，还要注意符合英语的表达方式、语言习惯，如使用第三人称、被动语态，省略可有可无的助词等。



四、关键词

关键词(keyword)是研究报告的文献检索标志，是表达报告主题概念的词或词组。关键词是论文信息最高度的概括，选择是否恰当，影响读者对文章的理解，关系到该文被检索和该成果的利用率。一篇论文的关键词通常有 3~8 个，位于摘要之后。

关键词的表达要符合规范，通常从专业词汇表中选用，如心理学名词审定委员会审定的《心理学名词》(2001)。如果专业主题词表中没有列出该词，或者为表达新理论、新技术等新出现的概念，则可以选择有关词汇作为关键词，而且尽可能选自权威的参考书和工具书，选用的词必须达到词形简练、概念明确、实用性强。既不要生造关键词，也不要对新现象或新发现生硬地套用既有的词汇。

关键词选用的关键是能够真正反映报告的主旨，例如，研究报告“青少年时间管理倾向量表的编制”（黄希庭、张志杰，2001）给出的关键词为“青少年、时间

管理、时间管理倾向量表”，就准确表达出了报告的中心内容特征。

五、前言

前言 (introduction) 是研究报告的重要组成部分，用以陈述研究问题、解释所要验证的理论、表明研究目的的重要性，提供研究问题背景说明。

前言在行文时可以用“引言”“问题的提出”“文献回顾”“研究假设”“背景假定”等标题，不同的研究报告前言所包含的内容有所不同，是否用小标题也可斟酌，但均需要讲清本研究所要解决的问题、研究目的、领域内的研究状况等内容。

(一) 研究的目的和问题

研究目的 (purpose of a study) 是说明所进行研究的必要性以及研究要探讨的方向，即进行此项研究有何价值及其理论意义和实践意义是什么。较常见的研究目的有验证理论、澄清过去研究的矛盾，或是提出解决问题的方法等。在撰写时，需要较详细地说明研究的背景及其重要性，对研究的目的可以只作简单明了的叙述。

研究问题 (research question) 是具体地指出要研究的问题所在，是针对研究目的所列出来具体的问题。通常一个研究目的会引申出数个研究问题。

例 15-2 研究报告“不同复述条件下命名难度对图形记忆的影响”研究目的表达 (丁锦红、林仲贤, 2002)。

复述对图形记忆肯定有促进作用，了解复述对图形记忆的影响规律将有助于认识图形信息的存储与提取特性。本研究的目的是通过比较自由复述、部分特征的言语命名复述和视觉表象复述三种实验条件下的图形再认成绩，探讨图形信息进入长时记忆过程中复述的特点以及图形信息转化对图形记忆的影响等。

例 15-3 研究报告“汉字阅读中的语音通达与表征 (I)”中研究问题的表达 (杨琿、彭聃龄、Perfetti 等 2000)。

本研究回答 3 个问题：(1) 语音激活是否发生在多个层次，即字的层次与声旁层次；(2) 多层次的语音激活是否在高频和低频字中都存在；(3) 语音激活的两个资源 (整字读音、声旁读音) 之间的交互作用是什么。

(二) 文献回顾

文献回顾 (literature review) 是对以往研究的述评，对相关理论和研究的说明与总结。文献回顾是一个较长的部分，在学位论文中表现得尤其突出。

文献回顾的目的是评价课题和发展课题。通过回顾文献论证当前研究问题的可行性，明确研究的问题能否通过本研究加以回答，提供此研究的理论基础和提出研究假设的依据。通过回顾文献，可以确定当前研究问题的意义性，论证研

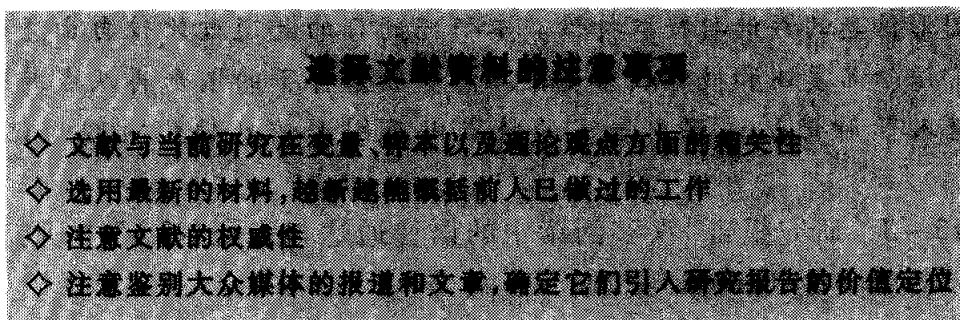
研究的深度或研究将要做出的贡献。文献回顾也表明研究者对于所研究领域的熟悉度和理解深度。

文献回顾无固定的写法，一般是先系统介绍基本的理论，然后才探讨有关的研究。在论述过程中，文献材料可以按与研究问题的关系组织，也可以按时间顺序组织。要避免援引一些无关紧要的文献，也不要面面俱到，应从复杂的材料中选择与自己的研究关系最密切、有代表性的材料，将它们组织起来，阐明以前研究与当前研究的关系，为读者提供一个理解研究问题的背景。对于无法找到直接理论依据的探索性研究，则必须陈述有关间接的理论或研究。

文献回顾不能流于叙述，而必须作评论。对有关的理论和研究的归纳、评价，既是在帮助读者理解文献，也是在论证作者自己的研究。因此，文献回顾过程中，要注意先前研究的不足和没有解决的问题，仔细检查那些研究结果不一致的方法，从而作出相应的评价。

在回顾研究的文献背景之后，应直截了当地提出研究问题。除了要说明问题是什么，包括哪些子题、问题的性质与重要性等之外，还要说明其在本领域的研究状况、可能的答案（本研究的假设）。所提出的每一个研究假设，均需要充分的文献支持，并表明逻辑推理过程。对于假设中涉及的变量、重要名词要加以界定，并给出变量的操作性定义。

研究问题的论述和研究假设的提出均需要文献支持，因此它们往往与文献回顾结合起来。事实上，一个好的问题陈述就是从研究文献到本研究问题与假设的自然过渡。



六、方法

研究方法（research method）提供用以检验假设的设计和程序，详细说明有关如何执行研究的细节。研究方法包括样本选取、研究架构、研究工具、研究步骤及数据处理等组成部分。

研究方法是用以评估本研究方法论最重要的部分，它使其他研究者在必要时可以重复所报告的研究。

研究报告的阅读对象都有一定的方法知识，这部分的描述不必过于细致，但

一定要提供给读者以下重要问题的解答：执行的是什么类型的研究；样本的选择方法是什么；样本的基本资料为何；资料是如何收集的；采用的什么程序；变量是如何测量的；用的是什么仪器或工具；是否具有信度与效度；研究设计中的伦理议题与特定议题是如何处理的。

（一）研究对象

研究对象 (participants) 有时也称为被试 (subjects) 或样本 (sample)，即明确指出“谁”是该研究的对象。

样本选取主要是说明所选样本的大小、特性（如性别、年龄、职业等）、样本选取的方法（如简单随机取样、分层随机取样、立意取样等）、实验研究分组原则及样本的来源。如有必要，还应指明被试是否自愿参加研究。以动物为研究对象，也要交代其基本的特征。

例 15-4 研究报告“责任、情感及帮助行为的归因结构模型”中对人被试的描述（张爱卿、刘华山 2003）。

被试包括管理者（企业领导、行政干部及高校管理人员）和被管理者（职员、工人及教师），共 238 人，主要来自于华中地区。其中管理者 143 人（男 84 人，女 59 人）年龄范围为 37.62 ± 7.86 。被管理者 95 人（男 27 人，女 68 人），年龄范围为 31.26 ± 6.27 。这些被试均有两年以上的工作经验。

例 15-5 研究报告“情绪应激体液免疫调节作用的影响因素”中对动物被试的说明（邵枫、林文娟 2001）。

三组实验共 90 只雄性 Wistar 大鼠，体重 250 ~ 300 克，鼠龄 3 个月左右，购于中国科学院遗传所动物中心。每只大鼠单笼喂养，控制室温 ($20 \pm 3^\circ\text{C}$) 和照明 (12 小时照明 / 12 小时黑暗，每天早 7 时开始照明)。适应期内所有动物自由饮食、水，适应期后动物定时饮水、自由饮食，饲料由北京市实验动物中心提供。

（二）研究工具

研究工具 (research instrument) 是在研究过程中所使用的仪器或者材料，包括实验仪器、测量工具及自制的材料或量表等获取数据资料的工具，以及数据处理的工具。

对于定型且众所皆知的工具可简单交代，例如，器械说明其厂牌、型号，公认量表说明其信度和效度。对于鲜为人知或自行设计的工具则需要详细说明。例如，自行设计的问卷，要说明问卷包含有哪几个主题，并将整个问卷放在附录里，供阅读者参考；研究者自制的量表，则要详细的说明整个编制过程，包括编制依据的理论，预试、项目分析及选题的过程，以及此量表的信度和效度等。

数据处理的工具主要报告研究所使用的统计分析方法以及统计程序。如果使用多种方法，应该分别对应交代这些方法所要解决的问题或处理的假设。使用公认的统计程序包，要说明其版本而不必交代算法，例如 SPSS10.0 等。

例 15-6 研究报告“回忆准备就绪程度的判断发展”对自行设计实验材料的描述（刘希平、唐卫海,2002）。

词表 3 组共 6 张：按词对组合方式不同，分成三种材料：即反义词对表，如大船 - 小船、坏人 - 好人等；动宾词对表，如关门 - 关灯、灭火 - 灭虫等；人为组合词对表，如月亮 - 刻字，尺子 - 抱羊等。

例 15-7 研究报告“航空模拟舱飞行行为的 EPIC 理论仿真”对所使用的自行设计仪器的说明（符德江、沈模卫,2001）。

实验装置采用浙江大学心理学国家专业实验室开发的飞机模拟座舱。模拟座舱的功能结构如图 15.1 所示。模拟座舱的实物模型由三个部分组成：(1) 飞行视景系统；(2) 动力电子系统，处理操纵杆和多功能显示器按钮的输入、仪表数据的输出和被试反应的记录；(3) 座舱辅助设施，如飞机座椅和座舱框架。飞行视景系统和动力电子系统由两台计算机管理，并通过 TCP/IP 网络协议进行通讯。

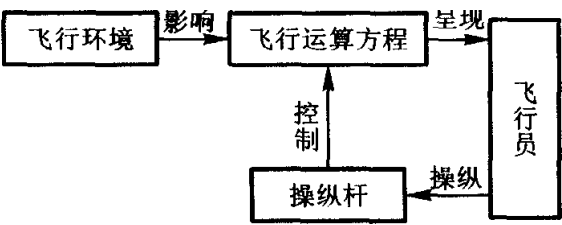


图 15.1 飞机模拟座舱的功能结构图
（采自符德江、沈模卫,2001）

（三）研究程序与步骤

研究程序（research procedure）是研究工作的具体流程。研究程序通常包括指导语和实施研究的具体步骤，说明如何引导被试使用实验材料或工具进行研究。对研究程序与步骤的描述可以使研究的重复检验成为可能，基于这一部分的描述，其他的研究者可以利用相同的程序进行重复验证。

为了让读者了解研究的方法，有必要将研究中的每一步描述清楚。研究程序与步骤的描述应该包括研究进行方式，数据收集方式，如观察法的时间取样方式、实验法的实验设计与实施、调查法的问卷分发回收情形，信度与效度检测方法，以及数理分析模式或统计方法。

七、结果

结果(result)是对本研究得到的数据进行分析，并报告所得到的结果。

（一）结果表达的内容与顺序

结果报告是将研究所得到的必要数据、典型案例、观察记录等，用统计图表、

照片结合文字，客观、准确地叙述出来。报告结果时，不讨论假设是否得到证实，也不必对此结果加以评论，更不要呈现不相关的资料。

结果表达一般按照研究假设的提出顺序，采用由总到分的方式组织叙述，即先对假说进行简要的重述，然后用数据描述结果，或对假说进行检验，最后说明接受还是拒绝假说。如果内容较多，应分段叙述，必要时可以加小标题。

（二）统计检验结果的表达

统计结果的表达要说明选择某种统计方法、统计参数的原因，尤其要表明结果的意义。例如，“方差分析的结果表明：动静态的主效应显著， $F(1,72) = 8.245, p < 0.01$ ，说明动态与静态条件下的平均错误率之间存在显著差异”（郭秀艳等 2004）。

在呈现推论统计结果前，要先呈现描述统计资料，即先将研究变量的平均数、标准差及样本量等给出，使读者了解各种基本资料。

（三）图表的使用

在结果报告中常使用图表，特别是在数据量较大时，统计图表可使读者对研究发现获得概括性认识。较常见的图有直方图、圆型比例图、交互作用图等。常用的统计表格有 t 检验表、方差分析摘要表、相关矩阵表、回归分析摘要表等。每个图表之后，都应该加以解释，并指出其中的重要特征，让读者了解其意义。但在说明图表的时候，应避免简单重复图表内的数据。

例 15-8 统计图（如图 15.2 所示）

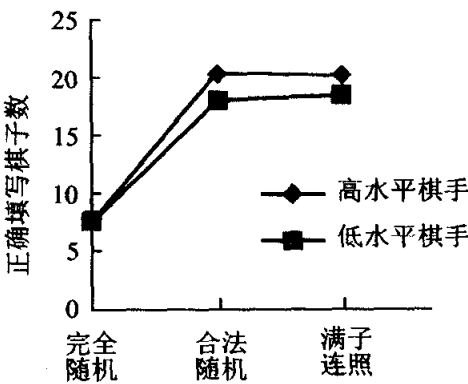


图 15.2 棋手对棋局的回忆成绩
（采自杜建政、杨治良,2002）

呈现统计图的注意事项

1. 统计图的标题要放在图的正下方。
2. 纵轴和横轴应有小标题。
3. 应标示出坐标轴的单位 and 刻度，横轴若是以名义变量来分组，则只要标示出各组的名称即可。

4. 纵轴的单位应由下而上依次增大，而横轴的单位则由左而右依次增大。
5. 纵轴和横轴若不是由原点开始，则应标示缺口。

例 15-9 统计表(如表 15-1 所示)

表 15-1 高低回报谨慎者在谈判过程中与对手分享信息的差别

信息分享	高回报谨慎的谈判者		低回报谨慎的谈判者		$F(1,164)$
	M	SD	M	SD	
直接	15.00	2.05	15.98	2.03	9.51**
间接	6.30	1.10	6.66	1.10	4.49*

注：* $P < .05$ ** $P < .01$

(采自张志学、韩玉兰，2004)

呈现统计表的注意事项

1. 表以居中的方式放在内文的中间。尽量排入一页，但不要太拥挤。
2. 标题必须能确切说明表格的内容，而且标题要放在表的上方正中央。
3. 标号按顺序编排，如表 1、表 2……标号和标题的文字间空一格。
4. 表格的行、列应有适当的表目。
5. 一般采用三线格，表内只有横线，无竖线。顶线和底线用粗线，其间的横线则用细线。
6. 表中的小数点上下对齐。表中的数值一般四舍五入到小数第 2 位，若不到整数，小数点之前要加上 0； p 值可列出小数到第 3 位以外，且小数前面不需要加 0。
7. 若统计检验达到显著水平，在表中的 F 值或 t 值等统计检验值右上方标注 *，并在表左下方加注。
8. 统计软件输出的表格通常存在冗余信息，要经过整理后呈现。

八、讨论

讨论(discussion)是研究者对研究结果进行评估，探讨结果支持研究假设或不支持研究假设的原因，对其理论意义或实践意义进行综合解释的过程。讨论使整个报告成为一体，是研究报告最重要的部分。

(一) 讨论的内容

讨论要基于本研究的结果，但不局限于本研究。讨论的内容可以从以下几个方面考虑：

1. 说明研究的结果，解释新出现的现象

基于心理学的原理和规则，结合本研究领域的知识对研究结果作出说明。如，讨论某种差异对于验证研究假设、说明有关理论的价值；从理论的角度讨论

意外发现、产生不一致的原因及其对研究的意义。

2. 综合分析本研究的结果来说明某种理论和可能的展望

以理论的洞察力和逻辑的推演来综合讨论研究结果的意义和价值，如各项研究结果和推论是否验证了某项理论，是否具有发展理论、形成新的理论可能，是否具有实践上的应用价值。

3. 与已有的观点和发现作比较，分析异同和原因

通过与已有的观点和发现的比较，提供给读者有关数据和发现的意义，理解异同产生的原因。在比较时，要围绕如何解释研究结果和发展理论，切不可为了比较而进行琐碎无力的理论比较，也不要做选择性的强调或派系之见。

4. 用相关领域的成果来解释本研究的结果和推论

用脑科学、计算机等有关基础学科的成果解释研究的结果和推论；用管理、教育、传播等有关应用学科领域来解释研究的结果和推论。在援引各相关学科成果解释心理学发现时，要避免漫无目的地解释，只有那些与研究数据、理论有密切联系的，才可以进行推测并展开讨论。

5. 研究的限制和进一步研究的建议

研究有时会发生超出研究者能力范围的一些问题，如实验操纵、抽样、被试者参与研究的动机、问卷回收率，等等。讨论这些研究的限制，可以让读者更深入地了解所得结果的意义。所有讨论的问题都应该是研究设计中的重要内容，如信度、内部效度、外部效度、内容和结构效度（Heiman, 1998）。

研究的结束往往是新研究的开始，研究者可以根据本研究提出进一步研究的建议。

（二）讨论的组织

并不是所有研究都包含以上五个方面的讨论，而是根据具体研究的特点展开讨论。讨论亦无固定的组织形式（写法），关键是突出理论深度和逻辑的力度。以下给出一种常见的组织方式（Myers & Hansen, 2002）。

在讨论的开始部分，先简单复述自己研究的发现，并解释研究的假设是否得到支持。例如：

当前的实验支持观看暴力影视会影响对暴力行为接受的假设。观看这些录像的被试比没有看暴力色彩的被试对暴力表现出积极的态度。

接下来，以前言中提及的论点为参考，论述与以前的研究成果的关系，是否与以前的研究一致。

这些结果与文献中报告的许多实验结果是一致的。例如，[引用]的研究表明……另外，在[引用]的研究中，相似的效应在被试……

如果，你的研究结果与其他人的研究结果不同，要解释产生该结果的原因什么：

当前研究的发现与 [引用] 等人的报告的发现不同。当前的研究发现.....；而 [引用] 发现.....对于这种不一致，最确切的解释是.....，[引用] 用的是一种不同的程序，这种程序会产生的结果会是.....

接下来说明研究的理论意义和实际应用，最后说明研究的局限和进一步改进的建议。

九、结论

结论(conclusion) 部分是将研究结果做一个总结，研究者在结论中重述研究问题，并且摘要叙述研究发现。由于是对报告做出摘要概述，因而有时可以总结为题。

结论须与题目挂钩，根据结果与假设之间的关系，简要叙述本研究最具意义的主要结论以及未来研究的方向。应用研究通常指出研究所得结论的应用范围，指明结果可否应用在教育、工业或商业等领域；理论性的研究则说明其在此领域中所具有的价值。研究者进行一项研究后，常常会体会到某些部分有深入研究的必要，这些可作为未来研究的建议提出。对于未来研究方向或研究限制，可以从理论的精确度，方法的替代性、代表性，资料的质量，结果的应用等方面来思考。

结论撰写通常使用简短的陈述句或判断句。例如：研究报告“儿童和青少年信息加工速度发展函数的研究”(林崇德、沃建中、於国荣，1997)所得出的结论是：

不同的加工任务，儿童反应变量与青年反应变量建立的函数关系是不同的。在句图匹配任务中，儿童的反应时是青年反应时的函数，指数函数能较好的描述这种共同的变化趋势，但在字母匹配任务中，这种线性关系较差；随着年龄的变化，信息加工速度的下降速率是根据不同的加工任务而变化的，简单任务（指选择反应和字母匹配）下降速度快，复杂任务（指句图匹配）下降速度慢。

十、致谢

致谢(Acknowledgement) 是对本研究的完成在学术方面有较重要帮助的团体和人士，在报告中予以书面致谢。致谢表达尊重他人的劳动、感谢他人的帮助，也表明作者恰如其分地评估自己的贡献。

致谢对象应是对报告有实质性贡献的人，包括指导者、技术协作者、提供特殊试剂或器材者、经费资助者和提出过重要建议者。不要泛泛地致谢，也不要只致谢名家、教授。写致谢前要征得被致谢者的同意，不能拉大旗作虎皮。有的学术期刊明确要求来稿致谢必须附被致谢者的签名。

致谢内容应简洁明了、实事求是。致谢应是真诚的、实在的，不要庸俗化。

致谢一般安排在参考文献之前，也可以将致谢辞置于篇首页末。学位论文还可以在适当的位置拿出专页写致谢文字。

补充讨论 15-2 站在巨人的肩膀上

科学的发展是一种接力跑，没有任何人漠视他人的成果而取得成功。只有尊重并吸取前人的成果，才可能具备能够超越前人的基础，许多大家都坦言自己是“站在巨人的肩膀上”才取得巨大的成就。现代科学研究也是一种团结协作的事业，没有任何人可以脱离他人而能够完成科学研究活动，成功的科研背后都有团队的支持。

不能深刻认识这个简单的道理，就很难取得成就和赢得他人的尊重。

十一、参考文献

撰写研究报告时，几乎都会引用（citation）他人的研究资料。参考文献（references）是撰写研究报告中所引用的有关期刊或图书资料。引用他人的观点、数据或材料在文中出现的地方予以标注，并在文末列出参考文献的工作称为文献著录。不管是直接引语还是间接引意，都需要进行文献著录工作。

对于学术报告，参考文献的著录是必不可少的工作。严肃的学术杂志不可能接受没有参考文献的投稿。著录文献指出所引述资料的正确来源，供读者评估时参证，显示所论有根有据，信而有征，有经验的读者，从作者标引的文献中就可以评估论文的起点和深度；同时，文献著录还可将自己的研究与他人的成果区分开来，维护学术忠诚与尊严，避免剽窃之嫌。从节省行文篇幅和有利于文献的计量分析的角度，文献著录也是必需的。

参考文献因放置位置而分脚注（footnote）与尾注（endnote）两种，脚注列在当页下端，尾注位置在报告正文之后，也称附注。文后注可使正文简洁，文意流畅，是研究报告最常用的注释方式。

参考文献多少没有硬性的限定，要视具体的情况而定，通常在 15 ~ 25 篇（400 ~ 600 字）之内为宜。参考文献的排列符号多以数字表示，且必须与文内一一对应。

参考文献的著录格式有一些共同的规范，因来源不同而有所差异。以下给出最常用的书籍、期刊等文献的著录格式。

1. 书籍 作者. 书名. 版本（初版不标）. 出版地点：出版社，出版年. 页码（选择项）

张厚粲 徐建平. 现代心理与教育统计学. 北京：北京师范大学出版社，2003

中国心理学会编. 心理学论文写作规范. 北京: 科学出版社, 2002

朱滢主编. 实验心理学. 北京: 北京大学出版社, 2000

Giles, D. Advanced research methods in psychology. NY: Routledge, 2002

2. 期刊杂志: 作者(超过3位时, 只著录3位). 篇名. 期刊(杂志)名称, 年, 卷(期): 页码

肖少北. 汉字学习与儿童分析能力和综合能力的关系. 心理学探新, 2002, 22(3): 33 - 36

吴兴曲, 杨来启, 李栓德, 等. 睡眠剥夺后大鼠海马生化及病理变化. 中国心理卫生杂志, 2003, 17(2): 77 - 79

Cook, S. R. A Note on Testing for Homogeneity among Effect Sizes Sharing a Common Control Group. Psychological Methods, 2004, 9(4): 446 - 452

3. 研讨会发表的报告: 作者. 题名. 会议名称, 会址(出版社), 会议年份. 页码(选择项)

邢红兵, 舒华. 语料库建设与汉语认知研究第三届华人心理学家学术研讨会. 北京, 1999

孟庆茂, 黄良东. 对价值观测验计分方式的研究. 海峡两岸学术研讨会论文集, 杭州: 浙江教育出版社, 1997: 81 - 89

4. 学位论文: 作者. 题名. 硕士/博士学位论文. 保存地. 保存者. 年份

于海琴. 小学生的亲子依恋与其同伴交往的关系研究. 硕士学位论文. 武汉: 华中师范大学, 2000

李伯约. 时间顺序标码的层次网络表征研究. 博士学位论文. 重庆: 西南师范大学心理学系, 2001

蔡永红. 教师职务绩效一结构及其影响因素研究. 博士学位论文. 北京: 北京师范大学发展心理研究所, 2002

5. 印刷中的论文或书籍: 在书籍期刊著录格式基础上标注“印刷中”

王娟, 王玉凤, 任园春. 注意缺陷多动障碍儿童平衡功能测定的病例对照研究. 北京大学学报(医学版), 2003, 35(印刷中)

6. 析出文献: 作者. 题名. 见: 编著者. 书籍名称. 出版地: 出版社, 出版年. 页码

林崇德. 儿童心理过程的发展. 见王甦, 林仲贤, 荆其诚主编. 中国心理科学. 长春: 吉林教育出版社, 1997. 477 - 541

周镐, 鲍碧君, 古茂盛. 小学儿童推理特点的调查研究. 见刘范主编. 心理发展的近期研究. 北京: 北京师范学院出版社, 1990. 116 - 124

7. 译著: 作者. 题名. 译者. 出版地: 出版社, 出版年. 页码(选择项)

珀文. 人格科学. 周榕, 陈红, 杨炳钧, 等译. 上海: 华东师范大学出版社, 2001

8. 数据库中引用的数据：作者.题名.网址，发表日期/检索日期

Grohol J M. Internet addiction guide. <http://psychcentral.com/netaddiction/>
1999

中国心理学会. 中国心理学会 2003 年工作总结. www.cpsbeijing.org/2003zj.doc 2004 - 1 - 13

9. 从报纸上引用的资料：作者.题名.报纸名称 年 - 月 - 日（版次）

黄光成. 单亲家庭学生心理亟待关注. 光明日报 2002 - 5 - 16

冯仕政. 应对非典疫情下的心理恐慌. 人民日报 2003 - 5 - 30(14)

十二、附录

附录(appendix)是正文的补充部分。因为受行文的限制，数学公式的推导、问卷、大型的表格、计算程序等信息不适宜放到正文，往往以附录的形式呈现。公开发表的研究报告只有在必要时才会刊印附录，但在评估研究（例如审稿）中，附录的内容往往是一个重要的组成部分。

一篇文章可以有多个附录，要按照顺序排列，并给出标题。附录中的图表、公式的序号应该与正文相区别。例如，正文中用“表 1、表 2、表 3……”排序，那么附录中就可以使用“表 A、表 B、表 C……”排序。

补充讨论 15 - 3 APA 格式

APA 格式(APA Style)是美国心理学会(American Psychological Association)的出版手册(Publication Manual)中有关投稿该学会所属期刊必须遵守的格式，也是社会科学领域学术期刊广为采用的格式之一。APA 的出版指南源于 1929 年刊登在 Psychological Bulletin 上关于投稿者准备稿件和期刊编辑审核稿件的写作指南。1944 年 APA 的编辑委员将其内容加以扩充，以鼓励年轻的学者从事专业写作，尤其是帮助第一次投稿的作者高效地准备符合期刊水平的稿件。1952 年正式以 Publication Manual 的名称出版，2001 年出版第 5 版。

APA 的出版手册中，从稿件的内容与组织、观念的表达、APA 编辑格式、稿件的准备与模板，到稿件的接受与出版、APA 期刊的政策与编辑对稿件的管理、参考书目等都进行了详细的阐释。其中在 APA 的编辑格式部分，对书目的写作与引用有详细介绍。

第三节 报告的交流

研究报告的交流主要有期刊发表和会议交流两种。

一、期刊发表

(一) 研究报告评估的标准

报告发表与否不但取决于学术质量，还受写作规范的制约。

评定一项研究所用的标准包括（中国心理学会，2002）：

- ◇ 研究的问题是否有意义，是否有创新之处。
- ◇ 研究的工具是否具有令人满意的信度和效度。
- ◇ 研究的设计是否合理。
- ◇ 研究假设是否得到了充分的检验。
- ◇ 研究结果是否真正反映了所考察的变量。
- ◇ 被试是否具有代表性。
- ◇ 研究的结论是否对以前的研究有所发展。

在研究设计和报告方面常出现的缺陷：

- ◇ 分拆出版。将一个报告分拆成几个内容重叠的报告。
- ◇ 分析不深入。报告简单的、没有实际意义的结果。
- ◇ 缺乏逻辑性。从研究设计到研究结果讨论，思路不够统一。
- ◇ 缺乏控制而使解释困难。研究中涉及的重要变量未加必要的控制。
- ◇ 没有报告效应的大小。
- ◇ 论述抓不住研究的重点。

(二) 投稿

如果研究达到了一定的学术水平，值得发表，而且报告撰写符合出版规格，就可以将研究报告投往有关的期刊，表达公开发表的意向。

在投稿前，特别是初学者，要按照研究报告评估的标准对研究报告进行评估。一定要通读自己的研究报告，作者不清晰的地方，别人也很难搞懂。

补充讨论 15-4 投稿前的检查要点

一、题目

1. 是否简明清楚。
2. 是否指出研究的对象。
3. 是否包含研究的重要变量。

二、中文摘要

1. 字数是否合适。
2. 内容是否包含研究目的、对象、工具、重要研究结果。

三、英文摘要

1. 时态是否用过去式。
2. 内容是否和中文摘要相对应。

四、前言 / 文献综述

1. 研究背景描述是否清楚。
2. 问题是否具有研究价值。
3. 重要理论及相关研究是否引用。
4. 研究理由是否充足。
5. 研究目的是否明确。
6. 研究问题是否具体可行。
7. 概念界定是否清楚。
8. 研究假设是否可加以验证。

五、研究方法

1. 研究设计是否清楚易懂。
2. 研究设计是否适当（实验的自变量如何操纵，因变量如何测量，无关变量是如何控制等）。
3. 抽样方法是否适当。
4. 研究对象的特性和样本是否说明清楚。
5. 研究工具是否描述清楚（如仪器型号，量表的信度、效度、分量表名称和题数是否说明清楚）。
6. 数据分析的方法是否适当。

六、结果

1. 统计图表是否清楚。
2. 在推论统计之前是否呈现描述统计资料。
3. 研究结果的说明方式是否中性客观。

七、讨论

1. 研究结果是否和过去的理论及研究比较。
2. 对研究假设支持 / 不支持的结果是否已进行原因分析。
3. 研究限制是否提出说明。

八、结论与建议

1. 研究结果摘述是否扼要。
2. 对于研究结果的应用是否提出说明。

3. 是否对未来的研究提出建议。

九、参考文献

1. 是否列出了所有文内所引用的文献。

2. 是否按序列出。

3. 是否有错误引用。

4. 是否按规定的格式呈现。

十、附录

1. 研究使用的问卷、量表是否都完整呈现。

2. 研究所用的实验材料是否呈现。

十一、撰写方式

1. 文字是否流畅。

2. 层次是否分明。

3. 各级标题是否清楚易懂。

4. 相关内容是否相互呼应（如研究的目的、问题、假设、数据处理等）。

除了学术水平和写作规范，投稿时还应该注意：

◇ 查看刊物以往发表的文章内容，搞清刊物宗旨、对象、学术水平，以及对文稿书写格式的要求等，有的放矢地写稿和投稿。

◇ 初学者最好向中级、初级刊物投稿，或者向面向青年的刊物投稿。

◇ 切勿一稿两投。投稿之后若不采用，可另作处理。

◇ 提供作者的姓名、所属单位、通讯地址、邮政编码、电子信箱、电话等，以便通讯联系。

◇ 也可以附言对稿件作些说明。

投往期刊杂志的研究报告进入评审流程。评审的结果通常是两种，同意刊出或不同意刊出。同意刊出的文章大多需要按照审稿人的建议进行修改。

如收到退稿通知，不要过分看重，几乎所有的初学者都会有此经历；亦不必气馁，应结合审稿意见冷静思考，找出稿件的不足。分析稿件内容、学术水平是否符合、达到投稿刊物的要求，论据是否充分，是否合乎逻辑，技术内容上有什么问题，写作表达是否符合要求，是否有理论和实践的价值等。一般说来，只要作者谦虚冷静，总会找出不足之处。假如遇到研究的价值、方法没有被理解的情况，一方面是要坚持原则；另一方面则要讲究写作方法。

二、会议报告

（一）口头报告

口头报告（oral presentation）是在会议等公共场合使用口头语言在较短的时

间内将研究公布。

口头报告主要通过视听通道传达信息，它除了保持书面报告的主要特征与要求外，还应着重突出简洁和生动的特点。听者每次能记住的信息是有限的，所以报告尽可能地简洁，内容必须围绕主要观点精心组织，不可分散离题；在提高信息传达效率的同时，应避免用单调、沉闷的语气，以维持听众的兴趣；应合理利用幻灯、图表等视觉辅助手段，以利于听众理解、记忆有关内容。

1. 准备

口头报告的准备就像排练一场戏剧，应该针对报告的特点和要求，分析听众、精心构思、反复演练。从报告内容、投影辅助设计、预设问题的回答等方面准备。

报告内容准备。口头报告一般包括 10 分钟左右的陈述和 5 分钟左右的讨论。因此，将正式的研究报告转化成可以口头汇报的形式时，需突出要点，叙述时要简洁有力。任何口头报告都要避免冗长，不要太多客套话，或做过多与主题无关的阐释，务必让听众可以在最短的时间内知道你所要表达的内容。如果听众在五至六分钟内无法将报告内容明晰，表明口头报告还没有构思好。

投影准备。在口头报告中最常用的信息呈现手段是投影。为了使投影增进交流而不是阻碍交流，必须对投影的内容和形式进行规划。

投影的内容应该突出以下特点：

- ◇ 简明：所有的影片条理清晰、表达研究的主要架构。
- ◇ 聚焦：每一张投影片表达特定的主题。
- ◇ 相关：每一张投影片均与整体有关联。
- ◇ 完整：报告结构应包括各个部分的内容。
- ◇ 正确：影片内容、数字、符号及拼写避免错误。

投影片在表达形式上，要注意：

- ◇ 每张投影片外观清晰、内容不可过多，字数控制在 7 ~ 11 行之间。
- ◇ 使用的图表总量不宜超过总量的三分之一，过多的图表会失去生动的初衷，增加听众的负担。

◇ 呈现复杂的图表，应把关键内容用不同的颜色或字体等方式加注醒目标示。

◇ 投影片的数量和放映顺序要根据报告的时间确定。每张投影片的放映需要 1 ~ 2 分钟 如为 20 分钟的报告，则 15 张左右的投影片应足够，如 30 分钟报告 则最多 20 张投影片。

准备可能被询问的问题。口头报告要回答听众提出的问题，因此应针对听众可能提出的有关研究目的、方法、结果等方面的疑问，逐一开列问题的清单，并根据本研究的状况逐一准备，以给予明确具体的答复。

2. 报告

正式报告是一种信息沟通的过程。为提高沟通的有效性，除了报告内容准确无误外，报告人必须在报告进行中随时以目光关怀听众，适当地运用肢体语言，引领听众进入所要表达的情境中。最忌讳不理睬听众的反应，如此不但无法引起听众的兴趣，同时也无法让听众全然相信报告的内容。

对于听众所提出的问题，应该围绕研究本身进行扼要回答。对于无法回答的问题，避免无谓的辩解，可以明确无法回答，也可以提出猜测性回答。

(二) 展报

展报 (poster) 亦称张贴，是将研究用简短生动、专业性的书面形式展出。张贴是一种高效的信息交流方式，受到越来越多会议的重视 (Solso & Maclin, 2002)，课程论文也可以采用这种形式。

1. 展报的基本要素

专业展报可以看成是常规研究报告的浓缩，其内容一般包括以下组成部分：

◇ 摘要：如果有摘要集，本项就可以省略；有的组织者可能明确规定不张贴摘要。

◇ 简介：总括研究的目标和理论基础，将研究的重要性和创新性讲清楚。

◇ 方法：说明研究的设计。

◇ 结果：结果可以用简表、图形、示意图、照片等呈现。

◇ 结论：简要陈述研究的发现。

◇ 建议：对研究发现的应用提出设想。

◇ 参考文献：只包括 2 ~ 3 篇最重要的文献。

◇ 致谢：说明对研究做出支持的基金、机构和个人。

2. 展报设计

展报设计是一种艺术，遵循以下的原则可以使展报更有吸引力：

◇ 简洁：删除不必要的细节，文字在 20 ~ 28 号，读者在 2 ~ 3 米之外就能够看到醒目的标题，清晰安排的图文。这样，读者很快就可以找到重点。

◇ 精干：少就是多，每一部分都是重要的。要留出一定的空白，缺乏空白会使眼睛和大脑劳累。

◇ 强烈的视觉对比：使用不同大小和比例的字体、图片和纸张。避免使用临近的颜色、不同的阴影，许多人对此不敏感，例如绿色和蓝色就容易混淆。

◇ 破除单调：例如可以包含图片以吸引注意力。

◇ 布局合乎逻辑：可以给每一部分编号来表达展报的走向，也可以使用箭头引导。

在使用各种技术突出展报可读性的同时，要注意展报的专业性，避免过分夸张。设计研究展报时，还可以考虑准备一定的分发品。分发品作为展出内容的

补充，可以提供展报不能提供的信息，如详细的数据、图表、参考文献等，供读者深入了解所报告的研究。

展报设计还应该考虑展板的大小，不同的学术会议提供的展板大小不同。一般要考虑最佳的展出空间，然后取舍、安排各个设计要素。图 15.3 是一种简单的纵向空间设计的展报示意图，图 15.4 是一种常见的横向空间的展报设计示意图。

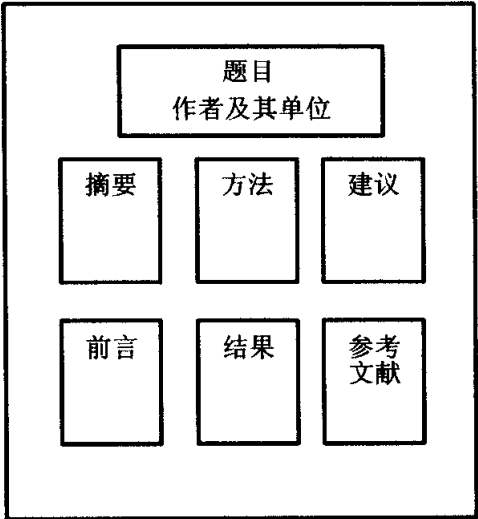


图 15.3 纵向布局的展报设计示意图

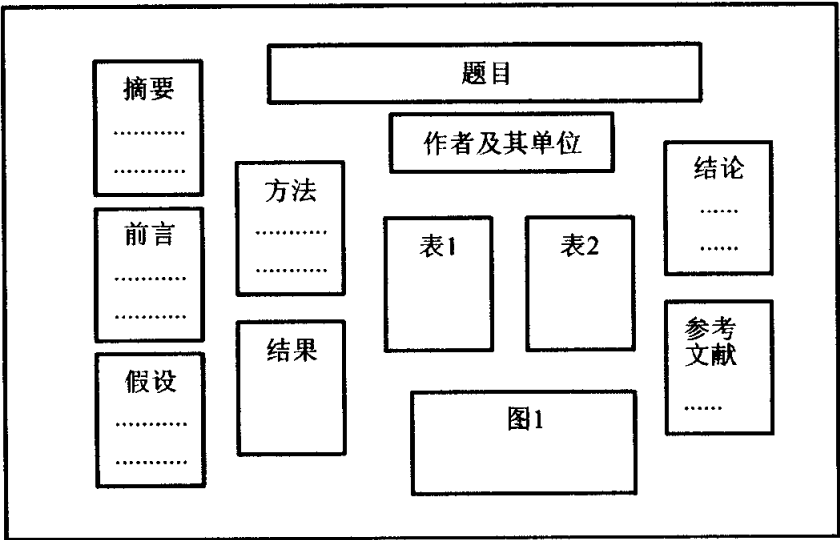


图 15.4 横向布局的展报设计示意图

3. 展讲

展讲是就张贴出的报告与读者交流。展讲是现场交流新思想，观众在参展

的作品中游动，阅读其中的内容，向研究者提问并作出评价。展讲是一种社会性活动，是结识研究兴趣相同的人的机会，展报主人的形象十分重要，也会影响到交流的效果。要衣着得体，从容、自然，随时准备回答来访者的问题。

►本章提要

1. 撰写研究报告是心理学研究最后的一个重要环节，规范及时的研究报告可以充分表现研究的价值。

2. 研究报告是一种应用型文体，按照不同的标准可以分成不同的类型。各种报告的写作都要遵循基本的规范，例如客观、准确和简洁，以有助于交流。

3. 题目是研究报告的名称，表明研究报告的内容指向。实验研究报告的标题通常包括研究的变量及其关系，非实验研究的报告通常会表明研究的主题。摘要是研究报告内容的简单概括。

4. 前言、方法、结果和讨论是组成心理学研究报告的最主要部分。前言是介绍问题的提出，研究的目的和设计，还包括回顾相关的文献，引导读者理解当前报告的研究。方法部分详细介绍研究的被试、操作细节，以备他人重复研究。一般在行文时分成几个小的部分，如被试、工具、程序。结果部分则客观地介绍本研究的发现。讨论是将研究的发现与已经存在的知识体系整合，表明当前研究的意义。

5. 参考文献是学术型报告的重要内容，它是在研究报告中引用的文章或书籍的列表。列出参考文献既是学术规范的要求，也是尊重他人劳动成果的体现。

6. 达到一定学术水平和写作规范的研究报告，就可以投稿发表。除了期刊发表，会议也是交流研究的重要形式。口头报告是在会议等公共场合使用口头语言在较短的时间内将研究公布，展报是将研究用简短生动的书面形式展出。

►本章关键术语

被试(subjects)

参考文献(references)

附录(appendix)

关键词(keyword)

脚注(footnote)

结果(result)

结论(conclusion)

口头报告(oral presentation)

前言(introduction)

讨论(discussion)

题目(title)

尾注(endnote)

文献回顾(literature review)

研究报告(research report)

研究程序(research procedure)

研究对象(participants)

研究方法(research method)

研究工具(research instrument)

研究目的(purpose of a study)

摘要(abstract)

研究问题(research question)

展报(poster)

样本(sample)

执行提要(executive summary)

► 复习与练习

1. 撰写研究报告的目的是什么？
2. 前言应该包括哪些不可缺少的信息？
3. 如何对研究结果进行讨论？
4. 有人认为遵照出版手册的格式撰写报告会妨碍创造力的发挥，你是怎么认为的？
5. 复印一篇新近发表的研究报告，或将自己的研究写出报告，并结合本章进行讨论。
6. 模拟进行一次学术会议，体会如何进行会议交流。

► 推荐参考读物

中国心理学会. 心理学论文写作规范. 北京：科学出版社，2002

► 在线资源

《心理学报》、《心理科学进展》刊物介绍与征稿细则，以及当期文章的目录、摘要和全文等信息可以访问 <http://journal.psych.ac.cn/>，在该网站还可以下载中国心理学会编《心理学论文写作规范》。

美国心理学会设计的研究报告撰写帮助软件 APA-Style Helper，可在 www.apastyle.org/stylehelper/ 下载学习版。

附录：研究报告样例

作者所属单位, 所在省市 邮编

内隐时间表征的实验研究

报告的题目

黄希庭 梁建春

作者姓名

(西南师范大学心理学系 重庆 400715) (重庆大学法学院 重庆 400044)

摘要

用两个实验五种作业，探讨内隐时间表征存在的可能性及其特点。结果表明：在表征动量中，(1)无论诱导图形内隐运动方向向左还是向右，测试图形与识记图形之间位置偏移的错误率在与诱导图形内隐运动方向一致的条件下比不一致的条件下大；(2)诱导图形内隐运动的一致性时间顺序消失之后，表征动量不存在；(3)在 100ms 的时间范围内，识记图形、测试图形之间的保持时距与记忆位置偏移呈线性相关；(4)在水平方向和垂直方向上，客体的概念和背景知识对表征动量无影响。研究结果肯定了表征动量中内隐时间表征的存在并具有方向性、顺序性、连续性和认知不可渗透性的特点。

中文摘要

关键词

内隐时间表征，外显时间表征，认知可渗透性，表征动量

关键词

分类号

B842

中图分类号

1 引言

一级标题

在时间的认知加工研究中，时间表征的研究令人困惑，主要原因为：(1)时间表征是时间认知加工研究的核心问题，涉及的时间现象复杂，较难作整体认识；(2)时间的认知加工研究最初以时间判断、时距估计等的研究为主^[1]，近年来时间认知研究与记忆研究相结合，时间表征的问题被逐步认识和了解^[2]；(3)尚未形成系统、完善的研究途径和实验方法，对该领域的探讨多为间接性的。Michon 提出时间表征可分为外显表征和内隐表征^[3]，外显时间表征是一种有意识的、概念性的时间结构，是对外界环境中时间关系的一种主观的表达方式，内隐时间表征是潜意识的、具有动力性的时间结构，是对外界环境的一种自动的、直接调节的时间模式，与内隐记忆有关^[4]。近年来对时间认知的研究，主要涉及外显的时间表征，内隐时间表征至今尚无人做过实验性探讨^[5]。

参考文献引用标志

该如何用实验来验证内隐时间表征的存在？内隐时间表征具有哪些特点和认知

结构？其认知加工方式是什么？按照 Michon 的设想，外显的时间表征具有认知可渗透性，通过各种时间概念来表达；内隐的时间表征具有认知不可渗透性，通过内隐记忆中的操作表达，不能用外显的命题或概念来表征。认知可渗透性是指概念和背景知识对认知加工的影响；如果概念和背景知识影响对某种作业的加工，则此种加工被认为具有认知可渗透性^[6]。基于此，可以设想，对内隐时间表征的实验研究，至少可以选择两种途径：(1)探索时间记忆表征的认知可渗透性，(2)采用内隐记忆的启动效应范型，进行研究。

本研究以表征动量的实验变式来探索内隐时间表征的性质。表征动量是类似于物理动量的一种现象，根据物理学的原理，由于惯性的作用，物体沿着运动轨迹呈持续运动，即使在阻力作用下也会继续移动一段时间，同样地，客体的心理表征也会在客体停止运动后沿着内隐的轨迹继续移动。表征动量的典型实验是，当某种图形以不同的角度相继呈现，构成一种内隐的旋转运动时，观察者的记忆会受到歪曲，使对角度方向的判断偏向于前者，其中图形的运动和运动轨迹是内隐的，观察者看到的是静止的图象^[7]。之所以选择此种实验范型作为内隐时间表征研究的切入点，因为表征动量与存在于时空世界中的物理动量一样，具有一定时空关系的表征系统的特征，是时间内化的产物^[8]。鉴于此，本研究用两个实验五种作业对表征动量中内隐时间表征存在的可能性及特点作初步探讨。

收稿日期：2001-09-11.

2 实验一

2.1 目的

二级标题

操纵测试图形与诱导图形之间位置偏移方向的一致性、顺序性以及两者之间的时距和偏移程度，以考察表征动量中内隐时间表征是否存在及可能的特点。

2.2 方法

数字

2.2.1 被试 本科生 52 名，男 30 名，女 22 名，被试分成 3 组，各完成表征动量实验中测试图形与诱导图形之间位置偏移方向的一致性（16 名）、顺序性（16 名）和连续性（20 名）三种作业。被试身体健康，视力或矫正视力正常，无类似实验经验。

2.2.2 仪器 586 多媒体兼容机，14 寸标准彩显和扫描器各一台，键盘用纸板覆盖，留 P、Q 两键作为反应键。

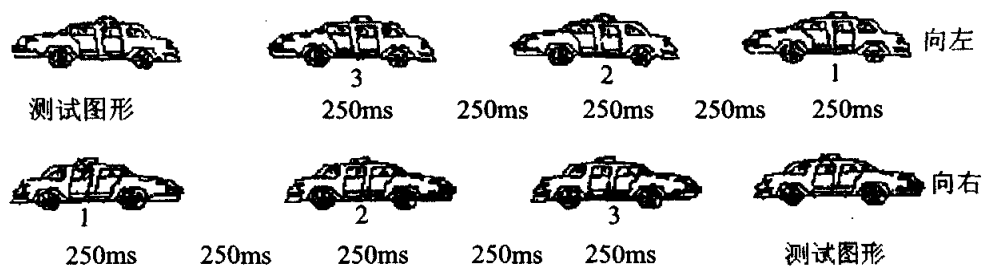
2.2.3 诱导与测试图形 用碳素墨水绘制两张外形相同的黑色小汽车线条图，规格各为 $4.74^{\circ} \times 1.59^{\circ}$ （视角），一张图车头向左，另一张车头向右（见附

乘号

图 1)，用扫描器输入计算机。微机屏幕下半面为黑色，上半面为灰白色，图形在

屏幕上下两半分界线上呈现，从左到右（或从右到左）顺序依次呈现四个相同的小汽车图形，前三个为诱导图形，其中第三个为识记图形，第四个为测试图形。

2.2.4 程序 作业 1：表征动量中客体记忆表征时间维度的方向性实验，采用 2×3 被试内设计，变量分别为诱导图形内隐运动方向（向左、向右）和测试图形与识记图形位置偏移方向的一致性（ $+0.2^\circ$ 、 0.0° 、 -0.2° ）。向右运动时，出现“嘟”的提示音，250ms 后呈现第一个图形，位置在屏幕中央垂线向左 3.6° 视角处，然后向右依次呈现第二个和第三个图形，第一个与第二个，第二个与第三个图形在屏幕上的间距都为 1.8° 视角，每个图形在屏幕上的保持时间及第一个与第二个，第二个与第三个图形信号间的时距为 250ms；向左运动时，图形呈现方式与向右相同，但方向相反（见附图 1）。由于前三个图形在屏幕上出现的时间及位置不同，分别构成向左、向右两个方向内隐的、诱导图形序列。



附图 1 向左和向右方向的小汽车图形呈现方式

无论向左或向右，识记图形的位置正好位于屏幕中央，测试图形出现后保持在屏幕上，直到被试作出反应，之后伴随“嘟”的提示音，又呈现下一轮图形。测试图形出现的位置与识记图形的位置偏移分为 $+0.2^\circ$ 、 0.0° 、 -0.2° 三种，“+”表示测试图形出现的方向与诱导图形内隐运动的方向相同（向前）；“-”表示方向相反（向后）； $+0.2^\circ$ 表示测试图形的位置与识记图形之间向前偏移 0.2° 的视角距离； -0.2° 表示向后偏移 0.2° 的距离； 0.0° 表示两种图形的位置相同。

被试端坐于微机前，面部离屏幕约 60 cm；认为测试图形与识记图形的位置相同，用右手食指按 P 键；认为不同，用左手食指按 Q 键。实验前告诉被试测试图形与识记图形位置相同和不同的次数各占 50%；在不同的条件下，方向相同（向前偏移）和方向不同（向后偏移）的次数各占 50%。每个被试共做 60 次实验。

作业 2：表征动量中客体记忆表征时间维度的顺序性实验，将作业 1 中第一、第二个诱导图形呈现顺序和位置互换，使诱导图形内隐运动的一致性时间顺序消失。前三个诱导图形向右呈现时，第一个图形首先出现在离微机屏幕中央垂线向左 1.8° 视角位置，第二个图形出现在离中央垂线向左 3.6° 视角位置，第

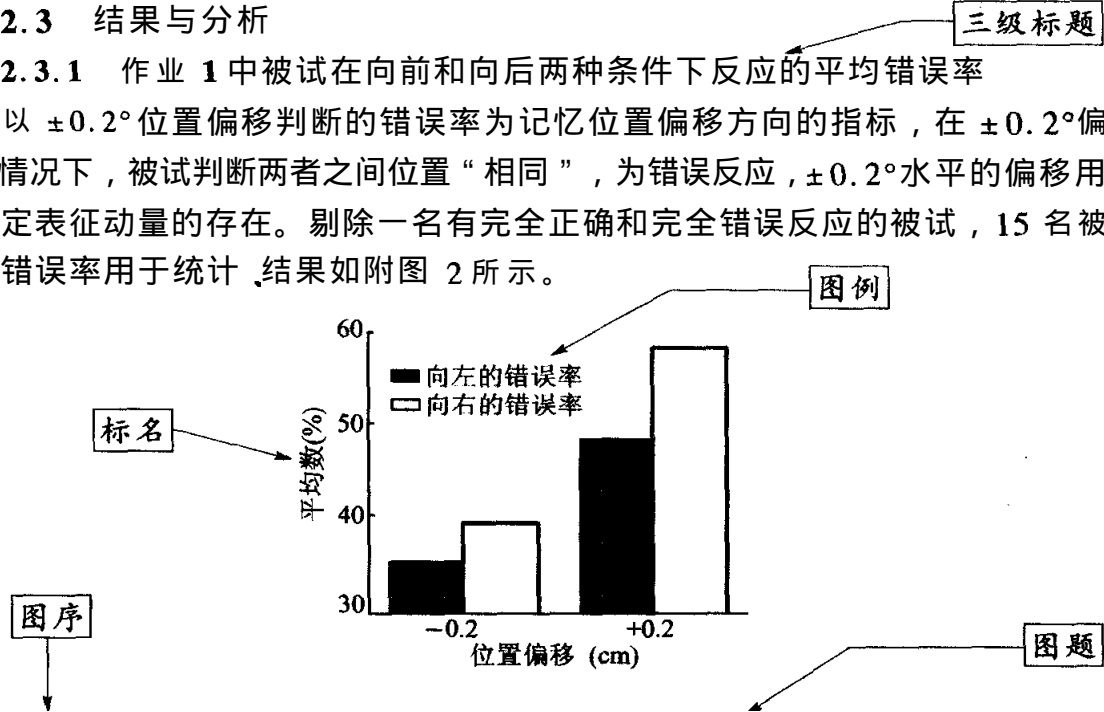
三个图形在屏幕中央；向左运动的情况与此相反。每个被试共做 60 次实验。

作业 3：表征动量中客体记忆表征的时间维度的连续性实验，将作业 1 中识记图形与测试图形之间的时间间隔由 250ms 改为从 16.6ms（微机上可实现的时距单位）到 149.4ms 范围的 9 种时距，分别为：16.6ms、33.2ms、49.8ms、66.4ms、83.0ms、99.6ms、116.2ms、132.8ms、149.4ms；同时将测试图形与识记图形之间的记忆位置偏移改为： -3.1° 、 -2.3° 、 -1.5° 、 -0.8° 、 0.0° 、 $+0.8^{\circ}$ 、 $+1.5^{\circ}$ 、 $+2.3^{\circ}$ 、 $+3.1$ 。九种，诱导图形的呈现时间及各图形间的时距仍为 250ms。其中一半被试完成从左至右的实验，另一半完成从右至左的实验。每个被试共作 324 次，每完成 60 次休息 3 分钟。

2.3 结果与分析

2.3.1 作业 1 中被试在向前和向后两种条件下反应的平均错误率

以 $\pm 0.2^{\circ}$ 位置偏移判断的错误率为记忆位置偏移方向的指标，在 $\pm 0.2^{\circ}$ 偏移的情况下，被试判断两者之间位置“相同”，为错误反应， $\pm 0.2^{\circ}$ 水平的偏移用来确定表征动量的存在。剔除一名有完全正确和完全错误反应的被试，15 名被试的错误率用于统计，结果如附图 2 所示。

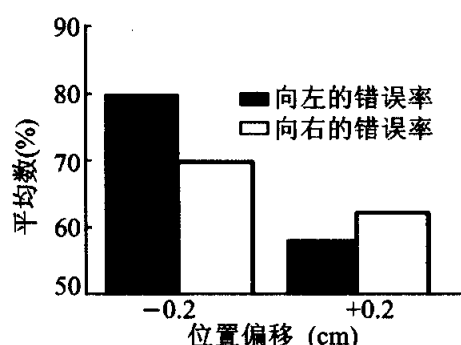


附图 2 作业 1 中向前和向后（ -0.2° 、 $+0.2^{\circ}$ ）条件下被试反应的平均错误率（%）

方差分析表明：客体内隐运动的两种方向（左、右）之间无显著差异（ $p > 0.05$ ），说明被试对向左、向右方向图形判断所依据的标准是一致的，被试的作业成绩未受被试的视觉或行为习惯的影响；测试图形与识记图形位置偏移方向的一致性（向前、向后）之间有显著差异： $F(1,15) = 6.27, p < 0.05$ 其中测试图形与诱导图形内隐运动方向一致（向前）的平均错误率为 53.13 大于内隐运动方向不一致（向后）的平均错误率 37.5 表征动量存在 说明客体的记忆表征具有一致性向前的方向。

2.3.2 作业 2 中被试在向前和向后两种条件下反应的平均错误率

剔除两名成绩有完全正确和完全错误反应的被试，14 名被试错误率成绩用于统计，平均错误率如附图 3 所示。



附图 3 作业 2 中向前和向后 (+0.2°、-0.2°) 条件下被试反应的平均错误率 (%)

方差分析表明：测试图形与识记图形位置偏移的一致性（向前、向后）之间无显著差异： $F(1,14) = 9.55, p > 0.05$ ，表征动量不存在，说明客体内隐运动的时间顺序的一致性改变对表征动量有影响，表征动量与诱导图形依次出现的时间顺序有关。

2.3.3 作业 3 中时间维度连续性的估计

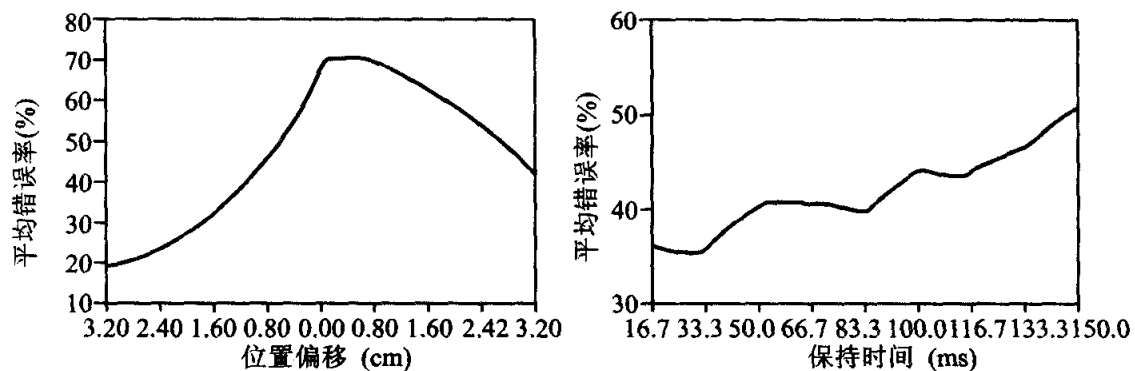
剔除两名有完全正确和完全错误反应的被试，18 名被试的反应成绩用于统计。完成对图形“小汽车”向左和向右内隐运动位置判断的被试各为 9 名。算出被试在每种位置偏移上作“相同”反应次数的百分比（错误率）为记忆位置偏移的指标（0.0°偏移位置上为正确率），附图 4 列出了 $-3.1^\circ \sim +3.1^\circ$ 9 种水平位置偏移作“相同”反应的次数百分比。

根据附图 4 所示的数据作出不同保持时距条件下记忆位置偏移与被试反应错误率之间的一元二次方程；求出每种保持时距条件下一元二次方程曲线的顶点作为每种时距条件下所估计的记忆位置偏移，求出的 9 个顶点所在的记忆位置偏移如附图 5 所示。

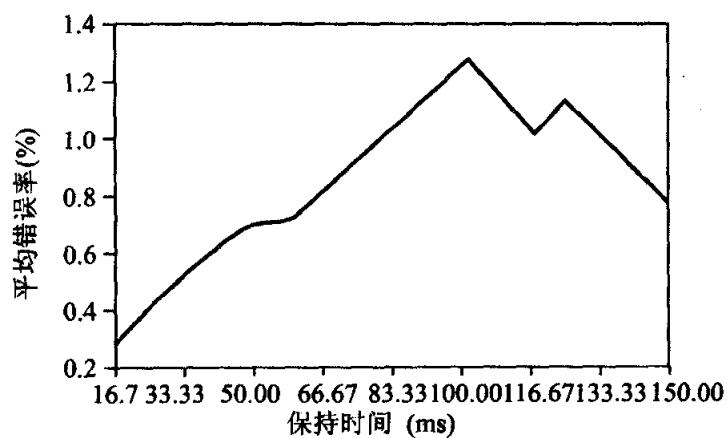
所估计的 9 种保持时距的记忆位置偏移均值为 0.79° ，范围从 0.26° 到 1.24° ，都为正偏移。从附图 5 可看出，所估计的识记图形与测试图形之间记忆位置偏移在 100ms 的保持时间以内，呈上升趋势，100 ms 后开始呈曲线下降趋势。将 100 ms 以内的保持时距与所估计的记忆位置偏移值之间建立一元回归方程如附图 6 所示，两者之间存在明显的线性关系 ($r = 0.959, p < 0.0004$)，说明客体记忆表征的时间维度具有连续性。

图引用标志

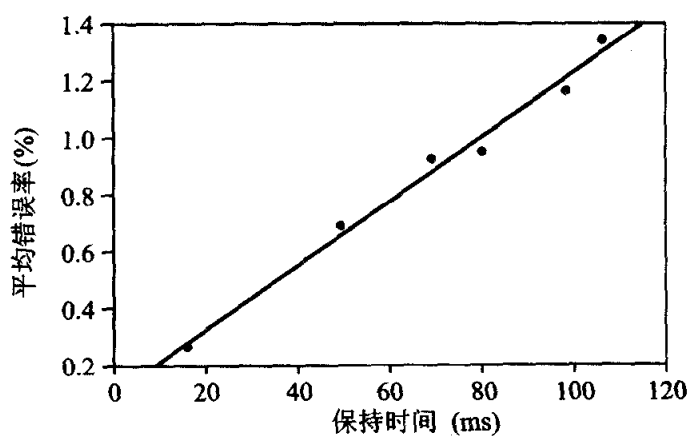
斜体



附图 4 不同记忆位置偏移和保持时间条件下的平均错误率 (%)



附图 5 不同保持时间条件下所估计的记忆位置偏移



附图 6 保持时间与估计的记忆位置偏移的线性回归图

3 实验二

3.1 目的

操纵水平方向和垂直方向上客体的相关概念和背景知识，以考察表征动量中内隐时间表征的认知可渗透性。

3.2 方法

3.2.1 被试 本科生 80 名,男 48 名,女 32 名,被试分成两组,各完成水平方向(32 名)和垂直方向(48 名)作业。被试身体健康,视力或矫正视力正常,无类似实验经验。

3.2.2 仪器 同实验一

3.2.3 诱导与测试图形 在预备实验中要求被试用顺序量表法对计算机屏幕上呈现六种客体的图形“小汽车、狗、牛、载重车、卡车、马”(规格为 $4.74^{\circ} \times 3.18^{\circ}$ 视角)运动速度的快慢作出判断,选出被认为跑得最快和跑得最慢的两种客体“小汽车”和“牛”用于本实验的作业 1;作业 2 为直径 2.74° 视角的黑色圆球状(边缘凹凸不平)图案一个。

3.2.4 设计与程序 作业 1 为 $2 \times 2 \times 5$ 设计:“小汽车”和“牛”两种代表客体运动速度的不同背景知识,客体内隐运动方向为向左、向右,测试图形的位置偏移参考相关文献分为: -0.7° 、 -0.2° 、 0.0° 、 $+0.2^{\circ}$ 、 $+0.7^{\circ}$ 五种^[9]。图形在水平方向上向左、向右呈现,实验操作程序同实验一。被试端坐于微机前,判断测试图形与识记图形的位置是否相同,认为相同,用右手食指按 P 键,认为不同,用左手食指按 Q 键。每个被试完成 120 次实验,每 60 次休息约 5 分钟。

作业 2 为 $2 \times 2 \times 5$ 混合设计,组内变量为诱导图形的内隐运动方向(向上、向下)和测试图形的位移(-0.7° 、 -0.2° 、 0.0° 、 $+0.2^{\circ}$ 、 $+0.7^{\circ}$) ;组间变量为一组被试(21 人)被告知图形是棉花,另一组被试(27 人)被告知图形是铁。图形在中央垂线的垂直方向向上、向下呈现构成内隐运动。向上运动时,出现“嘟”的提示音,250ms 后,第一个诱导图形的位置在离屏幕中央垂线向下 3.6° 视角位置,依次向上呈现第二个和第三个图形,第一个与第二个,第二个与第三个图形之间在屏幕上的间距都为 1.8° ,每个图形在屏幕上的保持时间及第一个与第二个,第二个与第三个图形之间的信号间的时距为 250ms 向下运动时,图形呈现方式与向上相同,但方向相反。每个被试完成 100 次实验,每 50 次休息约 5 分钟。

4 结果与分析

4.1 被试对图形“牛”和“小汽车”在各种条件下判断的平均错误率

以测试图形与识记图形之间位置偏移判断的错误率作统计指标,在两种图形有记忆位置偏移的情况下,被试仍判断两种图形位置“相同”,为错误反应。以

±0.2。的记忆位置偏移用来确定被试表征动量效应的存在，结果如附表 1 所示。

附表 1 被试对图形“牛”和“小汽车”在各种条件下判断的平均错误率(%)							
表序	客体	运动方向	偏移量(度)				
			-0.75	-0.20	0.00	+0.20	+0.75
栏目	小汽车	向左	34.89	69.27	15.62	79.68	20.84
		向右	23.96	67.18	20.83	73.43	30.73
表体	牛	向左	29.16	76.04	14.07	78.65	30.73
		向右	21.36	67.18	22.92	80.20	35.42

方差分析表明：在 ±0.2。水平上，被试对测试图形与识记图形之间位置偏移一致性判断的差异显著： $F(1,31) = 5.82, p < 0.05$ 说明表征动量存在 其他效应均不显著。其中，运动速度不同的两种客体之间； $F(1,31) = 1.20, p > 0.1$ 说明表征动量未受运动速度的背景知识影响；诱导图形的内隐运动方向： $F(1,31) = 2.61, p > 0.1$ ，说明在作业中，被试的左右视觉或行为习惯对表征动量没有影响。±0.7°的位置偏移是实验中特设的控制变量，这一偏移量比用于表征动量标准的 ±0.2°约大 4 倍，目的在于侦察被试在实验过程中是否意识到了实验意图而有意产生表征动量效应，如果被试严格按指导语做实验，在保持正确的前提下尽快反应，那么位置偏移量越大错误率应更小。统计表明：±0.7°位置上的平均错误率比 ±0.2°小(28.93vs73.95)，说明被试在实验过程中没有刻意产生表征动量效应，实验是客观的，表征动量是自动产生的。

4.2 被试对图形“铁”和“棉花”在各种条件下判断的平均错误率

以错误率作统计指标，±0.2°的图形记忆位置偏移用来确定表征动量的存在 ,48 名被试反应的平均错误率如附表 2 所示。

附表 2 被试“铁”和“棉花”在各种条件下判断的平均错误率(%)

客体	运动方向	偏移量(度)				
		-0.75	-0.20	0.00	+0.20	+0.75
铁	向上	31.55	70.24	17.27	75.59	41.07
	向下	41.66	82.74	18.46	76.78	35.13
棉花	向上	37.89	77.26	30.31	62.86	28.04
	向下	45.45	76.51	25.78	66.66	31.82

方差分析表明，在 ±0.2°水平上，测试图形与被试识记图形之间位置偏移的一致性有主效应： $F(1,48) = 8.88, p < 0.01$ ，说明表征动量存在，其他变量均无显著主效应，各变量间无交互作用。比较图形“铁”和图形“棉花”两组被试，在

$\pm 0.2^\circ$ 水平上表征动量无显著差异： $F(1,48) = 1.47, p > 0.1$ 说明铁和棉花的概念知识对表征动量没有影响；在 $\pm 0.7^\circ$ 水平上，平均错误率比 $\pm 0.2^\circ$ 水平低（36.67 vs 73.85），虽然方差分析表明测试图形、识记图形位置偏移的一致性主效应显著，但内隐运动方向不一致条件下的错误率比一致性条件下的错误率更高（39.14 vs 34.02），出现了负效应，说明实验是客观的，被试并未猜测实验意图而有意产生表征动量。

5 总的讨论

本研究借鉴表征动量的实验范型探索时间认知领域的一个新问题：内隐的时间表征。实验一的三种作业探讨的是内隐时间表征存在的可能性及其特点。作业 1 的实验结果显示，无论诱导图形内隐运动方向向左还是向右，测试图形与被试识记图形之间位置偏移的错误率在与诱导图形内隐运动方向一致（向前）的条件下比不一致的（向后）条件下更大，出现了表征动量现象，说明对客体图形位置的记忆受到了诱导图形内隐运动方向的干扰，客体内隐运动变化的记忆表征具有一致性（向前）的方向，但是客体记忆位置向前偏移的现象是否由诱导图形本身的空间或结构属性所造成，而与其出现的时间顺序无关呢？作业 2 在向左和向右的方向上改变诱导图形呈现的位置，造成其内隐运动时间顺序的改变，结果显示表征动量不存在，说明动量效应与诱导图形呈现的时间顺序有关，客体内隐运动变化的记忆表征具有时间顺序性。作业 3 探讨识记图形、测试图形之间的时距与记忆位置偏移之间的关系，发现在 100ms 的时间范围内，保持时距与所估计的记忆位置偏移之间呈线性相关，说明表征动量所反映的客体内隐运动变化的记忆表征中的时间维度具有连续性。

时间总是与一定的客体或事件运动相联系的，三种作业的结果肯定了客体内隐运动变化的记忆表征的时间维度与物理时间一样，具有方向性、顺序性和连续性的特点，客体内隐运动变化的时间记忆表征是客观存在的^[10]。问题是，这不是一种内隐的时间表征呢？在实验中，被试注视的是静止的图像，客体的运动和运动轨迹是由客体图形呈现的方向和位置不同而构成，是一种内隐的运动，且表征动量效应产生于极短的时间范围（16.6 ~ 100 ms）在此时间内 客体记忆表征的位置偏移受个体意识控制的可能性很小，如果被试在作业中有意识地对各种偏移位置进行比较加工，就不可能产生在 16.6 ~ 100 ms 的时间间隔内，客体记忆位置偏移随保持时间的延长而持续增大的结果。因此可以认为动量效应是自动产生的，表征动量中的时间记忆表征是内隐的。实验二从表征动量是否具有认知可渗透性的角度对实验一的结论作进一步验证。作业 1 通过预备实验在 6 种客体中选出运动速度不同的两种客体：“牛”和“小汽车”利用“小汽车”比“牛”跑得快的先占概念，在左右水平方向顺序呈现“牛”和“小汽车”的图形以构成诱导图形的内隐运动，发

现“小汽车”和“牛”两种图形之间表征动量并无显著性差异，说明表征动量未受客体运动速度的背景知识影响，表征动量具有认知不可渗透性。作业 1 采用的是完全不同的两种客体图形，有可能因客体本身的特殊性而得出特殊的结论。作业 2 对此不足而采用相同的呈现图形，变量的差异仅为被试对客体图形的概念知识的不同，在垂直方向上考察表征动量的认知可渗透性，结果发现被视为“铁”组和被视为“棉花”组被试之间表征动量并无显著差异，说明两种客体的概念知识并未对表征动量产生影响，表征动量具有认知不可渗透性。

$\pm 0.7^\circ$ 作为一种控制变量，比 $\pm 0.2^\circ$ 大将近四倍，作业 1 中方差分析其主效应不明显，且错误率均数比 $\pm 0.2^\circ$ 水平小，作业 2 中虽然测试图形与识记图形位置偏移的一致性主效应显著，但测试图形与诱导图形内隐运动方向不一致条件下的错误率比一致性条件下的错误率更高，出现了负效应，说明被试在实验中的反应是客观的，被试并未猜测实验意图而有意产生表征动量。综合实验结果，由于表征动量产生的时间极短，客体的时间记忆表征不受客体概念和背景知识影响，不受被试有意识的控制，是自动产生的，表征动量中客体记忆的时间表征是一种内隐的时间表征，具有方向性、顺序性、连续性和认知不可渗透性等特点。内隐的时间表征问题是一片亟待探索的领域，受实验方法所限，本研究仅探讨了表征动量中的内隐时间表征，内隐时间表征还具有哪些形式，其认知加工方式如何，这些问题将有待今后进一步研究。

参考文献

作者姓名

文章题目

句号

- [1] Huang XT. A cognitive approach to the temporal duration information processing (in Chinese). Journal of Southwest China Normal University (Natural Science), 1993, 18(2): 207 ~ 215

刊名, 年, 卷(期): 页码

(黄希庭. 时距信息加工的认知研究西南师范大学学报(自然科学版), 1993, 18(2): 207 ~ 215)

- [2] Huang XT, Zheng Y. Theories and experimental moderns of the memory of time (in Chinese) Psychological Science, 1995, 18(4): 201 ~ 205

(黄希庭, 郑涌. 时间记忆的理论及实验模型. 心理科学, 1995, 18(4): 201 ~ 205)

- [3] Block RA. Cognitive Models of Psychological Time. Hillsdale, NJ; Hove and London, 1990. 37 ~ 58

- [4] Macar F, Poulhas V, Friedman WJ. Time, Action and Cognition: Towards Bridging the Gap. Dordrecht: Kluwer Academic, 1992. 307 ~ 321

出版地: 出版社

- [5] Liang JC, Huang XT. A Cognitive study of time representation (in Chinese). Psychological Science, 1999, 22(1): 10 ~ 13

(梁建春, 黄希庭. 几种时间表征的研究概况. 心理科学, 1999, 22(1): 10 ~ 13)

- [6] Pylyshyn ZW. The imagery debate: analogue media versus tacit knowledge. Psychological Re-

view, 1988, (1):16 ~ 45

逗号

- [7] Freyd JJ, Finke RA. Representation momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1984, 10(1):126 ~ 132
- [8] Freyd JJ, Johnson JQ. Probing the time course of representation momentum. *Journal of Experimental Psychology: Learning, Memory and Cognition*, 1987, 13(2):259 ~ 264
- [9] Reed CL, Vinson NG. Conceptual effects on representation momentum. *Journal of experimental Psychology: Human perception and performance*, 1996, 224:839 ~ 850
- [10] Finke RA, Freyd JJ. Mental extrapolation and cognitive penetrability: reply to ranney and proposals for evaluative criteria. *Journal of experimental Psychology: General*, 1989, 118(4):403~408

序号

英文题名

EXPERIMENTAL STUDY ON IMPLICIT REPRESENTATION OF TIME

作者英文姓名

作者单位的英文名称

Huang Xiting (*Department of Psychology, Southwest Normal University, Chongqing 400715*)

Liang Jianchun (*Trade and Law College, Chongqing Normal University, Chongqing 400044*)

Abstract Using two experiments, the characteristics of implicit representation of time on representation momentum were investigated. In operation 1 of the experiment one, the orientation of implicit time was examined; In operation 2, the succession of implicit time by changing the order of the motoring figure was examined; In operation 3, the continuity of implicit time was examined by dividing the temporal duration and the departure between the test figure and the memorial figure into 9 levels. In experiment two, the cognitive penetrability of the representation of time in representation momentum was investigated. The results showed that the implicit representation of time has characteristics of orientation, succession, continuity, and in penetrability. It also implied that the representation momentum is one kind if implicit representation of time.

英文摘要

英文关键词

Key words implicit representation of time, explicit representation of time, cognitive penetrability, representation momentum

(采自黄希庭、梁建春, 2002)

参考文献

(汉语文献以拼音字母为序,英文文献以英文字母为序)

埃森克.心理学——一条整合的途径.阎巩固等译.上海市:华东师范大学出版社,2000

埃维森.统计学.吴喜之等译.北京市:高等教育出版社,2000

艾尔·巴比.社会研究方法.邱泽奇译.北京市:华夏出版社,2000

贝利现代社会研究方法.许真译.上海市:上海人民出版社,1986

伯恩斯坦、伯恩斯坦.统计学原理(上)——描述性统计学与概率.史道济译.北京市:科学出版社,2002

伯恩斯坦、伯恩斯坦.统计学原理(下)——推断性统计学.史道济译.北京市:科学出版社,2002

蔡程辉,庄永龙.世界优秀统计软件 SPSSv10.0 for Windows 实用基础教程.北京市:希望电子出版社,2001

蔡永红.教师职务绩效——结构及其影响因素研究.博士学位论文.北京:北京师范大学发展心理研究所,2002

陈平雁、黄浙明.SPSS10.0 统计软件应用教程.北京市:人民军医出版社,2002

陈世平.基础心理统计与 SPSS 应用.兰州市:兰州大学出版社,2003

陈向明.质的研究方法与生活科学研究.北京:教育科学出版社,2000

陈仲庚、张雨新.人格心理学.沈阳市:辽宁人民出版社,1986

丁锦红,林仲贤.不同复述条件下命名难度对图形记忆的影响.心理学报,2002,34(2):135~140.

董奇.心理与教育研究方法.北京市:北京师范大学出版社,2004

董奇、申继亮.心理与教育研究法.杭州市:浙江教育出版社,2004

杜建政,杨治良.随机棋局存在专家记忆优势效应吗?心理学报,2002,34(3):254~258

风笑天.社会学研究方法.北京市:中国人民大学出版社,2001

冯仕政.应对非典疫情下的心理恐慌.人民日报,2003-5-30(14)

弗里德曼等.统计学(第二版).魏宗舒等译.北京市:中国统计出版社,1997

弗林克尔、瓦伦.教育研究的设计与评估.蔡永红等译.北京市:华夏出版社,

2004

车文博. 弗洛伊德. 列奥纳多·达·芬奇和他对童年的一个记忆. 弗洛伊德文集第四卷. 长春市: 长春出版社, 1998

符德江, 沈模卫. 航空模拟舱飞行行为的 EPIC 理论仿真. 心理学报, 2001, 33(3): 225 ~ 230

格里格、津巴多. 心理学与生活. 王垒, 王甦等译. 北京市: 人民邮电出版社, 2003

郭秀艳. 实验心理学. 北京市: 人民教育出版社, 2004

郭秀艳, 肖辉彬, 周楚. 动态和静态条件下内隐时间的存在和特征. 心理学报, 2004, 36(3): 322 ~ 326

郭志刚. 社会统计分析方法——SPSS 软件应用. 北京市: 中国人民大学出版社, 1999

哈克. 改变心理学的四十项研究. 白学军等译. 北京市: 中国轻工业出版社, 2004

亨特. 心理学家的故事. 李斯译. 海口市: 海南出版社, 1999

洪小良. 社会调查研究. 原理与方法. 北京市: 华文出版社, 1998

黄光成. 单亲家庭学生心理亟待关注. 光明日报, 2002 - 5 - 16

黄希庭. 心理学实验指导. 北京市: 人民教育出版社, 1988

黄希庭. 心理学导论. 北京市: 人民教育出版社, 1991

黄希庭. 人格心理学. 杭州市: 浙江教育出版社, 2000

黄希庭. 简明心理学辞典. 合肥市: 安徽人民出版社, 2004

黄希庭, 梁建春. 内隐时间表征的实验研究. 心理学报, 2002, 34(3): 235 ~ 241

黄希庭, 张志杰. 青少年时间管理倾向量表的编制. 心理学报, 2001, 33(4): 338 ~ 343

黄希庭、凤四海、王卫红. 青少年学生自我价值感全国常模的制定. 心理科学, 2003, 26(2): 194 ~ 198

黄希庭、杨雄. 青少年学生自我价值感量表的编制. 心理科学, 1998, 21(4): 289 ~ 292

黄希庭、余华. 青少年自我价值感量表构念效度的验证性因素分析. 心理学报, 2002, 34(5): 511 ~ 516

黄希庭、余华等. 中学生应对方式的初步研究. 心理科学, 2000, 23(1): 1 ~ 5

江新, 荆其诚. 句法和语义在汉语简单句理解中的作用. 心理学报, 1999, 31(4): 361 ~ 368

金瑜. 心理测量. 上海市: 华东师范大学出版社, 2001

- 坎特威茨、罗迪格、埃尔姆斯. 实验心理学—掌握心理学的研究. 郭秀艳等译. 上海市: 华东师范大学出版社, 2001
- 柯惠新. 调查研究中的统计分析法. 北京市: 北京广播学院出版社, 1992
- 李伯约. 时间顺序标码的层次网络表征研究. 博士学位论文. 重庆: 西南师范大学心理学系, 2001
- 李德明, 刘昌, 陈天勇等. 加工速度和工作记忆在认知年老化过程中的作用. 心理学报, 2003, 35(4): 471 ~ 475
- 李则宣、李凌江. 创伤后应激障碍的生物学机制研究. 中华精神科杂志, 2003, 36(4): 254 ~ 256.
- 林崇德. 发展心理学. 北京市: 人民教育出版社, 1995
- 林崇德. 儿童心理过程的发展. 见王越, 林仲贤, 荆其诚主编中国心理科学. 长春: 吉林教育出版社, 1997
- 林崇德, 沃建中, 於国荣. 儿童和青少年信息加工速度发展函数的研究. 心理学报, 1997, 29(1): 43 ~ 50
- 林崇德、杨治良、黄希庭. 心理学大辞典. 上海市: 上海教育出版社, 2003
- 林汉生, 杨丽, 夏苏建等. 多选题答案的编码与 SPSS 中 SUBSTR 函数的使用. 中国卫生统计, 2001, 18(6): 371 ~ 373
- 林杰斌、刘明德. SPSS10.0 与统计模式建构. 北京市: 科学出版社, 2002
- 林南. 社会研究方法. 北京市: 农村读物出版社, 1987
- 刘希平, 唐卫海. 回忆准备就绪程度的判断发展. 心理学报, 2002, 34(1): 56 ~ 60
- 卢纹岱主编. SPSS for Windows 统计分析 (第 2 版). 北京市: 电子工业出版社, 2002
- 鲁尼恩. Fundamentals of Behavioral Statistics (9th ed) (心理统计影印版). 北京市: 人民邮电出版社, 2004
- 马斯洛. 自我实现的人. 许金声等译. 北京市: 三联书店, 1987
- 马斯洛. 动机与人格. 许金声等译. 北京市: 华夏出版社, 1987
- 蒙哥马利. 实验设计与分析. 汪仁官等译. 北京市: 中国统计出版社, 1996
- 孟庆茂, 黄良东. 对价值观测验计分方式的研究. 海峡两岸学术研讨会论文集, 杭州: 浙江教育出版社, 1997
- 孟庆茂、常建华. 实验心理学. 北京市: 北京师范大学出版社, 1999
- 孟祥芝, 周晓林, 吴佳音. 发展性协调障碍与书写困难个案研究. 心理学报, 2003, 35(5): 604 ~ 609
- 潘军、李焰. 美国贝克斯特测慌系统在我国法庭科学中的应用. 心理学报, 2001, 33(3): 276 ~ 283

- 彭聃龄, 邓园, 陈宝国. 汉语多义单字词的识别优势效应. 心理学报, 2003, 35 (5): 569 ~ 575
- 彭凯平. 心理测验. 理论与实践. 北京市: 华夏出版社, 1989
- 珀文. 人格科学. 周榕, 陈红, 杨炳钧等译. 上海: 华东师范大学出版社, 2001
- 钱铭怡、武国城、朱荣春、张莘. 艾森克人格问卷简式量表中国版 (EPQ - RSC) 的修订. 心理学报, 2000, 30 (3): 317 ~ 323
- 萨尔斯伯格. 统计, 改变了世界 (繁体版). 叶伟文译. 台北市: 天下文化出版社, 2001
- 邵枫 林文娟. 情绪应激体液免疫调节作用的影响因素. 心理学报, 2001, 33 (6): 543 ~ 547
- 舒华主编. 心理与教育研究中的多因素实验设计. 北京市: 北京师范大学出版社, 1994
- 斯皮格尔等. 统计学 (第三版). 杨纪龙等译. 北京: 科学出版社, 2002
- 汤旦林、柯惠新. 使人聪明的技术——生活中的统计观念与方法. 北京: 人民交通出版社, 1996
- 王重鸣. 心理学研究方法. 北京市: 人民教育出版社, 1990
- 王宏印. 跨文化心理学导论. 西安市: 陕西师范大学出版社, 1993
- 王娟, 王玉凤, 任园春. 注意缺陷多动障碍儿童平衡功能测定的病例对照研究. 北京大学学报 (医学版), 2003
- 王甦、汪安圣. 认知心理学. 北京市: 北京大学出版社, 1992
- 吴明隆. SPSS 统计应用实务——问卷分析与应用统计. 北京市: 科学出版社, 2003
- 吴兴曲, 杨来启, 李栓德等. 睡眠剥夺后大鼠海马生化及病理变化. 中国心理卫生杂志, 2003, 17 (2): 77 ~ 79
- 夏普. 社会科学统计学. 王崇德译. 北京市: 科学技术文献出版社, 1990
- 肖少北. 汉字学习与儿童分析能力和综合能力的关系. 心理学探新, 2002, 22 (3): 33 ~ 36
- 心理学名词审定委员会. 心理学名词. 北京: 科学出版社, 2001
- 邢红兵, 舒华. 语料库建设与汉语认知研究. 北京: 第三届华人心理学家学术研讨会, 1999
- 宣宾, 张达人. 输出方式和输出时间对数字广度的影响. 心理学报, 2004, 36 (3): 315 ~ 321
- 严辰松. 定量型社会科学研究方法. 西安市: 西安交通大学出版社, 2000
- 杨国枢、文崇一、吴聪贤等. 社会及行为科学研究法 (繁体版). 台北市: 东华书局, 1978

- 杨琿, 彭聃龄, Perfetti C A 等. 汉字阅读中的语音通达与表征 (I)——字水平与亚字水平的语音及其交互作用. 心理学报, 2000, 32(2): 144 ~ 151
- 杨治良. 实验心理学. 杭州市: 浙江教育出版社, 1998
- 叶浩生. 多元文化论与跨文化心理学的发展. 心理科学进展, 2004, 12(1): 144 ~ 151
- 由智勇、王洁贞. 音乐疗法辅助治疗慢性精神分裂症的 Meta 分析. 中国医学科学院学报, 2002, 24(6): 564 ~ 567
- 于海琴. 小学生的亲子依恋与其同伴交往的关系研究. 硕士学位论文, 武汉: 华中师范大学, 2000
- 袁方主编. 社会研究方法教程. 北京: 北京大学出版社, 1997
- 袁卫. 探索性数据分析. 数理统计与管理, 1999, 6 ~ 53
- 张爱卿、刘华山. 责任、情感及帮助行为的归因结构模型. 心理学报, 2003, 35(4): 535 ~ 540
- 张锋、戴海崎、陈雪枫. 心理与教育测量. 广州: 暨南大学出版社, 2002
- 张厚粲. 心理与教育统计学. 北京市: 北京师范大学出版社, 1993
- 张厚粲. 心理与教育统计学. 北京市: 北京师范大学出版社, 1998
- 张厚粲, 徐建平. 现代心理与教育统计学. 北京: 北京师范大学出版社, 2003
- 张力为. 体育科学研究方法. 北京市: 高等教育出版社, 2002
- 张敏强. 教育与心理统计学. 北京市: 人民教育出版社, 2002
- 张文彤. 世界优秀统计工具 SPSS 11 统计分析教程 (高级篇). 北京市: 希望电子出版社, 2002
- 张文彤. 世界优秀统计工具 SPSS 11 统计分析教程 (基础篇). 北京市: 希望电子出版社, 2002
- 张亚旭, 刘友谊, 舒华等. 中文句子中双音节兼类词句法分析历程初探. 心理学报, 2003, 35(4): 433 ~ 440
- 张增杰、黄希庭. 六至七岁儿童时间知觉的初步研究. 心理学报, 1963, 7(3)
- 张志学, 韩玉兰. 回报谨慎对谈判过程和谈判结果的影响. 心理学报, 2004, 36(3): 370 ~ 377
- 芝祐顺. 因素分析法. 曹亦薇译. 北京市: 人民教育出版社, 1999
- 中国心理卫生协会. 心理评估质量控制规定及从业人员道德准则. 中国临床心理学杂志, 2001, 9(1): 79 ~ 80
- 中国心理学会. 心理测验工作者的道德准则. 心理科学, 1994, 17(2): 2
- 中国心理学会. 心理学论文写作规范. 北京: 科学出版社, 2002
- 中国心理学会. 中国心理学会 2003 年工作总结. www.cpsbeijing.org/2003zj.doc, 2004

周镐 鲍碧君 古茂盛. 小学儿童推理特点的调查研究. 见刘范主编. 心理发展的近期研究北京: 北京师范学院出版社, 1990

周谦. 心理科学方法学. 北京市: 中国科学技术出版社, 1994

周文钦. 研究方法—实证性研究取向. 台北市: 心理出版社, 2002

朱滢、张力. 自我记忆效应的实验研究. 中国科学 (C 辑), 2001, 31(6): 537 ~ 543

朱滢. 实验心理学. 北京: 北京大学出版社, 2000

邹枝玲 施建农 辉梅等. 7 岁超常和常态儿童的信息加工速度. 心理学报, 2003, 35(4): 527 ~ 534

Allport, G. W., Vernon, P. E., & Lindzey, G. Study of values (3rd ed.). Boston: Houghton Mifflin, 1960 American Psychological Association. Ethical standards of psychologists. Washington, DC. Author, 1953

American Psychological Association. Ethical principles of psychologists and code of conduct. American Psychologist, 47, 1992, 1597 ~ 1611

Anastasi, A. Psychology Testing (7th ed). New York: Macmillan, 1997

Aron, A., & Aron, E. N. Statistical methods in psychology. New Jersey: Prentice - Hall, 1994

Asch, S. E. Opinions and social pressure. Scientific American, 1955, 193, 310 ~ 35

Atkinson, R. C., & Shiffrin, R. M. Human memory. A proposed system and its control process. In W. K. Spence, & J. T. Spence (Eds.), The psychology of learning and motivation. Advance in research and theory, Vol. 1. New York. Academic Press, 1968

Bandura, A., Ross, D., & Ross, S. A. Transmission of aggression through imitation of aggressive models. Journal of Abnormal and Social Psychology, 1961, 63, 575 ~ 582.

Bandura, A., Ross, D., & Ross, S. Imitation of film mediated aggressive models. Journal of Abnormal and Social Psychology, 1963, 66, 3 ~ 11.

Banister P., Burman, E., Parker I. et al, Qualitative Methods in Psychology: A Research Guild. Philadelphia: Open University Press. 1994, 3, 109 ~ 119

Becker, H. S. Writing for social scientists: How to start and finish your thesis, book or article. Chicago: University of Chicago Press, 1986, 147 ~ 148

Berenbaum, S. A., & Hines, M. Early androgens are related to childhood sex-typed toy preference. Psychological Science, 1992, 3, 203 ~ 203.

Berry, J. W., Poortinga Y. H., Segall, M. H., & Dasen, P. R. Cross-cultural

Psychology: Research and Applications. Cambridge University Press ,2002

Boring, E. G. A history of experimental psychology. New York : Appleton -- Century -- Crofts, 1950

Bouchard, T. , Lykken, D. , McGue, M. , Segal, N. , & Tellegen, A. Sources of human psychological differences: The Minnesota study of twins reared apart. Science, 1990, 250, 223 - 229.

Bryant & Yarnold. Principal components analysis and exploratory and confirmatory factor analysis. In: Grimm & Yarnold (Eds.) , Reading and understanding multivariate analysis. American Psychological Association Books, 1995

Camines, E. G. & Zeller, R. A. Reliability and Validity Assessment. Beverly Hills, CA : Sage, 1979

Campbell, D. T. & Fiske, D. Convergent and discriminate validation by the multi-trait-multimethod matrix. Psychological Bulletin, 1959, 56, 81 ~ 105

Carducci, B. J. , & Zimbardo, P. G. Are you shy? Psychology Today, 1995, 28, 34 ~ 40

Cherry, E. C. Some experiments on the recognition of speech with one and with two ears. Journal of the Acoustical Society of American, 1953, 25, 975 ~ 979

Cheung, S. F. , & Chan, D. K. S. Dependent Effect Sizes in Meta-Analysis: Incorporating the Degree of Interdependence. Journal of Applied Psychology, 2004, 89 (5). 780 ~ 791

Cohen, J. Statistical power analysis for the behavioral sciences (Rev. ed.). San Diego, CA. Academic Press, 1977

Cook, S. R. A Note on Testing for Homogeneity among Effect Sizes Sharing a Common Control Group. Psychological Methods, 2004, 9 (4). 446 ~ 452

Creswell, J. W. Research design: Qualitative and quantitative approaches (2nd ed). Thousands Oaks: Sage. 1994, 1 ~ 2

Denzin, N. K. The research act (3rd Ed), Englewood Cliffs, NJ: Prentice Hall. 1989, 237 ~ 241

DerSimonian, R. & Laird, N. Meta-analysis in clinical trials. Controlled Clin Trials, 1986, 7, 177 ~ 188.

Ellenberg, S. S. Meta-analysis. the quantitative approach to research review. Seminars Oncol, 1988, 15 (5) , 427

Feins, T. A. R. Meta-analysis. Statistical alchemy for the 21st century. J Clin Epidemiol. 1995, 48, 71 ~ 79

Field, A. P. Meta-analysis of correlation coefficients. A Monte Carlo comparison

of fixed and random-effects methods. *Psychological Methods*, 2001, 6(2), 161 ~ 180.

Freud, S. Analysis of a phobia in five-year-old boy. In J. Strachey (Ed.), Standard edition of the complete psychological works of Sigmund Freud, Vol. 10. London. Hogarth Press, 1909/1955

Freud, S. Leonardo da Vinci. A study in psychosexuality. Translated by A. A. Brill. New York. Random House, 1947

Gazzaniga, M. S. The split brain in man. *Scientific American*, 1967, 217, 24 ~ 29

George, E. P. B., William, G. H., & William, J. S. H. Statistics for experimenters: An introduction to design, data analysis, and model building. New York: John Wiley & Sons, 1978

Giles, D. Advanced research methods in psychology. NY: Routledge, 2002, 318 ~ 327

Giles, D. Advanced research methods in psychology. NY: Routledge, 2002

Glass, G. V. Primary, secondary, and meta-analysis of research. *Education Research*, 1976, 6(5), 3 ~ 8.

Glass, G. V. et al. Meta-analysis in social research. Beverly Hills. CA. Sage Publications, 1981

Goodwin, C. J. Research in psychology: Method and design (3rd ed). New York: John Wiley & Sons, 2002

Grohol J M. Internet addiction guide. <http://psychcentral.com/netaddiction/>, 1999

Guadagnoli, E., & Velicer, W. F. Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 1988, 2, 265 ~ 275

Gustafson, C. R., Garcia, J., Hawkins, W., & Rusiniak, K. Coyote predation control by aversive conditioning. *Science*, 1974, 184, 581 ~ 583

Hall, S. M., & Brannick, M. T. Comparison of two random-effects methods of meta-analysis. *Journal of Applied Psychology*. 2002, 87(2), 377 ~ 389

Harman, H., & Jones, W. H. Factor analysis by minimizing residuals. *Psychometrika*, 1966, 31, 351 ~ 368

Hedges, L. V & Olkin, I. Statistical methods for meta-analysis. London. Academic Press, 1985

Hedges, L. V. Distribution theory for Glass's estimator of effect size and related estimators. *Educational Statistics*, 1981, 6, 107 ~ 128

Heiman, G. W. Understanding research methods and statistics: An integrated introduction for psychology (2nd ed). Boston: Houghton Mifflin Company, 1998

Heiman G. W. Understanding Research Methods and Statistics. NY : Houghton Mifflin Company. 1998,627

Hoaglin, D. C. , Mosteller, F. , & Tukey, J. W. 探索性数据分析 . 陈忠琰、郭德媛译 . 北京市中国统计出版社 , 1998

Hudelson P. M. Qualitative research for health programmers. Geneva: Division of mental health WHO. 2 1994

Johnson, R. A. , & Wichern, D. W. 实用多元统计分析 . 陆璇译 . 北京市 . 清华大学出版社 ,1998

Kagan, J. Galen ' s prophesy : Temperrament in human nature. New York: Basic Books 1994

Kaiser, H. F. The application of electronic computers to factor analysis. Educational and Psychological Measurement 1960,20 141 ~ 151

Kaiser, H. & Caffrey J. Alpha factor analysis. Psychometrika 1965 30 1 ~ 14

Kangas, J. &Bradway K. Intelligence at middle age: A thirty-eight year follow-up. Developmental Psychology 1971,5,333 ~ 337

Kim J. O. & Mueller, C. W. Factor analysis: Statistical methods and practical issues. Beverly Hills CA: Sage Publications 1978

Kosslyn S. M. Pascual - Leone A. Felician O. Camposano S. Keenan J P. Thompson W. L. Ganis G. Sukel K. E. &Alpert N. M. The role of Area 17 In visual imagery :convergent evidence from PET and rTMS. Science 1999 284 167 ~ 170

Krathwohl D. R. Methods of educational and social science research: An integrated approach(2nd ed). NY : Longman. 1998,740

Kretschmer, E. Körperbau und Charakter [Physique and Character]. Berlin: Springer - Verlag 1921

Kuhn, T. S. The structure of scientific revolutions. Chicago. University of Chicago Press 1962

Lattin J. M. Carroll J. D. & Green P. E. Analyzing Multivariate Data (多元数据分析影印版) . 北京市机械工业出版社 , 2003

Lazarsfeld P. F. Problems in methodology. In R. K. Merton(Ed.). Sociology to-day. New York. Basic Books 1959

Loehlin J. C. Genes and environment in personality development. Newbury Park CA. Sage 1992

Maslow A. H. Motivation and personality. New York : Haper&Row.

Maxwell, J. A. Qualitative Research Design: An Interactive Approach, London: Sage. 1996, 5

Merriam, S. B. Qualitative research in practice: examples for discussion and analysis. San Francisco: Jossey-Bass. 2002, 12 ~ 13

Mezulis, A. H. , Abramson, L. Y. , Hyde, J. S. etc. Is There a Universal Positivity Bias in Attributions? A Meta – Analytic Review of Individual, Developmental, and Cultural Differences in the Self – Serving Attributional Bias. Psychological Bulletin, 2004, 130(5) , 711 ~ 747

Milgram, S. Behavioral study of obedience. Journal of Abnormal and Social Psychology, 1963, 67, 371 ~ 378

Milgram, S. Behavioral study of obedience. Journal of Abnormal and Social Psychology, 1963, 67, 371 ~ 378

Milos, J. Meta-analysis in medicine where we are and where we want to go. Journal of Clinical Epidemiology, 1989, 42(1) , 35 ~ 44

Mowbray, G. H. Simultaneous vision and audition. The comprehension of prose passages with varying levels of difficulty. Journal of Experimental Psychology, 1953, 46, 365 ~ 372

Moyer, C. A. , Rounds, J. , & Hannum, J. W. A Meta-Analysis of Massage Therapy Research. Psychological Bulletin. 2004, 130(1) , 3 ~ 18

Murphy, K. R. , & Davidshofer, C. O. Psychological testing: Principles and applications (3rd ed). New Jersey: Prentice – Hall, 1994

Myers, A. & Hansen, C. H. Experimental psychology (5th ed). New York: John Wiley & Sons, 2003

Myers, A. & Hansen, C. H. Experimental psychology (Fifth Ed). NY: Wadsworth. 2002, 462 ~ 464

Myers, A. , & Hansen, C. H. Experimental Psychology (5th ed.). Pacific Grove. Wadsworth Group, 2002

Pavlov, I. P. Conditional reflexes. London: Oxford University Press, 1927

Pearson, K. Report on certain fever inoculation statistics. British Medicine Journal, 3, 1904, 1243 ~ 1246

Petitti, D. B. Meta analysis, decision analysis, and cost-effectiveness analysis. methods for quantitative synthesis in medicine. New York. Oxford University Press, 1994

Ponterotto, J. G. & Grieger, I. Merging Qualitative and Quantitative Perspectives in a Research Identity. In Kopala, M & Suzuki, L. A. (eds) Using qualitative

methods in psychology. Thousand Oaks: Sage. 1999,55

Ratner, C. Cultural psychology and qualitative methodology: Theoretical and Empirical Considerations, NY: Plenum Press. 1997,3

Rorschach, H. Psychodiagnostics. A diagnostic test based on perception. New York. Grune & Stratton, 1942

Rosenhan, D L. On being sane in insane places. Science, 1973, 179, 250 ~ 258

Rosenzweig, M. R. , Bennett, E. L. , & Diamond, M. C. Brain changes in response to experience. Science American, 1972, 226, 22 ~ 29

Rosenzweig, M. R. , Bennett, E. L. , & Diamond, M. C. Brain changes in response to experience. Scientific American, 1972, 226, 22 ~ 29

Runyan, W. M. Progress in psychobiography. Journal of Personality, 1988, 56, 295 ~ 326

Sacks, H. S. , Berrier, J. , Reitan, D. et al. Meta-analysis of randomized controlled trials. The New England J Med, 1987, 316(8) , 450 ~ 455.

Sacks, J. J. , Krushat, W. M. , & Newman, J. Reliability of the health hazard appraisal. American Journal of Public Health. 1980, 7, 730 ~ 732

Salkind, N. J. Exploring Research (4th ed). New Jersey: Prentice Hall, Upper Saddle River, 2000

Salkind, N. J. Exploring Research (4th ed.). New Jersey. Prentice Hall, Inc, 2000

Selltiz, C. , Wrightsman, L. & Cook, S. Research Methods in Social Relations. New York: Holt, Rinehart & Winston, 1976

Seymour, S. Applied Sampling. Handbook of survey Research, edited by P. H. Rossi, J. D. Wright, & A. B. Anderson. New York: Academic Press, 1983

Solso, R. L. & Maclin, M. K. Experimental Psychology: A Case Approach (7th Ed). Boston: Allyn and Bacon. 2002, 152, 177

Sperling, G. The information available in brief visual presentations. Psychological Monographs, 1960, 74, 1 ~ 29

Spitzer, W. O. Meta-analysis. unanswered questions about aggregating data. J Clin Epidemiol, 1991, 44(2) , 103

SPSS Inc. SPSS10.0 Tutorial & Syntax Reference Guide. Chicago: SPSS Inc, 1990

Stevens, S. S. Handbook of experimental psychology. New York: John Wiley & Sons, 1961

Terkel, J. , & Rosenblatt, J. S. Maternal behavior induced by maternal blood plasma injected into virgin rats. *Journal of Comparative and Physiological Psychology*, 1968, 65, 479 ~ 482

Thacker, S. B. Meta-analysis. a quantitative approach to research integration. *JAMA*, 1988, 259(11) , 1685

Tucker, L. B. , Koopman, R. F. , & Linn, R. L. Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 1969, 34, 421.

Tucker, L. R. , & MacCallum, R. C. Exploratory Factor Analysis. Online factor analysis course. <http://www.unc.edu/~rcm/book/factornew.htm>. 1997, 2004 - 5 - 30.

Watson, J. B. , & Rayner, R. Conditional emotional responses. *Journal of Experimental Psychology*, 1920, 3, 1 ~ 14

Watson, J. B. , & Rayner, R. Conditional emotional responses. *Journal of Experimental Psychology*, 1920, 3, 1 ~ 14.

Watson, J. B. , & Rayner, R. Conditioned emotional responses. *Journal of Experimental Psychology*, 1920, 3, 1 ~ 14.

Witkin, H. A. The nature and importance of individual differences in perception, *Journal of Personality*, 1949, 18, 145 ~ 170

Wolins, L. , Responsibility for Raw Data. *American Psychologist*, 1962, 17, 658

Zimbardo, P. G. , Haney, C. , Banks, W. C. , & Jaffe, D. The mind is a formidable jailer: A Pirandellian prison. *The New York Times Magazine*, 1973, Section 6, 38