

第一编 绪论

现代心理学的每一项发展，都离不开心理学的实际研究和应用，而科学心理学的理论正是建立在心理学研究与应用的基础之上的。因此，心理学研究方法已经成为现代心理学的重要组成部分之一，学习和掌握各种心理学研究方法，对于发展我国心理学体系，丰富适合我国实际情况的心理学原则，促进心理学在社会主义现代化建设中的应用，有着十分重要的意义。

本编第一章“心理学研究方法的特点与发展”，介绍心理学研究方法的一般概念、作用和发展概况，特别是讨论现代科学方法对于心理学研究的影响和促进。进行心理学研究，必须有明确的目的和正确的理论指导。第二章“心理学研究的理论基础与思路”，讨论了心理学研究的理论假设与因果关系分析的基本条件，强调了充实的理论基础与理论构思对于心理学研究的重要性，并介绍了心理学研究的新思路和新途径。

第一章 心理学研究方法的特点与发展

第一节 心理学研究方法与科学方法论

一、科学研究的一般特点和原则

1. 科学研究的特点

心理学研究需要从理论或实际课题入手，系统地收集与分析数据资料，从而得出有意义的研究结论。心理学研究具有科学研究的基本特点。一般来说，科学研究是对研究对象之间关系的命题所进行的有系统的、有控制的实际研究。科学研究与常识或日常经验的主要区别，就在于科学研究的系统性和有控制的特点。

(1) 系统的理论框架与目的

科学研究都有比较系统的理论框架，有目的地通过实际调查与研究，检验理论假设的正确性，解决实践中遇到的各种问题；而人们的一般常识或经验，虽然也运用某些“概念”或“假设”，却缺乏系统性与周密性，更不可能严格地检验各种想法或观点。

(2) 控制的机制

科学研究都具有一定程度的控制机制。科学家们在研究中总是设法恒定或排除某些无关的因素，以便着重观察与分析有关对象的关键特征及其影响因素，找出事物发展的因果关系；而人们在日常体验中却很少会有意识地控制某些条件，系统地解释各种现象。

(3) 严密的分析

科学研究中总是有意识地、系统地寻求研究对象之间的关系，而对现象作出严密、深入的分析与解释；而一般常识中对于某些关系的说明，则往往是松散的、缺乏系统和比较表面的。

由此可见，科学研究有其鲜明的特点，能够解决常识无法解释的许多问题。了解科学研究的上述特点，有利于更有效地开展各种心理学研究。

2. 科学研究的目的是与一般过程

进行科学研究，需要有明确的目的。科学研究的目的是主要包括三个方面。

(1) 描述对象的状况

科学研究的目的是就研究对象的现状作出描述与说明。以心理学研究来说，这方面的课题常常是：儿童的认知发展处于什么水平？学生的学习成绩怎么样？职工的工作积极性高不高？等等。这些问题都是要说明对象的现状。这类研究的层次一般比较低，可以为进一步的研究打下基础。

(2) 解释对象的活动过程

科学研究的又一目的，是对研究对象的活动过程与特点作出解释。例如，儿童认知发展包含哪些阶段、影响各阶段发展的因素是什么？学生的学习成绩为什么提高得比较快、差生学习困难的原因何在？职工为什么对自己的工作感到满意、影响职工积极性的因素有哪些？这方面研究的层次比前一类研究提高了一些，能对事物发展的因果关系作出更进一步的分析与解释。

(3) 预测对象将来的发展

科学的另一方面目的，是根据描述与解释的结果，预测在采取某种措施或创设一定条件以后，对象状况可能发生的变化，或者依据现有的测量指标，预测一定时间间隔以后对象的发展。例如，

增加游戏活动对于儿童认知发展的影响；教学改革新措施对于提高学生解题能力的作用；企业民主管理制度对于增强工作积极性的实际意义；等等。这些都属于较高层次的研究目标。

为了描述、解释和预测对象，就需要建立系统的理论。在心理学研究中，我们并不是试图说明每一种单独的心理特征或行为，而是努力就许多不同行为的关系和原因作出一般的解释，即寻求具有普遍意义的心理学理论。这里，我们把理论定义为一组相互联系的概念与命题，它通过对变量之间关系的系统、具体说明，来描述、解释和预测各种现象。

心理学各个分支的研究虽然各具特点，却都具有科学研究的一般过程。从辩证唯物论的认识论来看，科学研究实际上是这样的一般过程：从各种事实中归纳出一定的理论，即从个别或特殊的知识中概括出一般性的原则、规律和定理；根据这种理论演绎出各种有关假设，即从一般的知识出发，对特殊或个别现象作出推论和假设；以进一步的事实验证这些假设，从而进入科学研究的高一级水平。图 1-1 为科学研究的一般过程。

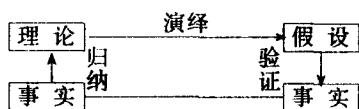


图 1-1 科学研究的一般过程

上图表明，科学研究是一个整体循环，是一个不断从事实到进一步事实的过程。在研究中，既要注意运用归纳方法，不断从各种特殊现象中总结出一般性的理论；又要重视演绎方法，从理论作出有关的假设，并在实验中加入验证。充分认识科学研究的一般过程，有利于在研究中抓住本质的方面。

3. 辩证唯物论是科学研究的指导

进行科学研究，总离不开科学的方法，没有客观而正确的科学

方法，就不可能获得可靠和有效的实验数据或其它感性材料。可见，科学方法对于科学研究是很重要的。但是，可靠的实验数据和感性材料并不是科学研究的最后目的，科学研究是为了透过众多的数据和资料，分析和认识对象的本质，总结出事物发展的规律，从而认识世界和改造世界。因此，科学研究就离不开理论思维，只有从实验事实 and 材料出发，通过理论思维发展理论，才能认识对象的本质，并进一步检验认识的正确与否，这就是心理学家在科学研究中应该坚持的辩证唯物论的认识论。

在科学方法论的发展过程中应该提到的是英国的弗兰西斯·培根（1561—1626）和法国的勒奈·笛卡儿（1596—1650），他们都试图对科学方法进行全面分析。培根是“英国唯物主义和整个现代实验科学的真正始祖”（马克思和恩格斯：《神圣家族》），他以唯物主义和创新精神，开创了唯物主义经验论。培根的主要著作《新工具》，强调了归纳法的重要性，他对归纳的三种方法：求同法、差异法和共变法作了详细论述，克服了简单枚举法的局限性。他认为，任何科学都应是实验的科学，揭示了实验方法的特点和巨大作用。笛卡儿是唯理论的创始人，他的《方法论》一书，特别重视演绎法和数学方法的作用。他片面地认为，只要从正确无误的公理出发，用类似数学的方法进行论证，就可以把自然界的显著特征都演绎出来。可是，他们两位的思路都有着一一定的局限性。19世纪中叶以后，自然科学迅猛发展，实验方法广泛采用，科学方法与哲学方法的分界也日益明显。辩证唯物论的认识论，严格地以自然科学和社会科学的成就为基础，越来越成为科学研究的指导。

近年来，我国社会主义现代化建设和改革的日益深入，给心理学各个领域的研究提出了各种迫切需要解决的新课题。与此同时，心理学的许多领域，正在进入“科学的、系统的和全面发展的时代”。多方面的心理学理论和大量的研究成果，需要在实践中进一步加以

检验和整理。这些研究课题和任务，都对心理学研究方法提出了新的、更高的要求，也越来越迫切地需要更高层次的理论指导。恩格斯的《自然辩证法》正是提出了这种在心理学科学研究中必不可少的指导原则。《自然辩证法》中有关科学研究方法的精辟论述和重要思想，对于更好地掌握心理学研究方法，具有十分重要的现实意义。

二、心理学研究方法的作用

方法论也叫方法学，是关于方法的理论和学说。科学方法论则是有关所有科学方法的理论，它以方法为研究对象，通过方法研究，改进、加强和发展科学方法，为科学研究提供最佳的研究方案和方法。学习和掌握心理学研究方法，不仅能增加研究的知识、开阔研究的思路，而且能更有效地评价以往的研究、应用各种心理学原理。

1. 增加心理学研究的知识

心理学研究方法本身就是心理科学体系的重要组成部分，因此，不了解心理学研究方法，就没有全面了解与掌握心理学的体系。学习与掌握心理学研究方法，可以使我们增添心理学的知识，充分了解各种心理学原理的研究基础。实际上，世界各国心理学界，都把心理学研究方法的学习作为心理学教学方案中不可缺少的方面。

2. 开阔心理学研究的思路

心理活动是一个十分复杂的过程，包含着各种各样因素的作用与影响。因此，必须学习与掌握科学的研究方法，从而系统、周密、深入地研究各种心理现象。由于心理现象的动态与多因素特点，所以要求我们在进行心理学研究时采用多种研究方法与手段。学习心理学研究方法，有助于我们开阔研究的思路，从各种不同角度、运用不同手段，考虑与制定更有效的研究方案。

3. 提高评价研究的能力

一项研究的效果与质量如何，这是科学研究中人们共同关心的

问题。研究的理论与实际意义，在很大程度上取决于研究设计的周密性和研究方法的严格程度。学习和掌握心理研究方法，将能够大大提高我们从方法论角度评价研究的能力，这不仅有利于对以往研究的效度和意义作出评定，而且有助于设计与实施更有效的新研究。

4. 加强心理学的应用

说心理学的发展在于它的应用，这是一点也不过分的。学习和掌握心理研究方法，将能够加强心理学科学原理的实际应用。我国社会主义现代化建设事业，正向我们提出大量的急需解决的心理问题，各行各业的同志都越来越认识到学习和应用各种心理学原理的重要性。心理研究方法正是为广泛深入地应用心理学原理，提供了有效的工具，进而调动人们的积极性、促进现代化建设的各项工作。

三、心理学研究方法的内容与分类

1. 心理学研究方法的主要内容

心理学研究方法主要包括以下一些方面。

(1) 研究方法的发展与理论思路

这部分内容有心理学研究的特点、方法的作用与分类、方法的历史发展、研究的主要范式、研究的理论思路和理论模型的构建等。这些内容提供了心理学研究方法的总体框架。

(2) 研究的设计与评价

这部分内容包括研究的类型、设计步骤与研究的计划；研究效果的总分析与评价；研究的取样方法与研究效度的分析；实验与准实验的设计，等等。这些内容既是心理学研究设计的纲要与方法，也是对研究进行评价的标准。

(3) 心理测量的理论与方法

进行心理学研究，需要作各种心理学测量，采集各种数据与资料。因此，心理测量的理论与方法就是心理学研究方法的重要内容，其中包括经典与现代的心理测量理论、心理测验的方法和心理物理学方法。还有各种常用的用于收集心理学数据的特殊方法，例如，访谈法、问卷法、观察法、口语报告法、社会测量法和心理分类法等。

(4) 研究数据的分析与研究结果的报告

心理特征复杂多样，心理活动千变万化，需要用强有力的统计分析手段，进行系统、深入的数据分析，寻求心理发展与变化的基本规律。各种心理统计方法是心理学研究方法的重要组成部分。除了一些基本统计方法，还包括因素分析、多元回归分析等多元统计方法。对于统计分析所得到的结果，还需要运用心理学理论进行各种分析与讨论，并按照一定的格式，作出科学的、有意义的研究结果报告。研究报告与论文的撰写，也是心理学研究方法的重要内容。

2. 心理学研究方法的分类

关于心理学各种研究方法的分类，有不同的看法。目前比较一致的分类，把心理学研究的具体方法分成六大类：观察法、实验法、模型法、测量方法、统计方法和其他研究方法。图 1-2 表示出这六类方法的分类图解。

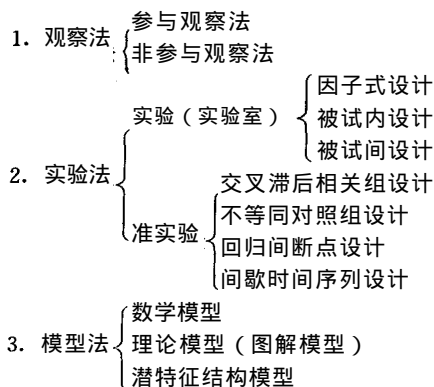




图 1-2 心理学研究方法的分类

第二节 心理学研究方法的发展

一、早期心理学研究方法

19 世纪初以前的一个相当长的时期里，由于经验自然科学还没有发展到能够解释自然现象及其内在联系的阶段，因此只能依靠哲学的抽象思辨方法来研究心理现象，从哲学上对心理现象进行一般性的描述和推论。这个时期所作的分析和描述，还不足以称之为心理学科学研究。由于生产和科学技术的发展，实验方法和定量研究方法在各个科学领域中迅速推广，以 1879 年冯特在德国莱比锡建立世界上第一个心理实验室为标志，心理学发展成为一门独立的科学，心理学科学方法也得到新的发展。费希纳（Fechner, G. T. 进行了独立的心理学实验，提出了心理物理法 1860），揭示了客观刺激量和主观感觉量之间的函数关系；艾宾浩斯（Ebbinghaus, H. 运用实验方法和无意义音节，测定了记忆和遗忘之间的数量关系，提出了著名的遗忘曲线 1885）；比纳（Binet, A. ）和西蒙 Simon, T. 运用测验方法，编制了第一个智力量表 1905）等等，都是心理学科学方法的成果。

19 世纪末、20 世纪初，心理学在发展中形成了许多学派，人们也把这个时期称为“学派时期”。各学派采用了不同的研究程序和方法。这个时期的心理学研究，集中于通过经验控制、从外部验证来确定因果关系，即研究者运用过去的经验来控制背景变量。各学派以不同的研究方法，研究不同的领域，提出了各自的理论。有关心理现象的经验研究，强调能够观察和测量的感觉过程。构造主义心理学派的代表人物冯特，主张用客观的方法去研究心理问题。他利用不同的方法研究感觉运动过程和认知心理活动，并提出了统觉的概念来描述观察者选择与构造其意识体验的注意“焦点”。为此，他

还提出了内省的方法，即训练观察者，让他们报告自己意识体验的“元素”。构造主义学派运用内省方法，仔细地采集和分析了自己和其他人有关心理过程的报告，对注意和意识的结构进行了全面的研究。尽管内省方法在与实验方法结合在一起时确实有一定的价值，但它并不能对心理过程作出明确的定量分析。同时，它的报告受到研究者自身理论构想和主观解释的很大干扰，因而具有较大的局限性。

以华生(Watson, J. B.)和斯金纳(Skinner, B. F.)为先后代表的行为主义学派，认为冯特等的内省方法是无效的方法，并且否认内部心理过程有研究的必要。行为主义学派主张采用实验方法，研究可观察、可测量、可控制的外显行为。行为主义学派在排斥内省法的同时，利用了巴甫洛夫和桑代克用于动物研究的条件反射等方法。尽管行为主义否认意识和认知过程的主张渐被弃绝，但在研究方法上强调客观观察与实验方法的观点，至今在心理学研究中颇有影响。

另一个重要的心理学学派，是以韦特墨(Wertheimer, M.)及其学生为代表的格式塔学派。他们既不同意冯特、铁钦纳等的内省方法，也反对华生等的行为主义方法。格式塔学派集中于知觉组织的研究，注重于分析和描述知觉结构而不是其过程。在他们的许多研究中，都采用了现象学方法。例如，有关似动现象的研究，把两个相距不远、呈现时间间隔为1/10秒到1/30秒的两个视刺激物，看成一物在移动。他们在这些研究的基础上，建立了格式塔心理学，总结出有关知觉的一系列规律。

此外，以弗洛伊德(Freud, S.)为创始人的精神分析学派，采用了临床分析方法，就心理异常和潜意识进行了系统的研究。机能主义学派则通过内省和客观观察，研究“心理过程”的特点与规律。

以上学派时期的各种研究，加深了我们对于心理活动的认识，推动了心理学研究方法的发展，提高了研究的水平。图1-3说明了各主要学派的研究领域和研究方法。

学 派	研究程序和方法	研究领域
构造主义	内省方法	意识、经验、心理结构
机能主义	内省和客观观察 (非正式观察)	心理过程的功能
精神分析	临床分析法	心理异常、潜意识
行为主义	实验方法 客观测定	外显行为
格式塔学派	现象学方法、主 观报告、行为测定演示	知觉组织、主观体 验强调知觉、记忆、思维

图 1-3 学派时期的研究程序和方法

二、心理学研究的主要范式

随着现代科学技术和心理学的发展，心理学家越来越需要一些新的研究方式与思路。在心理学研究的实践中，逐步形成了两种主要的研究范式：大样本范式与小样本范式。

1. 大样本研究范式

1925 年，费歇尔 Fisher, R. A.) 在其经典著作《研究工作者的统计方法》中，把大样本研究范式引进了心理学研究，成为心理学研究与分析方法的重大突破。大样本范式以随机化程序选择大量被试，对他们进行较少量的心理学测试，从而检验与分析有关的心理特征与发展过程。费歇尔的贡献在于，他把以下三项程序创造性地引入了心理学实验研究：

- (1) 运用随机化程序选择被试；
- (2) 对被试进行实验控制；
- (3) 运用统计方法检验研究的结果。

大样本研究范式的基本思想是：一项研究所采用的被试越多，统计检验就越有效。大样本研究范式大大提高了心理学研究的效率和水平，很快被心理学家作为重要的科学研究范式，在心理学研究方

法的发展中起了十分积极的作用，成为现代心理学研究的主要范式之一。

2. 小样本研究范式

虽然大样本研究范式在心理学研究中很有用处，却也存在一些局限性。例如，大样本范式并不适合于许多临床心理研究，因为被试的情况往往比较特殊；在不少情况下，由于条件的限制，研究者只能对少数被试或者动物进行实验研究；在大样本条件下，不容易作追踪测量和深入的分析，等等。因此，心理学研究中出现了小样本研究范式。

小样本研究范式的基本思想认为，对少量的被试进行大量的心理测定，能够对其心理活动展开有效的研究。小样本研究范式是由斯金纳在其名著《有机体的行为》（1938）中提出的。他强调心理学必须使用自然科学方法，认为小样本研究范式应作为心理学研究的主要方式，从而丰富了心理学研究方法。

根据小样本研究范式，研究应包括三个阶段：

- （1）确定所研究对象的行为基线或反应方式；
- （2）操纵自变量，观察和记录因变量的变化；
- （3）撤除自变量，继续监测因变量的情况。

小样本研究范式已逐步为许多心理学家所采用，它与大样本范式相比，更严密地监测了对象的行为，更系统地测量了自变量的效应，并得以更细致地进行“被试内”比较和分析。但是，许多研究也表明，小样本研究范式的外部效度不易达到要求，并容易受被试系统误差的影响，因此，对小样本范式应作更多的验证性研究。

三、定性研究方法

心理学研究还可以采用定量和定性两种方式。所谓定量研究，是指从数量化的思路出发，运用心理量表或其他手段，对心理特征及

心理活动作定量测定，并对所测得的数据进行严格的统计分析。所谓定性研究，则是采用书面的、言语的和观察的方式，取得有关心理特征与行为表现的描述性资料，并对这些资料作出定性的分析。

在心理学定量研究方法发展的同时，定性研究方法也得到了不断的改进和发展。特别是最近一些年来，定性研究方法也非常受人重视，并被作为定量研究方法的必要补充，成为许多研究课题的重要手段。

心理学家认识到，当我们把心理活动和过程转换为某些统计数字时，实际上失去了不少有关心理过程主观性质的信息（例如，反应时）而且不少心理现象和活动都还只能用定性方法研究。因此学习和运用心理学定性研究方法也是掌握心理学研究方法的重要方面。

目前最流行的定性方法有两种：个案分析法与文献综合法。个案分析法是指选择少量典型案例进行全面、深入的观察与分析。常用的方式是进行“参与观察”，即让研究者参加到被试的工作或活动环境中，通过与被试的深入交往，系统地、不引人注意地采集研究数据。个案分析法往往能够比较深入地了解 and 把握研究对象的心理特征及其发展过程。

文献综合法是指从各种书面的、口头的方式获得有关研究对象的资料，包括日记、书信、传记、简历和自由访谈的记录等等。这种方法常常可以更深入地研究无法直接观察的方面和内容，例如，对青少年犯罪问题的研究、对人们一生心理状况和发展情况的研究，等等。

近年来，在管理、儿童、教育和认知心理学等许多分支的研究中，许多人主张以定性和定量研究相结合的方式从事更为深入细致的分析和研究，进一步提高研究的效度。

四、心理学研究方法发展的新特点

心理学研究正在出现引人注目的新趋势，表明了心理学研究方

法的许多新特点。概括起来，主要有这样两个方面：研究的现场化与多分支特点；研究手段的综合化、数学化和计算机化。

1. 研究的现场化与多分支特点

随着心理学研究应用性的日益加强，正从传统的理论与实验室研究，转变为各种形式的现场研究，在现实的情境与条件下控制和观察心理活动、测定和记录整个心理过程。通过现场研究，不但进一步验证了实验室条件下获得的某些研究结果，而且解决了实际生活与工作中大量需要解决的问题。在我国心理学研究中，有关工效学标准的工程心理学研究、有关人员测评的劳动人事心理学研究、有关企业计算机应用与系统开发的计算机心理学研究和有关人工智能的认知心理学研究等，都是十分有意义的研究项目，这些研究大多数都在现场进行。

同时，心理学家认识到，心理过程包含着各方面心理特征的交互作用并受到各种因素的影响，只从心理学某一分支，是不可能全面、准确地解释和预测心理现象的。因此，必须运用多个心理学分支的原理与方法开展研究，从而相互补充和促进，推动心理学的各类研究。

2. 研究手段的综合化、数学化和计算机化

现代心理学研究越来越强调测量的可靠性（信度）和研究的有效性（效度），这就要求使研究手段更为综合化、数学化和计算机化。所谓“综合化”，是指在心理学研究中尽可能采用多重设计和多种测量方法。对于同一种心理过程，人们已经不再满足于用单一方法和单一研究设计进行研究，而是倾向于运用多种方法和综合性的设计。例如，对于同一课题，采用纵向研究和横向研究相交迭的聚合式设计；对同一组变量，先后或同时用观察法、访谈法、问卷法和实验法进行研究，并对不同方法所取得的结果进行相互比较和验证。即使是在严格控制的实验室实验中，也倾向于在用实验仪器测定某些

心理或生理心理指标的同时，以问卷法或者口语报告法，了解被试的主观体验，从而取得更为全面、深入的资料与信息。

心理学研究方法的数学化是十分引人注目的。前面曾提到，费歇尔把随机化程序和统计检验方法引进心理学研究，使心理学研究方法出现重大的进展。从那时以来，数理统计方法和其它数学方法在心理学研究中应用日益广泛。50年代以前，比较流行的是单因素的或者一对自变量与因变量之间关系的研究和分析。50年代以后，随着电子计算机在心理学研究中的应用，人们越来越多地采用多元分析方法，弥补了单因素分析方法的局限性。70年代以来，一些社会科学专用统计软件包的开发和运用，使多元分析方法在心理学研究中日益流行，其中较常用的方法有：复方差分析、多元回归分析和因素分析等。就连“研究文献总结”这一传统使用定性方法的领域，也开始采用数理统计等数学方法。这些方法的采用，使心理学研究和分析水平出现了重要的突破与发展。与此同时，心理学研究方法出现了计算机化的趋势，计算机技术就像统计分析方法那样，成为心理学家必须具备和掌握的技能。计算机除了帮助进行研究数据的处理与统计分析以外，还被用于心理学实验操纵和控制以及进行计算机模拟，从而使心理学研究方法达到了现代科学的水平。

第三节 现代科学方法与心理学研究

一、信息方法在心理学研究中的应用

现代科学的发展，不断为心理学研究提供新的理论、方法和手段。特别是20世纪40年代以来，电子计算机和控制论的发展和运用，使得信息方法、控制论方法和系统方法等现代科学方法在心理学各个领域得到日益广泛的应用。

香农(Shannon, C. E., 1948) 在贝尔系统技术杂志上发表的文章“通讯的数学理论”的文章,标志着信息论的诞生。信息论是由数理统计方法来研究信息处理与传递的科学,是研究通讯与控制系统中信息传递共同规律的通讯理论。次年,韦弗(Weaver, W.)和香农在《科学美国人》上发表了一篇介绍性的文章,使信息论思想从工程领域扩展到其它各个领域。

1949年,米勒(Miller, G. A.)和弗里克(Frick, F. C.)发表了“统计行为论和反应顺序”一文,第一次明确论证了信息论与心理学的关系,并提出了对“行为组织”进行定量分析和对事件发生顺序作出模式的方法。米勒等的文章标志着心理学研究中开始运用信息测量。从50年代开始,心理学家作了许多尝试,设法运用信息技术解决心理学问题。从目前的研究情况和理论进展来看,一方面,信息论在提出某些心理学问题和分析某些心理学数据方面有很大的价值,并且被广泛用于解释许多心理现象和构建心理学理论;另一方面,人们也认识到,信息论并不能为所有心理学问题提供现成的解决方法。因此,信息论更多地被作为心理学研究的思路和分析工具。

信息方法在认知心理学、工程心理学以及近年来发展起来的计算机心理学的研究中有比较广泛的应用,我们通常介绍的是信息论基础和对一组事件的信息的计算方法,称为单变量信息论。计算公式如下:

$$H = \log_2 N \text{ (N 个相等可能性的事件)}$$

$$H = \sum_i p_i \log \frac{1}{p_i} = - \sum_i p_i \log p_i$$

近年来,心理学家日益重视双变量甚至多变量的信息运算。由于实际生活中的复杂情景和多种变量影响心理活动,因此在考虑两组或两组以上事件的时候,信息论的真正功效就更显示出来了。例

如，以一组刺激表示机器状态，显示器显示机器的开闭，另一组事件是操作者的反应。这种情况下运用的就是双变量信息论，其中一个重要部分涉及到刺激事件和反应事件之间的关系，因此，它对于心理活动的信息计算与分析就更有实际意义。图 1-4 为 S、R 两组事件的信息关系图。

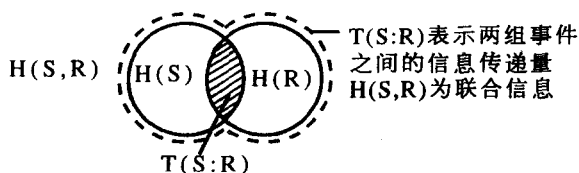


图 1-4 两组事件的信息关系图

S 与 R 的信息传递量为： $T(S:R) = H(S) + H(R) - H(S, R)$ 。

由于许多工程设计的目标之一是增加所传递的信息量，这也成为这方面心理学研究的目的之一。运用双变量信息方法，可以提高显示器和控制器以及其它设计的质量。

例如，有一显示器显示系统状态 1、2、3、4，操作者必须打开 1、2、3、4 阀的控制器作出反应。操作试验的结果如表 1-1 所示，我们可以通过传递信息量的计算，决定控制器设计的结果。

表 1-1 某控制器操作试验的结果

系统 状态 显示 (S)	反应(R)				
	阀 1	阀 2	阀 3	阀 4	Σ
1	20	5	0	0	25
2	5	15	5	0	25
3	0	5	15	5	25
4	0	0	10	15	25
Σ	25	25	30	20	100

根据表 1-1 可以求得刺激信息量 $H(S)$ 、反应信息量 $H(R)$ 和联合信息量 $H(S, R)$ 分别为：

当显示信号相等概率出现时，

$$H(S) = \log 4 = 2 \text{ bits} ;$$

$$H(R) = .25 \log 1/.25 + .25 \log 1/.25 + .30 \log 1/.30$$

$$+ .20 \log 1/.20 = 1.9855 \text{ bits} ;$$

$$H(S, R) = .20 \log 1/.20 + .50 \log 1/.50 + \dots$$

$$+ .15 \log 1/.15 = 3.1086 \text{ bits} ;$$

$$\text{所以 } H(S : R) = 2.0 + 1.9855 - 3.1086 = .8769 \text{ bits}。$$

由于最大传递信息量实际上不会超过 $H(S)$ 与 $H(R)$ 中较小的一个，而实际传递信息 $H(S : R)$ 却只有 .8769bits 为最大传递信息量的 44%。因此，这一试验结果表明，应该重新设计该控制器，以便增加传递信息量。

二、控制论方法在心理学研究中的应用

维纳 (Wiener, N.) 所创立的控制论，是研究生物、机器、社会中通讯和控制的共同规律的理论，是一门新的横向科学，对于心理学研究具有重要的方法论意义。控制论分析方法是以前某种标准来侦察系统状态的偏离，并采取步骤来矫正系统行为以消除误差，即通过反馈指导行为。

控制论方法已被广泛运用于心理学研究，特别是应用心理学的研究。管理心理学中的组织控制模型的核心思想是反馈控制；工程心理学中的行为控制理论运用拉普拉斯 (Laplace) 变换，把控制论方法应用于控制器和监视器的作业活动分析，使对控制器的研究从静止地研究控制器设计本身提高到研究整个控制作业的动态过程，进而从较低水平的感知运动研究向较高层次的认知研究发展。

基本的反馈控制系统包含图 1-5 所表示的成分。控制论正是撇

开了对象的物质与能量的具体形态，着重从信息方面研究系统功能和系统的动态变化过程。

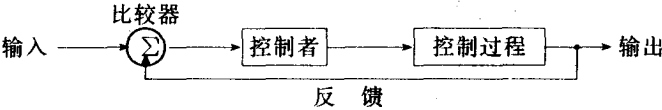


图 1-5 反馈控制系统的基本成分

组织控制模型是建立在基本控制模型的基础上的。它包含着某种信息反馈、决策和控制过程，使组织按照其目标和计划发挥功能。

图 1-6 表示了基本控制模型。

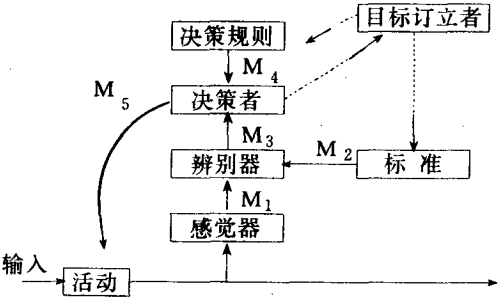


图 1-6 基本控制模型

从图 1-6看，活动的输出被感受以后，把信息提供给辨别器，在那儿与活动的标准进行比较。根据比较的结果，决策者运用一组决策规则，决定所应发送的校正反馈（ M_5 ），对活动作出控制。组织管理正是发挥着上述控制功能。

三、系统方法在心理学研究中的应用

奥地利生物学家贝塔朗菲（Bertalanffy, I., 1945）运用逻辑和数学的最新成就，建立了适用于一切综合系统或子系统的一般系统论。所谓系统，是由许多相互联系和相互作用的各部分组成的整体或实体。系统方法则是一种把对象放在系统的形式中加以考察的方

法。它从系统的观点出发，着重从整体与部分和与环境的相互联系、相互作用和相互制约的关系中综合和精确地考察对象，找到解决问题的最佳方法。因此，从很多方面来说，系统方是一种思路，是一种认识和分析整体及其不同水平之间相互关系的能力。

在研究方法方面，以系统方法使复杂的系统层次化，并把不同的理论或技术应用于不同时间的问题或问题的不同部分与水平。目前，在心理学研究中应用较多的是认知系统的分析、人一机系统的研究、组织的社会技术系统分析、人与计算机界面系统以及管理信息系统的研究。实践证明，从系统观点对心理过程进行研究和系统方法的采用，正使得许多心理学研究从零碎地、孤立地和局部地看待和分析心理学活动，提高到系统的、整体的、多方面因素相互联系的水平，从而提高了研究的效率。当然，系统方法也受到某些批评，主要认为它是描述性的、缺乏严格的逻辑定义，同时，预测功效比较低，等等。可是，随着系统理论的应用和发展，系统方法和理论正变得更为准确和具有预测能力。

与系统方法密切有关的，是所谓列联方法（contingency approach）。如果说系统观点为心理学研究提供了较为宏观的图式并包含了较高水平的概括化，那么列联方法则是一种较为具体的分析手段。它强调了系统之间关系的具体特征与模式，重视在变化的条件和具体的情况下研究对象的活动过程。根据“对具体情况作具体分析”这一列联观点，在心理学研究中就应同时考虑和分析几方面的因素和变量及其相互关系，从而发现在各种条件不同组合下的共同作用和效应。心理学家越来越认识到，人的心理活动是由多种因素和变量共同作用所决定的。因此，列联方法和系统方法一样，已成为心理学研究中的重要方法和思路。例如，在领导行为的心理学研究中，弗德勒（Fiedler, F. E.）从列联的思想出发，同时考虑了工作任务结构、领导岗位权力和上下级关系等三种因素的九种不同

组合的条件；提出了有关领导行为的权变模型。

四、计算机在心理学研究中的应用

电子计算机在 40 年代的兴起和运用，为心理学和其他学科的研究和发展开辟了新的广阔的道路。现在，计算机就像统计方法那样，在心理学研究中得到越来越普遍的运用，成为所有心理学家不可缺少的知识和技能，也成为心理学研究方法的重要组成部分。

1. 计算机的基本用途

在心理学研究中，计算机主要有三方面的用途。

(1) 研究数据的处理

这是计算机目前在心理学研究中的最主要用途，除了自编程序以外，主要运用一些专用统计软件包（SPSS、BMDP 和 SAS 等）进行有关研究数据的收集、贮存和统计分析等等。

(2) 心理学实验的操作与控制

近些年来心理学研究的新发展之一，是在心理学实验中运用计算机。特别是实验中计算机的实验操作和联机控制，大大提高了心理学实验的精密度和可靠性，也使得能够对许多复杂心理过程进行实验和分析。

(3) 心理过程的模拟和模型化

运用计算机模拟技术，可以建立有关各种心理过程的理论模型。心理模型主要有两大类：功能模型和结构模型。前者就像工程师的“黑箱”只重视输入和输出，并不了解输入/输出之间的情况，后者则强调箱子的内容。计算机心理模型又有其特殊的形式，它们实际上是一些使计算机在相似条件下类似于人或动物的方式从事作业的计算机程序。由于计算机心理模型的编程原则与产生模拟行为的原则相同，因而可以通过模拟过程，深入系统地了解行为的信息加工过程。它不仅可以检验理论的效力及其意义，而且可以用简化的形

式提出理论。

2. 计算机在心理学各领域研究中的应用

计算机正日益广泛地被应用于心理学各领域。这里主要介绍计算机在认知心理、临床心理、心理语言学等研究中的应用情况。

(1) 计算机在知觉实验中的应用

知觉实验中常常需要运用某些复杂而有严格要求的图形或结构作为实验材料，例如螺旋线图 and 立体图形，如果以徒手作图，就很难达到实验要求。因此，计算机作图已成为十分有效的途径。

- a. 螺旋线图：不少知觉实验以螺旋线图研究运动的后效应，但由于螺旋线的曲率恒定地变化，因此很难作出。用计算机根据数学方程作图，就比较容易。例如阿基米德螺旋线 ($r = a\theta$)
- b. 灰色旋转图：有的视觉实验需要旋转产生均匀明度变化的图形，这就是要求转盘的黑白比例十分不同。这种实验材料就需要由计算机按照反曲规律作出。
- c. 心理旋转图：三维心理旋转的实验研究要求成对地、旋转一定角度地呈现一些由立方体组成的立体结构。这些严格的图形都是由计算机作出，并照相后作为实验材料。谢泼德 (Shepard, R. N.) 等的著名研究使用了这种材料。
- d. 60 年代初，贝尔电话公司实验室的专家在研究深度知觉时发明了计算机画立方图形的方法，可以使除向两眼呈现的映象之间的立体差异线索之外不存在深度线索。60 年代后期 朱利斯 (Julesz, B.) 利用计算机设计出随机点立体图 (又称 Julesz 图) 他用这种实验材料发现了中央眼视觉的机制在立体知觉方面作出重要贡献。而 Julesz 图成为研究知觉的重要工具。

(2) 计算机在心理语言学研究中的应用

从 60 年代起，人们开始利用计算机进行心理语言学的研究，主要运用计算机进行自然语言对话的实验，研究语法结构。这种方法控制了许多无关因素，提高了实验效率。此外，在心理语言学研究，还进行了大量的计算机联机实验，以研究句法知觉的特征和机制。例如，著名语言学家乔姆斯基（Chomsky, N.）曾经编制了 BASEBALL 等程序进行研究，以期完善他的理论；认知心理学的先驱米勒（Miller, G. A.）也曾运用各种程序，在心理语言学方面作了许多计算机联机研究。

（3）计算机在临床心理学研究中的应用

临床心理学研究及临床治疗中，都已广泛地运用了计算机。计算机被直接用于临床信息与数据的收集与贮存，成为临床心理诊断的组成部分。例如，60 年代以来，韦克斯勒成人智力测验（WAIS）已用于这类计算机诊断；许多地方用计算机作自动临床访谈，让病人与计算机对话，了解与记录其心理症状及病史。人们还用计算机直接分析所测量的数据，或通过一系列问题，判断病人所遇到的心理障碍、情景障碍、社会障碍以及各种心理异常或精神病症。有一种自动诊断程序是以贝叶斯统计方法作为基础，利用先前的大量案例，求出新病例诊断正确的概率，其诊断的准确性比较高。此外，许多研究用计算机分析与记录临床 EEG 等生理指标，并通过计算机实施行为治疗，进行自动监测和反馈，取得了很好的效果。

第一章思考题：

1. 心理学科学研究有哪些特点？如何更好地发挥辩证唯物论对心理学研究的指导作用？
2. 学习和掌握心理学研究方法有何作用和意义？
3. 学派时期的心理学研究采取哪些方法？它们各有何特点？

4. 心理学研究的大样本范式与小样本范式各有何优点和局限性？举例说明这两种范式在心理学研究中的实际应用。
5. 现代科学方法对于心理学研究有何实际意义？试举例说明。
6. 计算机在心理学研究中有哪些主要用途？

第二章 心理学研究的理论 基础与思路

第一节 心理学研究与因果关系分析

一、因果关系分析的意义

我们在第一章里谈到，心理学研究与其他科学研究一样，其目的是要描述、解释和预测心理特征和心理活动过程，寻求心理活动的规律，即心理过程的因果关系。因此，心理学研究的一项中心任务，就是对各种变量与因素进行因果关系分析。典型的心理学实验，往往要求恒定或控制某些情景条件与变量，以便使所研究的因果关系能明确地显现出来。在这个意义上说，实验就是检验因果关系的手段。通过不同的实验设计，创设各种进行因果关系分析的条件，操纵或控制原因变量（又称为自变量），观察或测量结果变量（又叫因变量）的变化，从而确定自变量与因变量之间的因果关系。

心理学研究中，经常以因果关系分析的水平作为衡量研究质量的标准。一般来说，如果一项研究的变量及其因果关系明确，就认为该研究具有较好的效度，其层次水平也比较高。相反，如果一项研究中变量定义不明确、变量之间的因果关系含糊不清，就表明这项研究因果关系分析的水平很低，研究的质量和效果也比较差。

因果关系分析的水平，实际上涉及到研究所包含的控制程度，即排除或恒定无关因素、操纵关键变量的程度，在心理学研究中，“控制”可以包含三种含义：一是指研究对于外部因素和实验情景的控制

制能力，例如，运用消音室排除无关噪音的影响在各种心理学研究中，都不同程度地具有这种控制能力。二是指研究对于自变量的控制程度，例如，在特定时间改变自变量的水平，观察其效应。三是指研究设计过程中的控制成分，通过研究设计的总体安排、研究程序的规范化以及测量的可靠性与有效性等方面，排除影响因果关系推论的各种因素。可见，心理学研究从设计到实施的各个环节，共同决定了其控制的程度和因果关系分析的水平。

二、有关因果关系分析的若干理论

因果关系分析问题，一直是哲学界与心理学界最感兴趣的课题之一。人们就因果关系分析的条件和标准，进行了大量的讨论与研究，提出了各自的理论，从不同角度，为心理学研究的构思与因果关系分析，提供了有益的依据。

1. 休谟的理论观点

英国哲学家休谟（Hume, D. 1711—1776）有关因果关系的理论观点，在心理学发展史上有着很重要的影响。休谟特别强调了因果关系推论的三个条件：（1）一个原因和它的结果，在空间和时间上是常相接近的，原因和结果要是隔开了距离或时间，就难以作出因果关系推论；（2）一个原因常前于它的结果；（3）一个原因与其结果之间，必须有一种必然的联系，无论何时得到这种结果，原因总是同时存在的。这些条件中，他认为最后一条是最重要的。

2. 穆勒的因果关系推论的标准

穆勒（Mill, J. S. 1806—1873）对于因果关系的看法与休谟不完全相同，在某些方面可以说是前进了一步。穆勒认为，因果关系推论取决于三条标准：（1）一个原因在时间上总是先发生，然后才是结果；（2）所假定的原因与结果之间相互关联；（3）其他可能的因果关系解释都被排除。其中特别是第（3）条标准，显然有着较

深刻的含义，指的就是运用“控制”的手段，主动地操纵原因变量，突出显示所假定的因果关系。为此穆勒提出有关因果关系分析的三种方法：一致法（原因存在就将出现结果）、差异法（原因不存在也将没有结果）和共存法（上述两种关系如果满足，就能排除其他解释，作出有力的因果关系推论）。这也是穆勒对于心理学研究方法论的重要贡献。

3. 波普的理论

现代科学哲学家波普（Popper, K. R.）有关因果关系的理论（1959），对于心理学研究方法论产生了深远的影响。波普强调了否认所假定的因果关系的重要性。他认为，由于研究者不可能构建出所有的备择假设，在研究中设法否认某种因果关系的预测，要比证实这种理论预见的收获更多。根据波普的“否认”思想，需要在心理学研究设计中着重于分析各种有可能否认变量之间关系的因素或条件，使得有关因果关系的命题经得起多种检验与否认。

4. 科林伍德等的活动理论

科林伍德（Collingwood, R. G.）等许多科学哲学家提出所谓“活动理论”，以此解释因果关系。他们认为，在日常语言里，因果关系的概念蕴含着“操纵”。原因被操纵以后，才会使结果发生；原因与结果都是人的活动。活动理论有关因果关系的观点，包含着强烈的动机与意向的成分。这种理论所强调的因果关系定律应包含可操纵原因的思想，对于心理学研究方法论显然具有显著的意义。

5. 海德的认识论观点

与活动理论密切有关的，是一种“进化认识论”的观点，其中比较有代表性的，是海德（Heider, F.）的归因理论。按照进化认识论的看法，人类具有强烈的寻求因果关系的心理倾向；人类的进化总是伴随着某种探索环境的需要，通过对于环境的操纵和改造，获取有关因果关系的新的知识。海德因此提出著名的归因理论（1944，

1958),认为人们在活动中不断地作出因果关系的解释与推论,以此理解、控制和预测他们的环境以及随后的行动。海德的归因理论正在心理学各领域的研究中产生显著的影响,并对于心理学研究方法论和理论构思等诸方面的发展,起了重要的促进作用。

三、心理学研究中的因果关系分析

恩格斯在谈到因果关系问题时,曾精辟地指出:“单凭观察所得的经验,是决不能充分证明必然性的。……必然性的证明是在人类活动中,在实验中,在劳动中”,如果能够造成现象之间的一定的顺序,那么这就等于证明它们的必然的因果联系(《自然辩证法》第549页—550页)恩格斯在这里提出了验证心理活动因果关系的根本途径,这就是人类的活动和实验。通过实验操纵和活动,检验与寻求因果关系的基本规律。

根据恩格斯的上述思想,并且吸取其他学者有关因果关系分析的理论中的有益成分,可以提出心理学研究中因果关系分析的一些原则。

1. 因果关系的可检验性

心理学研究中进行因果关系分析的意义,在于它能够受到检验。这就要求在心理学研究中注意控制情景条件,对变量之间的因果关系进行多次的不同方式的检验。根据这一原则,单一研究总是具有一定的局限性,因为它只能对因果关系作出局部的、单一的验证。

2. 因果关系的层次性

因果关系分析具有不同的层次水平。在较低的具体水平上尚不确定的关系,有可能在较高的复杂水平上具有意义。因此,高层次的因果关系定律常常取决于许多低层次的具体规则和其他条件。

3. 因果关系的复杂性

一定的结果很可能是由于多种原因共同造成的,而这些原因并非总是以简单相加的方式发生作用,却往往具有交互作用。因此,交

互作用分析对于确定因果关系，有着重要的意义。

4. 因果关系的动态特点

严格控制的实验室条件表示了某种封闭系统，而现场研究则包含着开放系统。在这两种条件下，因果关系具有不同的普遍意义。在心理学研究中应重视因果关系的动态特点，注意在现实的、不确定的条件下研究因果关系规律。在一定的条件下，因果关系表现为可逆的、相互转换的特点。

上述四方面的原则，将是我们在确定研究构思和技术路线时的重要理论依据。

第二节 心理学研究的理论构思

一、心理学研究的理论思路

在心理学研究中，理论思路是至关重要的。心理学研究的理论思路，是指整个研究的理论出发点和分析问题的基本路线。只有运用充分而周密的理论构思、遵循清晰的理论思路，才能在心理学研究中把心理学普遍原理与具体条件结合在一起，更有效地解释、预测心理活动，并在解决实际问题的同时，总结出更具有普遍意义的心理学规律。正如著名心理学家勒温（Lewin, K.）指出的，没有一个比好的理论更加实用的。如果忽视理论的指导和总结，满足于就事论事的讨论或者经验式的照搬，那么心理学研究就将是十分肤浅和局限的，心理学就很难说会有真正的发展。

在许多心理学研究中，人们正更多采用自由观察和接近现实的模拟，更重视在现实情景中考察人的心理活动及其特点。这样，心理学研究中面临更为错综复杂的情景和多种因素的交互影响与作用。在这种情况下，心理学研究中的理论思维就尤其重要。恩格斯

在谈到这个问题时明确地指出：“经验自然科学积累了如此庞大数量的实证的知识材料，以致在每一个研究领域中有系统地依据材料的内在联系把这些材料加以整理的必要，就简直成为无可避免的。建立各个知识领域互相间的正确联系，也同样成为无可避免的。因此，自然科学便走进了理论的领域，而在这里经验的方法就不中用了，在这里只有理论思维才能有所帮助。”（《自然辩证法》第 465 页）这样，恩格斯不但强调了相同领域内对材料加以系统整理的必要，而且提出了一个重要思想，认为需要在不同知识领域之间建立相互的联系，这种联系又只有通过理论思维来实现。

1. 心理学研究的一般理论观点

在心理学研究中，人们常常有意无意地运用某种一般的理论观点，主要是静态观点和动态观点。静态的观点把心理学研究看成一种提供系统知识的活动，即发现新事实、增加新信息的活动；这种观点还把研究作为一种对所观察的心理现象进行解释的方式。因此，静态的观点强调现有的知识、理论、假设或原则，并设法在此基础上增加新的知识。动态的观点则与此不同，它承认现有知识的重要性；但主要把它作为进一步科学研究与理论发展的基础。这种观点又称为启发式的观点（*heuristic view*），它强调理论和相互关联的概念图式，并以此开展进一步研究。

采取静态观点还是动态启发式观点进行心理学研究，这会直接影响整个研究的理论思路 and 实际效果。以静态的观点从事心理学研究，就比较忽视对以往研究的总结和理论基础的构建，缺乏清晰的理论思路，把研究看成“前无古人，今无同类”的孤立活动。其结果往往是满足于对某些心理现象的描述和解释，既忽视自身的理论发展，又不重视同一研究领域中有研究之间的比较与沟通，因而只能缓慢地积累零碎、局部的事实。采取动态启发式的观点则不一样，在心理学研究中比较重视理论基础的构建，并在以往研究的

基础上形成进一步研究的明确思路，同时，把研究作为一系列相互关联的研究中的一个部分，并注重对新的研究结果的理论总结和确定新的研究课题。虽然许多人都懂得应该采取动态的启发式的理论观点，但是，在实际研究中做到这一点，并不是容易的事。在我们学习和运用心理学研究方法的过程中，应该清醒地把握自己的理论观点，从而更有效地完成研究课题。

2. 归纳与演绎

心理学研究与其它科学研究一样，其最重要的基本成分是数据和理论。数据是观察与测定的结果，理论则是对于结果的解释。在心理学研究中是从数据出发提出理论还是从理论出发寻求数据，这实际是归纳与演绎两种研究思路。

采取归纳思路，主要关心的是数据，数据是一个技术性术语，相当于平时所说的事实。可是，数据是通过系统观察而得到的信息，而事实却往往是偶然获得的或者是由一般性观察得到的。数据有许多重要的特征，就心理学研究而言，数据的变异程度、可靠性、可重复性等都是我们特别关心的。研究工作的基本任务之一就是设法获得能满足要求的数据并证明数据具有一定的特征。通过心理学研究，从一定的数据中概括出一般的原理。采取演绎思路的心理学研究则强调理论解释的重要性，而数据主要用来决定某些备择解释的正确性，研究者的主要精力在于提出理论解释。在心理学研究中完全采用归纳或演绎思路的情况是很少的，而常常需要在研究的不同阶段或不同的研究中侧重于某一种思路。

3. 心理学研究的理论假设

建立明确的假设，这是理论构思的重要组成部分。忽视或缺乏明确的假设，是研究构思或思路不清的表现，也不可能取得良好的研究成果。假设是有关两个或两个以上变量之间关系的猜测性陈述。研究的理论假设必须明确指出变量之间关系和变量的测定方式，从

而提供检验所假定关系的基础。例如，某教育心理学研究假设：小组式学习能够提高学习成绩。这一假设说明了变量“小组学习”和变量“学习成绩”之间的关系，并且可以通过变量的测量，检验所假定的关系。因此，这个假设满足了理论假设的要求。有的心理学研究，虽然也假定了变量之间的关系，但没有规定变量的测量方式，结果就很难通过研究进行假设的检验。这种假设就不符合理论假设的要求，在心理学研究中，应避免提出这类假设。

理论假设是进行心理学研究时不可缺少的重要工具。假设是从理论和其它假设演绎而来的，因而它实际上是一种理论工具。假设又是一种增长知识和发展理论的有力手段，离开了假设，研究的科学性就会受到很大的影响，研究甚至会盲目的。

二、相互作用心理学的理论思路

传统心理学研究中，主要采取了三种理论思路和概念模型，即特征论模型、心理动力模型和情境论模型。特征论模型只重视人们的个性与智力特征；心理动力模型则主要强调人的“动态心理场”。这两种模型尽管各自的测量方式、数据类型和数据分析方法不同，却都强调以人的自身因素决定其行为，在研究思路，都把注意力集中在基本的个性变量及其相互关系以及这些变量在一定环境条件下的作用方面。情境论模型则主要通过环境的考察来寻找决定人们行为的主要因素；这方面的研究集中于从外部刺激的类型和强度出发，分析行为反应的一般规律。上述三种思路不仅提出了不同的行为决定因素，而且运用了不同的研究策略。

70年代中期以来，心理学研究中逐渐形成一种较新的思路，即相互作用心理学，其基本思想，集中于行为发生过程中人与环境的多向性交互动作用，并且认为，在这种交互作用中，人和情景是相互依存而非单方面所决定。

相互作用心理学的基本观点有四个方面：

(1) 个体的行为是他们与所处情景之间多方向的连续的交互作用的结果；

(2) 个体在这一交互作用过程中是有意识的和主动的；

(3) 在交互作用的人的方面，认知和动机因素是行为的基本决定因素；

(4) 在交互作用的情景方面，情景对于个体所包含的心理学意义是重要的决定因素。

相互作用心理学的新思路，为真正从个体和情景两方面对心理过程作出描述、解释和预测，提供了理论和方法方面的基础，特别是对情景研究提供了较系统的分析手段，分别从“情景内交互作用”和“情景间交互作用”考察情景的心理学意义。目前，相互作用心理学已从个性研究，发展到社会心理、发展心理、教育心理、临床心理和工业心理等心理学分支，成为心理学研究的基本理论思路。

三、测量理论

心理学研究的重要理论基础之一是心理测量理论，即对心理特征和心理活动过程的数量化测定与分析的理论原则。心理学的测量并不是某种随意的定量操作，而是有严格的理论假设与原则的，忽视这一点，就不可能准确、有效地进行心理测量和分析，甚至有可能使心理学研究误入歧途。

经典的测量理论与方法，起源于本世纪初斯皮尔曼 (Spearman, C. E.) 的早期研究，至今已有七八十年的历史。这一理论的基础，是认为心理测量所得到的分数包含两大部分：反映真实心理特征的“真分数”和由测量误差造成的“误差分数”。应该指出，经典测量理论形成体系，还是四五十年代的事情。古利克森 (Gulliksen, H. 1950) 的著作《心理测验的理论》和洛德 (Lord, F. M., 1968) 的

《心理测验分数的统计理论》一书，对经典测量理论作了全面、系统的阐述和分析，成为测量理论发展的重要里程碑。与此同时，吉尔福特（Guilford, J. P. 1954）再版了其经典著作《心理测量学方法》，介绍和总结了心理学研究中最常用的一些心理测量方法。南纳利（Nunnally, J. C. 1978）再版了其《心理测量学理论》，总结了经典测量理论的新进展。

随着心理学研究广泛、深入的开展，人们对测量理论提出了越来越高的要求，经典心理测量理论的缺陷也日益明显。其主要问题，是按照经典测量理论求得的心理测量参数与指标，会由于研究样本的不同而变化，反应出测量结果的不稳定性；同时，经典测量理论忽视了被试对于测量项目的不同反应模式。这些问题都在很大程度上影响了心理学研究的质量。因此，人们对经典心理测量理论的假设提出了各种疑问，并通过研究实践，提出了新的心理测量理论——潜特征理论，充实了心理学研究的理论基础。潜特征理论中比较具有影响的是“项目反应理论”，它指出了个体的项目反应类型与某些个体特征之间的关系，可以说比经典心理测量理论前进了一大步。由于计算机的应用，使得潜特征理论更广泛地运用于教育心理，儿童心理、临床心理、管理心理、劳动人事心理、工程心理以及其它许多心理学领域。我们将在第七章“心理测量理论与测量方法”中详细介绍和讨论潜特征理论及其应用情况。

第三节 理论模型的构建

一、理论模型的一般问题

在心理学各个领域的研究中，人们日益重视理论模型的构建。在大量实际研究的基础上，总结和建立相应的理论模型，这将大大促

进整个心理学理论体系的完善和发展。事实上，在心理学不少领域，还十分缺乏理论建设，这就在相当大的程度上限制了心理学研究和应用的广度和深度，影响了某些心理学分支的发展。在其它一些领域，虽然提出了若干理论模型，却由于其中的概念不确切或者表述过于“灵活”，因而并不具备应有的预测能力。正是这些原因，理论模型的构建成为心理学研究的重要任务。

1. 什么是理论

理论可以粗略地定义为一组解释某些现象的相互关联的陈述。陈述数目越少，能解释的现象越多，则理论越有功效。在心理学研究中，理论主要有两种功能。第一，理论起着组织数据的作用，它提供了一种框架，使得能够系统地、有顺序地表示各种数据，从而帮助我们组织研究的结果；第二，理论发挥着预测功能，它使我们得以对其它未曾获得数据的情境作出预测。预测的准确性愈高，理论就愈好。在心理学研究中，许多理论都还是以比较松散的语言方式表述的，因而即使在相同情境下，同一理论也可能作出不同的预测。尽管心理学家正越来越多地通过数学方法和计算机模拟，对理论作出更精确的表述，心理学的理论建设与其它许多更为成熟的学科相比，仍是一项艰巨的富有挑战性的任务。这里，不仅需要大量脚踏实地的心理学研究，而且需要心理学家具有更丰富的创造性和想象力，以及更扎实的方法论基础。

2. 心理学理论构建的类型和方式

在心理学的研究与发展中，形成了四种理论构建的类型，它们都运用了“数据”和“解释”，但各自有不同的侧重。这四种理论构建类型或方式是：归纳式理论、演绎式理论、机能式理论和模型。图2-1表示出这四种理论构建的方式。

(1) 归纳理论

归纳理论以数据为基础，把数据反映的实际关系组织为理论原

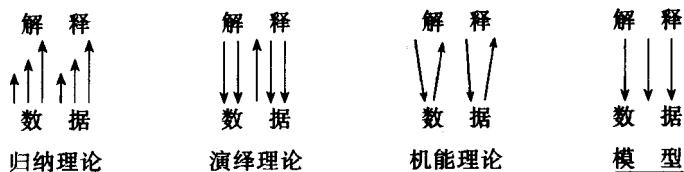


图 2-1 心理学理论构建的四种方式

则。斯金纳（1956）有关强化与学习的研究，就是以这种方式提出操作条件反射的理论的。用斯金纳自己的话来说，“我从来不靠建立一种假说来研究问题。我从来不推演定理，或予它们以实验的检验。”（斯金纳，1956年，第227页）

（2）演绎理论

演绎理论从理论解释出发，寻求数据以检验理论的预测。如果数据与预测的情况不一致，则对理论加以修正，并作出新的预测，直至形成更有力的理论。由于这种比较经典的理论构建方式在物理、化学等学科比较成功，因而颇受心理学家重视。赫尔（Hull, C. L.）有关学习与动机的理论，运用了演绎式理论构建方式。赫尔与斯金纳虽然都是行为主义者，两人的研究方法却完全不同。赫尔强调理论的重要性，常常在先验的基础上假设一种理论，然后用实验证据加以检验，并推演出结论。

（3）机能理论

机能式理论综合了归纳与演绎两种成分，既强调数据，又重视解释。这种理论构建方式更为普遍，比较典型的例子，是费斯汀格（Festinger, L.）有关态度改变的认知失调理论。这个理论认为，当一个人具有两个不一致的认知元素时，就存在着认知失调；失调元素越重要、占比重大，引起的失调也越强烈；人们倾向于对认知或行为作出调整，以减少失调所造成的心理紧张。费斯汀格的这一理论，成为社会心理学中最有影响的理论之一。

（4）模型

模型式理论构建近年来比较流行。特别是在认知心理学中更多采用了这种方式，其中影响很大的是以所谓“黑箱”图说明信息加工理论。模型实际上是一种类比，因此，它能够适应其预测与实际数据之间较大的差异。由于这个原因，我们在图 2-1 中没有标出从数据至解释的任何箭头。心理学理论的类比主要建立在物理、电子学、化学和计算机技术的基础之一，目前最流行的则是计算机类比。模型被用于指导数据的收集并作为进一步研究的框架，因此，模型的进一步发展，将越来越类似于机能理论或演绎理论。模型的另一特点是其范围比较有限，多数模型只与特定的情境联系在一起。模型的局限性在于难以把几个模型结合在一起对一些情境作出解释；滥用模型还会影响理论的系统结构和一致性。

3. 理论的评价

理论的评价并不存在绝对意义上的好坏之分，常常是理论的某一部分不正确而其余方面仍能应用。因此，我们便经常说某理论在相当程度上被数据所支持，保留理论得不到新数据支持的可能性。在某种意义上说，理论并不能被证明，而只能被否定；与理论不一致的数据最终会发展这一理论。我们对于理论的评价则主要为了决定现有理论中最好的理论并进一步发展理论。

在心理学理论的评价中，可以采用以下的标准。

(1) 简洁性

一项理论所包含的陈述越少，简洁性就越高。简洁的理论能以较少的概念解释较多的结果，因而具有较高的功效。

(2) 精确性

一般来说，包含了数学方程或计算机程序的理论，比只运用了松散的语言陈述的理论，精确性更高。心理学的许多理论在这方面比较欠缺。

(3) 可检验

对于心理学理论的评价，需要特别强调可检验性，不能被检验的理论自然就不可能得到发展。乍一看来，这似乎是理论的优点，因为这样就不可能证明理论的不正确。但科学工作者则持相反的看法。如果理论不可能得到检验，也就不可能进行评价，这样的理论对于我们是没有用处用的。

二、理论模型的准确性

理论模型的准确性在很大程度上取决于研究本身的效度。在心理学研究中，十分关心各种心理特征与行为在不同情境条件下的变化，并设法解释和预测心理活动的变化与发展，这些变化又称作“变异”。心理学研究建立的理论能解释的变异量越大，其准确性也越高。可是，心理学研究中常常会遇到各种没有能加以解释的变异，因而影响研究的质量和理论模型的构建。研究表明，产生没能解释的变异的情况主要有以下三种：

(1) 在研究中运用了无关的或关系很小的自变量，因而不能解释因变量的变异；

(2) 在研究中没有考虑较多的、正在起作用的变量，所以没能作出比较准确的理论解释，在这种情况下，需要同时考虑多个变量；

(3) 虽然在研究中正确地选择与变量，却由于对变量的测量不准确而产生了差误变异。差误变异是自变量无法加以解释的。

为了在心理学研究中减少或消除无关因素的干扰，需要尽可能运用随机化程序，并设法同时研究多个变量。此外，还可以通过某些简化假设，使研究集中于较少量的关键变量。实际上，当我们把研究与理论联系起来，并设法构建理论模型时，正是采用了某种简化图式。由于在许多情况下，难以在研究中真正运用随机化程序，所以，简化图式就成为主要的方法。

举例来说，如果我们准备研究大学教育对于学生毕业后的职业

成就的作用 那么除了学生是否完成大学学业这个变量以外 还需要考虑学生的学习能力和社会经济方面的外部条件等变量对大学学习的影响,而且能力与外部条件又会直接影响职业成就的获得。这样,可以选用上述四个主要变量 作出一个尝试性的理论模型 图 2-2)。

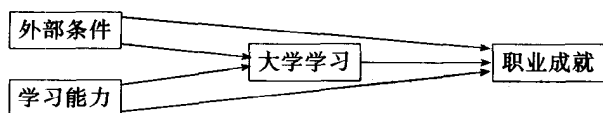


图 2-2 大学教育与职业成就关系的模型

作为研究的任务,是要把大学教育的作用或效应从其它变量的效应中分离出来,这就要考虑许多问题:是否应该控制“外部条件”与“能力”变量?从模型中各变量之间的不同关系,能作出什么推论,模型中的理论构思和时间顺序,是否足以假定单向的因果关系?应作何补充?各变量的测量中可能会出现什么样的误差?各变量间是否还有更复杂的关系?测量上述四种变量对于职业成就的作用,可以采用何种较好的方式?等等。这些问题并非一般常识所能解决,而是需要运用一定的方法和原则。这就是接下去要讨论的理论模型构建的重要基础。

三、因果模型的构建

心理学研究的重要任务之一是发现、解释和预测变量之间的因果关系;理论模型构建的基本方面也是表述这种因果关系,或叫做因果模型。可是,变量之间的因果关系只有在实际活动中才能被证明,因此,准确表述因果关系的前提,是进行严格的心理学研究。这里介绍与讨论因果模型构建过程中需要运用的概念和条件。

1. 运用外源变量

计量经济学研究提出,把变量区分为外源(exogenous)和内源(endogenous)变量(相应于自变量与因变量),有利于因果模型的构

思，这种思想已经被心理学家所接受和采用。通常，可以把变量看成一个一般的线性关系系统，表示为以下的一组方程：

$$\begin{aligned}x_1 &= a_1 + b_{12}x_2 + b_{13}x_3 + \cdots + b_{1k}x_k + e_1 \\x_2 &= a_2 + b_{21}x_1 + b_{23}x_3 + \cdots + b_{2k}x_k + e_2 \\&\vdots \\x_k &= a_k + b_{k1}x_1 + b_{k2}x_2 + \cdots + b_{k,k-1}x_{k-1} + e_k\end{aligned}$$

其中， e_i 表示了那些被省略的变量的作用，这些 e_i 项往往是相互关联的，在模型中 e_i 项一般作为干扰因素。

我们一般从理论上把外源变量看成恒定的因素，只有当它被作为自变量时，才认真地考虑它们的效应。这时，认为任一内源变量的变化是由于外源变量和其余内源变量联合作用的结果。尽管如此，对于理论模型的构建来说，未知的或未能测定的因素仍然过多。因此，就遇到所谓“变量识别”（identification）的问题。解决的方法之一，是假定变量系统中的一部分系数等于零，即假定某些变量并没有发生显著的作用；此外，在建立单向因果模型时，运用某些简单的假设，也能够避免“识别”问题。

2. 一般递归系统

如果变量系统中在时间 t 上被作为内源变量的任何变量都不能同时作为另一内源变量的原因和结果，那么这种系统就称为“一般递归系统”。在递归系统的方程中，至少一半的系数被假定为零。以下是某递归系统的方程组：

$$\begin{aligned}x_1 &= e_1 \\x_2 &= b_{21}x_1 + e_2 \\x_3 &= b_{31}x_1 + b_{32}x_2 + e_3 \\&\vdots \\x_k &= b_{k1}x_1 + b_{k2}x_2 + \cdots + b_{k,k-1}x_{k-1} + e_k\end{aligned}$$

在递归系统中，排除了所有交互式因果关系（即同时被作为原因和后果），例如，变量 x_1 只依赖于外部变量的总和 e_1 ； x_2 只依存于 x_1 ，等等。一般来说，研究中可以把某些因变量作为较早时间点上的自变量或外源变量。可见，一般递归系统更适合于对变量作出因果解释。

3. 因果模型的条件

建立因果模型，需要具备一定的条件，这些条件实际上也是心理学研究中发展各种理论的基础。在构建模型时，应该注意使模型满足这些条件。

（1）规范的理论陈述

因果模型应该运用比较正式的理论陈述，这一条件包括尽可能使理论陈述严密，在提出某一效应时，使之包括与这一效应有关的所有原因或主要的原因；能够表达较复杂因素之间大的因果联系和与之有关的中介变量；提出初步的变量间因果关系的结构图，并对模型的內源变量、外源变量和干扰项作出规定，建立函数方程。图 2-3 为某结构模型图的例子，表示了对作业质量与数量的期望、角色知觉和焦虑状态等变量之间的因果关系。假定这些变量之间的关系是线性和可加的，可以推导出以下函数方程：

$$Y_1 = \beta_{y_1x_1}X_1 + \beta_{y_1x_2}X_2 + d_{y_1}$$

$$Y_2 = \beta_{y_2y_1}Y_1 + d_{y_2}$$

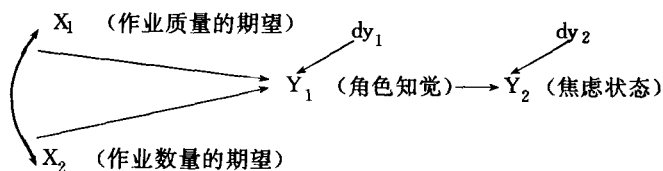


图 2-3 结构模型的形式

（2）因果关系的理论假设

为了建立良好的因果模型，需要明确地假定和说明原因变量是如何影响结果变量的，即原因如何产生效应，包括必要的中介变量及联系。理论假设使较大的因果联系更具体明确了，例如在上一例子中，可以对 $Y_1 \rightarrow Y_2$ 的因果联系作出进一步假设：角色知觉 $Y_1 \rightarrow$ 对失败可能性预期 $\rightarrow$ 对消极后果可能性的预计 $\rightarrow$ 焦虑状态 Y_2 。

(3) 明确的因果顺序和因果方向

这里，因果顺序是指原因与后果的时间顺序。在构建因果模型时，需要具体考虑这一因果顺序，并对此作出假设。同时，还应考虑相应的测量时间。关于因果关系的方向，需要运用一般递归系统的知识。在心理学研究中，有许多非递归性的关系和周期性递归关系。但是，在因果模型的构造中，往往先运用简单的递归模型。

(4) 自抑式函数方程

有关因果模型的理论认为，只有当函数方程是自抑式 (selfcontained) 时，才能代表某种因果关系。所谓自抑式方程，指的是内源变量的所有有关的原因变量都包括在方程中。在上例中， X_1 与 X_2 均不与 d_{y1} 项相关，因此，该方程是自抑式函数方程，以函数关系 $Y_1 = f(X_1, X_2, d_{y1})$ 表示了因果关系。自抑式函数方程的条件是在方程中所明确包含的原因变量与误差项不相关。

(5) 明确的理论边界

所谓理论边界，是指理论所说明的函数关系能够成立的前后关系与情境（被试类型与情境特征等），即说明理论的普遍化程度。当模型的参数和变量均为线性时，如果方程中有关原因和结果的关系并不同时依赖于其它变量，则认为模型满足了边界条件。可是，如果函数关系受到某些缓冲变量或干涉变量（见第三章）的影响，则认为模型边界不清。因为这时候，干涉效应意味着某些变量之间的交互作用，无法明确地解释 $X \rightarrow Y$ 的直接因果联系。

(6) 模型的稳定性

稳定性是理论的基本条件之一，它指的是所假定的因果关系结构在特定的时间间隔中具有一致性或不变性。因此，要求模型参数的测定值在相同时间中保持一致，使结构模型在合理的时间间隔中具有足够的稳定性，从而具有所要求的普遍意义。

构 0 T d (在) T j 0 g 0 0 0 r g 1 0 . 5 6

合度的检验，通常采用 χ^2 检验。在这一阶段，如果检验表明函数方程以外的参数值都没有达到显著水平，即数据与模型参数没有显著差异，那么认为数据与模型具有良好的拟合度。

70 年代以来，一些统计学家运用电子计算机。在传统因素分析的基础上，提出了新的方法——验证型因素分析（confirmatory factor analysis）。根据这种方法，可以从实际研究数据，确定所假定的理论模型中各因素之间的实际参数，寻求数据与模型的最佳拟合。具体的运算是通过称为“LISREL”的专用软件包进行的，分别计算测量模型和结构模型。测量模型用以检验因素指标的信度和与理论构思的关系；结构模型则提供了各因素之间的结构关系或因果关系。因此，验证型因素分析方法为心理学的理论发展和模型的精密化及数学化，提供了有力的工具，成为心理学研究方法发展的又一里程碑。

第二章思考题：

1. 因果关系分析的意义是什么？在心理学研究中，因果关系分析有哪些特点？因果关系验证的关键条件是什么？
2. 理论思路在心理学研究中的重要性何在？心理学研究所采用的理论思路与观点，各有哪些特点？
3. 相互作用心理学有哪些主要观点？为什么说它为心理学研究提供了理论基础？
4. 心理学理论的构建主要有哪几种方式？它们各有何特点？理论评价的标准是什么？
5. 因果模型的构建，应满足哪些条件？如何构建因果模型？

第二编 研究的设计和评价

进行心理学研究，需要有周密的设计和准备。第二编“研究的设计和评价”，着重介绍和讨论各种研究设计的类型、计划与评价问题。研究类型的选择，取决于研究课题的性质与要求；研究设计过程中，需要考虑研究的每一步骤的特点，从而作出全面的计划。研究评价的重要标准，是研究的效度，在研究设计阶段就考虑和注意影响效度的各种因素，这将在很大程度上提高研究的质量。由于研究一般都不可能对研究对象的总体进行，而只能选取一定的样本。因此，研究的取样就成为研究设计的重要一环，它直接会影响研究结果的准确性和可应用性。心理学研究大量地在实验室和现场进行，从研究一开始就需要重视实验室和现场的条件与特点。本编的最后一部分：第六章“实验和准实验”，专门讨论了现场研究的特点，以及心理学研究的最重要类型——准实验和实验室实验的设计。

第三章 心理学研究的设计

第一节 研究的类型

一、自然观察

自然观察是心理学研究中常见的研究类型，它通过对自然发生的活动的系统观察或调查，分析和研究有关的心理特征、心理过程以及主要的影响因素。自然观察有两个显著的特点：一是研究的对象是在自然发生与发展中的现象，并非人为地创设的情景；二是对研究对象不施加任何干预或操纵某些因素，使所进行的观察尽可能不引人注意而影响正常活动。实践表明，自然观察比较适宜于在研究初期进行，从而为进一步的研究打下基础。自然观察有许多变式，其中最常用的是个案研究和抽样调查。

1. 个案研究

个案研究 (case study) 是心理学中历史最长的研究形式之一。学派时期弗洛伊德的心理分析理论，就是对一些临床个案观察和分析的成果。一般来说，个案研究是对于单一案例或有限数目的事例所进行的全面、深入和详尽的考察和研究。个案研究还比较重视研究成果对于样本所属总体的普遍意义。

心理学研究的许多课题都可以采用个案研究的方式，例如，对学校中差生的诊断和辅导、对后进职工的教育与引导、对新型劳动组合方式效果的评价、对工作环境的改造，等等。个案研究是一种

自然观察，它既可以作定性的考察，又可以进行定量的分析。个案研究的主要方法是运用观察法，目前最多采用的是参与观察法，即由观察者或研究者直接参与所观察和研究的活动或事件。本书第八章将详细介绍和讨论这种方法。

2. 抽样调查

个案研究通常只考察少数被试，而且这些人员往往不能代表较大总体的情况。例如，弗洛伊德的理论基本上是以患神经症的维也纳家庭主妇的个案为基础的，因而很难代表各类正常人。这样，常常需要对某一地区人员的较大随机样本进行调查，以便获得更多的信息。抽样调查在心理学的多数领域运用得很少，这主要是由于通过准确的选择，可以从小样本获得对于整个总体具有普遍意义的研究结果，不必采用抽样调查作大样本分析；另一个原因是，抽样调查的结果一般是描述性的，而心理学研究却日益重视解释和预测。抽样调查的关键环节是取样的代表性和调查问卷设计的质量。我们将在以后的章节中讨论这些问题。

3. 对个案研究和抽样调查的评价

个案研究和抽样调查都属于自然观察，特别适用于研究的早期阶段，从而为进一步的比较严密的研究提供依据。社会心理学中最重要的理论之一——弗斯汀格的认知失调论，就是在费斯汀格等有关群体预测的著名个案研究结果的基础上，进行实验研究而建立的。这项个案研究通过对群体活动的参与观察表明，人们在致力于某项行动却发现行动结果与自己的预期不一致时，最容易产生认知失调。

个案研究和抽样调查的一个主要弱点，是这些研究一般是描述性的，不容易作出有关因果关系的推论；个案研究的另一个弱点是难以重复研究，再现原有结果，并且容易受研究者现有的理论假设和其他主观因素的影响，从而出现偏向；此外，个案研究一般取样

比较小，限制了研究结果的可应用性和普遍意义。

近年来，不少心理学研究中采用了一种个案研究的变式，称为“偏差个案分析”，它通过对两个背景相似而结果不同的案例的比较与分析，找出造成差异的原因，较好地克服了个案研究不易作因果关系推论等问题。我们曾在有关计算机系统开发的心理学研究中（1986），运用偏差个案分析法，对背景情况相似而计算机应用效果不同的企业进行了个案研究，总结出促进计算机应用的若干心理学策略。实践表明，进行个案研究，要求研究者具有较高的综合分析能力，善于充分利用所获得的数据与资料，使个案研究成为更有价值的研究手段。

二、相关研究

相关研究是根据与研究对象有关的各种因素之间的相互关系就对象特征和行为作出解释和预测的研究。最常用的研究指标是皮尔逊积矩相关系数（Pearson product-moment correlation coefficient）。此外，等级相关系数、偏相关与复相关系数，也在相关研究中普遍使用。在选用研究指标时，应根据研究的目的，认真考虑研究中变量的性质、数据的特征、各变量之间关系等方面的条件。应特别注意研究中的原始数据是否满足相关分析的基本统计假设。

相关研究的适用范围比较广。在心理学研究中，可以运用相关研究，考察企业中职工的工作技能与生产效率之间的关系，或者分析学生的学习兴趣与学习成绩之间的关系，等等。一般来说，在研究初期多用相关研究，以便发现和了解有关变量之间的基本关系，进而开展更深入、严格的研究。当研究涉及多个因素的复杂关系时，可以采用偏相关或复相关的方法。

相关系数本身并不能说明因果关系，这是在进行相关研究时应特别注意的。例如，许多心理学研究证明，工作满意感和工作成绩

之间存在较高的正相关，可是究竟是工作满意感引起好的成绩，还是由于高成效产生了满意感和积极的情感呢？这就不能光靠相关研究来解决。人们因此进行了实验研究，证明当奖励和成绩的关系较紧密时，高工作成绩往往导致工作满意。可见，实践活动决定了人们的心理状态。

又比如，有一项研究发现，头围与词语回忆的成绩之间的相关为 $r = +0.70$ ，是不是能够说头围大，记忆的词语就多呢？进一步的研究表明头围大并不一定是记忆能力高的原因，一些其它的因素实际上在起作用，其中比较重要的变量是年龄，人们的头围随着年龄的增长而加大，记忆能力也随着年龄的提高而改进。年龄这第三个变量是头围与词语回忆能力之间高相关的原因。可见，其它因素的干涉，使得相关研究的解释更为困难。因此，进行相关研究时，还应注意其他变量的影响和作用，避免被表面的高相关所迷惑。例如，有一项研究，在一家较大的企业中收集了有关某种饮料与胃病发病率的数据，样本中包括操作工人、班组长、车间管理人员等不同工作职务水平的人员。研究结果的相关分析证明，喝饮料和胃病发病率之间有显著相关。因而，研究者认为喝饮料导致了胃病，并建议取消这种饮料。可是，进一步的研究和分析却发现，最初的相关分析并没有考虑到第三个变量——职务水平。研究表明，随着职务水平的提高，喝饮料量也增加，同时，工作的心理紧张程度也提高了。喝饮料和胃病发病率的关系并没有像相关系数所表明的那样高，实际上存在的关系是：职务水平——心理紧张——胃病发病率。这一研究说明了相关研究的局限性，克服这一弱点的办法，是进行实验性研究和多变量分析。图 3-1 表示了“喝饮料——职务水平——胃病率”之间的关系。整个来看，似乎喝饮料的程度与胃病率之间有较高的相关，但实际上却是职务水平在起作用。

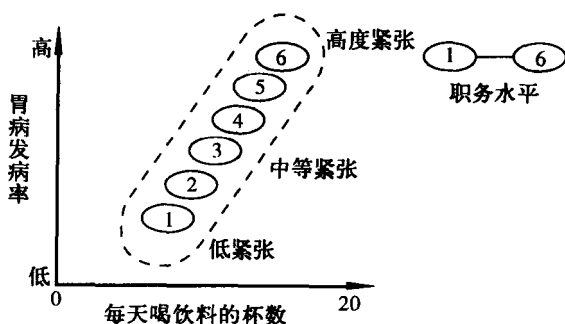


图 3-1 喝饮料——职务水平——胃病率之间的关系

三、实验和准实验

实验和准实验是心理学研究的最重要类型之一。实验是对环境进行系统的操纵，从而观察这种操纵对于行为的效应的研究形式。实验中运用自变量和因变量来说明所操纵的环境变量和所观察的行为，而使不感兴趣的环境变量保持恒定。实验的另一个特点是实验者必须进行两种水平的变量操纵，有时甚至是以操纵与不操纵作为两种水平。因此，只是有因变量和实验者控制的自变量，并不能算是一项实验。

实验的主要优点是能够对研究变量进行严格的控制，从而决定自变量的变化是否引起了因变量的差异，作出有关心理现象的因果解释。实验的基本条件之一，是被研究对象的随机化和实验程序的某种随机化，从而使各组被试的条件尽可能相等，保证实验的周密性和严格性。实验还具有简化各种心理因素之间联系的作用，比其它一些方法和研究形式更为经济、可靠。最常见的实验图式如图 3-2 所示。更严格的实验设计可以采用三个被试组，更好地控制和操纵变量，图 3-3 表示了这类实验的设计。

实验组	RO_1XO_2
对照组	$RO_3 O_4$

R:随机化

$O_1、O_3$:事前测定

$O_2、O_4$:事后测定

X:实验操作

图 3-2 常见的实验设计

实验组	RO_1XO_2
对照组 1	$RO_3 O_4$
对照组 2	$R XO_5$

R:随机化

$O_1、O_3$:事前测定

$O_2、O_4、O_5$:事后测定

X:实验操作

图 3-3 较严格的实验设计

为了加强研究的效度和提高研究结果的普遍意义，心理学家越来越重视现场研究。60年代以来，在心理学的各个领域，都日益增多地采用了“准实验”这一研究类型。准实验是在现场条件下经常运用的研究设计，它与实验的主要区别在于，准实验中没有运用随机化程序进行被试选择和实验处理。准实验中不必拆散原有的学习或工作小组，而是直接以原小组作为研究中的实验组或对照组。最常用的准实验设计有“不等同对照组设计”和“间歇时间序列设计”等。在第六章中，将就实验与准实验设计进行详细讨论。

四、发展研究

发展研究是从心理活动或事件的过程研究它们的发展变化、影响事件发展的各种因素及其相互关系。发展研究多半是描述性的，着重于说明和解释心理活动与事件发展及主要因素之间的关系和条件。当我们需要考察某种教学方法或管理措施的效果时，就可以运用发展研究。例如，在学校中进行发展研究，分析某种教学方案的实施中，学生学习技能与知识的掌握和运用过程；还可以在企业的考察现有管理制度对于职工心理特征和积极性的影响，分析职工思想状况的变化，从而作出有意义的解释，提出相应的改进措施，等等。

发展研究既涉及到描述一定情境中心理、社会和组织等各种因素的相互关系，也试图说明这种关系随时间发展所出现的变化。由于发展研究以发展的观点考察对象，因而更容易发现和把握心理活动的动态特点和发展规律。近年来，发展研究中运用了高级统计分析技术，进一步提高了研究的效力。

1. 发展研究的类型

发展研究主要有三种类型：纵向研究、横切面研究和趋势研究（又称为预测研究）。

（1）纵向研究 longitudinal study

纵向研究是从时间纵深的发展过程来考察研究对象。它是在一段时间中进行的，短至几周，长到几年、几十年。纵向研究正为心理学界所日益重视，无论是教育、儿童还是管理心理学的研究中，都越来越多地采用纵向研究的设计。例如，为了考察青工劳动技能的变化情况，选取一批青工为对象，在一年时间内每月进行一次观察和测定，并通过综合分析，找出影响技能发展的主要因素。

进行纵向研究，应特别注意使研究对象尽可能保持原有样本，避免更换被试。纵向研究的图解如图 3-4。对于研究结果主要采用简单回归和多元回归等统计分析方法。

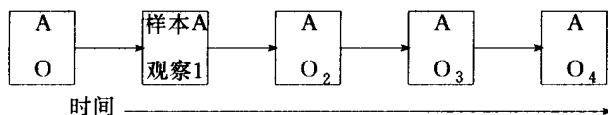


图 3-4 纵向研究设计

（2）横切面研究 (cross-sectional study)

横切面研究是指在较短时间间距中对不同的对象样本进行观察和测定，从而对各类对象作出总体分析的研究类型。这种研究，便于随时了解人们的学习或工作态度以及其他心理特征。研究的特点是按照职业、职务、性别、年龄、文化程度等因素抽取不同的对象

样本，并根据研究的目的和要求，对各样本作出相同的观察或测量，然后对结果进行分析和比较。这种研究设计的图解如图 3-5。

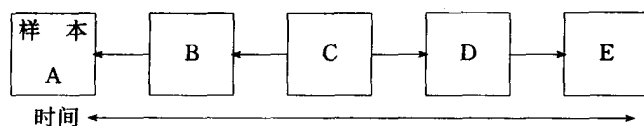


图 3-5 横切面研究设计

横切面研究在教育心理、儿童心理和管理心理研究中都常常被采用，可以在较短时间里有效地了解和把握学校各类学生的心理特点、不同年龄阶段儿童的认知和语言能力、或企业各组织层次中各类职工的工作态度、需要和技能等特征。横切面研究的结果，一般采用方差分析（ANOVA）等方法进行统计分析。

（3）趋势研究（trend study）

趋势研究又称预测研究，它是在有关对象的先前情况和资料的基础上，建立某种预测模型，并在进一步的多次观察与测量中作出检验。在比较严格的趋势研究中，还需要证实预测模型与实际发展情况（数据）的配合度。趋势研究在儿童发展心理学研究、人事预测与人力资源计划研究等方面，都是很有价值的研究设计。

趋势研究的模式如图 3-6 所示。趋势研究的时间跨度应根据研究目的而定，在较大时间跨度的条件下，会介入许多无关因素，影响和干扰研究的结果。因此，为了提高研究的准确性，趋势研究不宜采用过大的时间跨度。

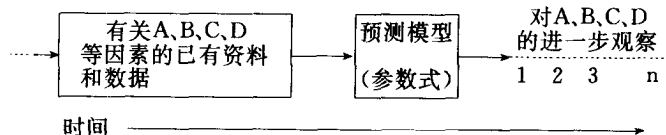


图 3-6 趋势研究的模式

趋势研究主要有这样一些步骤：收集与分析有关对象特点的已有数据和资料；通过理论构思，建立有关对象发展趋势的预测模型，

并确定模型中各个因素的测量指标；在对象发展过程中对各有关变量或因素作进一步观察与测定，检验预测模型。

2. 对发展研究的评价

发展研究的三种主要类型各有其优缺点。

纵向研究在一段时间里对研究对象进行了周密考察，因而适合于确定引起某些特征变化的因素，建立某种因果关系。纵向研究往往不是单一测定或观察的结果，它的结果比较可靠。纵向研究的缺点是比较耗费时间和研究资金，被试样本不易保持稳定不变，研究设计不够灵活。此外，还容易在研究过程中受到其它因素的影响。

横切面研究可以同时研究较大样本，研究效率比较高，成本较低，费时也较少。但是，横切面研究是对不同对象在某些时间点或同一时间的考察，因此较难确定因果关系，它的取样程序也较为复杂。

趋势研究的实际意义大，但要进行这种研究，需要相当水平的研究方法方面的知识，各阶段的要求比较高。这种研究特别强调可靠的基础数据和较全面的理论构思。趋势研究的结果容易受某些未测因素的影响。

近年来，人们主张采用一种把纵向研究与横切面研究结合在一起的交迭式研究设计，即在纵向研究图式中分段进行横切面研究。这种发展研究称为聚合式研究，它既克服了纵向研究的缺陷，又保持了横切面研究的许多长处。例如，图 3-7 为聚合式研究的一种图式，可以在 3 年时间内完成对 3—14 岁儿童心理发展的研究，而如果按单纯纵向研究，至少需要 6 年时间；同时，这种设计还可以在每年获得四个年龄组的横切面数据，从而在较短时间里取得成果。

聚合式研究固然比较复杂，测量程序的要求也比较高，但只要条件允许，都应采取这类设计。许多研究证明，聚合式研究设计在有关心理特征发展、个体差异、特别是智力、能力和个性的研究中，

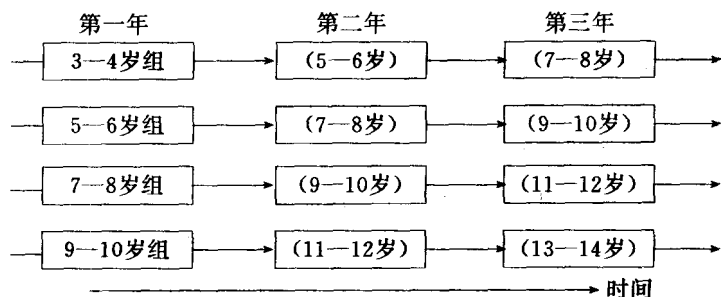


图 3-7 聚合式研究的图示

具有很高的应用价值，已经在发展心理、教育心理、管理与劳动人事心理和普通心理等领域的研究中运用。

第二节 研究的步骤和计划

心理学研究都包含比较严格的步骤，需要遵循比较周密的计划，这样才能达到预期的研究目的。一般来说，心理学研究包括以下步骤：课题与变量的选择，研究设计的确定，研究方法和程序的制定，研究的实施，研究结果的分析，研究结果的总结与报告。

一、研究课题与变量的选择

1. 课题的选择

研究课题的选择是研究的第一步，需要认真确定研究的目的和问题。这里所谓“问题”，是指某种矛盾或令人困惑的情境，常常可以帮助决定随后的研究的方向；研究的“目的”，则是要运用一定的方法去解决那些问题。在心理学研究方法中，我们把研究的问题又称为研究课题。

(1) 课题的类型

研究课题的类型很多，它可以是验证某种理论，或者为已经建立的理论中未曾回答的问题寻找实际依据；可以是澄清某些矛盾的

研究结果，矛盾的结果往往是由于研究设计中的问题造成的，因而可以由此入手；与此有关的课题类型，是更正错误的研究方法，针对研究的测量、分析与统计中的差错，进行新的研究；更多的课题，是解决实际生活或工作中的问题。“实际问题”并不意味着研究可以缺乏理论基础而是指问题更为直接和紧迫，目前我国心理学研究更多面临的也正是这类课题。

（2）课题选择的标准

选择课题时应该充分了解该研究领域以往研究的情况和趋势，还应考虑研究者的能力和专业知识、研究者对课题的兴趣、研究的独特性和新颖性，以及研究的费用和代价。实际上，人们常常选择最迫切需要解决的、具有理论意义和实际意义的课题。

判断研究课题的优劣，有几条标准：所研究的问题应该涉及两个或两个以上变量之间的关系；所研究的问题应该尽可能明确具体；具有通过实际研究解决问题的可能性。

在提出研究课题时，还必须考虑研究的假设。假设比问题进了一步，它提出了回答和解决问题的尝试性的策略。

2. 概念和构思

概念是把不同来源的几种成分结合为单一观念的一种看法，例如圆形是人们观察到各种圆形物体而形成的概念，从球、橘子、轴承、轮子等物体概括出圆的特征。心理学中的满意、成就、动机、认同等等都属于概念。

构思则是由于特定科学研究目的而有意识提出或采用的概念。因此，构思是研究者想象中的不能独立地观察到的行为维度。例如，管理心理学中“公平”这一构思，就需要从一些问题推论而来。

构思是科学理论发展的基本成分，理论都涉及构思，而不是对具体的、可观察变量的陈述。在心理学研究中，同样一个词，如动机、满意感、成就、绩效等，都可以作为概念、构思或变量。例如

“动机”可以是概念，指有目的的行为，说明指向一定目标的内驱力；它也可以作为构思，成为动机理论的一个元素；它又可以作为变量，以“动机分数”来测量。

3. 常见的变量

(1) 自变量、中介变量和因变量

自变量是另一变量变化的原因，因变量是自变量作用的结果，这两种变量一般可以直接或者间接地观察。中介变量是与有机体有关的不可观察的过程或状态，用以帮助解释自变量与因变量之间的关系，也常作为假设的构思，从可观察到的测定作出推论。例如，奖励办法（自变量）引起学习成绩（因变量）的变化，可是我们又认为，有一种中介变量能解释学习成绩变化的原因，这就是学习动机。新的奖励方法增强了学习动机，从而提高了学习成绩。如图 3-8 所示。

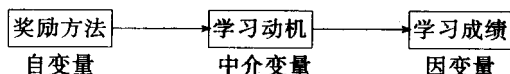


图 3-8 自变量-中介变量-因变量

自变量是由实验者操纵的变量，上例中改变奖励方式（自变量），观察与记录学习成绩（因变量）的变化。当自变量的变化引起了因变量的变化，就认为因变量处于自变量的控制之下。如果自变量无法改变或控制因变量，我们称之为“虚无结果”（null results）。对于虚无结果可作几种解释；一种可能是自变量选择有误，这时出现虚无结果是正常的；另一种可能是实验者没能有效地操纵自变量，因而还不足以显示出自变量的效果；还有一种可能是因变量有缺陷，不能反映出自变量的作用。常见的问题是“底层效应”（floor effect）和“顶层效应”（ceiling effect）。底层效应是指被试在因变量上的测量得分普遍很低，或者在某项测验中都没有能通过；顶层效应则是指被试在测量中得分都相当高，或者某项测验的结果几乎完

美无缺。这两种情况都无法反映自变量变化所产生的效应，也说明因变量的测定缺乏应有的辨别力和灵敏度。

还有一类变量称作“控制变量”（control variable）这是指实验期间保持恒定的某些潜在的自变量。对于任何实验，这类变量的数目比较多，使某些外源变量保持恒定，是最直接的实验控制方法。在心理学实验中，往往包含多个自变量或因变量，需要按照研究目的和实验设计原则，确定其中的控制变量。

应当指出的是，一项研究中的自变量可以是另一研究中的因变量；在同一研究中，相同的变量可以在一种分析中作为自变量而在另一分析中作为因变量。图 3-9 表示出这种关系。在图 3-9A 中，工作满意度在研究 1 里被作为因变量，而在研究 2 中却当成自变量。在图 3-9B 中，研究者在分析“工作努力”与“工作绩效”的关系时，把“工作努力”作为自变量，而以“工作绩效”为因变量。在同一项研究中，在进行第二层次的分析时，则把“工作绩效”看成了自变量。分析的角度不同，变量的性质也会发生变化。

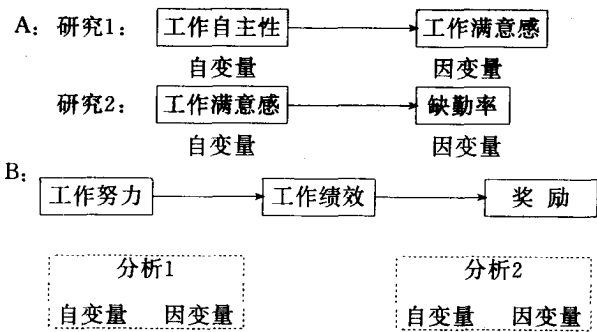


图 3-9 不同研究中的变量关系

(2) 干涉变量

当一个变量的系统变化可能改变其它两个变量之间的关系时，这个变量就称为干涉变量。在这种情况下，干涉变量在很大程度上

决定了那两个变量之间的关系。例如，工作知识对于努力与绩效之间的关系来说，就是一个干涉变量。在工作知识丰富的条件下，努力容易导致高绩效；相反，要是缺乏应有的工作知识，工作再努力也难以取得好成绩。在心理学研究中，干涉变量的作用是很值得重视的。在确定研究变量时，应该充分考虑到那些可能影响自变量与因变量之间关系的主要干涉变量。

4. 变量的操作定义

任何讨论、对话或者交流，都需要对所运用的术语有一致的定义。在科学研究中，更需要有严格的技术性定义，最普遍的办法是对所研究的构思作出操作定义（operational definition）。变量的操作定义是有关构思的可观察指标的具体陈述。在心理学研究设计中，对变量作出操作定义，是十分重要的步骤。这就是把心理学的构思或概念（如动机、情绪、焦虑、参与等）定义为可测定、可重复的具体指标。例如，以愿意参加某项活动的程度作为动机的操作定义；以皮电反应作为焦虑的操作定义，等等。研究中变量操作定义的质量，将直接影响研究的可重复性、结果的可检验性以及研究的普遍意义。作出操作定义包括两个程序：（1）确定与心理学构思有关的具体测量指标；（2）在实际研究或实验中产生并观察该变量指标的变化，进行测定与检验。

二、确定研究设计

确定研究设计这一阶段，主要考虑和完成三个方面的工作：按照研究设计的目的和控制原则，确定具体的研究问题和控制变异的方法；根据研究设计标准，检验和决定研究方案；制定具体研究计划。

1. 研究设计的目的和控制原则

研究设计的基本目的有两个方面：一是回答所研究的问题：二

是控制研究中的变异量。前者是所有研究活动共同的目的，后者则是指对研究中的实验变异、外源变异和误差变异进行控制。通过研究设计，确定收集和分析研究数据的方式。因此，研究设计实际上具有一定的控制机制，这种控制的统计学原则是：

(1) 使系统变异最大，主要指研究假设中变量的变异，通常采用选取较异质样本、提高测量的辨别力和加强统计功效等办法；

(2) 控制外源的系统变异，包括对研究结果有影响而研究者又不感兴趣的变量的变异，可以在研究设计中对各种变量作出明确规定，特别是自变量、因变量、控制变量和干涉变量，注意恒定可能引起外源变异的因素与条件；

(3) 使误差变异最小，即把随机变异和测量误差减少到最低限度，注意排除影响研究效度的因素（见第五章），采用具有高信度与高效度的测量工具和量表，在研究取样与实验安排中尽可能运用随机化程序。

2. 研究设计的标准

在进行研究设计时，应注意使设计方案达到下列标准：

(1) 研究设计回答了研究问题并准确地检验假设；

(2) 研究设计通过不同程度的随机化，较严格地控制自变量；

(3) 使研究具有一定的内部效度和外部效度。关于研究的内部效度和外部效度，是心理学研究方法在最近二三十年中最显著的重要进展之一，第五章将专门讨论研究的效度。

3. 研究的计划

制定研究的计划，这是研究设计的重要一环，也是研究工作顺利进行的保证。在制定研究计划时，不仅需要认真考虑研究的理论基础、具体课题和研究方法，而且应该充分预计可能遇到的问题或困难。在研究计划的制定过程中，还应广泛听取同行的意见与建议，交流研究的构思，讨论研究的具体方法，从而完善整个研究计划。

研究的计划应包括以下三个方面：

(1) 研究目的和理论框架：说明研究的具体目的和问题，提出研究的理论框架和假设（国内或国外的理论），指出有关理论对于本研究的意义；

(2) 以往研究的考察：说明与该课题有关的早先的知识与经验，总结以往的研究和尚待解决的问题，使本研究建立在以往研究的基础上；

(3) 研究的设计与方法：指出研究的设计，作出主要变量相互关系的图解，确定研究总体及取样方法，说明收集数据（测量）和统计分析的方法。

三、研究的方法和过程

1. 研究的方法

研究的方法是研究设计的组成部分，在确定了研究的目的、理论框架和有待解决的问题以后，就需要慎重地选择相应的研究方法。在心理学研究中，切忌忽视研究的实际需要而盲目追求“高、精、尖”研究方法的倾向。为用方法而使用某些高等统计或新型实验设计，只能使研究偏离正确的方向。在确定研究方法时，需要考虑以下四个方面的方法：

(1) 研究被试的取样方法：明确研究的总体、样本的大小、取样的手段等；

(2) 实验处理的方法：自变量的操纵和控制方式，实验处理的实施、仪器的配备和程序的安排等；

(3) 收集数据的方法：确定测量指标及其测定方法，设计相应的量表，规定数据的记录方式等；

(4) 数据的统计分析方法：根据数据的特征，选择研究数据的统计分析方法和检验方式。

2. 研究的过程

在研究过程中应注意按照研究设计行事，并随时记录研究中发生的情况。在有条件的情况下，应尽可能在实施正式研究之前进行“可行性研究”。在小范围内测试指标和检验研究设计，进而改进研究方案。此外，在研究过程中，可以根据实际研究情况和条件，及时修正研究的设计。但是一般应在研究之前充分考虑实验和现场研究中的各种可能发生的情况和条件，避免不必要的失误。

四、研究结果的分析和总结

在收集到各种研究数据以后，就需要对结果进行统计分析。研究结果的分析应该与研究的理论构思联系在一起，针对研究目的，作出适当的定性或定量分析。心理学研究中常常涉及众多的变量与因素，包含了十分复杂的关系，因此，多元统计分析成为心理学研究的重要工具。多元分析方法很多，心理学研究中常用的有：多元回归分析、因素分析、判别分析、典型相关分析、聚类分析和复方差分析，等等。多元分析方法也是心理学研究方法的重要组成部分，本书第四编中将作较系统的介绍和讨论。

通过统计分析并确定变量之间的关系以后，就可以作出研究的结论。在作出结论的过程中，应该注意以下问题。

1. 单一研究的局限性

单一研究往往不能决定研究假设的真伪，因为单一研究在研究方法、策略和条件等方面都会有一定的局限性，需要有多重研究的印证，在不同条件下运用多种方法，检验变量之间的关系。

2. 作结论时的决策差误

当我们从研究样本的数据推论总体状态时，会出现统计决策差误。常见的差误有两种类型。

(1) I类差误：从样本的统计结果推论诸变量之间存在某种关

系而实际上没有关系。

(2) I类误差：从样本推论变量间没有关系而实际上存在关系。因此，在作出研究结论时应注意避免上述决策性误差。

3. 研究的结果与理论的发展

理论可以说是一组相互联系的构思、概念、定义和命题，它通过具体说明变量之间的关系来表达对实际现象的系统看法，从而达到解释和预测现象的目的。因而，心理学研究的结果应该是对某种理论命题的检验，除了解决或解释实际问题以外，还应特别重视理论上的建树，促进心理科学的发展。

4. 有待进一步研究的问题

在对研究结果进行总结时，还需要发现和判别所进行研究的不足之处或问题，特别应注意在研究结果的基础上，提出进一步研究的问题。对于没有显著差异的结果或“虚无结果”，应防止置之高阁的“存档问题”(filing problem)，因为从虚无结果中往往能够发现研究中的缺陷和有待研究的新问题，甚至可能总结出重要的新发现，这在心理学研究的历史上是常有记载的。

第三节 研究的元分析

一、元分析的发展和特点

近年来，心理学研究中日益重视研究的评价，以此作为以往研究的总结和进一步研究的基础。在心理学研究方法的体系中，出现了“研究评价”(research evaluation)这一新领域。研究评价的重要方法之一，就是格拉斯(Glass, G. V.)1976年提出的“元分析”(meta-analysis)。元分析是对已有研究结果的总体分析，也有人称为“总分析”。元分析运用了测量和统计分析技术对分别进行了一些

实验或研究进行量化的总结，找出一组相同课题研究的结果所反映的共同效应。因此，元分析是对原有分析的进一步分析，它已成为总结和评价研究的有效手段，被认为是研究方法方面的重要革新。

1. 元分析的发展

尽管人们早就有对已有研究结果进行综合分析的思想，但是多半是进行一些定性的总结或归纳，缺乏比较客观和系统的方法与理论。从 50 年代开始，研究结果的分析中逐步采用了越来越严格的定量分析手段。罗森塔尔（Rosenthal, R.）1976 年对几百项有关实验者期望效应的研究的结果进行了综合分析。他所用的方法，与格拉斯（1976）提出的元分析方法非常相似。自从元分析正式提出以来，在教育心理、儿童心理、临床心理、社会心理、管理心理、人事心理和普通心理等各领域得到了越来越广泛的应用，促进了各领域的心理学研究。

然而，应该看到，元分析目前还处于尝试阶段，在应用中也有了一定的局限性。例如，在把多种不同结果综合为单一研究时的倾向性、收集研究文献时的倾向性、选择研究时采用的标准、在分析中只讲数量而忽视研究的不同性质与条件，等等，这些都会造成误估，影响元分析的效果。

2. 元分析的特点

元分析是对许多实际研究的综合性结果的统计分析，它主要有三方面的特点。

（1）元分析是一种定量分析

元分析是一种定量分析手段，它运用一些统计方法，从大量的研究数据中组织和选取出相似的信息，进而综合出单从个别研究不能获得的效应。

（2）元分析是一种全面的评价

元分析的重要特点之一，是并不因为研究的质量而排除于评价

之外。特别是对于那些由于研究的设计、测量和实施方面的问题而未被发表的研究，都可以通过元分析而得到全面评价，包括分析其方法上的弱点，考察这些弱点与研究结果的关系。

(3) 元分析寻求普遍性的结论

元分析的又一特点是，它寻求某种概括性的、简化的结论。元分析有时被误解为旨在寻求“平均效应”，而实际上，元分析的必要条件是对研究的特征和结果进行分析，包括研究的问题、假设的提出、变量的定义与测量、取样和数据分析等。通过元分析，比较不同的研究，概括出普遍性的结论。

二、元分析的步骤

元分析是一项细致的工作，它主要包含四个基本步骤：对以往研究文献的检索、对研究的分类与编码、研究结果的测定、分析与评价效果。图 3-10 表示了这四个步骤。

1. 以往研究的检索

元分析的第一步，是对相同课题的以往研究进行文献检索，找出供分析用的各项研究。在进行检索时，应注意检索的可靠性（即信度）在具备计算机文献数据库时，应在不同数据库之间作出比较，避免“检索偏向”。在国内尚无研究文献数据库

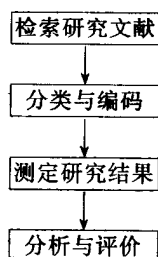


图 3-10 元分析的步骤

的情况下，可以从“心理学文摘”（Psychological Abstract）以及各类心理学杂志中进行检索。研究的检索还应防止“发表偏向”，即由于发表机会的不均等或者发表内容的倾向而造成研究文献中的偏向。大量的研究出现在心理学学位论文中，而论文的发表往往要经历一段时间，甚至多年不能发表。因此，在以往研究的检索中，需要注重对各类学位论文研究的查寻，使所检索到的研究具有足够的

代表性。

2. 研究的分类和编码

在检索工作完成以后，就需要对各项研究进行分类和编码。研究的分类可以按照研究的特点，并考虑到研究的信度和效度、所研究问题的特征以及方法论方面的特征。特别是需要注意研究设计中变量的类型、被试特点、样本大小、随机化程度等。研究的编码可以采用两种方式，一种称为“方法编码”，以发表日期、设计单盲还是双盲（即实验者与被试都不知道研究设计的意图）、研究的效度、测量指标的特征等，对研究进行编码；另一种称为“内容编码”，这是以实验者职业、被试特征、实验处理类型以及结果测量类型等作出编码。例如，同样是运用了心理疗法，需要分清疗法的类型，是催眠疗法、认知疗法、心理动力疗法还是认知行为疗法；对心理治疗结果的评估，是采用了个性评定、紧张测定，还是运用了自尊测量或是心理顺应评级。对不同的疗法和测量方法，都应作出不同的编码。

3. 研究结果的测定

元分析的关键一步，是对研究的结果进行统计显著水平的检验和效果量（effect size）的测定。元分析中最经常采用的效果量计算公式为：

$$\text{效果量} = \frac{\text{实验组的因变量平均值} - \text{对照组的因变量平均值}}{\text{对照组的方差}}$$

效果量又用 Δ 表示，在存在多个因变量时，需要计算出平均效果量。在计算效果量时，标准差的选择是十分关键的一环。当实验组与对照组的标准差接近或相差不大时，可选用对照组的标准差；如果上述两个标准差相差十分悬殊时，应采用两组标准差的平均值；当同时存在多个实验组时，应选用对照组的标准差。

4. 分析与评价

对效果量的进一步分析与评价，常常可以采用方差分析、非参

数统计等方法，也可以运用最小乘方法等。对效果量的解释，不能直接作出大、中、小的比较或评价，而需要以正态分布为基础作出相对的解释。例如，某类研究分别对儿童“活动过度”问题进行药物治疗试验，并就某种新药的效果作出评估。元分析求得效果量为 $-1.00\sigma_x$ ，对这一结果，可以用图 3-11 中的分布作出解释与评估。以服新药组儿童与服用安慰剂的儿童进行比较，认为在服用安慰剂的儿童中只有 16% 的人（阴影部分）比服新药的儿童的一般效果更少活动过度。换言之，绝大部分服安慰剂的儿童都比服新药的儿童的一般水平更为活动过度，从而表明新药具有相当的实际效果。

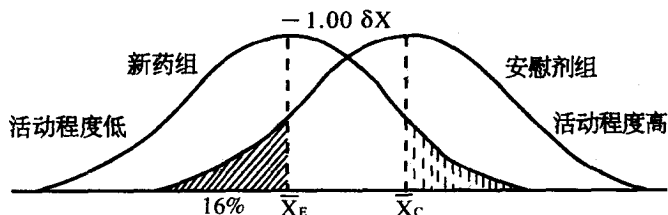


图 3-11 对新药组效果的评价

对于图 3-11 的结果，还可作出另一种解释，即认为新药组中只有 16% 的儿童（带点部分）比安慰剂组儿童的一般情况表现得更为活动过度，大部分的儿童服药后减轻了活动过度的症状。

三、元分析的应用

元分析有许多变式，各有其特点。下面结合元分析的应用，介绍其中两种方法。

1. 有关心理疗法效应的元分析

施莱辛格 (Schlesinger, H. J.) 等 1978 年对 13 项心理疗法研究进行了元分析，检验了心理疗法对于气喘病的效应。他们总结了以下一些特征：作者姓名、疗法类型、被试平均年龄、治疗时间、对

对照组的性质、治疗和效果测定之间的时间间隔和因变量的性质，等等。这一元分析采用了最常用的计算方法，即：

$$\text{效果量} = \frac{\text{心理治疗组的因变量均值} - \text{对照组均值}}{\text{对照组的标准差}}$$

$$\left(\Delta = \frac{\bar{X}_{\text{实验组}} - \bar{X}_{\text{对照组}}}{S_{\text{对照组}}} \right)$$

这一元分析表明，在 13 项研究共 22 个因变量上，参加心理治疗的组与对照组相比。具有 $.86\sigma_x$ 的平均效果量（心理治疗的被试高于对照组均值的 $.86$ 标准差）。以外，研究者分别按疗法类型、病人年龄和治疗时间等计算了效果量，获得了进一步的有关不同心理疗法的效应。

元分析结果证明，接受治疗的患者中大约 80% 的人发病情况优于未经治疗患者的一般发病情况，心理治疗对气喘病具有较明显的疗效。

2. 有关学校班级大小和学习成绩之间关系的元分析

1980 年，格拉斯与史密斯（Smith, M. L.）一起对 1900 年至 1979 年间的 77 项有关班级规模与学习成绩关系的研究进行了元分析。他们收集了诸如学习科目、教学时间、被研究的学生人数、教师人数、班级大小、实验控制类型、成绩测定等资料。他们把班级大致分成较大和较小规模的一些对子，并使各对的成绩之间的平均差异标准化：

$$\Delta_{S-L} = \frac{\bar{X}_S - \bar{X}_L}{\hat{\sigma}}$$

较小（S）班级与较大（L）班级相比的成绩均值之差与“班级内”标准差之比为总效果量。这项元分析是对十几个国家在 70 余年间对 900,000 多学生的 77 个研究的各种结果作出的。元分析获得 725 个 Δ ，证明较小班级的成绩比较大班级的成绩高出 $.1$ 的（成绩）标准差，60% 的效果量为正值，说明小班级成绩略优于大班级。

为了进一步明确班级大小与成就的关系，还必须考虑效果量的大小和不同班级条件下效果量的变异程度。这一元分析中，班级规模和学习成绩的关系被假定为呈对数曲线。元分析的总的结论是，当其它条件均等的情况下，班级越小，学习成绩越高。这一元分析中，还按照研究质量，把所分析的研究分为“严格控制”和“不良控制”两类，并总结出这两类研究中班级规模与学习成绩的分等级的关系。图 3-12 表明了这一元分析的结果。

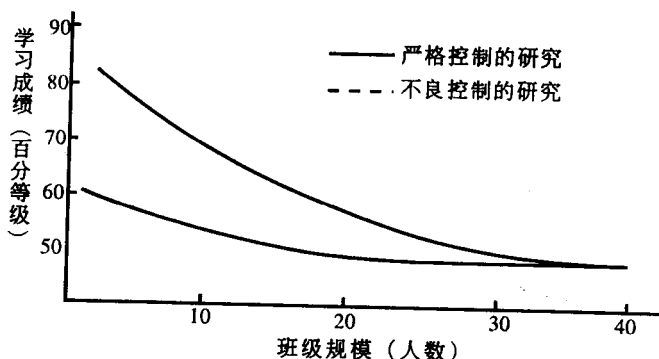


图 3-12 不同质量的研究中班级规模与成绩的关系

以上有关元分析应用的实例，说明了元分析的基本方法和应用价值。目前，在心理学各个分支的研究中，元分析越来越普遍，分析方法和手段也越来越精细和完善。

第三章思考题：

1. 发展研究的三种类型各有哪些特点？举例说明发展研究的实际意义。
2. 在进行相关研究时，应注意哪些问题？实验具有哪些基本成分和功能？

3. 心理学研究设计包括哪些主要步骤？完整的研究计划应包括哪些方面的内容？
4. 在研究结果的分析与总结中，应注意哪些问题？如何看待“虚无结果”？
5. 元分析有何特点？它包含哪些基本步骤？
6. 举例说明元分析结果的解释和评价。元分析对于心理学研究有何意义？进行元分析应注意哪些问题？

第四章 研究的取样

第一节 取样的理论

一、取样在研究中的重要性

心理学研究都是针对特定总体进行的，例如，研究儿童的心理发展过程、职工的工作积极性、教师的个性品质，等等。由于我们不可能对整个儿童、职工和教师的总体进行研究和调查，因此，研究的取样问题就成为心理学研究的关键环节，即确定研究对象的总体并决定如何选取具有足够代表性的样本。

在取样问题上，人们的意见并不是很一致的，总会认为样本越大越好，最好能以总体进行研究，因此，在例如人口调查的工作中就宁愿耗巨资、动员大批人员、花费很大时间作“全员调查”。而实际上，只要运用取样理论，抽取有充分代表性的小样本，即可取得反映总体特征（如年龄、性别、职业、文化程度等的分布特征）的可靠资料与数据。在心理学许多领域的研究中，也存在这类情况。其实，早在本世纪初，统计学家们就曾经有过热烈的争论，讨论究竟能不能用少于总体总数的样本来代替总体进行研究。以后，人们在事实面前逐步接受了取样的思想，并提出了一系列适合于各种实际情境的取样方法。实践证明，取样程序不但能节省人力与物力，而且能显著地提高研究的效率。

事实上，取样还能使研究的结果更为准确、有效和可靠。在

“ 全员调查 ” 中，由于大量人员参与调查活动和大规模的数据处理，操作性的差误和数据加工方面的差误急剧增多，从而使研究的准确性与可靠性明显降低。这类差误又称为“ 非取样差误 ”（nonsampling error）。在选取代表性样本进行研究时，在很大程度上排除了可能造成非取样差误的因素，并有可能对样本作出更为细致的测试与分析，提高研究质量。

在许多情况下，人们希望只对一部分样本（人员、产品等）进行测试，因为测试难免会造成损耗，还会使被测对象受到“ 污染 ”，不利于作进一步的研究。因此，取样还能起到减少损耗、保护总体的作用。在心理学研究中，特别需要防止测量或实验处理的污染。智力测验、态度调查或是新异的实验处理等，都可能使被试产生“ 敏感反应 ”，影响进一步的测试。解决的办法之一就是选择若干个相似样本，分别接受测量或处理。可见，学习和掌握取样的理论与方法，具有很重要的实际意义。

一项研究的设计，包含着一系列相互联系的方面：研究的理论框架、收集数据的方法、加工数据的途径以及样本设计等等。因此，取样是科学研究设计的重要部分，它直接关系到研究的效度，特别是外部效度。以缺乏代表性的样本所进行的研究，其结果的可应用性是十分有限的。不少心理学理论或研究结果之所以至今颇有争论而且应用甚少，其中一个原因，是由于与这些理论有关的研究在取样方面存在严重的缺陷，使这些研究难以重复或验证。由此可见，良好的取样在研究中有重要意义。这里涉及到几方面的问题，所研究的是什么总体？需要多大的样本？样本的主要特点是什么？如何选取这种样本？如何使样本具有足够的代表性？如何减少取样的误差？等等。这些都是本章所要讨论和解决的主要问题。

二、总体和样本的特征

1. 总体与样本

总体 (population) 是指具有某些共同的、可观察特征的测量值的完整集合，它也可以是个体或者某种客体的总合。在心理学研究中，总体是研究者感兴趣的研究对象的总合。例如，某实验心理研究中被要求记忆某字母列表的所有人；某儿童心理研究涉及的所有学龄弱智儿童的学习成绩；某社会心理研究感兴趣的所有具有选举权的公民，等等。在研究中，我们往往根据研究的兴趣和目的规定研究的总体，其特征和范围也随目的和要求的变化有所不同。可是，在实际研究中，无法对整个总体进行研究，而必须从总体中选出一些样本作为研究的具体对象。样本 (sample) 是总体的子集，它是从总体选取的有限数目的成分或元素，用以代表总体。我们在研究中以样本作为基础，通过统计推论，作出有关总体的结论。

2. 样本的代表性

样本的基本条件是能够代表它所属的总体。所谓代表性，并不意味着样本是其总体的缩影，不要求样本具备总体的所有特征，而是在统计意义上获得的。因此，只能在经过透彻的研究以后，在有限总体条件下达到样本的代表性。常用的方法之一是通过随机取样，例如用随机数表方式从总体选取样本，使之具有代表性。在心理学研究中必须明确地规定所研究总体的基本特征，当样本具备相似于总体特征的特点时，就认为获得了“代表性样本”。通常，我们只能把研究结果的普遍意义限于所取样的那个总体，并且在一定意义上把结果推广到具有相似特征的其它总体，这是在从事心理学研究时应充分注意的问题。

在进行研究时，可以通过可行性研究对样本的代表性进行检验，可以抽选额外的样本，与原样本的特点和结果进行比较；也可以在

样本中加入额外元素或成分，检查其统计测定值的变化；还可以随机选取若干样本，看每次是否都有相似的结果。

三、取样的置换性和随机化

1. 取样的置换性

取样置换 (sampling replacement) 是取样过程中常用的概念。所谓取样置换，是指取样过程中已取样的样本成分是否返回总体，继续参加取样，这在随机取样中特别重要。取样置换可以分为两类：置换取样和非置换取样。当所抽选的样本成分在抽选后继续返回总体参加进一步抽选时，称为置换取样；当已抽选的样本成分不再返回总体参加下一步取样时，称为非置换取样。假定某研究需要对 200 名学生的总体取样。如果采用随机取样，每位学生被抽选的机会为 $1/200$ 。这时，如果在选取一名学生以后，把这名学生又交还总体，那么，抽选第 2 名学生的机会仍然是 $1/200$ 。可是，如果在选取第 1 名学生以后，并不交还总体，那么剩下学生被抽选的机会就会增大。因此，取样的置换性会在很大程度上影响样本的特征，并使从样本估算的总体参数的准确性下降。一些研究表明，非置换性随机取样避免了在样本中出现重复选择的成分，因而比置换性取样更准确地估算样本指标。

2. 样本空间

与取样置换性有关的一个概念是“样本空间” (sample space)，指所有的抽样可能性。样本空间的概念在取样中也十分重要。假定某研究需要从 20 名儿童的总体抽取一个 2 名儿童的随机样本，抽样的可能性有 190 种，即抽选 2 名儿童样本的概率为 $1/190$ ，这时，我们说样本空间为 190。如果要从 4 名学生中抽选 2 名，就有 6 种可能： (a,b) 、 (a,c) 、 (a,d) 、 (b,c) 、 (b,d) 、 (c,d) ，样本空间为 6，在非置换性取样中，样本空间是变化的。

3. 随机化

随机化 (randomization) 是指研究中的客体、被试、实验处理或分数等的每一成分被抽选的概率均等。随机化的思想是费歇尔提出的,并成为当代科学研究领域中最有意义的成就之一。这一思想对于改进实验过程和对实验结果的统计推论的重要性,是怎么评价都不会过分的。在研究的取样中,应该尽可能采用“随机化原则”,即在取样中采用随机程序,使总体中每一成分被抽选的机会均等。这样,具有不同特征的总体成分(或成员)在被抽选以后,将在长远的意义上平衡和抵销“相反”的特质。

第二节 取样的程序和方法

一、取样的一般程序

一项研究究竟选择什么样的总体和样本,这取决于研究者所要达到的目的和使研究结果具有普遍意义的范围。例如,研究结果将对全国学习心理学的学生具有普遍意义,那么,样本就必须包含全国所有心理学学生这一总体的特征,不仅应选自全国各类地区和学校、不同民族和社会经济水平及背景,而且需要包括心理学各种专业的学生,并考虑与心理学学习成绩有关的其它因素。通常,为了节省时间、精力和经费,研究者选择一定范围和大小的总体作为研究的对象,从而选取较小的样本。在这种情况下,其研究结论就不应该超越那一总体,除非有证据表明,这一总体具有许多与另一更大的总体相似的特征。

我们决定了总体的基本特征以后,下一步面临的问题就是如何选取能够准确地反映总体特征的样本。取样过程实际上包括三个方

面：(1) 规定总体，对总体作出明确的定义；(2) 选取样本，运用适当取样方法抽取样本；(3) 统计推论，从样本的统计数据估算出总体的有关参数。传统的有关取样的讨论主要涉及前两方面，却很少考虑统计推论问题。而样本的统计推论却是取样所不可缺少的一步，它关系到总体参数的可靠性、取样的误差乃至取样的效果和实际意义。

样本可以是有偏向的或者无偏向的。有偏向的样本会高估或低估总体参数。例如，采用自愿被试所获得的结果往往会高估总体参数。只有采用无偏向样本，统计推论才能得出符合实际的参数值。

“选取样本”是取样程序中的中心环节，它包括决定取样方法、确定总体参数估算值及其标准误差、选定置信区间和决定样本的大小等一系列工作。选取样本这一步，对于从样本对总体特征参数的推论的效度是至关重要的。

二、几种主要的取样方法

取样的方法很多，这儿介绍几种主要的方法。

1. 简单随机取样法

最简便的取样方法之一是简单随机取样 (simple random sampling) 在随机取样中，总体中每一个体被抽选的机会是均等的。运用随机取样方法所选取的样本称为“随机样本”。

随机取样的一种方法是采用“抽彩法”。假如我们需要从 1800 名学生中选取 250 人的随机样本，就可以制作写有 1800 名学生姓名或学号的小圆卡片，并放入容器充分混合，然后随便抽取 250 张卡片，以此作为随机样本。“抽彩法”的关键在于使圆卡片充分混合，而做到这一点是不容易的，因此这种方法在研究中用得并不多。

随机取样的另一种方法是“随机表取样法”。这就是运用严格制作的随机数字表，随意地“进入”包含总体数目的随机数字区，选

取所需要的样本数目。随机表取样法的关键是“随意”进入随机数字区，从而确定一个随机的“起点”。

随机取样又分为两类：(1) 置换性随机取样，或者叫“非限制性随机取样”；(2) 非置换性随机取样，或者叫作“简单随机取样”。

在研究者对所研究的总体中各类个体的比例不了解的情况下，简单随机取样是最好的取样方法，能够选出具有代表性的样本。但是如果样本比较小，简单随机取样仍可能发生偏向，比如，从 1000 名学生中随机抽取 20 名学生，就有可能抽到的全是男生。为了减少这种偏向，可以加大样本，另一个办法是运用随机取样法的一种变式——分层随机取样法。

2. 分层随机取样法

分层随机取样法 (stratified random sampling) 是先把总体分成若干层次或子总体，然后独立、随机地从每一层次选取样本。所谓层次，是指不同类型的个体。例如，在总体中常常存在男、女两个不同层次。因此，要运用这种方法，必须大致了解总体中不同个体类型的比例，从而为分层提供基础。

分层随机取样与简单随机取样相比，具有三方面的优点：

(1) 样本更具有代表性，由于分层取样是在了解总体特征的条件下进行的，所选取样本的代表性比较高；

(2) 参数估算更为准确，研究证明，取样误差只出现在“层次内”而很少发生在“层次间”。由于分层取样使得每一层次内变异程度减少，取样误差也随之减少，这就使参数估算更准确；

(3) 取样更为灵活，分层随机取样使得可以对各层次采取不同的抽取方式和比例，使取样更加灵活。

分层随机取样可以分成“比例分层取样”和“非比例分层取样”两种方法。前者按每一层次个体在总体中所占比例，决定样本中该层次个体的数目；后者则不按上述比例取样，而是根据研究者

对不同层次个体的研究兴趣和侧重程度不同，确定不同的比例。

某项研究着重考察工作态度方面的年龄差异，需要从 24,000 人的职工总体中选取 500 人的样本进行比较系统的分析。研究者采用了“比例分层取样”的方法，先确定总体的层次以及每一层次的人数；然后求出每一层次人数与总体人数的相应比例值，以此作为层次抽选的系数；最后，在各层次完成分层取样。这项研究的抽选情况见表 4-1。

表 4-1 比例分层取样的实例

总体和样本的特征	被试的年龄分组			
	小于 30 岁	30—50 岁	大于 50 岁	总 数
总体中每一层次的人数	10,000	8,000	6,000	24,000
各层次人数与总体人数的比例	.42	.33	.25	1.00
每一层次中应选取的人数	210 ($500 \times .42$)	165 ($500 \times .33$)	125 ($500 \times .25$)	500

可是，在许多情况下，所面临的取样任务复杂得多，比例分层取样往往容易混淆许多干涉变量和因素，影响样本结果的实际效度。例如，有关青少年心理发展特点的研究，需要抽取 5% 的中学生。从理论上讲，可以抽选 5% 的学校，但实际上并不如此简单，因为中学的规模、地区、教育质量等都很不一样。如果只是简单地抽选 1/20 的学校，而不考虑学校的规模差异的话，那么样本中的大型中学的数目就会很少，而小型中学却可能很多，从而使样本出现偏向。因此，应采取非比例分层取样的办法。在这项研究中，分别以 2%、5% 和 8% 的比例，对小型、中型和大型中学进行分层抽样，获得了比较理想的“无偏”样本。需要注意的是，对于非比例样本，还应以各层次数目与总体数目的比例值对结果进行加权。

3. 系统取样法

系统取样法 (systematic sampling) 是从总体中取一随机起点, 从该起点开始选取每 K 项元素 (个体或分数等) 直至取满所需要的样本量。 K 值一般以样本量与总体量的比值来确定。例如, 在 2,000 人的总体中选取 250 人的样本, 样本与总体之比为 $1/8$ 在系统取样时就从随机起点开始分别选取第 8 名、第 16 名……以此类推。

系统取样法比较简便, 它能在总体整个范围内有系统地抽取样本, 因此样本更为准确。为了避免出现偏向, 要求以随机数字表决定第一样本元素 (取样起点)。此外, 要是在总体中存在周期性波动或变化, 系统取样所得的样本就容易发现系统性偏差, 这是在运用系统取样法时应该注意的。

4. 聚类取样法

当我们在研究中不知道或不很了解总体的特征时, 可以采用聚类取样法, 即先把总体分成若干群类, 然后在各群类内进行取样。这种方法成本比较低, 但参数估算的准确性不如简单随机取样等方法高, 取样误差较大。为了提高取样精度, 必须把总体细分, 使群类之间尽可能同质。聚类取样法与分层取样法很不相同。聚类取样法总是设法尽可能减少群类间的差异、尽可能加大群类内的差异; 分层取样法则尽量加大层次间的差异而缩小层次内的差异。相比之下, 聚类取样法更为有效。

5. 非随机取样

随机取样可以避免研究者的主观意志的影响, 获得比较有代表性的样本。但有些研究为了某种研究目的或由于研究条件的限制, 不能运用随机取样方法, 而需要从总体选择具有某方面代表性的个体成分或根据一定的主观判断抽取样本。这种取样方法称为非随机取样 (nonrandom sampling), 又称作目的性取样或计划取样。非随机取样由法国莱珀雷 (Leplay) 首创, 在市场调查和各种经济指数编制

中常常采用。消费心理研究中在分析物价变动对于消费者的心理影响时，往往采用非随机取样法，选择有代表性的几种商品，以此反映整个市场的物价变动情况，确定适当的物价指数。非随机取样法主要有两种：（1）定额取样法：按特定的标准（职业、年龄、文化程度、性别等）确定取样定额，这一方法与分层取样很相似，但主观成分较大；（2）判断取样法：根据研究人员的需要或方便，以主观判断决定所需样本，这一方法的代表性很难确定，适用于试测和可行性研究。

6. 方便取样法

方便取样法（handy sampling）是以志愿者或容易招聘的人员作为样本，而很少考虑样本的代表性。以这种方法所得样本进行的实验，其普遍意义就比较小，这是在研究工作中需要注意的。当我们用方便样本获得某些令人兴趣的结果时，必须用随机样本或其它有代表性的样本，检验实验结果对于特定总体的普遍意义。

除了上述六种主要的取样方法以外，还有一些其他的取样方法，例如：顺序取样法、穿透重复子样本法等等。这些方法在心理学研究中不常用，但在市场研究和企业管理研究中十分重要。在从事心理学研究时，应根据研究的目的与要求、经费与人员条件、研究对象的特点等情况，选用适当的取样方法。

第三节 取样的误差和影响因素

一、取样的误差

在进行取样时，必须考虑到取样的误差。取样误差是难以完全避免的，我们的要求，是尽可能减少取样中的误差，并使之保持在研究所允许的范围内。

1. 样本均值的分布

在讨论取样误差时，必须明确一些有关的统计概念。在心理学研究中，我们把测量分数或个体的总体特征称为参数（parameter）；把分数或个体的样本特征叫做统计值（statistics）。因此，整个分数总体的平均值是一个参数，而样本的平均值是一个统计值。在许多统计书籍中，通过不同的符号把这两类数据区分开来，例如，以 N 表示总体分数的数目，以 n 表示样本分数的数目。其它一些常见的符号有：总体均值， μ ；样本均值， $\bar{x}$ 或 $\bar{M}$ ；总体方差， σ^2 ；样本方差， S^2 ；总体标准差， σ ；样本标准差， S 或 SD 。等等。

样本均值的分布（distribution of sample means）是讨论取样误差时需要涉及的重要概念。例如，我们拥有一个 1,000 个分数的正态分布总体，并且知道其均值为 $\mu=50$ 标准差为 $\sigma=15.8$ （分数值为 0—100）。让我们用 1,000 张纸片分别写上这些分数并放入某容器，然后从容器中抽取 10 张纸片作为一个样本，并写下样本的分数和均值。此后，将纸片返回总体，并将容器内的纸片充分混合。这样，一共抽取 96 个样本，得到 96 个样本均值。表 4—2 表示出这 96 个样本均值的分布，左边为均值落入的区间，右边为落入该区间的均值数目。

可以看到表 4—2 中数据的频次分布十分对称，96 个样本均值的平均数为 49.99，十分接近总体均值（50）。可是，应该注意到，样本均值的变化很大，小至 37.8，大到 62.3。因此，尽管是从相同总体随机抽取的样本，其均值却十分不一样。这样，当我们在实验中获得两个非常不同的样本均值时，就需要区分这两个均值来自相同分布还是属于两个不同的分布。从上例中，即使是随机取样，样本均值也十分不同。因此，决定样本均值的分布和检验测量的信度就显得格外重要。

表 4-2 样本平均值分布 (实际结果)

区 间	频 次
62.0—63.9	1
60.0—61.9	1
58.0—59.9	3
56.0—57.9	7
54.0—55.9	9
52.0—53.9	12
50.0—51.9	15
48.0—49.9	15
46.0—47.9	13
44.0—45.9	9
42.0—43.9	6
40.0—41.9	3
38.0—39.9	1
36.0—37.9	1
	96 样本

2. 均值的标准误差

均值的标准误差 (standard error of the mean) 是指样本均值分布的标准差。上例中, 样本均值的标准误差为 5.01。均值的标准误差表示出样本统计值与整个总体结果的差异。标准误差与一组分数的标准差有点相似, 但标准差 (SD) 表示分数对于其平均值的离散程度, 而标准误差则是样本平均值对于总体平均值的离散测度。

均值的标准误差说明了样本均值分布的变化程度, 或者说, 任何特定的样本均值发生误差的可能性, 标准误差大就表明较大的变异, 而小标准误差则说明特定样本均值很可能十分接近实际的总体均值。因此, 均值的标准误差是研究中非常有用的指标。

均值的标准差误 (σ_x) 与总体标准差 (σ) 及样本观测值的数目 (n) 密切相关, 可以表示为:

$$\sigma_x = \frac{\sigma}{\sqrt{n}}$$

但是, 在实际研究中, 总体标准差往往很难获得。因此, 我们采用样本标准差作为估计值。这样, 均值标准差误的计算公式就是:

$$S_x = \frac{S}{\sqrt{n}}$$

小样本时, 采用 $n-1$ 代替公式中的 n 。例如, 我们选取了 100 位学生的样本, 并测得该样本智商分数的标准差为 20, 则可求得其均值的标准差误为: $S_x = 20 / \sqrt{100} = 20 / 10 = 2.0$ 。这一结果表明, 样本智商的平均水平与其所在总体的智商平均水平之间的典型差异为 2 个智商点。

当研究者关心的是比例问题时, 可以计算“比例的标准差误”, (standard error of the proportion) 为:

$$S_p = \sqrt{\frac{P(1-P)}{n}}, \text{其中 } \sqrt{P(1-P)} \text{ 为比例的标准差。}$$

例如, 从某 100 人的样本测得对某项措施持积极态度者的比例为 0.60, 则比例的标准差误为:

$$S_p = \sqrt{\frac{.60 \times (1-.60)}{100}} = \sqrt{\frac{.60 \times .40}{100}} = .049$$

此外, 还可以计算“百分数的标准差误”, 上例中, 持积极态度者占样本人数的 60%。则其百分数的标准差误为:

$$S = 100 \times \sqrt{\frac{P(1-P)}{n}} = 100 \times \sqrt{\frac{.60 \times (1-.60)}{100}} = 4.9$$

显然, 我们要求均值的标准差误尽可能小, 因为, 它表明了在以样本均值代表总体均值时所出现的误差。从均值标准差误的公式可以看到, 随着样本 (n) 增大, 均值的标准差误将减小。在表 4-2

的实例中 当选用 50 个分数作为样本时，均值的标准差误从 5.01 减少到 2.23 当选用 $n=100$ 的样本时， $\sigma_x(S_x)=1.59$ 。

二、样本的大小和影响因素

1. 样本的大小

一项研究所用的样本究竟以多少大为好？这是每一位研究者面临的问题。一种比较流行的观点认为，只有采用大样本，才能从一定研究数据作出有效的结论；或者认为样本越大越好。方法论方面的研究却表明，当样本增大到一定规模以后，更大的样本对提高研究准确性的作用就不很显著；同时，当样本太大时，容易引起统计上的某种敏感效应，增加统计决策错误（如 I 类错误）的可能性，因此，不少研究者不主张运用过大的样本。

统计学的原理告诉我们，均值的标准差误和样本大小之间的关系是，在足够大的样本条件下，可以在一定精确度范围内估算总体的均值。这就是说，如果可以用总体 σ 单位表示所要求的精确度，就能够推算出所需要的样本大小。

举例说明，为了使样本均值 M 有 99% 的可能落入真均值 μ 的 0.1σ 的范围内，应该选取多大的样本呢？即 $(|M-\mu| \leq .1\sigma)$ 的概率 = .99。如果我们假定样本分布是近似正态的，即 99% 的置信区间在 $M-.1\sigma$ 和 $M+.1\sigma$ 范围之内，因此， $.1\sigma = 2.58\sigma_M$ 。当样本较大时 ($n=30$ 以上)，可以用 $\sigma/\sqrt{n}$ 代表 σ_M 。因而， $.1\sigma = 2.58\sigma/\sqrt{n}$ 所以 $\sqrt{n} = 25.8$ ， $n=665.64$ ，可见 需要有 666 个个体或观察值的样本，才能保证样本均值估算超出 $.1\sigma$ 的概率只有 1%。

2. 影响样本大小的一些因素

样本的大小取决于许多因素，其中比较重要的有三种因素：

(1) 研究的准确性

准确性是指样本均值与总体均值的一致程度。某些研究表明，选

取较大的样本，并不一定显著地提高研究的准确性，只有当总体相对小的情况下，样本大小才较明显地影响参数估算的准确性。

(2) 研究的经费和成本

在决定样本大小之前，应考虑为取得和运用较多样本所需的成本和费用。例如，访谈研究往往比问卷研究需要更多的时间、精力和费用，因此常常只能选取较小样本。

(3) 总体的同质性

在确定样本大小以前，还必须考虑总体的同质性或异质性，一般来说，总体越是同质，就越有条件采用较小样本；总体越是异质，就越需要选取较大样本。对于不同的取样方法，总体同质性这一因素的影响有所不同。例如，在采用简单随机取样法时，为了达到确定的准确度，所需样本大小随总体同质性下降而增大。在用分层随机取样法时，各层次已尽可能同质，因而可以选取较小样本。

此外，在进行取样时，应该看到取样的某些局限性，从而保持清醒的头脑。取样的作用取决于是否能够运用样本数据准确地估算总体参数。在正确地运用适当取样方法的基础上，还必须准确地选用统计方法，对样本统计值作出认真、严格的测算，并对总体参数作出推算。取样的目的是为了研究总体而不是个体，但样本统计值却只是有关总体的估计而并不包含很多有关个体心理过程的具体信息，这也是在研究工作中需要加以注意的问题，特别应重视采取某些控制误差和补充信息的措施。在对样本数据作出解释时，应防止绝对化倾向，以免对样本信息给予过分的解释，使取样程序更好地为研究的目的服务。

第四章思考题：

1. 总体与样本各具有哪些特征？什么叫取样的置换性和样本

空间？

2. 取样的一般程序包括哪些步骤？在取样的各个阶段，应注意哪些问题？
3. 讨论几种主要取样方法的特点和优缺点？聚类取样法与分层取样法有什么异同？
4. 什么叫均值的标准差误？如何减少均值的标准差误？
5. 如何推算所需样本的大小？影响样本大小的因素有哪些？

第五章 研究的效度

第一节 研究的内部效度

一、研究效度的概念

心理学研究的选题、构思、设计、实施与评价，是一组相互影响、密切联系的科学活动。在此过程中，人们总会面临各种问题，特别是需要对以下问题作出判断和决策：

(1) 所研究的两个或多个变量之间是否存在一定的关系？尤其是，研究的自变量与因变量之间是否有关系？

(2) 如果所研究的变量之间存在一定关系，这种关系是否因果关系？

(3) 如果变量之间可能存在因果关系，它所包含的具体的因果关系构思是什么？

(4) 假定变量之间的因果关系构思明确，这种因果关系对于其他的人员、背景条件和时间的普遍意义如何？

以上四类问题相互关联，都涉及到研究的有效性，即研究的效度。1957年，社会心理学家坎贝尔 (Campbell, D. T.) 第一次明确地提出研究的效度问题。十年以后，他在《研究的实验和准实验设计》(1966) 的专著中，借用了测验领域的“效度”概念，把研究效度作为设计与评价各种研究的理论框架，并系统地应用于研究和实验的设计。1979年，坎贝尔与库克 (Cook, T. D.) 一起出版了心理

学方法论的重要专著《准实验研究 现场背景的设计与分析》更为全面、系统地讨论和分析了研究效度及其在实验设计中的意义。现在，有关研究效度的理论和方法，已为心理学家所普遍接受，成为研究与实验设计的关键，也成为心理学研究方法发展的里程碑。

根据研究效度的理论，前面所提到的四类问题，实际上与四种研究效度有关。研究中的自变量与因变量之间的关系属于内部效度的问题；在变量之间判定因果关系，这是统计结论效度的问题；因果关系的构思性质，涉及到研究的构思效度；变量之间关系的普遍意义，则是研究的外部效度问题。可见研究效度在研究的设计、实施与评价等活动中的关键意义。在这一节里，着重讨论研究的内部效度；在本章第二、三节中，将分别讨论研究的外部效度、统计结论效度和构思效度等问题。

二、内部效度的定义和条件

1. 内部效度的定义

效度常常是指一项测验测到所要测量的东西或达到某种目的的程度，也有人叫作真实性。研究的内部效度是指在研究的自变量和因变量之间存在关系的明确程度。如果自变量和因变量之间关系并不由于其他外部变量的影响而变得模糊不清或复杂化，那么这项研究就具有内部效度，这时就可以得出结论，说某种实验处理或条件确实具有某种效果。坎贝尔（1957）曾经把内部效度定义为“实验的刺激确实造成特定情况下的某种显著差异”。要做到这一点，就必须在研究或实验的设计和实施时，注意排除干扰因素。

2. 内部效度的条件

研究的内部效度并不会自动形成的。内部效度的获得，主要是通过认真细致的变量选择和准确周密的研究设计。一项研究之所以有内部效度，首先是由于正确地选择了研究的自变量和因变量，这

样才有可能作出有关变量之间因果关系的陈述。如果研究中没有明确的自、因变量，就很难保证研究的内部效度。正如瑞典著名应用心理学家鲁本诺维兹（Rubenowitz, S.）指出的，要是研究中缺乏真正的自变量，其内部效度就值得怀疑（1984）。内部效度的另一个重要条件是周密的实验或研究设计。这主要是指在进行实验设计时，明确地意识到各种影响内部效度的因素，运用多种设计措施，控制这些影响因素，突出自变量与因变量之间的关系。在选用不同研究类型时，应注意各种研究类型的弱点，善于利用它们的优点，为探测和验证各种变量之间的关系服务。

三、影响内部效度的因素

在心理学研究过程中，影响内部效度的因素比较多。归纳起来，主要有以下一些因素。

1. 历史因素

历史因素可以分为前摄历史因素和后摄历史因素两类。前摄因素包括被试固有的和习得的差异，比如性别、工龄、经验、身高、体重、态度、个性、运动与心理能力等等。如果对照组与实验组在这些因素上差异很大，则会影响所测得的两组间差异，降低效度。在研究中控制这些因素的方式是对被试进行随机化选择。后摄历史因素是指在研究测量以外可能影响自变量和因变量之间差异的因素。这常常是由于研究中未控制因素的干扰，使因变量发生变化，而不是实验处理本身所造成的差异。例如，有一项研究希望检验某项工作丰富化方案对于提高职工积极性的作用，但是，在方案试行期间，一部分职工增加了工资，因此混淆了工作丰富化方案对于积极性的影响，使积极性（因变量）变化的原因复杂化了。这一实例中，增加工资就是一种后摄历史因素。后摄历史因素可通过恒定其它事件的方式加以控制。

2. 选择因素

这实际上就是前摄历史因素，是由于选择被试的程序不适当而使各被试组之间一开始就存在系统性差异，影响内部效度。当研究中不可能对被试作出随机选择和分配时，最易受选择因素的影响。

3. 成熟因素

这是指在研究期间，由于被试自身的生理与心理发展、成长（技能、知识、经验等）或变化而引起的系统变异，因而随着时间的推移而产生某种系统效应，降低研究的内部效度。成熟因素具有动态的性质，控制的方法是使被试选择与分组尽可能随机化，并设立对照组。

4. 研究被试的更换和淘汰

实验和研究期间，被试由于事故、生病或研究调整等原因而被更换或停止参加研究，这种因素会明显地影响对研究结果的比较和分析，大大降低内部效度。在大家熟悉的“霍桑实验”中，就存在被试更换这种严重影响研究效度的问题。此外，在纵向研究中更容易出现这类情况。被试的更换和淘汰并不是统计方法可以弥补的，因此，在研究中应特别注意使被试保持稳定和一致。

除了被试的更换与淘汰以外，被试本身的质量，也是影响研究效度的一个重要因素。研究表明，实验研究中常见的有四种被试：合作型、消极型、评价担忧型和忠实型。合作型被试倾向于在研究中设法帮助实验者发现和证实研究假设，这种“合作”往往会带来偏向；消极型被试倾向于抵制研究的计划与安排，常常与研究者的“对着干”，这类被试不利于对研究假设的检验和有效地开展研究；评价担忧型被试则对参加实验抱着疑虑，担心别人总是在评价自己，因而总是设法避免消极的评价，盲目猜测研究目的，以“理想”方式表现自己。这三类被试都不是心理学研究所要求的被试人员。第四种被试能按研究要求行事，没有前三类被试的缺陷，被称为忠实型

被试，是研究中理想的被试。

5. 测量和测试手段

这一因素包括两个方面。

(1) 测量：当研究运用事前与事后两次测量时，事前测量会影响事后测量的结果，影响研究的内部效度。常见的有练习效应、敏感效应和选择性效应。这种因素一般通过设立无事前测定组加以控制。

(2) 测试手段：这是指由于测评者本身的疲劳、单调感、注意分散以及其它主观因素和测试工具本身的变化，使测定和评级的精确性受到影响，降低研究的效度。这种因素与研究设计的关系不大，主要靠研究者或测评者加强基本训练和严格测试手段，消除影响。

6. 统计回归效应

数理统计学的基本原理告诉我们，在进行重复测量时，初测时获高、低极端分数者会在重测时倾向于向平均值偏移，即发生“统计回归效应”。在心理学研究中，常常通过事先测试选择极端分数的被试参加研究，这样就会由于统计回归效应而混淆再次测试的结果所反映的实验效应，从而降低研究的内部效度。这种情况在教育心理、儿童心理、临床心理、社会心理和工业心理等领域的研究中是比较常见的。控制的办法是在研究中把极端分数者单独分成对照组和实验组，注意它们的结果差异，进而作出有关自变量效应的正确结论。

7. 实验处理和程序

实验处理和程序本身的一些特点会影响研究的内部效度，其中特别值得重视的有以下几方面：

(1) 有关实验处理和程序的信息的扩散或交流，使对照组实际上得到了实验处理或受其影响；

(2) 实验处理的实施包含着其它好处（如资助），产生补偿性等同效应；

(3) 由于实验处理而使实验组十分引人注目或炫耀，产生补偿性竞争。

上述几方面的因素都会抵消、降低或夸大实验处理的效应，严重损害研究的内部效度。

8. 多种研究条件和因素的交互作用

许多研究对不同条件或因素的效应进行比较，可是在研究过程中却往往会由于测试程序、因素控制和实验安排等方面的原因，造成多种条件和因素之间的交互作用，从而影响效度。

上述八种主要因素，都是在研究中应特别注意控制的，最有效的途径是采用随机化程序。在心理学研究中具有较高内部效度的设计之一，是著名的“所罗门（Solomon）四组设计”，它通过随机化程序和设立四个不同处理的实验或对照组（图 4-1），保证了研究的内部效度。

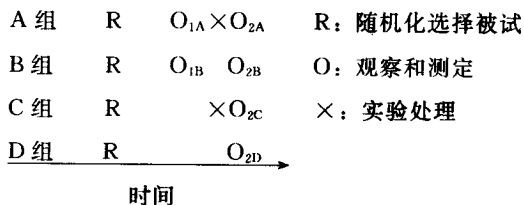


图 5-1 所罗门四组设计

第二节 研究的外部效度

一、外部效度的定义和条件

1. 外部效度的定义

研究的外部效度，是指实验和研究的结果能够一般化和普遍化到其他的总体、变量条件、时间和背景中去的程度，即研究结果和

实验效果的普遍性或可应用性。一般认为，内部效度是外部效度的必要条件，但不是充分条件。内部效度低的研究结果就谈不上对其他情景的普遍意义；可是内部效度高的研究，其结果却不一定能够一般化到其他总体和背景中去。在最近的一些有关研究效度的讨论中，不少心理学家认为，外部效度比内部效度重要。从目前的研究发展趋势来看，这种看法是有道理的。

2. 外部效度的条件

外部效度的要求是使研究能够代表真实世界的情况。可是，心理学研究或实验一般总是包含许多人为的因素，因此，单一的研究难以在被试、变量和背景等方面保证其代表性，必须在使研究尽可能模拟现实情景的基础上，通过多个相互关联的实验，以不同的研究条件寻求具有普遍意义的结论。这种多重实验手段，是获得外部效度、提高研究结果可应用性的重要条件。

二、影响外部效度的因素

影响外部效度的因素主要有八个方面。

1. 总体效度

这是指研究结果对于研究者最终感兴趣的总体的普遍意义。研究样本的代表性在很大程度上决定了研究的这一总体效度，高总体效度才会提高研究的外部效度。最好的办法是明确规定研究的总体，并从这一总体随机地抽取样本。

2. 自变量和因变量的定义和测试方式

研究中必须明确地对自变量和因变量作出定义和描述，否则就会影响研究的外部效度。同时，还应规定自、因变量的测试方式，注意提高测量的效度。这样，不但明确了实验效果，而且增加了研究结果的可重复性，从而提高外部效度。

3. 对测量的敏感化

在我们讨论内部效度时，已经提到这一因素。由于在研究中运用了事前测量，使被试对事后测量发生敏感效应，从而影响了研究结果；或者，由于事前测量而干扰了实验处理的作用。在这种情况下，研究的结果很可能不能应用于或一般化到那些没有受过相似事前测定的对象中去，而只能应用于已作过相似事前测定的样本，因此，降低和损害了研究的外部效度。这方面的研究表明，认知性的事前测量和两次测量之间较长的时间间隔，可以减轻这种敏感化效应。此外，事后测量也会产生敏感效应，使被测者在测量中学到实验期间没有获得的东西，这样的研究结果就只能一般化到接受事后测量的样本。为了解决测量的敏感化问题，应尽量少用容易产生敏感效应的测量，尽可能采用“不引人注目的测量”(unobtrusive measures)，例如，间接观察行为或参与观察等。

4. 多重实验的干扰

当被试多次接受实验处理或同时参加多个实验时，就会取得某些实验经验，产生敏感效应，这样既会影响随后研究的效率，也会很大程度地影响研究结果的普遍性。事实上，建立“实验学校”或“基地”，对同样对象反复多次地进行心理学或教育实验研究，就特别容易产生多重实验的干扰，大大削弱研究的外部效度。

5. 特异性效应

这是指由于实验和研究采用了新异的方法、特殊的材料和情景，或者使被试改变已习惯了的活动、工作方式和环境，因而产生暂时的新异或接触性的反应，影响研究的效度。

6. 实验者效应

这是指实验者本身的个性特征、动机、情绪或其它细微而无意的行为影响被试，使得把实验目的、结果的期望等也在无意中传递给被试，因而限制了研究结果的外部效度。人们早就注意到了这种效应，建议运用标准化的指导语和实验程序，并采用“双盲”实验

(被试和实验者都不知道研究的真正意图)，以提高研究的外部效度。

7. 研究与实际情景的相似性

研究与实际情景的相似性，是外部效度的关键因素之一。一般来说，这种相似性越高，研究结果的可应用性也越高。此外，研究中所用变量的强度和范围，也应该尽可能接近研究结果所要应用的实际情况，这样就能在很大程度上提高研究的外部效度。

8. 被试选择和实验处理的交互作用

这是指研究和实验被试的各种特征会使实验处理的效果具有特定的含义，影响实验和研究结果的普遍性。这时，被试选择和实验处理之间，实际上存在着交互作用。例如，被试的自愿性（义务参加实验）、年龄与成熟、实验处理或测定的时机，等等。这些因素都比较容易造成被试选择和实验处理这两方面之间的交互作用，降低研究的外部效度。

上述八种因素，有时单独存在，有时同时存在多方面的影响，都是我们在考虑研究的外部效度时所要认真对待的。

第三节 统计结论效度和构思效度

一、统计结论效度

1. 统计结论效度的定义和条件

统计结论效度是有关决定实验处理效应的数据分析程序的效度检验，它并不涉及系统性偏向的来源问题，而是研究误差变异源和如何适当运用统计显著性检验的问题。例如，当研究样本小的时候，如果依赖于统计显著性水平作出推论，是比较危险的。这时候，应该运用功效分析（power analysis），看一看一定的样本大小、变异程度和 α 水平上能够检验出多大的效应，因为在小样本时，样本成分和

测量都波动比较大，具有不稳定性。这就是统计结论效度所涉及的问题。

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，包括数据的量表特征（顺序量表或等距量表数据）、数据分布、信度和效度等，数据质量差的研究是谈不上统计结论效度的；二是统计检验的假设，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假设，一项研究中统计检验假设不明或者被违犯，就会显著降低统计结论效度。

2. 影响统计结论效度的因素

(1) 统计功效低

统计功效又称作统计检验能力，在统计学上定义为不犯Ⅱ类错误的概率 $(1-\beta)$ （有关系却接受虚无假设）。当样本小而且 α 值定得较低时，犯Ⅱ类错误的可能性就增加；如果这时 α 值定得较高，又容易犯Ⅰ类错误（无关系却拒绝虚无假设）。这些都会降低统计功效，影响研究的统计结论效度。

(2) 违犯统计检验的假设

大多数统计检验都要求满足一定的假设，然后才有可能对数据分析的结果作出有意义的解释。因此，在采用不同检验时，应了解其特定的基本假设并解决有关的问题。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；许多统计检验方法要求数据呈正态分布，应注意数据的这一特征，以免违犯统计检验的假设，影响研究的统计结论效度。

(3) 测量信度低

低信度的测量会夸大估计值的标准误差。测量的标准误差的公式为： $Se = S \sqrt{1 - r_{xx}}$ ，其中 S 为标准差， r_{xx} 为测量信度。可见，信度（ r_{xx} ）降低以后，测量的标准误差会明显增大。由于标准误差在

作出统计差异推论时起着关键性作用。因而，测量信度低就降低了统计结论效度。提高测量信度的办法是运用具有高相关项目的较长测验，或者采用较大分析单位（例如用小组均值代替个人分数）。

（4）实验处理实施的可靠性

由于研究者之间的差异或同一研究者在不同时间采用了不同方式实施实验处理，使实验处理不够标准化，这会增大差异变异并减少发现真实差异的可能性，进而降低统计结论效度。因此，应尽可能使实验处理实施标准化。

（5）研究背景中随机的无关因素

这是指实验处理之外的一些特点和因素对研究结果的影响。因此应尽可能排除无关变异源，并使被试集中注意于实验处理，减少背景变量的突出性和影响。

（6）被试的随机异质性

研究中被试往往会在主要因变量方面有较大的差异，有的被试比较容易受到实验处理的影响，还有的被试却不容易接受实验的效应，这就有可能降低研究结论的意义，既会影响研究的统计结论效度，也会损害研究的外部效度。对于被试随机异质性的问题，可以通过选择相对同质的被试（这可能影响外部效度）和采用“被试内设计”（参阅第六章）的办法加以解决。这两种方法都有利于减少差异变异量，提高统计结论效度。

二、研究的构思效度

1. 构思效度的定义和条件

研究的构思效度涉及到研究方案和测量指标的理论构思及其操作化的问题，即理论构思及其转换的有效性。为了使研究具有较高的构思效度，首先要求研究的理论构思结构严谨、层次分明，形成某种“构思网络”，并对研究中的自变量与因变量作出严格的定义。

其次，针对研究构思的特点，给予明确的操作定义，并确定相应的测量指标。心理学研究中常常包含着复杂的、多维度的理论构思，因此，改进研究的构思效度，是研究中的一项十分重要的任务。

2. 影响构思效度的因素

(1) 对于研究构思的解释

心理学研究需要对其理论构思有明确的说明。如果在对研究构思作出操作定义之前，对构思的基本特点缺乏明确的概念分析和解释，就会影响研究的构思效度。

(2) 单一操作的偏向

心理活动总是包含着多方面的因素，心理特征也具有系统与层次的特点。因此，心理学研究的构思一般都具有多维性。当研究中只用一种尺度和样本来表示这种多维复杂构思时，就会产生单一操作偏向，并且容易受到无关因素的影响。这些都会降低研究的构思效度。

(3) 单一方法的偏向

理论构思的多维性，要求有一种复合的测量，这就是在研究中选用多种指标，运用多种方法，从不同角度分析所假设的关系。如果在对构思进行测量时，只采用单一方法，或者在量表设计中使用单一陈述和单一项目，就会出现单一方法偏向，削弱研究的构思效度。目前，研究者普遍主张用多种方法研究相同理论构思，比如分别用访谈法、问卷法和观察法研究群体心理结构，就容易取得比较理想的研究结果。

(4) 实验者期望和被试猜测

实验者的期望会使研究数据产生偏向，这样就会混淆实验处理的因果效应。此外，被试有时会猜测实验者的期望和需要，进而影响自己的行为，干扰真正的实验效应。这种期望或猜测，都不利于研究的构思效度。有的研究曾经认为，在实验中，被试可能因为处

于特殊研究条件而变得更为合作和努力，人们称之为“霍桑效应”。可是，这种所谓“霍桑效应”，并没有广泛的现场或实验室研究的证据。

(5) 构思水平之间的混淆

当理论构思包含几种水平时，某指标在不同水平上会出现不同关系，这就会影响研究的构思效度。例如，当自变量只在中等强度时才对因变量发生效应，而在高、低强度时却关系很弱。在这种情况下就可能产生构思水平间的混淆。因此，在操纵自变量时，应在其不同水平上，测量因变量的相应的不同结果，确定变量和构思在不同水平上的真正关系。

(6) 不同实验处理的交互作用

这是指被试经历多种实验处理而引起交互作用，损害构思效度。这种情况在实验室研究中很普遍，在现场研究中较少发生。这时，我们难以确定单一实验处理的效应。此外，研究中的事前或事后测定也会对实验处理发生交互作用，影响研究效度。

三、四种研究效度之间的关系

研究的内部效度、外部效度、统计结论效度和构思效度，是相互联系和相互影响的。

统计结论效度实际上是内部效度的特例，它们都涉及研究本身的因果关系和统计检验的可靠性。构思效度则与外部效度有一致之处，即它们的基本点都在于作出概括化和结果的普遍性。如果某一关系对于特定的总体、背景和时间都成立，那么这也明确了研究的构思效度。构思效度与外部效度之间的主要差异是：前者所指的目标总体较难具体确定，而后者则往往要求一般化到实际上存在的特定总体、背景和时间等条件。

四种研究效度的相对重要性，主要取决于研究的具体目的和要

求。一般来说，可以在保证研究内部效度和构思效度的情况下，提高统计结论效度和外部效度。可是，从影响研究效度的因素可以看到，用于提高某一种研究效度的措施，很可能会降低另一种研究效度。因此，研究者需要明确地确定不同效度的优先顺序，在不同的措施之间作出适当权衡，从而避免不必要的效度损失。影响研究效度的因素多种多样，但大致可以分成四类：与被试有关的因素；测量与具体研究方法方面的因素；研究构思与程序设计因素；实验条件与不同因素的交互作用。在决定了各研究效度的重要程度以后，就可以从被试、测量方法、构思与程序以及交互作用等四个方面入手，改善和加强研究的效度。

第五章思考题：

1. 什么是研究的效度？四种研究效度的关系如何？
2. 研究的内部效度和外部效度分别受到哪些因素的影响？如何控制或避免这些因素？
3. 所罗门四组设计有何优点？这一设计中的 A、B、C、D 四个组分别起了什么作用？
4. 什么是研究的统计结论效度？提高统计结论效度的关键是什么？
5. 研究的构思效度与外部效度的关系如何？怎样加强研究的构思效度？
6. 以有关研究效度的理论作为框架，分析和评价一项研究设计实例。

第六章 实验和准实验

第一节 现场研究的特点和条件

一、现场背景的特点

心理学的生命力在于应用，而传统的实验室方法很难准确、充分地反映实际生活和生产活动中人的心理过程的许多边界条件和背景因素。因此，在心理学研究中，无论是教育、儿童和临床，还是管理、社会和工程等各分支，现场研究日益成为主要方式。它不但进一步验证了实验室条件下的某些结果，而且更广泛深入地了解实际生活中众多的心理现象。

现场研究的背景十分不同于实验室条件，由于是在实际工作、学习和生活情景中进行研究，因而包含了各种复杂因素。现场背景的下列特点是心理学研究中必须认真考虑的。

1. 边界特点

(1) 变量强度

在现场背景中，某些重要的自变量可以显示出较高的强度，比如应激情景；而在实验室背景中，由于伦理和其它原因，某些自变量的强度受到限制。

(2) 变量变化范围

在现场背景中，许多变量的变化范围远远大于实验室条件，例如，群体规模、管理幅度、组织结构集中化程度、复杂性和时间广

度等因素和变量的变化范围，都比实验室中更大。此外，跨文化现场研究的范围更是如此。

(3) 变量操纵的频次和持续时间

由于在实验室条件下研究时间往往比较短，人们常常有能力缓冲较短时间中的各种刺激，而且，实验室中不易提供包括各种反应阈限的条件。正因为如此，许多心理效应在实验室中没有被察觉，而在现场情景中却十分明显。某些小刺激会由于反复出现或持续较长时间而产生令人兴趣的心理效应。变量操纵的频次多和持续时间较长，这是现场研究的又一边界特征。

2. 结构特点

(1) 时间结构

许多心理过程都是在一定的自然时间广度上展开的，现场研究情景为我们提供了长时间结构和自然时间单元，而这些特征对于真正掌握心理规律又是十分重要的。

(2) 自然单元

这是指在现场条件下，研究者可以寻求特定环境下人们行为的自然单元，即自然发生的过程，这就避免了实验室条件所造成的人为背景与结构的影响。

(3) 复杂性

现场背景比实验室条件要复杂得多，许多情况在实验室里不可能模拟，因此，现场背景提供和保持了研究复杂心理过程的条件。

以上三种结构特点，都是在实验室研究中为了控制条件而常常设法消除的，因而使研究的效度受到影响。

二、现场研究的条件和局限

现场研究具有实验室研究所不可能有的一些特点，同时，也受到现场的某些条件的限制。

1. 现场研究的情景效应

在现场研究条件下，情景是开放的、动态的，往往会产生某种情景效应。例如，在有关态度和态度改变的研究中，情景效应就特别强烈。此外，社会背景、政治、文化和经济形势等也会明显地影响研究，这是在现场研究中需要重视的问题。

2. 研究的代表性

研究的代表性主要有两方面：研究样本的代表性和实验处理或程序的代表性。我们知道，实验室研究往往有可能通过随机化程序，使样本具有代表性。可是，在现场条件下，常常不能运用随机化程序。企业、学校和医院等部门都有正常的工作和学习秩序，不可能随便改变。因此，现场研究中经常选择现成的车间、班组和班级作为样本，研究的代表性就会受到影响。但是，实验室研究所实施的实验处理却带有很大的人为性，其代表性就比较低。现场研究则可以实施比较符合实际条件的实验操作或处理，其代表性就比较高。例如，不少曾经在实验室研究中得到证明的社会心理效应，在现场条件下却观察不到，这主要是情景和实验处理的代表性问题。从这一点看，许多现场研究的结果都可以较有效地应用于实际。

3. 现场研究中的因果关系

现场研究中包含着众多的因素，又不易对变量作出有效的控制或操纵。因此，现场研究结果较难确定变量之间的因果关系，需要运用不同方法、在不同时间对变量之间的关系进行多次检验。如果不同方法和时间的研究结果比较一致，就能为确定因果关系提供充分的依据。

此外，多数现场研究花费多、代价高，而且不易有多组被试参加研究，这些都是它的一些局限。

三、现场研究和准实验

从现场研究的背景特点和条件可以看到，它有许多优于实验室研究的地方。由于现场研究结果的可应用性比较高，它在心理学实际应用中正发挥日益重要的作用。目前，现场研究中最多采用的是准实验设计，不过准实验并不一定都是在现场进行的。

我们在第三章“研究的设计”中曾经提到，准实验的一个重要特点是没有运用随机化程序，即在被试的选择与编组、实验处理的分配等方面，都不是随机安排的。

准实验设计的种类比较多，常见的有五种类型：（1）交叉滞后组相关设计（cross-lagged panel correlational design）；（2）回归间断点设计（regression-discontinuity design）；（3）不等同对照组设计（nonequivalent control group design）；（4）间歇时间序列设计（interrupted time series design）；（5）重复处理设计（repeated treatment design）。这五种准实验设计各有特点，适合于不同的研究课题和条件。本章第二节将详细讨论前四种准实验设计。第五种设计，即重复处理设计，对相同被试组重复进行多次“测定—实验处理—测定”的程序，与第一章中介绍的小样本范式相似，如图 6-1 所示。教育心理、管理心理和工程心理的一些现场研究中都曾采用这种设计，但是这种设计由于周期性成熟、实验处理敏感和猜测效应等因素的影响，使研究的效度存在不少问题，这些是在研究工作中应加以注意的。

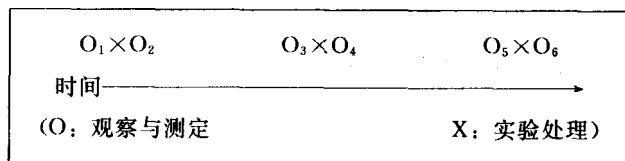


图 6-1 重复处理设计

第二节 准实验设计

一、准实验设计的一般问题

准实验设计的概念，是心理学家坎贝尔与库克（1966, 1979）在讨论研究效度问题时首先提出来的。这种设计适应了心理学理论与研究发展的实际需要，因而在近二十年来得到了日益广泛的应用，成为心理学研究的主要实验设计方案。库克和坎贝尔于 1979 年所发表的《准实验研究：现场背景的设计与分析》这一专著，为准实验设计提供了充实的理论基础，并且系统、全面地提出了准实验的设计原则和统计分析方法，使准实验设计的理论与方法日臻完善。

准实验设计的迅速发展，是与心理学理论与方法论的新进展密切有关的。特别是以下三个方面的新动向，对于准实验设计理论与方法的发展，产生了直接的、重要的影响。

1. 研究方法论重点的转移

70 年代以来，心理学研究方法论的重点出现了引人注目的转移。心理学家日益重视研究的“生态效度”，即要求研究能对现实生活和背景中的心理过程具有更大的意义。因此，心理学研究逐步从实验室里人为的实验向现场研究或接近现实的模拟转移，日益现场化。这种研究方法论的重点转移，对实验设计提出了新的、更高的要求，从而促进了准实验理论与方法的发展。

2. 传统实验的局限日益明显

传统的实验室实验，有一整套实验设计原则。然而，这些设计原则的效用是建立在严格的随机化程序的基础上的，只有实现随机化程序，才能有效地排除那些影响研究效度的因素。但是，传统的实验室实验中却常常无法保证随机化的实施。这种情况使传统实验

的局限日益明显。同时，经典的小样本范式在研究效度方面也面临许多问题。因此，心理学研究需要一种新的实验理论与范式，用以弥补传统实验的局限。准实验理论与方法显然适应了这种需要。

3. 日益重视研究的质量

在对传统实验的局限提出批评的同时，心理学家对于那些非实验性的研究越来越感到不满，日益重视研究的质量，希望有比较严格的实验设计模式来代替原有的既缺乏理论又很少有数据的一般调查。这种趋势也在相当程度上推动了准实验设计的应用与发展。

准实验设计不需要运用随机化程序，也没有采用非常人为的严格的控制方法，因而适合于更广泛的研究目的。准实验作为一种有效的研究手段，并不局限于在现场背景中进行，也可以在某些模拟实验室里实施。可以认为，准实验设计正在成为心理学研究中最有前途和最具有应用价值的实验设计。

二、交叉滞后组相关设计（CLPC）

心理学家在 40 年代末就注意到，当我们对相同的两个变量进行重复测量时，实际上提供了有关这两个变量之间因果关系的某种信息。坎贝尔（1963）把 16 重（16-fold）交迭变换与经济学中常用的滞后相关联系起来，研究连续变量的特征和相互关系，建立了交叉滞后组相关法。交叉滞后组相关的准实验设计正是在这一方法的基础上形成的，这种设计的基本原理是，通过交叉滞后相关系数的比较，找出交叉滞后相关差异的方向，如图 6-2 所示 然后 根据差异的方向，确定变量之间的关系。

交叉滞后组相关设计，要求在时间 1 对两个变量的关系作出测定和相关分析，再在时间 2 作相似测定和分析，求得同步相关 $r_{A_1B_1}$ 和 $r_{A_2B_2}$ 在图 6-2 中均为 $r = .50$ ； $r_{A_1A_2}$ 和 $r_{B_1B_2}$ 为稳定性相关，它们实际上是重测信度。更令人感兴趣的是 $r_{A_1B_2}$ 和 $r_{B_1A_2}$ 相关，特别是当交叉

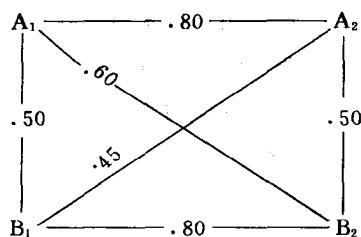


图 6-2 交叉滞后组相关设计

滞后相关之间有显著差异时，具有因果关系的意义。在同步相关稳定的情况下，如果 $r_{A_1B_2} > r_{B_1A_2}$ ，则 A 引起 B，反之，B 引起 A。

交叉滞后组相关设计中需要考虑 A、B 测量本身的信度，还应考虑 A、B 之间关系随时间发生的时间损耗。另外，变量本身的变化率也是重要的因素之一，两者变化率不同时，容易发生错误推论。

交叉滞后组相关设计，需要满足三条基本的假设：（1）交叉滞后组相关所表示的因果关系不随时间的推移而变化；（2）同步相关与稳定性相关都尽可能一致，如果出现不一致，应有合理的解释；（3）交叉滞后组相关中包含了主要的变量，并已作出相应的测量。

由于在实际研究中，上述假设不容易得到基本的满足，因此，对于交叉滞后组相关设计结果的解释，应持谨慎的态度。尽管如此，这种准实验设计在某些方面优于普通的相关设计，并具有一定程度动态的性质，而且较好地处理了心理学研究中几乎都存在的差误和独特变异问题，因而具有较好的内部效度。需要指出的是，交叉滞后组相关设计虽然比简单相关更充分地说明了因果关系，却仍然不能完全证明因果关系，因此，还需要其他研究的补充和验证。

交叉滞后组相关设计在心理学许多领域得到应用。其中比较著名的是埃伦（Eron, L. D.）等（1972）所进行的长达 10 年之久的追踪研究。这项研究考察了三年级儿童对于暴力或凶杀电视片的偏爱与他们的攻击性行为之间的关系，并在此十年以后作了重新测定，

求出了交叉滞后相关系数。结果表明，儿童早期生活中观看暴力或凶杀片，很可能加强成年初期的攻击性行为。埃伦等（1982）还作了进一步的研究，支持了上述结果。

三、回归间断点设计

这种准实验设计通过实验处理与事后测量回归线的间断点的特征，确定准实验处理的主效应。常见的回归间断点设计都包含着事前与事后测量，用以考察某些实验性措施（例如，奖励）的效果。如果奖励对生产和工作成效有作用，那么，奖励前后所作测定结果的回归线就会出现间断点，这是因为奖励以后的工作成效高于先前记录的结果。

图 6-3 表示了一项回归间断点设计的准实验的结果。在这一实验中，研究者采用了连续的等距量表对工作成效进行了事前与事后测定。在事前测定以后，对成效分数达到一定标准的被试（职工）发给奖金，然后测定奖励的作用，结果表明奖励是有效果的，因为回归线出现间断点。

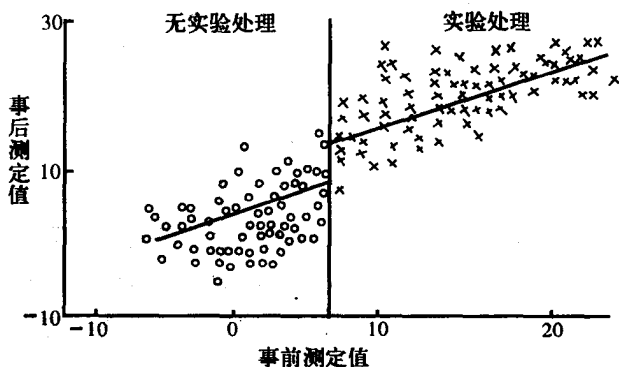


图 6-3 回归间断点设计的准实验结果

西弗(Seaver, W. B.)等(1976)曾经采用回归间断点设计，

样本，应尽可能从同一总体抽选，或者选用相似样本。如果被试条件过分异质，需要采用匹配程序，对被试进行必要的匹配和分组。但是，匹配程序有可能造成统计回归效应，并且容易在各因素之间引入微妙的交互作用，影响研究的效度。此外，在不等同对照组设计中，很可能发生补偿性竞争和被试历史因素等效应。这些都是在运用这种准实验设计时应加以重视的。

由于上述问题，不等同对照组设计的统计分析程序，比较注意使被试选择所造成的差异与实验处理的效果区别开来。常用的方法是基础方差分析（ANOVA）、协方差分析（ANCOVA）、“匹配型方差分析”和“得益分数方差分析”等，这些方法分别适用于不等同对照组设计的不同变式。

在对不等同对照组设计的研究结果进行分析和比较时，应该注意一些可能引起偏差和误估的因素。在进行基础 ANOVA 分析时，常常会由于实验处理和被试特征差异之间的相互作用而混淆实验处理的真正效果。这使得基础 ANOVA 的解释有一定局限性，其发现真正差异的能力也比较低。ANCOVA 分析比基础 ANOVA 提高了一步，在分析中包含了事前测定，从而调节了组间差异，减少了误差变异。可是，在 ANCOVA 中，事前测定的测量误差、行为特征的不稳定性以及行为结构的变化等因素，容易使分析结果出现偏差。匹配型方差分析适合于采用了被试匹配程序的不等同对照组设计，在这一分析中，应特别重视匹配程序所带来的系统性偏向和可能发生的回归效应。得益分数方差分析集中于检验“事后测定值—事前测定值”所得的得益分数，这一分析的弱点在于不能检验实验处理与事前测定的相互作用及影响。

五、间歇时间序列设计

间歇时间序列设计是在实施实验处理前后的一段时间里对研究

对象作出多次重复观察或测定，并通过对整个时间序列的测定结果的比较，确定实验处理的效果。最基本的间歇时间序列设计包括一个实验组和实验处理前后的多次观察，如图 6-5 所示。

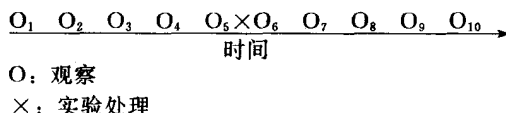
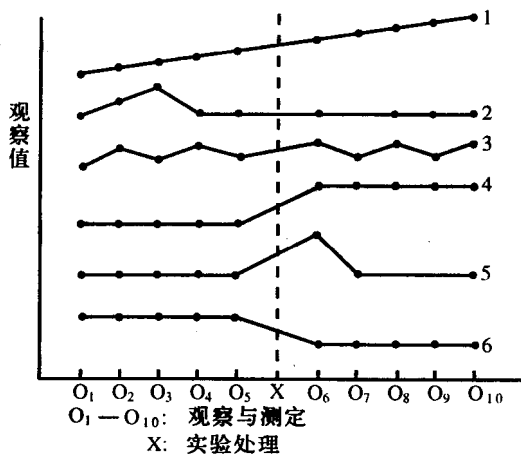


图 6-5 基本的间歇时间序列设计

如果我们要研究奖励制度对工作积极性的效果，可以在实施该制度前后分别定期进行几次工作积极性测定，从而比较各阶段测定结果的差异，决定奖励制度的有效性。研究表明，间歇时间序列设计的主要优点，是能够在实验处理以前评估被试成熟的趋势，并在较长的时间跨度上检验自变量与因变量之间的关系，有利于作出比较肯定的因果关系判断。这种准实验设计要求包含一定数量的观察与测定，从整体发展趋势上分析实验处理的效果。在采用这种设计时，应特别注意周期性因素的影响（季节、态度、情绪等方面的变化），同时，应尽可能使被试保持稳定，防止被试更换或淘汰的情况发生。

这种设计也带来一些影响内部效度的因素，例如，实验处理实施之后可能发生的历史因素、多次测定中测量操作程序的变化，等等。为了控制这些因素，可以把时间序列设计与不等对照组设计结合在一起；或者可以转换或撤除实验处理，观察各组的变化，这些都是时间序列设计的一些较有效的变式。

采用间歇时间序列设计的准实验，结果的分析应特别谨慎，需要从实验处理前后的整个发展变化趋势来评估其效应，而不能只看与实验处理相邻的前后两次观察值的差异。图 6-6 表示了间歇时间序列设计可能获得的六种结果，可以看出，前三种（1-3）结果都不能说明实验的效应，而后三种（4-6）结果的变化可以归因于实验处理（X）。其中，第四种结果说明实验具有稳定的正效应；第五种结



O_1-O_{10} : 观察与测定

X: 实验处理

图 6-6 间歇时间序列设计的六种可能的结果

果表现出实验处理的短暂的作用；第六种结果则表明了负效应。

间歇时间序列设计的准实验结果的统计分析，需要对实验处理前后的一系列观察值作出检验和比较。除了简单易行的 t 检验以外，研究者建议采用“自回归整合移动平均模型”（ARLMA）及有关的技术进行时间序列分析，并作出模型，这种方法可以较准确地决定实验处理的实验效果。由于 ARLMA 方法比较复杂，需要较多数学知识，在这里就不作详细介绍 ARLMA 为 autoregressive integrated moving average models，也有人缩写为 ALMA）。

第三节 实验室实验设计

一、实验室实验的一般问题

实验是心理科学研究的最重要方法之一，它可以在实验室进行，

也可以在现场背景中展开。这里主要讨论实验室实验的一般问题。

1. 实验室实验的特点

实验室实验具有三个主要特征：

- (1) 实验者创设情景并操纵事件或变量；
- (2) 通过控制手段确认变异源，包括实验操纵和统计控制两种方式；
- (3) 实验室环境中能最有效地运用实验控制策略。

这里，有必要区别实验室和实验这两个容易混淆的概念。实验室是一种研究背景；实验则是一种研究策略。实验室实验有许多优点，其中主要的优点是可以运用各种与实验策略有关的控制手段；同时，有条件较准确地确定变量之间的因果关系。

2. 实验室实验的形式

实验室实验主要有两种形式：标准实验室实验和模拟实验。标准实验室实验是在严格控制条件下操纵自变量和测定因变量，从而观察和确定实验处理的效应。这种标准实验室实验往往集中于自变量的特定刺激特征。模拟实验是在实验室条件下模拟各种实际事件和情景，观察和确定行为的特征和过程。模拟实验又分为“自由模拟”和“实验模拟”。自由模拟法创设了尽可能现实的复杂环境并允许被试自由地按一定规则行事，整个实验集中于考察实验参数和被试行为之间的交互作用 and 影响，实验模拟法克服了标准实验室实验中过分人为的因素，也弥补了自由模拟法对变量缺乏控制的弱点。它既对自变量具有相当程度的控制，又保持了自由模拟的某些复杂的情景特征。因此，实验模拟设计对所有实验信息事先程序化并能包含大量的变量，它还使被试较大程度地体验到真正介入了实验任务。在心理学领域的研究中正越来越多地采用实验模拟法。

3. 实验室实验的程序

实验室实验应遵循一定的程序和策略，其中比较重要的有以下

几点：

(1) 有关实验程序的描述和说明，包括变量的操作和测定方式、实验目的说明、被试选择程序以及实验步骤，等等。

(2) 训练被试，通常条件下，被试并不熟悉实验环境，因此，需要对他们进行一定的训练。例如，让被试体验实际情景、参加若干非模拟性的活动，等等。

(3) 实验背景的设置，背景条件的设计、实验用品的配备、观察记录仪器的安装等都应考虑到使被试能够更大程度地介入实验活动。

二、被试间设计和被试内设计

在心理学实验设计中，有两种常用的分配被试的方式，一种方式是把不同的被试分配到不同的实验条件或水平，称为被试间设计；另一种方式把相同的被试分配到不同的条件或水平，称为被试内设计。图 6-7 表示这两种分配被试的设计。图中，8 位被试分别或同时分配到不同的条件或水平，形成两种不同的设计。实验设计的基本要求之一，是使外在的或非控制的变异减少到最低程度，从而提高实验的内部效度，应按照实验的具体要求和条件，选择适当的实验设计。

被试间设计			被试内设计		
被试	条件 1	条件 2	被试	条件 1	条件 2
	1	5	1	1	1
	2	6	2	2	2
	3	7	3	3	3
	4	8	4	4	4
			5	5	5
			6	6	6
			7	7	7
			8	8	8
被试间设计			被试内设计		

图 6-7 被试间设计和被试内设计

1. 被试间设计

这种设计比较安全、保守。在被试间设计中，两种或多种实验条件之间不可能通过被试相互“污染”。可是，在采用这种设计时，应使两组被试的条件尽可能相似，可以通过匹配程序或者随机分组程序，做到这一点。

2. 被试内设计

许多实验心理学家更喜欢采用被试内设计，因为在这种设计中，每一位被试在不同实验条件下与自己相比较，不需要分组，实验效率比较高。但是，被试内设计很有可能产生实验处理的污染，两种或多种条件可能相互影响，降低研究效度。一般来说，可以通过随机化或平衡法，克服与抵消潜在的干扰效应。

在决定一项研究是采用被试间设计还是被试内设计时，需要考虑各方面的条件和因素。在被试的个体差异比较大的情况下，宜选用被试内设计；当实验处理可能造成污染时，应采用被试间设计。另外，还要分析实验中的情景差异和可能产生的影响。不少研究表明，把被试间设计与被试内设计结合在一起，可以获得更好的实验效果。例如，对实验中的一部分容易引起迁移效应的变量实行被试间设计，而对另一些变量运用被试内设计。

三、因子式实验设计

因子式实验设计是比较复杂的实验设计类型。由于这种实验包含着多个因子和因子的多种水平，因而适合于研究复杂情景下因子的交互作用。这种实验设计还能够在不同程度上表现现实生活中各种因素之间的复杂关系，具有一定的生态效度，成为心理学研究中重要的实验手段。

1. 因子式实验设计的一般问题

因子式实验设计是较多采用的较严格的实验设计。所谓“因

子”，是指构成实验设计的基础变量，它不同于“因素分析”里所说的“因素”。因子主要有两种类型：顺序型和称名型（非顺序型），每一类型又具有不同水平。称名型因子可以是能够操纵的变量（例如不同教学方法），也可以是不能操纵的类别变量（例如性别）。对于这后一种类别因子，并不能加以实验控制，因为不同类别属于不同总体，不是实验者可以分配或操纵的；而可操纵的称名型因子则可以从单一总体作出随机分配或控制，在实验设计中至少应有一种可操纵的因子。

在进行因子式实验设计时，应注意以下几方面的问题。

（1）因子与水平的组合

当实验包括多个因子而每一因子又具有多种水平时，就应该注意因子与水平的适当组合。例如：在研究三种教学方法与四种奖励方式之间的关系时，就形成教学方法和奖励方式两种因子不同水平的组合（ 3×4 ）。每一种包括实验者可操纵因子的组合称为“处理”（treatment）。如果我们以男、女两种水平的类别因子结合上述两种不同水平因子进行因子式实验设计，就可以得到 24 种因子与水平的组合（ $3 \times 4 \times 2$ ），其中包含着 12 种处理（ 3×4 ），被试被随机分配到每一种“处理”条件。

（2）随机化

我们知道，实验者对实验效果的判断往往带有人为的偏差，而造成误估。因此，为了避免在不同处理效果的比较中造成偏向，并使各处理效果之间差异的统计检验具备有效的基础，即获得统计与概率论所要求的随机化样本，应该把被试样本随机地分配到不同处理条件。简易的随机化，是采用投钱币或骰子的方式，但最好的办法是使用随机数表，这种表在大多数统计书中可以找到。

2. 因子水平之间的交互作用

因子式实验设计中，常常会发生因子之间的交互作用。这时，对

A 因子某一水平上效果的解释就会与 B 因子特定水平的情况密切相关。因子水平之间的交互作用可以分为顺序性和非顺序性两类。图 6-8 为顺序性交互作用，图 6-9 为非顺序性交互作用。

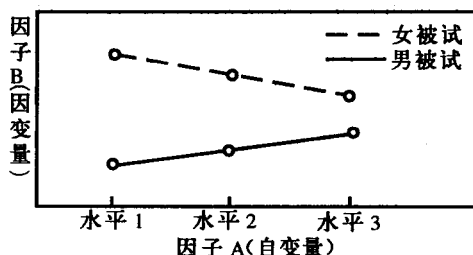


图 6-8 顺序性交互作用

图 6-8 表明，在发生顺序性交互作用时，女被试在因子 A 的各个水平上都优于男被试，但两者的差异，在不同水平上有顺序性的系统变化。图 6-9 的非顺序性交互作用则不同，各水平之间不存在这种变化。非顺序性交互作用表明，某因子特定水平的效应取决于另一因子的不同水平。有的研究者还把因子的主效应称作零阶交互作用，而把两因子交互作用称为一阶交互作用，三因子为二阶交互作用，等等。

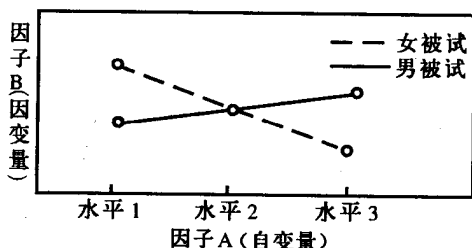


图 6-9 非顺序性交互作用

因子之间的交互作用往往比主效应更为重要和具有意义。人们可以根据这种交互作用，对因子不同水平在具体条件下的作用和效应作

出全面深入的分析和了解，高阶交互作用往往比低阶交互作用提供更多信息。

3. 被试间因子式实验设计

被试间因子式实验设计是让不同被试分别进入不同因子条件，从而比较实验的结果。在设计时，需要逐步形成有关因子效应的假设，并确定因子的水平。下面以一项记忆实验为例，说明这种实验的设计过程。

心理学家发现，心理映象在心理活动中起着很重要的作用，例如，心理映象能提高记忆能力，增加所记忆的词汇量。可是，心理映象是不是对所有类型的词汇都具有同样的效应呢？人们提出，有些词（如厨房、桌椅等）比较形象化，容易建立心理映象；另一些词（如命运、标准等）则比较抽象，难以形成心理映象。前一类词汇称为高映象词；后一种词汇属于低映象词。因此，研究者作出假设：创设心理映象，对于高映象词比对低映象词的记忆效应更大。这样，可以进行一项两个因子的实验：心理映象策略和词汇特征。每个因子都取两种水平，形成 2×2 的因子式实验设计：

创设心理映象 (策略) $\times$ 高映象词 (词特征)
不创设心理映象 (策略) $\times$ 低映象词 (词特征)

把被试分别分配到上述四种条件，测量他们的记忆成绩，并运用方差分析，检验心理映象策略与词特征的“主效应”和各因子之间的交互作用。

4. 被试内因子式实验设计

被试内设计常常能节省被试并控制个体差异。由于被试在记忆能力方面差异比较大，因而有关记忆的实验多用被试内的设计。心理学家经常用字谜游戏作为任务，研究记忆和问题解决能力等课题。人们认为，心理映象也会影响解字谜任务的效率，高映象词的字谜可能比低映象词更容易解决。字谜任务的另一个因素是任务本身

所提供的线索。一种称为“结构线索”，对词的文字结构提供暗示；另一种叫做“语义线索”，暗示答案词的语义类别。于是，可以作出这样的研究假设：结构线索更有利于低映象词的记忆，而语义线索则对高映象词更有帮助。由此作出 2×3 的被试内因子式实验设计：

高映象词
 低映象词

(词特征) $\times$ 结构线索
 语义线索 (线索类型)
 无线索

这里，“词特征”因子设立两种水平，而“线索类型”因子包含三种水平。为了防止不同线索类型之间的相互影响，要求运用随机化或平衡化程序，使被试随机地或有相等机会接受不同任务。有人曾用上述设计进行实验，获得图 6-10 的结果，表明“无线索”条件对两类词（特别是低映象词）都不利；“语义线索”对高映象词的字谜更有帮助；“结构线索”对两种词都有积极效应；任务线索可以提高解题效率，但两种线索类型之间存在非顺序性交互作用。这类因子式实验也主要采用方差分析（ANOVA 等）对结果进行统计分析。

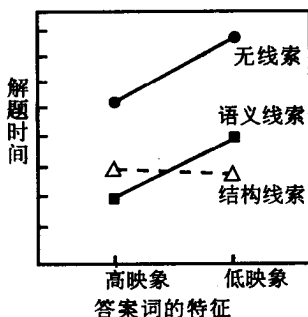


图 6-10 2×3 因子式实验结果图示

四、小 N 实验设计

小 N 设计实际上就是小样本范式的实验设计，在心理物理学研

究和操作条件反射的研究中经常被人采用，也是其他许多实验心理学研究常用的设计。

小 N 设计主要有三种类型：反转型设计、扩展式反转设计和多重基线设计。反转型设计又称为“ABA 设计”（A、B 表示不同实验阶段），其基本思路是：先取得被试行为的基线值；然后引入自变量，并观察被试的行为变化；最后，把实验条件反转到原来基线状态，再一次测量行为。有的研究为了使结果更有把握，采用包含四个阶段的反转设计，即 ABAB 设计。

小 N 实验像一些被试内设计的实验一样，常常会遇到“实验处理污染”的麻烦，使得难以采用一般的反转型设计。例如，如果在 B 阶段引入的实验处理具有长时效应，就无法作反转检验。在这种情况下，可以运用“扩展式反转设计”。这种设计使用多个自变量，并创设多个基线阶段。比如，有一种 ACABCBCB 的扩展式反转设计，其中 A 为基线阶段，B、C 分别包含着不同的自变量。这种设计能以考察不同自变量的效应，并在一定程度上控制了实验处理污染。可是，在自变量具有强污染效应的情况下，扩展式反转设计就无能为力了。这种条件下需要改用“多重基线设计”图 6-11 表示出这种设计的框架。多重基线设计的特点是，在引入自变量之前，不同的行为（或被试）具有不同的基线期；而且，这种设计既可用于被试内设计，也适合于被试间设计，从而能克服前面提到的“污染”问题。因此，被试间多重基线设计适用性更强。

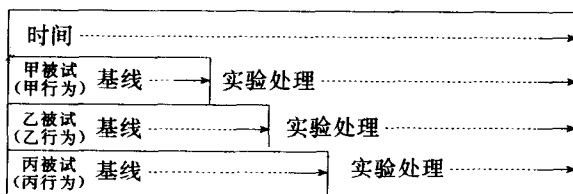


图 6-11 多重基线设计图示

五、拉丁方和希腊拉丁方设计

拉丁方设计原来主要用于农业研究，例如不同土壤配方的试验。

60年代以来，心理学研究中也常常采用这种设计，其主要目的是为了控制顺序效应。可是，近些年来的研究表明，拉丁方设计本身，往往带来不少影响研究效度的因素和统计检验分析方面的问题，特别是使实验处理之间的交互作用复杂化。人们认为，在心理学研究中，应该非常谨慎地运用拉丁方设计，防止滥用。

传统的拉丁方设计为图 6-12 所示。实际上，这种拉丁方设计使每一种条件（A、B、C、D）都按相同顺序依次排列，这样非但没有能消除顺序效应，反而人为地强化了顺序效应。为此，研究者提出了一些新的拉丁方设计，其中效果比较好的是“平衡式拉丁方”设计，使每一条件相等频次地先于或跟随于其它条件。图 6-13 表示了平衡式拉丁方设计的两个例子。

		顺 序			
		一	二	三	四
被 试 数 目	1	A	B	C	D
	2	B	C	D	A
	3	C	D	A	B
	4	D	A	B	C

图 6-12 传统的 4×4 拉丁方设计

		顺 序 A			
		一	二	三	四
被 试 数 目	1	A	B	C	D
	2	B	D	A	C
	3	C	A	D	B
	4	D	C	B	A

		顺 序 B			
		一	二	三	四
被 试 数 目	1	A	B	D	C
	2	B	C	A	D
	3	C	D	B	A
	4	D	A	C	B

图 6-13 两种平衡式拉丁方设计

在实验条件数目为偶数时，平衡式拉丁方设计比较容易；当实验条件为奇数时，每一被试必须在每一条件中测定两次，才能获得平衡式拉丁方。例如，当采用五种实验条件时，可作出图 6-14 的设计。

		方格 1					方格 2				
被 试	1	A	B	E	C	D	D	C	E	B	A
	2	B	C	A	D	E	E	D	A	C	B
	3	C	D	B	E	A	A	E	B	D	C
	4	D	E	C	A	B	B	A	C	E	D
	5	E	A	D	B	C	C	B	D	A	E

图 6-14 五种实验条件时的拉丁方设计

希腊拉丁方设计是以正交方式结合了两个拉丁方，使每一因素的各种水平交错在一起。图 6-15 为 3×3 的希腊拉丁方设计。

	1	2	3
I	Aa	Bc	Cb
II	Bb	Ca	Ac
III	Cc	Ab	Ba

图 6-15 3×3 的希腊拉丁方设计

这一希腊拉丁方设计的各单元并不重复，每一种实验处理只作了一次观察和测定，因此，希腊拉丁方设计比拉丁方设计更严格地控制变异源。

由于上述设计的某些局限性，无论是拉丁方设计，还是希腊拉丁方设计，在心理学研究中都常常只是作为整个实验设计的一个部分，而不是单独作为某实验的完整设计。

在本章的讨论中，我们强调了这样的思想，实验设计是完成整个实验的重要一环，在实验设计中应同时考虑到实验结果的分析问

题，这样才能真正发挥实验设计的效益。正如许多心理学家指出的，高超的实验设计必须与实验结果的分析能力密切联系在一起。

第六章思考题：

1. 现场研究有哪些重要特点？这些特点对心理学研究的意义是什么？
2. 准实验设计的主要类型各有何特点？在设计时应注意哪些问题？
3. 讨论因子式实验设计的特点和类型。设计一项因子式实验。
4. 比较“被试间设计”和“被试内设计”两种设计，说明它们各适合于什么实验条件。
5. 多重基线小 N 实验设计有何特点？
6. 拉丁方设计的目的是什么？平衡式拉丁方设计有何特点？

第三编 测量理论和方法

心理学科学研究离不开对心理活动的定量测定和定性分析，这就不但需要系统的心理测量理论作为基础，而且要求运用准确的采集数据的方法。在这一编中，具体介绍和讨论心理测量理论和几种常用的数据采集方法。在心理测量理论中，经典测验理论仍然在心理学研究中颇为流行，斯皮尔曼的分数模型则是这种理论的典型。随着测量要求的提高和研究实际的需要，现代测验理论也得到了很大的发展，其中最重要的是潜特征理论和项目反应理论。测量的信度与效度是心理测量的基本特点，研究者的一项重要任务，就是设法提高信度与效度，使心理测量更为准确、可靠。心理测验是一种比较系统的心理测量，但正确有效地设计与使用心理测验，却不是一件容易的事。这里，特别需要把测验与治疗或教育结合在一起。心理学研究中较常用的方法是观察法、访谈法和问卷法。近年来，由于人们在研究中普遍重视认知过程及其特点，口语报告法越来越多地得到应用，并且不断地精细化，成为心理学许多领域的重要研究手段。测量理论和方法的应用，大大提高了心理学研究的科学水平，促进了心理学理论的发展。

第七章 心理测验理论与测量方法

第一节 经典和现代测验理论

著名心理学家勒温曾经指出：“没有什么比好的理论更加实用的。”心理测量理论涉及许多重要课题：信度理论、效度模型、预测策略、效用理论、量表理论、测验设计、潜特征模型、反应定势与风格、心理物理测量、测验的项目分析，等等。在心理测量理论和测量技术的发展中，主要有两方面重要的内容，这就是经典测验理论和现代测验理论。

一、经典测验理论

经典测验理论（classical test theory）及其方法，起源于斯皮尔曼在本世纪初有关智力和智力测验的著名研究，至今已有七八十年的历史。目前流行的一些心理测验和有关测量信度与效度的方法，都是以经典测验理论作为基础的。经典测验理论包括一些不同的理论模型，其中影响最大的是斯皮尔曼提出的分数模型。

1. 斯皮尔曼的分数模型

斯皮尔曼的分数模型提出，任何心理测验所得到的分数（ X ），都是由反映对象稳定的心理特征的“真分数”（ T ）和由随机因素造成的误差分数（ e ）这两种成分所构成。斯皮尔曼的分数模型表示为：

$$X = T + e$$

这个分数模型依据三个基本假设：（1）模型中的真分数（ T ）是

稳定的，对相同对象多次重复测量的误差分数（ e ）呈正态分布；（2）真分数和误差分数是相互独立、互不相关的，两者之间没有显著相关（ $r_{Te} \approx 0$ ）；（3）真分数和误差分数是以简单可加的方式结合在一起。

以后，其他心理学家对斯皮尔曼的分数模型作了各种修正和补充，促进了经典测验理论的发展。例如，对分数模型中误差分数作进一步分析，把测量误差分为“不稳定误差”（ es ）和“不一致误差”（ ec ），提出新的测验理论模型：

$$X = T + es + ec$$

不稳定误差与测量过程中的情景和时间等因素的影响有关，由疲劳、焦虑和记忆的偶然性波动等引起；不一致误差主要与测量程序本身的差异有关，由重复测量时测验操作不同而引起。不一致误差一般与被试的个体差异无关。

2. 经典心理测验理论的主要局限和问题

随着心理学研究的深入，经典心理测验理论暴露出一些局限和问题。人们因此对分数模型理论的假设提出了疑问，并试图建立更切合心理测量实际的测量理论。经典心理测验理论主要有以下局限和问题。

（1）测量结果特定于具体测验

经典心理测验理论的问题之一，是测量所得的结果随测验项目不同而变化很大。这使得在研究中难以获得有关心理特征或心理活动过程的普遍性测量，不同测验的结果也无法直接比较，因而限制了心理测验潜力的发挥。

（2）测量参数特定于具体样本

测量参数主要指测量的信度、效度、测验的辨别力与难度等指标。经典心理测验理论的又一问题，是按照这个理论所确定的测量参数，都特定于具体的样本，随着样本的变化而不同。这样，两个

样本时就会得到两种测验难度与辨别力参数，如果研究包含多个样本，就很难决定真正的测验难度与辨别力，也无法取得统一、稳定的测量信度系数和效度估计，从而造成测量问题，影响研究的质量。

（3）忽视项目反应的模式

经典测验理论对所有被测者提供了笼统的测量标准误（standard error of measurement），在很大程度上忽视了个体差异以及测量中各种随机因素的影响。同时，经典测验理论是在整个测验的水平上分析测验结果，忽视了不同的项目反应模式，混淆了相同测验分数所包含的不同性质的特征。研究表明，相同数目的正确反应很可能是由于不同反应模式的结果，而这种反应模式恰恰反映出真正的心理特征或者某种心理定势。例如，同样的测验分数，有的人是由于某种反应定势而答对题目，有的人只是答对了低辨别力的项目，有的人则能反映出真实的心理特征。

由于上述局限和问题，经典测验理论没有能够准确地解释和解决许多测量问题；同时，在心理学研究中又日益强调以单一项目作为基本测量单位，并建立多种测验项目库。因此，人们不断寻求新的测验途径，使得心理测量理论在停滞了一个时期以后，出现了新的进展，这就是现代测验理论的形成、发展和应用。

二、现代测验理论

现代测验理论克服了经典测验理论的不足方面，实际上运用了贝叶斯统计的思想。现代测验理论中最重要的是潜特征理论（latent trait test theory）以及潜特征理论最近的发展，即项目反应理论（item response theory）。

1. 潜特征测验理论

这类理论起源于本世纪 20 年代中期瑟斯顿（Thurstone, L. L.）有关“心理年龄量表”的研究。40 年代起，莫热（Mosier）和古特

曼 (Guttman, L.) 等也在这方面作了许多研究, 促进了测验理论的发展。近年来, 随着电子计算机技术在心理学研究中的广泛运用, 使潜特征测验理论得到了迅速发展、应用和完善。

潜特征理论认为, 测验成绩在不同程度上反映出人们的心理特征。这些特征是不能直接观察和测定的, 所以称为潜特征。这一理论的主要目的是确定测量结果与那些不可观察的特征之间的关系。

潜特征测验理论有三个重要概念。

(1) 潜特征空间的维度

如果单一能力或特征可以说明测量结果, 则称为单维, 否则为多维。如果我们对项目进行预选, 往往可以增加单维的可能性, 人们把对测验项目的因素分析后第一因素方差与第二因素方差的比率作为潜空间维度的指标。

(2) 局部独立性

局部独立性是假定被测者对某一测验项目的正确回答的概率不受其他项目成绩的影响, 可以用 X^2 检验法检验项目反应的局部独立性。

(3) 项目特征曲线

项目特征曲线是项目分数对于能力 (θ) 的回归线, 它说明了从具有特定潜特征的总体随机选取的受测者作出特定反应的条件概率。例如, 第 i 项目的条件概率为:

$$P_i(\theta) = \frac{1}{1 + e^{-a_i [D(-b_i)]}}$$

其中 a_i 、 b_i 为项目参数 (辨别能力和难度);

D 为量表因素, 常定为 1.702。

项目特征曲线并不受样本的影响, 有关的参数独立于受测者的能力分布。因此项目特征曲线及其参数具有“不变性”这是潜特征理论最吸引人的特点之一。在单维测验中, 项目特征曲线又叫做

“测验特征曲线”；而在多维测验中，称为“测验特征函数”。

2. 项目反应理论 (IRT)

项目反应理论是潜特征测验理论的新发展，也是目前应用最多的。这个理论指出了个体某种反应类型的概率与个体某些特征之间的关系，它的基本思想与潜特征理论其他许多模型是一致的。

项目反应理论主要有以下几个基本特征。

(1) 测量指标与反应者的潜特征 (θ) 联系在一起

这里所谓的潜特征，是指不可观察的统计构思，而不是独立存在的心理学或生理学的实体。例如，认知测验所测定的“认知能力”、态度问卷所测定的“工作满意感”等都是一种潜特征。项目反应理论能够确定实际测量指标与这类潜特征的概率关系。

(2) 用数学函数确定项目特征参数

项目反应理论是以数学推导所得的函数作为基础的，有的运用“累积正态分布函数”，有的则采用了对数函数。它以三个重要参数描述项目特征曲线：项目辨别参数 (a_i) 即曲线的斜率，斜率越大，项目的辨别力越高；项目难度参数 (b_i)，这是项目正确反应概率 $P(\theta) = .50$ 位置所对应的能力量表值；猜测参数 (c_i)，如果在能力量表值为零时仍然有正确反应，就认为存在猜测因素，项目以能力量表值为零处的“正确反应”概率值作为项目猜测参数，这一参数对于包含多重选择的项目特别重要。图 7-1 是某测验三个项目的特征曲线，说明了上述三个参数在特征曲线中的关系。

图 7-1 中的纵坐标表示对测验项目作出正确回答的概率，横坐标则表示能力量表，这里的能力已作为潜特征，以各项目在正确反应概率为 .50 处的能力水平，作为项目难度参数。可以看到，项目 1 和项目 2 的辨别力都比较高，因为能力水平稍有差异，特征曲线所对应的概率值就会有明显的变化，从而可以从正确反应的概率灵敏地反映受测者的能力水平。图中项目 3 的斜率小得多，因而辨别力

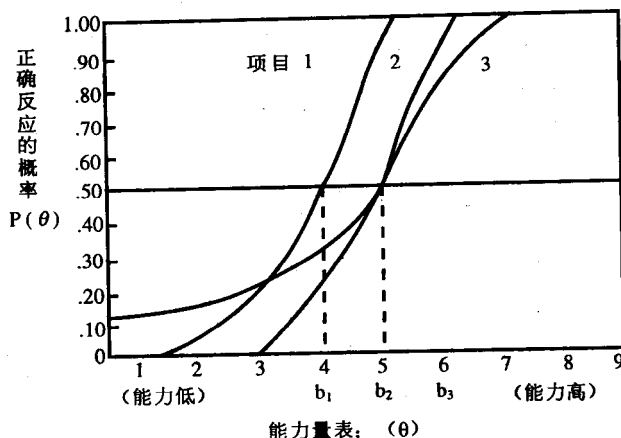


图 7-1 三个项目的项目特征曲线

也比较低。项目特征曲线在 $P = .50$ 处对应的能力水平表明，项目 2 和项目 3 的难度相同，而项目 1 难度较低。项目 3 在能力 (θ) 为零值时仍有一定的“正确反应”概率，表明在这个项目上有猜测。理想的测验需要使项目辨别力尽可能高，所包含的项目应既有高难度的，也有中、低难度的，但尽可能没有猜测因素。当猜测因素不可避免时，可以设法进行校正，确定猜测的允许范围。项目参数的估算可以采用迭代逼近法或递次逼近法，目前已有各种计算机程序。

(3) 通过项目信息函数来推算测量的信度和标准误

对于每一个测验项目，都可以计算“项目信息函数”，信息函数大的项目能提供高信息量。这样，就能在测验或量表设计中选择信息量最高的项目。研究表明，项目特征曲线的斜率越大，信息函数也越大。项目信息函数以所有项目参数作为基础，表示出测验项目对于不同能力水平的测量效力。项目反应理论通过项目信息函数来推算测量的信度和测量的标准误，为各种测量指标提供了稳定的测量参数。

此外，项目反应理论不象经典测验理论那样在整个测验的水平上建立测量模型，而是在项目水平上作出分析，并运用定量化的项目与能力参数，对项目反应作出定性的解释，从而大大提高了测量的精确度和心理学意义。

三、现代测验理论在心理学研究中的应用

现代测验理论在心理学研究中正在得到日益广泛的应用，其中应用比较多的课题有测验项目的分析、成绩测评量表和态度问卷的设计、测验项目库的设计，以及社会心理气氛和管理作风的研究，等等。这里介绍一些应用的实例。

1. 测验项目的分析

在测验项目的分析中，可以采用“项目-测验回归”的方法来确定项目难度和项目辨别力。这里以一个 10 项目的测验为例，测验项目的反应分为正确和错误两种。表 7-1 为 10 个项目中两个项目测定的正确反应比例。

表 7-1 某测验项目分析的数据

总 分		1	2	3	4	5	6	7	8	9	10
正 确 比 例	项目 A	.50	.10	.20	.25	.40	.60	.75	.88	.95	1.00
	项目 B	.00	.00	.00	.00	.00	.05	.20	.40	.82	.95

表 7-1 数据的含义是，测验总分分别为 1—10 分的被试答对项目 A 和项目 B 的概率。例如，测验总分为 8 分的人当中，有 88% 的人答对了项目 A 而只有 40% 的人通过了项目 B。根据表 7-1 提供的数据，运用“项目-测验回归”方法，可以作出图 7-2 的近似的项目特征曲线。从图 7-2 可以看到，项目 B 的斜率大于项目 A，项目 B 具有更高的辨别力；项目 A 在概率为 .50 处对应的能力水平低于

项目 B，因而难度较低，并且可能包含猜测。

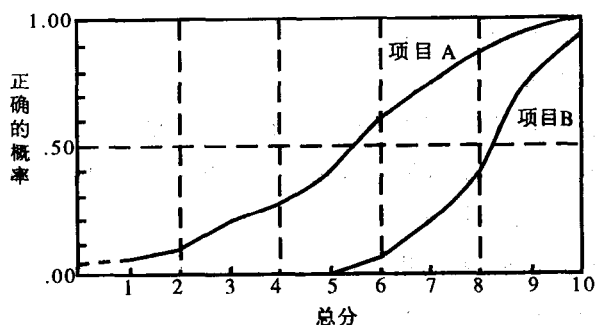


图 7-2 项目 A 和项目 B 的项目—测验回归结果

这种项目测验分数的回归图，运用了项目反应理论，得到近似的项目特征曲线。它不仅把项目难度和辨别力等信息结合在一起，而且使得能够完整地理解项目成绩与总分数之间的关系。

2. 成绩评估量表的设计

现代测验理论被广泛地应用于作业成绩评估研究和评估量表的设计。其中最常用的是潜特征理论的拉希模型 (Rasch model)。这个模型假定所有项目都具有较高的辨别力，认为项目辨别力的差异和猜测因素在一定条件下是可以忽略的。因此，拉希模型只使用一个难度参数，简化了项目分析程序，成为比较实用的测验模型。

在运用现代测验理论进行评估量表设计时，首先设计较多数目 (150 个左右) 的备择项目，并以五点量表记分，对被评估的对象进行试测；然后，对试测数据进行聚类分析或其它类似的统计分析，找出四个维度，并确定难度参数。在分析时，把五点量表分解成若干个 0 或 1 的两分小项目。例如，在某项目上答 1 或 2—5，答 1—2 或 3—5，等等。这样每个项目可以分解成四个两分小项目，并作出小项目的特征曲线。如果曲线 1—3 在正确反应概率 $P=.50$ 水平上基本上距离相等，而曲线 3、4 之间的距离大得多，这就表明 3—4 的

评估增量比较大，需要进行修正，使之等距。这类分析在各种量表的设计中都可以进行，能够使量表设计和项目库的配备标准化和科学化，具有较高的应用价值。

3. 态度问卷的设计

潜特征测验模型对态度问卷的设计有很重要的意义，有人以计算机模拟产生数据来编制态度量表。这方面的应用研究证明，潜特征分析具有传统测量方法所没有的优点。例如，它可以对每一态度特征的测量提供精确的指标；它还可以编制出一组项目，使得被测者可以在回答问卷时省略不了解的项目或无关项目而不影响整个问卷测量结果的质量。

近年来，项目反应理论在心理学研究和各种考试试题与评估表的设计中得到日益广泛的应用。特别是专用计算机软件的开发，进一步促进了这一理论的推广与发展，对各种参数模型的估算亦更为精确。项目反应理论还成为心理学研究中进行“潜结构分析”的重要理论模型。

第二节 测量的信度和效度

一、测量信度和效度的概念

测量是心理学研究中的一个重要环节，是进行各种心理学科学分析的基础。心理测量的可靠性和准确度，也就成为研究者十分关心的问题。例如，在对从事某项工作时心理负荷的测量中，首先要求所测量的指标能真正反映心理负荷的高低；其次，希望测量所得的数据有一定的可靠性，如果重新测试，能获得比较一致的结果。这些就是测量的信度和效度问题。

1. 测量的信度

测量的信度即可靠性，它是测量反映被测特征真实程度的指标，有人称之为测量的准确性，也有人把信度定义为测量的一致性，有关测量信度的比较标准的定义为：测量信度是指测量结果反映出系统变异的程度。这里所说的系统变异是指个体在心理特征方面的真实个体差异。高信度的测量很少受到随机因素或事件的干扰，能够准确无误地测量出人们的心理特征和各种心理过程。

测量信度的估算，以相关系数（ r_{xx} ）作为指标。相关系数 $r_{xx}=1.00$ 时，标志着测量完全可靠，当然这种情况在心理学研究中很少会出现。一般来说，当 r_{xx} 在统计上达到显著水平，就认为具有较高的信度，否则，就表明信度比较低；当 $r_{xx}=0$ 时，说明完全不可靠，信度为零。根据经典测验理论的分式模型，信度系数可以表示为总变异量中系统变异量所占的比重，即：

$$r_{xx} = \frac{S_x^2 - S_e^2}{S_x^2} = 1 - \frac{S_e^2}{S_x^2}$$

其中 S_x^2 为测量的总变异量， S_e^2 为误差变异量。可见，测量误差越小，信度也越高。

2. 测量的效度

测量效度是指心理测量的有效性，即测量到的是不是所要测定的心理特征。如果一项语言能力的测验测得的竟是机械能力，那么这种测验就完全没有效度。效度的定义方式也比较多，严格地说，效度涉及到从测量分数可以适当地作出何种有关心理特征的推论。因此，效度是指从心理测量作出推论适合度。推论总要依照一定的标准。所以，效度也是指测验与某种外部标准之间的关系程度。

效度的指标比较复杂一些，有的也用相关系数表示，往往是测量分数与一定外部标准之间的相关。效度按评估目的和用途不同，可分成几种类型。常用的有内容效度、同时效度、预测效度和构思效度。70年代以来，一种新的效度形式——聚合效度与辨别效度，在

心理学研究中应用得越来越多，成为重要的效度类型。

测量的信度与效度并不是完全独立的，高信度的测量并不能保证具有高效度。比如，用温度计测定操作工的体温时，多次测量结果很一致，因而信度很高；可是用这支信度很高的温度计去测量操作工对于职业病的态度，那就效度很低，应该说是完全无效。可见，信度是良好效度的必要条件，却不是充分条件。

二、测量信度的评定

在进行心理测量时，需要对测量的信度和效度作出评定，从而检验测量的准确性。实际上，我们要求每一项心理测验都必须在其使用手册中说明测量的信度、效度和常模，也要求心理学研究报告或论文中能表明所用测量的信度系数和效度的评估结果。

在评定测量信度时，需要区分两种测量误差：（1）测验内误差变异，主要由项目的取样、数目、难度以及项目间的相关程度与一致性等因素所造成；（2）测验间误差变异。测验在实施时间与程序、内容差异以及记分方面的因素会引起这类变异。这样，我们就能对信度系数的真正含义作出解释与分析，使信度指标为研究提供更多信息。

评定测量信度的方法比较多，常用的有重测信度、等同信度、折半信度和内部一致信度等。

1. 重测信度 (test-retest reliability)

重测信度是在一定时间间隔中运用同一测验作重复测量所得到的信度系数。例如，为了检验“操作技能测验”的信度，可以用这一测验对职工操作技能作两次测定，这两次测定间隔一定的时间，然后计算两次测验分数之间的相关，从而得到信度系数。重测信度也称为稳定性系数，因为它说明了使用相同测量工具作重复测定时个体分数的稳定性。在重复测量时，由于受测者在第一次测验中取得

了经验或熟悉了测验内容，使得第二次测验成绩往往与第一次结果比较一致，因此，重测信度一般都偏高。同时，重测的时间间隔也会影响测量信度的评定。如果时间太长，在这一时期出现的各种无关因素就会干扰和影响信度评定；如果时间过短，又会由于练习效应对测验项目的记忆而高估信度。研究表明，重复测量的时间间隔应掌握在 30 天左右，在这种情况下，重测信度比较准确。

重测信度的主要缺陷是无法满足测量项目结构上的要求，并可能出现严重误估现象。当测量项目相互之间毫无关系或内容杂乱无章时，或者，当测量中存在严重的反应定势和偏向时，重测信度仍可能相当高，而实际上测量信度很低。此外，对被试进行两次同样的测试，往往不容易取得他们的合作，在实施时有不少实际困难。这些问题都是在使用重测信度时应引起重视的。

2. 等同信度（又称平行效度）（**equivalent-form**）

等同信度是以两组等效测验分别作出测量来评定信度系数。它使用了相同难度、内容和形式而具体项目不同的两种测验，弥补了重测信度由于采用完全相同的测验所带来的缺陷。但是，等同信度的评定对于测验设计的要求比较高，在实际工作中也不易得到两组等效测验，因此在应用方面就受到一定限制。

等同信度也象重测信度那样，要求进行两次测验，这使得受测者容易因疲乏或单调而影响参加测验的积极性。而且，要求他们抽出时间参加两次内容相似的测验也比较困难。这些都会不同程度地影响信度评定的效果。为此，人们更多采用一次性的信度评定程序，这就是折半信度和内部一致信度。

3. 折半信度（**split-half method**）

折半信度是在测试以后对测验项目按奇项、偶项或其它标准分成两半，分别记分，以两半分数之间的相关系数作为信度系数（ r_{xx} ）。但是，这样得到的信度系数只表明一半测验的信度，还需要

根据斯皮尔曼-布朗公式 (Spearman-Brown formula) 确定整个测验的信度 (r'_{xx}):

$$r'_{xx} = \frac{2r_{xx}}{1+r_{xx}}$$

例如, 某项管理能力测验的折半信度为 $r_{xx} = .50$ 因此 整个测验的实际信度为:

$$r'_{xx} = \frac{2 \times .50}{1 + .50} = .67$$

需要指出, 评定折半信度时, 需要测验项目的奇偶项的难度、内容等特征相似, 因此并非所有测验都能计算折半信度。

4. 内部一致信度 (internal consistent reliability)

内部一致信度是目前比较流行而且效果较好的信度评定方法。它不需要把测验项目分成两个部分, 而是从测量构思层次化入手, 使测量项目形成一定的内部结构, 并以内部结构的一致性程度, 对测量信度作出评定。

内部一致信度主要有两种: 库德-理查森 (Kuder-Richardson) 的公式 20 (简称 KR_{20}) 和克劳伯克 (Cronbach) 的 α 系数。

(1) KR_{20} 系数

KR_{20} 的公式为:

$$r_{KR20} = \frac{K}{K-1} \left(1 - \frac{\sum pq}{S^2} \right)$$

式中 p 是测验中通过某项目的比例, q 为没有通过该项目的比例 ($q = 1 - p$), $\sum pq$ 是 p 、 q 之积的总和; S 为测验总分数的标准差; K 是整个测验所包含的项目数目。

例如, 某项机械能力测验中有一个三项目的子测验, 这三个项目试测时正确回答的比例分别是 .80、.70 和 .50, 该子测验总分数的标准差为 $S = .90$, 需要评定这一子测验的信度。

根据 KR_{20} 公式, 可以先求出 $\sum pq$:

$$\Sigma pq = .80 \times .20 + .70 \times .30 + .50 \times .50 = .62$$

因而，内部一致信度为：

$$r_{KR20} = \frac{3}{3-1} \left(1 - \frac{.62}{.90^2} \right) = \frac{3}{2} \left(1 - \frac{.62}{.81} \right) = .35$$

这个子测验的信度为 .35。

(2) 克劳伯克的内部一致性系数

克劳伯克 (Cronbach, L. J.) 于 1951 年提出一种内部一致性系数—— α 系数，以此作为测量信度的指标。这个指标准确地反映出测量项目的一致性程度和内部结构的良好性，成为心理学研究中目前使用最广泛的信度指标。克劳伯克 α 系数的运算公式是：

$$r_{\alpha} = \frac{K}{K-1} \left(1 - \frac{\Sigma s_i^2}{s^2} \right)$$

式中， K 为测验所包含的项目数目，当测验包含若干个子量表时， K 为子量表的项目数； s_i 为各项目的标准差， s_i^2 即第 i 个项目的方差； s 是整个测验或子量表总得分的标准差， s^2 即总得分的方差。当测验项目都是两分项目（是或否、对或错等）时，用通过各项目的比例表示方差，上述公式就与 KR_{20} 公式一样，因此， KR_{20} 系数是 α 系数的一个特例。

内部一致性系数的计算比较方便，关键是测量项目之间具有较好的内部结构基础，子量表的组合比较明确。例如，某测评表包括四个子量表，每个子量表又包含 5 个项目。对于这个 20 项目的测评表，需要分别对四个子量表估算信度系数。假定其中一个子量表的 5 个项目得分的方差分别是 2、3、3、4、6，子量表总得分的方差为 45，那么这个子量表的 α 信度系数就是

$$r_{\alpha} = \frac{5}{5-1} \left(1 - \frac{2+3+3+4+6}{45} \right) = \frac{5}{4} (1 - .40) = .75$$

需要指出，当项目内部结构不良或内容十分异质时，不宜使用 α 系数，而应采用其他方法（如等同信度等）。

三、测量效度的评定

在确定心理测量的信度以后，需要进一步对测量效度作出评定，从而决定测验或量表是否符合心理学研究的科学要求。由于信度是测量本身的自我比较，而效度却是有关测量与外部标准之间关系的评价，因此，效度的评定比信度估算更为复杂。

1. 内容效度

内容效度是指测验项目在多大程度上表示了所要测定的特征范畴。比如，如果我们要从一批待业人员中挑选机床操作工，所用的测验项目就应该能表明机床操作能力的范畴。

内容效度的评定主要通过经验判断进行，可以请一些熟悉该测量内容的人员来评判，确定测量项目与所需测量的内容范畴之间关系的密切程度。当然，这里又要注意评判的一致程度和可靠性。内容效度评定的一个常用指标是“内容效度比”，缩写为 CVR (content validity ratio)，它的计算公式是：

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$$

式中 ne 为评判中认为某项目很好地表示了测量内容范畴的评判者人数；N 为评判者的总人数。这个公式表明，当认为项目内容适当的评判者不到半数时，CVR 为负值，如果所有人都认为内容不当，CVR = -1.00；当认为项目合适和不合适的人数对半时，CVR 值为零；而当所有评判者认为项目内容很好时，CVR = 1.00。

在运用内容效度时，应比较谨慎，因为测量构思常常包含着多个方面，不容易就其内容作出全面的评判；同时，项目与内容范畴的关系有不同的程度，需要作具体的分析。当有关专家对测量的内容范畴意见不一致时，应采用其他的方法进一步检验项目，例如，对

项目的构思效度作出评定。此外，内容效度的评定主要是为测量项目设计服务的，需要与测量的构思、项目的内部结构等方面结合在一起。研究表明，通过加强项目内容同质性、采用多重评判等方法，可以提高测量的内容效度。

2. 预测效度

预测效度是指测验结果能够预测人们将来行为的程度。在心理学研究中，常常需要对某类人员进行一定的测评，并希望根据测评结果预计和推测受测者将来的表现和工作成绩。因此，需要在测评之前，确定测验或测评表的预测效度。预测效度实际上是测验结果与一段时间以后的行为表现（预测标准）之间的相关程度。测验分数与以后的学习或工作成绩之间相关越高，就认为测量的预测效度越高。这里，我们把测验项目内容作为预测指标，而把学习成绩或工作表现作为预测标准（又称为效标）。

在评定预测效度时，可以在测量对象中选取一个有代表性的小样本进行试测。例如，为了评定小学生认知能力测量的预测效度，可以先在学期初选取一个有代表性的小学生样本进行测试，取得认知能力分数；然后，在经过一个学期的学习以后，对这些孩子的学业成绩和学习表现作出评估，通常以学科成绩和教师对这些学生的评定作为评估的依据；求出认知能力得分与学业评估分数之间的相关，这就是认知能力测量预测效度的指标。这里，认知能力测量作为预测指标，学业评估分数是效标。

预测效度对预测指标和效标的信度，有较高的要求，因此属于最严格的效度指标。在预测效度的评定中，应注意选用敏感、有效的预测指标（如能力、知识等）和可靠的效标；还应选择和掌握适当的时间间隔，以免时间过长而受到其它因素的干扰。

3. 同时效度

在许多情况下，不可能或者不便于等待一段时间再评定测量的

预测效度。因而，人们就用测验分数与当时行为表现或工作实绩之间的相关作为效度指标，称为同时效度。同时效度实际上是间隔时间为零时的预测效度，因此同时效度总是高于预测效度的。如果一项测验的同时效度很低，那么就可以认为，这项测验的预测效度将更低。

同时效度与预测效度中所用的效标不应该是受测者自己评判的，而应以上级部门或心理学家所作出的有关受测者工作实绩的评定作为效标。工作实绩或其它效标的评定内容应通过职务分析和测验评价来确定。还需要指出，根据以往研究的结果，测量的效度水平一般在 .30— .60 的范围。这些都是在效度评定中应加以注意的。

4. 构思效度

构思效度是与一定测验的理论基础的有效性有关的。心理学研究中所用的测验，无论是能力测验、动机测评，还是工作负荷测验、疲劳测定和紧张程度测评，都是建立在一定的理论假设和构思的基础上的。构思效度就是要检验测验是不是真正测量了我们所提出的理论构思（如能力）。因此，对于理解测验结果的含义来说，构思效度是最重要的效度指标之一。

构思效度的评定比较复杂。首先需要对项目和测验（测量）的理论构想作出明确的说明，包括项目的结构、测量的总体安排以及项目与项目之间的关系等；然后，需要确定测量指标的信度，这是取得效度的前提条件；运用因素分析等高等统计方法，从测量数据中离析出若干基本因素或构思，以此检验测量的总的构思效度。

5. 聚合效度与辨别效度

坎贝尔和菲斯克 Fiske, D. W.)于 1959 年提出两种新的效度：聚合效度和辨别效度 (convergent validity and discriminant validity)，从而能用系统的相关证据对构思效度作出推论，成为构思

效度的新形式。

聚合效度是指运用不同测量方法测定同一特征或构思时测量结果相似的程度，即不同测量方式应在相同特征的测定中聚合在一起。辨别效度则是指在运用相同测量方法测定不同特征或构思时，辨别不同特征的程度，即不同特征的测量结果之间不应有高相关。

聚合效度和辨别效度的概念主要基于以下的考虑。

(1) 测验总是一个“特征—方法”的单元，特征的测量总离不开一定的方法。因此，任何测量结果分数的变异中都有一部分应归因于特定的方法，而不只是所测量的特征本身。有时候，两个测验可能只因为测试方法相同而出现高相关，而不是由于测了相同特征。

(2) 只有通过相似性和差异性的联合测定方法，才能准确地描述和解释测量结果。这就是说，必须说明测验测量了什么，没有测量到什么。因此，需要运用聚合原则和辨别原则。聚合原则认为，对相同特征的两种（或多种）测量之间应呈高相关；辨别原则指出，对不同特征的测量，即使运用相似方法，测量结果之间也没有高相关。

为了评定测量的聚合效度与辨别效度，坎贝尔和菲斯克提出“多特征—多方法”矩阵法（简称 M—M 矩阵法），这种方法要求我们至少有二种不同特征，每一种特征又至少以两种方法作了测量。表 7-2 为 M—M 矩阵的一个实例（multitrait-multimethod matrix）。

多特征—多方法矩阵中表示了相同特征的测量信度、相同特征不同方法测量所获得的效度、同样方法测量不同特征的相关系数，以及不同方法测量不同特征的相关系数。

从 M—M 矩阵，如何决定聚合效度和辨别效度呢？有四条评判标准：

表 7-2 表多特征—多方法矩阵

	特征	方法1			方法2			方法3		
		A ₁	B ₁	C ₁	A ₂	B ₂	C ₂	A ₃	B ₃	C ₃
方法1	A ₁	(.89)								
	B ₁	.51	(.89)							
	C ₁	.38	.37	(.76)						
方法2	A ₂	(.57)	.22	.09	(.93)					
	B ₂	.22	(.57)	.10	.68	(.94)				
	C ₂	.11	.11	(.46)	.59	.58	(.84)			
方法3	A ₃	(.56)	.22	.11	(.67)	.42	.33	(.94)		
	B ₃	.25	(.58)	.12	.43	(.66)	.34	.67	(.92)	
	C ₃	.11	.11	(.45)	.34	.32	(.58)	.58	.60	(.85)

注：()：信度系数；○：效度系数；

△ 不同特征—同一方法的相关系数（异特征—同方法）；

△ 不同特征—不同方法的相关系数（异特征—异方法）

(1) M—M 矩阵中的效度系数达到显著水平，（这是聚合效度的证据）足以作为进一步检验效度的依据；

在这一条标准满足的条件下，可以检验其它三条标准：

(2) M—M 矩阵中的效度系数大于同行或列邻近的“异特征—异方法”三角中的相关系数；

(3) M—M 矩阵中的效度系数大于“异特征—同方法”三角中的相关系数；

(4) 所有异特征三角中的相关系数模式相同，包括“异特征—同方法”和“异特征—异方法”。

在上一实例中，上述四条标准并没有全部得到满足，因此认为测量还没有完全具有聚合效度和辨别效度。

除了以上几种效度以外，还有一种称为“表面效度”。严格地讲，表面效度并不是真正的效度评定，而是受测者对于测验的一种评判，表明了受测者对测验的态度。可见，表面效度是与受测者的动机密切相关有关。如果受测者认为某项测验质量不佳或者有其它缺陷，这自然会影响测量的效果。因此，表面效度也是整个测量成效的重要方面。

在学习和掌握心理学研究中有关测量理论和方法的内容时，应该把效度的评定看成一个评价过程。它涉及到确定测量指标、建立效标、了解测验的理论构思和付诸实际测试等许多重要方面；它包括了实际工作需要分析、职务分析、作业要求分析、理论构建、取样、预测指标与效标选择、数据采集、假设检验和效度系数评定等一系列步骤。因此可以认为，测量效度的评定是一个理解、掌握、完善和实践心理测量的工作过程。

四、影响测量信度和效度的因素

测量的信度和效度受到许多因素的影响，我们在“经典和现代测量理论”这一节里已经指出，以经典测验理论为基础的信度和效度的参数评定，特定于测量样本和项目特征，因此，样本的特点（例如，同质或异质）和项目的特征（例如，难度或辨别力）都在很大程度上影响了测量的信度和效度，这正是传统心理测量理论的缺陷。除此以外，还有些因素会影响测量的信度和效度，这些因素也是我们在选用或设计测验或测量时应该注意的。

1. 影响测量信度的因素

（1）测量项目的数目

项目的数目也是指测验的长度。研究证明，在一定的限度内，一个测验或一次测量中所采用的性质相同的项目越多，测量信度也愈高。研究还表明，随着测验项目的增加，该测量的标准误差呈减小

的趋势。

(2) 测量的程序

测量程序对测量信度的影响甚大，包括测量的指导语、程序安排、测试场地的环境、多组测量之间的间隔时间，等等。这些方面如果不标准化或者缺乏控制，就可能降低测量信度。为此，需要在测量时运用某种实验控制，以便提高信度。例如，在速度测验时采用自动报时信号、对测量评判人进行训练以减少个人偏向、采用明确一致的指导语，等等。

(3) 测量项目的质量

测量项目的质量会直接影响测量的可靠性。比较重要的方面是项目的选择，这也是取样问题，应选用或设计具有代表性的项目（代表测量的特征）和内部一致性高的项目。

(4) 测试人和受试者

测试人的表现常常会影响测验时的气氛和受试者的反应定势。测试人的评分偏向也会降低测量信度；受试者的身心健康状况、动机、态度、注意力等的变化，也影响测量的信度。受试者样本的特征（如同质和异质）对信度的影响也比较大，异质样本的分数全距较大，测量信度将高于同质样本的测量结果。

2. 影响测量效度的因素

上面所提到的影响信度的因素也在不同程度上影响测量的效度。对效度影响比较大的因素有：

(1) 测验的组成

测验或测量的选材与构思、测验的长度、项目的难度、辨别力和编排方式等因素都会影响测量的效度。

(2) 测量程序

测量程序的质量对于测量效度是至关重要的，其中主要包括测量程序设计、测验手册的配备和实施、程序的标准化、指导语、场

地的布置、时间的限制等。

(3) 效标的特征

选择适当的效标，是测量效度的前提条件。效标的信度、效标与测量指标（如预测变量）之间真正相关的程度等都会影响测量的效度。

(4) 样本与受试者特点

研究中所用样本的代表性会影响效度，同一个测验对于不同对象的测验功能并不一致，效度也因此不同。此外，样本的异质性会使样本分数的全距增大，如果其它条件均等，异质样本会提高效率系数。受试者的兴趣、动机、情绪和态度等也决定了测验结果的有效性，对效度的影响也很大。

在对测量信度和效度进行评定时，应特别注意以上这些影响因素，努力设法排除或控制可能造成误差的原因。同时，从上述因素，也提供了提高测量信度和效度的线索。研究表明，在心理测量的设计和实施过程中，从上述诸方面采取改进措施，就能显著地加强测量信度和测量效度。

第三节 心理测验

一、心理测量和测验的一般问题

在具体讨论心理测验及其设计和编制问题之前，我们先谈谈心理测量的一般问题。心理测验和测量既有相同的意义，又不完全是同义词。测量的范围比较广，涉及心理学许多研究领域的度量问题；测验却不同，它主要用于评估个体的某些心理特征。测量分数的性质也比较广泛多样，比如用心理物理法测定听觉阈限，测量分数是以物理量来表示相应的心理量；测验则用测验分数直接表示心理量。

测量总是记分的；而测验却并不一定记分，有的个性测验只用语言来描述个性的特征。因此，可以说，并不是所有测量都是测验，也并非所有测验都是测量。

1. 心理测量的水平

心理测量是以不同量表方式进行的，常用的有四种量表方式，也叫做心理测量的水平：称名量表、顺序量表、等距量表和等比量表。

(1) 称名量表

又称类别量表，即以测量对象的类别方式记分。例如，在教育心理学研究中比较男、女学生的学习成绩，以男、女类别编码记分（如 0 或者 1）；在管理和人事心理学研究中以干部职工分类，或者以不同领导作风和方式评定干部，等等。

(2) 顺序量表

顺序量表既没有相等单位，又没有绝对零，只是要求评定人在若干个备择项目中按照一定标准排出等级次序。例如，在奖励方式的测评中，要求按照自己对于几种备择奖励形式的偏爱程度排出先后顺序。

对于以称各量表和顺序量表取得的数据，只能采用非参数统计方法进行分析和结果处理。

(3) 等距量表

等距量表有相等的单位，但没有绝对零，因而其测量水平比顺序量表提高了一步。等距量表从间距相等的分数点对心理特征作出测量，常用的是 5 点和 7 点量表。为了避免评判人的偏向，往往不在量表上写数目，而以括号代之。记分时则以 5 点或 7 点记分。

(4) 等比量表

等比量表更进了一步，它既有相等单位又有绝对零，因而属于最高测量水平。例如，对操作工或驾驶员的反应速度的心理测量就是采用了等比量表。

2. 心理测验的种类

心理测验是主要借助量表对心理特征和行为进行观察和描述的系统心理测量。心理测验有许多类型，例如，速度测验（测验时有严格的时间限制）、能力测验（没有严格的时间限制）、团体测验（同时给一组受试者测验）、个体测验（一次只给一个人测验）、纸笔测验（以笔、纸、书写方式进行，不操作客体或设备）、作业测验（通过操作特定客体或设备而作出反应），等等。

常用的心理测验分成两大类：

（1）最大成绩测验

这类测验用以测量人们的“最大”作业成绩，从而对他们所具备的能力作出评定。这种测验包括各种智力测验和技能测量，还包括成就测验。成就测验的内容集中于某一具体任务或者课题。另一种测验是能力倾向测验，用于预测受试者在某种职业训练课程或专业中是否能取得成功。这种测验往往包括一些子测验，例如，工程能力倾向测验用于工程工作成效的预测，包括几个部分，分别测量一般心理能力、机械与空间推理能力、数学能力等。

（2）典型反应测验

这类测验不是测量人们能够干些什么，而是测量人们在特定情境中的典型行为或者感受，主要包括个性、习惯、兴趣、焦虑、情绪等。

3. 影响心理测验的因素

心理测验成绩的好坏，受到许多因素的影响，除了受测者本身的知识 and 能力等特点以外，还有以下一些因素。

（1）测验的形式和练习因素

测验本身的特点和形式，以及测验情境的特征都可能影响测验成绩。测验中，有的人比较善于利用这些特征和线索；还有的人由于接触过某些测验而熟悉测验的特征；多次测验，会产生练习效应

而混淆测验结果的意义，这在速度测验和重复测验中尤其明显。

（2）焦虑和动机

对测验动机过高、期望过高或者对测验成绩过分关心，都可能产生焦虑，出现不安、焦躁、胃溃疡等消极的心理与生理反应，进而影响测验成绩，在一些涉及情感和动机的测量中，焦虑的影响更大。

（3）主观倾向和反应定势

反应定势是指受测者以某种习惯的方式对测验项目作出反应，使得我们无法正确判别受测者的真正能力水平。在心理测验中主要有这样一些反应定势：

- a. 速度与准确性定势：有的人倾向于慢而仔细地答题，另一些人常常快而粗略地作出反应；
- b. 认可定势：对正误或是否题不肯定时，倾向于选择“正确”或“是”的答案；
- c. 位置参照定势：有的人对等距量表的某些位置比较偏爱，另一些人则倾向于每题变换一种量表位置；
- d. 投机定势：也是一种猜测定势，倾向于探究题中线索，猜测答案。

此外，社会认知和评定中的各种消极效应会产生各种主观倾向，比如趋中效应、宽大效应和晕轮效应等。趋中效应是指在评级时倾向于作出“中等”区域的评级，而很少作出高分或低分的评定；宽大效应是指评分倾向于过高或者过低，前者又叫正宽大效应，后者称为负宽大效应；晕轮效应则指的是由于对受测者的一般印象而形成某种稳定的评级倾向。这些主观倾向都对测验效果有很大的影响。

（4）测验实施中的因素

测验实施过程中的各种因素也会影响测验成绩，例如，实施方式、测验主持人的特征、答卷格式、记分形式、指导语的标准化程

度、测验场地的布置、环境条件等。

在进行心理测验和评价测验结果时，应该注意以上这些因素，设法控制或消除这些因素，并在解释结果时把这些因素考虑进去。

二、心理测验的应用及其效果

心理学研究广泛地采用了各种类型的心理测验。例如，一般智力测验、能力倾向测验、标准化成就测验、兴趣测验、个性测验、情感测验和动机测验等。这里就心理测验的应用及其效果作一些讨论。

1. 一般智力测验和能力倾向测验

这类测验中应用比较多的有韦克斯勒成人智力量表（WAIS）、旺德利克（Wonderlic）人事测验、贝内特（Bennett）机械理解测验、斯坦福-比纳（Stanford-Binet）智力量表以及其他一些测验。应用研究和实际经验证明，一般智力测验对于实际学习和工作成效的预测效度并不高，能力倾向测验的效度略高一些；某些婴儿智力测验几乎没有预测效度。智力测验的效果很受文化背景的影响，在移植国外测验时应十分谨慎。此外，一些有关创造性的测验往往只反映出受试者的某些思维特点，并不能作为创造性的有效测量。还有一些研究说明，传统测验的结果很大程度上取决于项目内容和测验任务的特点，我们已经在有关经典和现代测验理论的讨论时指出了这一缺陷。

2. 标准成就测验

标准成就测验主要用于教育领域，因此其应用效果很受教育内容、教育背景和教学方式等因素的影响。实际应用情况表明，阅读测验和数学测验比语言测验和社会研究测验效度高得多，因为前者更少受到内容和课程设置方面因素的影响。在运用标准化成就测验时，应该仔细地考察现有课程的内容和教学方案，确定测验的目的和方式。目前，标准化成就测验在许多国家的学校中运用，人们认

为成就测验比一般智力测验有更高的效度和可应用性，因而主张尽量少用一般智力测验而采用标准化成就测验。成就测验主要用于三种目的：(1) 检查和控制教育质量；(2) 评价课程设置和培养计划；(3) 对学生进行个别诊断和评估。

3. 兴趣测验

兴趣测验中最有名的是斯特朗-坎贝尔(Strong-Campbell)兴趣问卷表(SCII)和库德(Kuder)的职业兴趣调查表(KOIS)(或称职业兴趣量表)，在国外应用得较广泛。SCII包括124种职业兴趣量表，组成六个职业主题。例如，研究型、艺术型和企业型等，要求受试者用“不喜欢”(D)、“无所谓”(I)或“喜欢”(L)迅速表示自己对于325种活动的偏爱与否。KOIS测验要求对许多由三种活动组成的项目，表示最喜欢和最不喜欢的活动。

实际应用情况证明，兴趣测验在单独使用时实际价值并不大，但是，如果把兴趣测验与能力倾向测验结合在一起，能够在职业选择、人员挑选和技术培训等方面提供比较有效的依据。

4. 个性测验

传统个性测验主要采用自我报告和投射技术，比如，明尼苏达多项个性量表(MMPI)、加州心理量表(CPI)、罗夏克墨迹测验和主题统觉测验(TAT)等。MMPI主要用于临床测量，以心理变态者为对象，这个量表在运用中对结果作错误解释的情况很严重。CPI量表用于正常成人和青少年，应用情况比MMPI理想一些。这类自我报告式的个性测量很容易发生误估，因为受试者往往倾向于把自己描述得比较理想。研究表明，这种个性测验对于在学校、职业或生活方面的成功并没有多大的预测价值，主要用于诊断某些症状。

投射测验是许多人所熟悉的，其记分和解释程序都很复杂，难度较高，不是一般人员能够胜任的，应由受过专门训练的临床心理学家或精神病学家使用。投射测验主要用于心理治疗，但是，这方

面的研究证明，这种测验的信度比较低，而且，在区别正常人和病人的诊断，以及学术与职业预测等方面，其效度一般都比较低，因此，不易用于学校、企业等处作为心理测量工具。

现代心理学家研究认为，个性的类型学只是一种比较原始和静止的个性划分，并不能说明人类个性的真正特点和动态发展。只有把工作和生活等实际情境联系在一起考虑，才能真正理解人的个性特征。因此，人们主张针对不同的学习或职业要求与特点采用适当的“情境测验法”进行直接观察，对个性特征作出评定。

5. 情感和态度的测量

这方面的测量多半使用各种问卷量表或访谈法，关于这些问卷量表和访谈程序的问题，我们将在下一章里专门讨论和介绍。这里需要指出，这方面测量的主要局限和问题，是受试者根据需要而虚构情况的倾向、设法作出符合社会要求的反应的定势、项目用词的语义问题等。这些问题都是在情感和态度测量问卷的设计中应该注意的。

运用心理测验，是一项严肃和要求较高的任务。在选择测验过程中，应特别注意各个心理测验的基本理论假设和构思方面的特点。对不符要求的测验，需要作必要的修正。心理测验的使用，需要有专门的训练，严格按照测验规程行事，防止误用甚至滥用。还应重视测验本身在效度和文化背景适应性方面的局限。

陈立教授曾经在“我对测验的看法”一文中论述了心理测验应用中的许多值得重视的问题，他特别强调了“测验只能由有过专门训练的人来用”的道理。由于智商分数等可能引起的严重问题，他“主张测验一定要和治疗或教育结合”，帮助我们发现问题和解决问题，使心理测验能够真正为我们的教育和整个社会服务。

三、心理测验的项目分析

在我们编制测验或者对原有测验进行检验时，都需要对测验的项目作一番分析，这就是项目分析。项目分析主要是用来决定每一项目的难度和辨别力的。当我们根据一定的理论构思设计出一套测验项目时，除了对项目的内容即内容效度进行精心地检查以外，还需要对这些项目的编排、用词、数量、语义等方面进行分析和修正，然后，可以把这一套测验付诸试测。试测的目的是了解受试者的想法，他们对项目的理解以及在测试过程中的反映。这时候，我们就可以对测验进行项目分析了，以此改进项目和测验的质量。

1. 项目的准备

这是项目准备的准备阶段。我们已谈到，除了内容效度，还要检查和确定项目的另外两个特征：难度和辨别力。项目难度是指所测的受试组中正确回答该项目的百分数；项目辨别力则是项目能够把不同知识或技能的人区分开来的能力。这两个特征之间有一定的关系，当项目难度过高或者过低时，难度因素会很大地影响辨别力，使得项目不能辨别不同水平或特点。因为在这种情况下，要么要所有受试者（无论能力高、低）都答对项目，要么全都答错，看不出好坏。

项目分析有不少方式，其中有的适合于标准化测验的分析，有的以程序化的电脑软件作出分析。这里讨论的是一种比较简便的、适合于心理学和教育研究一般目的的项目分析方法。这种办法可用于较大样本的测验工作，也可用于一个班或一个组的小样本测试。在测试中，要求受试者回答每一个项目，在时间安排上，应让绝大多数人有足够的时间回答问题；项目的编排，在难度上递增，即把较易的项目排在前面，把较难的安排在后边，这些都是在项目准备阶段应该注意的。项目分析并不能弥补项目的计划、编制和测试等过

程中造成的不少缺陷，因此，做好这些工作，是有效的项目分析的可靠保证。项目分析的结果可以为进一步的计划、编制和测试提供重要的反馈信息。

在项目和测验的试测过程中，应注意以下问题。

(1) 试测的对象，应从正式测验所要应用的总体中取样，注意其代表性。通常以 400 人为宜，也有人主张，可以用较小的样本，比如 100 人左右。

(2) 试测力求正规化，使得实施过程与将来的正式测验时的情况尽可能相似。

(3) 试测时，应使受试者有足够的答题时间，使结果更为可靠。

(4) 在试测过程中，随时记录受试者的反应，包括答题的一般进度、题意不清之处等情况。

2. 项目分析的基本步骤

在对测验结果评分以后，就可进行项目分析，主要有以下几个基本步骤。

(1) 把测验按分数高低排成顺序，清点测验份数 N 。

(2) 以 N 乘 .27 并去掉小数，取整数 n ，比如， $N=30$ ， $n=30 \times .27=8.1=8$ 。

(3) 从测验的高分部分从最高向低取 n 份最好的卷子，比如， $n=8$ 则取出 8 份，作为高分组，以同样方式从最低向高取得低分组。

(4) 分别求出高、低分组中答对项目的比例 P_H 、 P_L ，每一个项目都求出这一比例

$$P_H = \frac{\text{正确回答的数目}}{n} \quad (\text{高分组})$$

$$P_L = \frac{\text{正确回答的数目}}{n} \quad (\text{低分组})$$

(5) 求出项目难度 P ， $P = \frac{P_H + P_L}{2}$ 。

对于这一难度系数的解释需要考虑项目的机遇水平。例如，某

项目的试验中，高分组有 70% 答对，低分组有 30% 答对，则这个项目的难度为：

$$P = \frac{.70 + .30}{2} = .50 \text{ (或 } 50\%)$$

P 值越高，难度越低；P 值越低，难度越高。P = .50 为中等难度。但是如果这是一个两种选择的项目（是否题或正误题），那么，P = .50 实际上就很难说明受试者在这一项目上的情况，因为如果随机答题的话，也可以得到 P = .50 的正确率。因此，对于难度 P 值应该按照不同的项目特点进行“机遇校正”（correction for chance），可以用这一公式：

$$P' = \frac{KP - 1}{K - 1}$$

式中 P' 为校正后的正确答案的比例；P 为原比例，K 为每一的备择答案数目。

例如，一个 5 种备择答案的项目的原难度比例为 .80 则这个项目的实际难度为：

$$P' = \frac{5 \times .80 - 1}{5 - 1} = .75$$

(6) 求出项目辨别力 D， $D = P_H - P_L$ 。上例中 $P_H = .70$ $P_L = .30$ ，所以，项目辨别力 D 为：

$$D = .70 - .30 = .40$$

辨别力指数越大，表示个别项目的反应与测验总分的一致性愈高。研究表明，项目辨别力在 .35 以上时，就认为该项目有相当高的辨别能力；如果辨别能力在 .20 以下，则表示辨别力比较低。

项目难度与辨别力的关系可以用图 7-3 表示，可以看到当难度在 .50 时，即一半受试者能答对该项目，潜在的项目辨别力最高；而当项目太容易或者太难时，这个项目就无法评估出个体差异（无辨别力）。

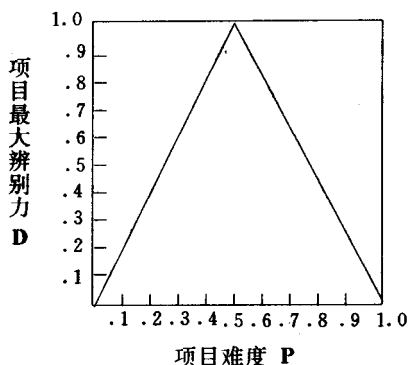


图 7-3 项目难度 (P) 和项目最大辨别力 (D) 之间的关系

在项目分析的基础上，我们应该保留中等难度的项目；对低难度和太高难度的项目，可以进行修改，或者删去该项目。一般可以保留难度水平在 .25— .75 范围的那些项目，对其它项目作进一步修改。如果从辨别力考虑，可以取 $D \geq .20$ 的项目。

第四节 心理物理方法和 信号检测法

心理学研究的许多课题涉及物理刺激与心理感受之间的关系，信号闪光频率、照明亮度、温度、颜色等，都是研究中经常碰到的物理量。心理物理方法就是用以研究心理量和物理量之间关系的一种方法，创立一百多年来，已成为心理学研究中十分重要的测量与实验方法。

心理物理方法是由费希纳 (Fechner, G. T. 1801—1887) 创立的。1860 年，费希纳在大量有关感觉阈限的测量和研究的基础上，发表了他的经典著作《心理物理学纲要》，为心理物理方法和理论的发展，奠定了基础，并对心理学研究方法和实验心理学的发展产生了

深远的影响。费希纳在有关感觉阈限的研究中，设计并采用了三种方法：（1）最小变化法；（2）恒定刺激法；（3）平均误差法。我们把这三种方法称为经典心理物理方法。这些方法一直在心理物理学实验中广泛应用。本世纪 50 年代以后，经典心理物理方法开始暴露出一些局限，一些新的更好的方法被运用于心理物理学研究，其中最有意义的是坦纳（Tanner, W. P. Jr.）和斯韦兹（Swets, J. A.）于 1954 年提出的信号检测理论和方法。信号检测法正在成为实验心理、工程心理和其他心理学领域中很有前途的研究方法。

一、经典心理物理方法

1. 最小变化法

最小变化法又称作极限法，这是让被试通过一系列尝试，对强度渐增或渐减的刺激作出反应，检验有关这一刺激的感觉阈限。例如，研究者选用某种频率的声音，使它在从一点也听不见到完全可以听见的范围内由强变弱或由弱变强。每一次尝试中，要求被试在一定时间间隔内报告是否听到了声音，并记录他们的反应。尝试分成渐增系列和渐减系列。在渐增系列中，声音强度一开始根本无法听见，使被试作出否定反应，然后逐次增强声音强度，直至被试报告“听见”为止；在渐减系列中，第一次尝试的声音强度足以被检测到（听见），然后逐次使强度慢慢减弱，直到被试报告“听不见”为止。需要注意的是，每一尝试系列应在不同的刺激强度上开始，以免被试习惯或熟悉系列中感觉转折的区域。当被试的反应由肯定向否定、或者从否定向肯定转变时，该刺激值就相应于被试的感觉绝对阈限。如果各系列的转折刺激值略有不同，就以多次刺激值的算术平均数作为绝对阈限。最小变化法还可以用来测量差别阈限，以比较刺激的变化与一标准刺激作对照，以渐增法或渐减法测得差别阈限，具体做法就不在这里介绍。

2. 恒定刺激法

恒定刺激法又叫正误法，其主要特点是以被试在判断中正确与错误或者肯定与否定的频次与比例来决定感觉阈限。通常的做法是，挑选一组强度不同的刺激，其中最强烈刺激应是每次呈现几乎都能为被试感觉到的，最弱刺激每次呈现几乎都感觉不到，其他的刺激则要求强度在最强与最弱刺激之间约 5—7 个间距相等的强度等级。这些刺激以随机方式呈现给被试，每种强度的刺激呈现次数相等，并且不得少于 10 次，要求被试报告是否觉察到所呈现的刺激，以被试作出肯定反应的比例来计算绝对阈限。可以根据实验结果，以刺激强度为横坐标、肯定反应比例为纵坐标，作出曲线图，取 50% 肯定反应的刺激强度值作为阈限值。在测定差别阈限时，需要确定一个标准刺激，并以随机方式呈现不同强度的比较刺激，比较刺激的决定方式和前面一样。实验时，要求被试作出“大于”、“等于”或“小于”标准刺激的反应，以此计算差别阈限。一般来说，每个比较刺激都与标准刺激比较 100 次，两者相继呈现，其中标准刺激在前和在后的顺序各 50 次。图 7-4 为某实验用恒定刺激法测量差别阈限的结果。图中 PSE 为 50% “重于” 反应相对应的刺激重量值，PSE

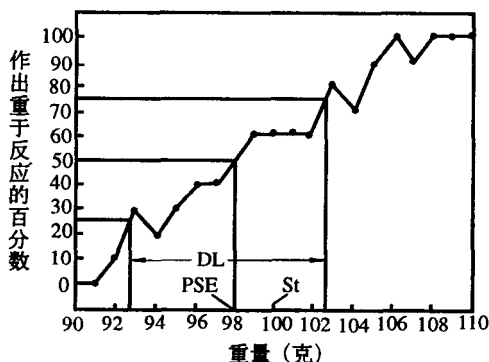


图 7-4 恒定刺激法测定重量差别阈限

称为主观相等点，是不肯定间距的中点。St 为标准刺激，DL 是 25%、75% “重于” 反应所对应的刺激值的距离，即差别阈限。

3. 平均差误法

平均差误法又叫做调整法。它的特点是让被试调节比较刺激，使之与标准刺激相等或匹配，并以匹配的差误来计算感觉阈限。这种方法由于被试能自己主动调整比较刺激，容易引起兴趣，因此在很大程度上避免了其它心理物理法常常造成的单调感。可是，调整动作本身有可能干扰被试的判断，应在实验中加以注意。有一研究用平均差误法测量缪勒—莱尔错觉的程度，标准刺激线长 25 厘米，要求被试调整比较刺激长度，使两者看起来相等，共作 8 次尝试（其中由短调长和由长调短各一半次数）。结果见表 7-3，其中 $\bar{X}$ 为平均值，PSE 为主观相等点，CE 为常误，也就是错觉量。

表 7-3 用平均差误法测定缪勒—莱尔错觉

尝试次数	由长调短 (D) (cm)	由短调长 (A) (cm)
1	26.5	—
2	—	26.0
3	—	25.5
4	25.5	—
5	—	25.0
6	27.0	—
7	25.5	—
8	—	27.0

(标准刺激长度 = 25cm)

$$\bar{X}_D = \frac{104.5}{4} = 26.125,$$

$$\bar{X}_A = \frac{103.5}{4} = 25.875$$

$$PSE = \frac{\bar{X}_D + \bar{X}_A}{2} = 26.00,$$

$$CE = PSE - St = 26 - 25 = 1.0$$

二、心理物理方法的新发展

心理物理学理论和方法并不只是感知心理学研究的手段，而是可以应用到心理测量的许多方面。事实上，本世纪以来的大量有关研究表明，心理物理方法不断在发展、完善，并开拓新的应用领域。在心理物理方法的新发展中，有四个富有意义的里程碑。这就是富勒顿-卡特尔原则（Fullerton-Cattell Principle）、瑟斯顿比较判断定律、赫尔森（Helson）适应水平理论、斯蒂文斯（Stevens）幂函数定律。前两者都涉及心理物理测量的分布方式，为心理测量提供了精细的统计技术；赫尔森以适应水平理论对心理物理学测量作出新的解释；斯蒂文斯则进一步发展了经典的韦伯和费希纳定律。其中，比较判断定律、适应水平理论和幂定律，对于心理学研究方法论的影响更大。

1. 比较判断定律

在一项心理物理学实验中给被试呈现两个刺激，要求他判断两者的差异，如果被试觉得这两个刺激完全不同，就会作出刺激 A 多于刺激 B 或者 B 多于 A 的反应。在两个刺激确实不同的情况下，被试将总是以同样方式作出反应。但是如果被试从来感觉不到差异，他的反应也会十分一致。然而，当两个刺激的性质有相似之处或它们的离差相互重叠，就需要对两个刺激差距的实际大小作出评估。经典心理物理方法只能对此作出间接的评定，瑟斯顿就此提出著名的比较判断定律：

$$X_1 - X_2 = Z_{ab} \sqrt{\sigma_a^2 + \sigma_b^2 - 2r_{ab}\sigma_a\sigma_b}$$

式中 Z_{ab} 为正态离差， σ_a 、 σ_b 分别为两个刺激均值（ $\bar{X}_1$ 、 $\bar{X}_2$ ）的标准差， r_{ab} 为 $\bar{X}_1$ 与 $\bar{X}_2$ 两个分布之间的相关， X_1 和 X_2 也可以表示刺激的量表值。这个定律不仅可用于解决经典心理物理法的一些问题，而

且对于各种量表技术的运用具有价值。

2. 幂函数定律

费希纳的理论与方法虽然在心理物理学实验中广为应用，却也暴露出一定的局限性。费希纳的对数定律 ($S=c\log I$) 认为心理感觉量是刺激量的对数函数，当刺激强度以几何级数增加时，感觉强度以算术级数增加。但实际上，两者往往并非这种简单的关系，而且，费希纳以间接方法制作量表的依据也受到越来越多的批评。斯蒂文斯在多年研究的基础上提出一个不同的定律，即幂函数定律： $S=bl^a$ （其中， S 为心理量， a 、 b 为常数， I 为物理量）。这个定律为心理物理方法的进一步应用，提供了新的思路。

三、信号检测法

信号检测论以统计决策论作为基础，由坦纳和斯韦兹首先应用于人的知觉研究。信号检测法也因此在各种心理学实验中被采用。由于信号检测理论和方法可以在心理测量中确定反应倾向，并使所测得的被试辨别力和他的操作水平区分开来，因而成为心理物理学发展中的重要突破。

根据信号检测理论，控制信号检测的有两种基本心理过程：

(1) 建立感觉印象，信号或刺激到达感受器时形成感觉印象，一般取决于信号强度；(2) 决策判断，感觉印象不足以引起肯定反应，还需要对感觉印象量作出判断，如果这一量值高，决策判断将产生肯定反应，否则就可能作出否定反应。在信号呈现时，总是有随机干扰因素——噪音存在。因此，在信号检测实验中，存在两个分布，一是噪音分布，二是信号加噪音的分布（图 7-5），两者常常重叠在一起。决策判断总是有一定标准，决定该作出何种反应。正确地检测到所呈现的信号称为击中，只有噪音时判断为有信号则称作虚报。当在检测信号时比较冒进，击中与虚报的比率都可能很高；如果检测

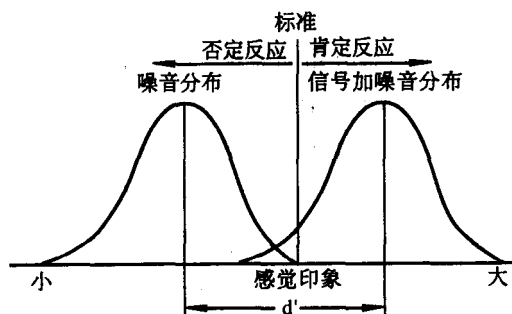


图 7-5 信号检测中的分布与标准

时比较保守，击中率和虚报率就都会很低。根据不同的检测策略，可以作出一条“接受者操作特点曲线”（receiver-operating-characteristic, 简称 ROC 曲线，见图 7-6），曲线的斜率表示出被试所采用的标准。斜率较小时，说明所用的是冒进标准；斜率较大时，则表示出保守标准。ROC 离图中对角线的距离也表明了噪音分布和信号加噪音分布之间的距离，两个分布间隔较大时，ROC 会偏向左上方；分布较近时，ROC 则接近对角线。因此，ROC 可以告诉我们信号检测中感觉过程和决策判断过程的特点。这里， d' 又叫敏感性或辨别力，指 ROC 与对角线的距离；决策判断过程的标准又表示为 β 。信号检测法用 d' 和 β 来决定被试的感受水平或阈限，以及整个检测过程的

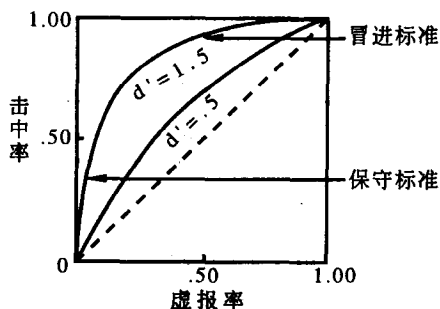


图 7-6 接受者操作特点曲线

心理特征。

d' 的计算运用了正态分布的原理。某研究得到表 7-4 的数据，怎样求出 d' 呢？从数据已知击中率 $P(y/SN) = 28/32 = .875$ ，虚报率 $P(y/N) = 14/68 = .21$ 。从图 7-7 可以看到，噪音分布中平均值零点至 β 的区域为 $.50 - .21 = .29$ ，查正态分布表可得 $t = .8 = \beta$ 。

再看噪音加信号分布 (SN) 由于击中率为 $.875$ ，因而 β 至 SN 零点的区域为 $.875 - .500 = .375$ ，查正态分布表可得 $\beta \approx 1.15$ 。这样， $d' = .8 + 1.15 = 1.95$ 。

表 7-4 某信号检测研究的数据

		条 件	
		信号加噪音 (SN)	噪 音 (N)
观察者反应	肯定 (y)	28 (击中)	14 (虚报)
	否定 (n)	4 (漏报)	54 (正确拒绝)

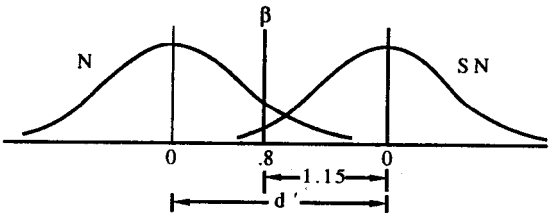


图 7-7 计算 d' 的图解

信号检测论最初应用于人的听觉检测和辨别，后来被用来研究工业检验、医疗诊断、警觉行为和信息检索系统等方面的课题。信号检测法的功效正在越来越多的心理学研究中显现出来。

第七章思考题：

1. 为什么说潜特征理论比经典测验理论前进了一大步？
2. 什么是项目信息函数？如何应用潜特征测验理论进行量表设计？
3. 比较计算测量信度的不同方法。测量的信度和效度的关系如何？
4. 影响信度和效度的因素各有哪些？如何提高信度和效度？
5. 讨论主要几种效度的评定方法。如何运用这几种效度？
6. 在心理测验的使用中应注意哪些问题？
7. 如何进行心理测验的项目分析？
8. 心理物理法和信号检测法各有什么特点？在心理学研究中如何应用这两种方法？

第八章 访谈法、问卷法和观察法

第一节 访谈法与问卷法的特点和类型

一、访谈法的特点和类型

访谈法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，特别是在心理学应用研究中，我们常常用访谈法了解人们的态度、看法、感受和意见，从而对他们的各种心理特征和活动进行研究。有的人对访谈法以及下面要谈到的问卷法有些成见，总觉得这些方法不如实验法“严格”和“科学”，其实这是一种传统的观念。现代心理学认为，心理现象是与物理、化学等现象完全不同的，因此必须有其特殊的研究手段，访谈法和问卷法就是这样一种手段；其次，心理学的研究访谈和问卷测量，不同于日常生活中的“谈天”或者一般的交谈和“民意测定”，它有一整套设计和编制的原则，更重要的是它具有特定的科学目的。正是由于这些原因，人们对于某些经典的心理实验的批评日益增多（例如，效度问题），主张在心理学研究中更多采用谈话、问卷、口语报告等方法，弥补以反应时间表示各种心理特点而造成的严重缺陷。事实证明，谈话、问卷、口语报告等方面所提供的有关心理活动过程的信息，比起传统心理实验中最多采用的反应时，错误率等指标所包含的信息，更为

丰富、完整和深层化。这里，并不是说，实验法已经过时，而是要强调访谈法等方法在心理学研究中的特殊意义。一个重要的前提就是，访谈法建立在心理测量理论和严格的方法设计原则基础之上，并在研究中得到正确的应用。

1. 研究访谈的性质

早在 20 年代，有人就把访谈法定义为“有目的谈话”。根据不同目的，可以把访谈分成各种类型，有用于心理治疗的“精神病学访谈”，有用于人事决策和评价的“人事访谈”，等等。我们所讨论的访谈法称为“研究访谈”(research interview)，它不同于前面提到的精神病学访谈。精神病学访谈常常是由病人发动的，为了获得某种精神上的解脱，医生的目的是改变其行为。研究访谈则是由研究者（访谈者）发动，目的在于采集与研究课题有关的信息。这种访谈集中于研究的特定内容，以期系统地解释、描述和预测一定的心理特征或者活动。研究访谈的结果往往不象心理治疗那样直接影响被访者，而是有比较长远和广泛的利益。

心理学研究中，研究访谈用得很普遍。例如，管理心理学中的双因素动机理论就是通过访谈法进行研究而建立的；在一些心理学实验研究中，也通过访谈法向被试了解心理体验、感觉、信心和自尊等方面的反应，与其它实验指标结合在一起，取得比较全面可靠的结果。目前，在心理学各个领域（除了动物心理研究）的研究中都越来越多地运用了访谈方法。

2. 访谈法的类型

研究访谈按照提问和反应的结构方式不同可分为不同类型，所谓“结构方式”是指事先对问题及其回答方式规定的程度。“有结构”方式是一种有指导性的、正式的、事先决定了问题项目和反应可能性的访谈形式；“无结构”方式则是一种非指导性的、非正式的，自由提问和作出回答的访谈形式。表 8-1 为访谈法的分类。

表 8-1 访谈法的分类

		谈话项目特点	
		无结构	有结构
反应的 可能	无结构	无结构访谈	半结构访谈 I
	有结构	半结构访谈 II	有结构访谈

从表 8-1 可以看到，除了无结构访谈和有结构访谈以外，还有两种半结构式的访谈：（1）半结构访谈 I，要求被试自由地回答预定的访谈问题，也可以用讨论的方式作答；（2）半结构访谈 II，按有结构的方式回答无结构问题。

在研究中采用何种类型的访谈，这要看研究的目的、访谈者对被试人情况的熟悉程度，以及被访人的能力和其它心理特征。

3. 访谈法的优缺点

访谈法有不少优点：它比较灵活、谈话双方都可以随时改变方式，有利于“捕捉”和了解新的或深一层次的信息；访谈法适用面广，可用于不同问题，既可用于成人，也适合于儿童，并可对各种类型的人进行（包括文盲）；访谈法能够比较有效地收集人的态度、知觉、意见等方面的资料，并可进一步探究各种信息和材料，还可以随时观察谈话过程中的行为表现。了解被访人的动机、个性和情感特点；访谈法容易建立主客双方融洽的关系，消除顾虑，使被访人坦率直言，提高结果的信度和效度。

访谈法也有一些缺点，主要是对访谈结果的处理和分析比较复杂，要求由专门的人员进行；访谈人的价值观、信念和偏向也会影响被访人的反应，必须事先进行适当的访谈技术训练；访谈工作比较花费时间和精力，因此代价比较高。

在运用访谈法时，应十分重视访谈设计，注意访谈法可能出现的偏差和口头报告的局限性，并对访谈人（主持者）进行严格而规范的训练。

二、问卷法的特点和类型

问卷法不但是目前心理学研究中最经常应用的研究方法之一，也是经济学、社会学、管理科学和各种教育研究的重要方法。特别是在心理学一些现场应用研究中，由于现代化的生产条件、正常的教学秩序和计划、繁忙的商业活动等限制了进行现场实验的可能性，因此，问卷法的作用就越来越重要。管理心理、教育心理、儿童心理和消费心理等方面的心理学研究，都逐渐以问卷法作为主要研究方法，就是专门研究人机关系和仪表设备设计的工程心理学和工效学专家，也已经认识到问卷方法在许多实验研究中的功效了。

1. 问卷法的性质

问卷法是通过书面形式、以严格设计的心理测量项目或问题，向研究对象收集研究资料和数据的一种方法。它主要采用量表方式，进行定量化的测定；也可以运用提问方式，让受试者自由地作出书面回答。问卷法总是为一定的研究目的服务的，因此，必须根据研究的理论框架和心理量表及问卷设计的原则进行严格的设计和编制。问卷法适用的研究问题也很广泛，采用问卷法，可以系统地了解人们的满意感、基本需要、学习和工作动机、工作紧张、工作负荷、工作疲劳、群体气氛、领导作风、价值观和态度等等。

问卷法在心理学研究中有特殊的意义，它的科学性和价值，不应该由于某些人的滥用或误用而受到怀疑。我们学习和掌握问卷法的设计和使用方面的知识和原则，正是为了在心理学研究中真正发挥问卷法的重要作用，把它作为研究的有效手段。

2. 问卷法的类型

问卷法象访谈法一样，可以分为有结构和无结构问卷，问题提问的方式决定了问卷的结构程度。下面是三种结构程度的问题：

(1) 结构性强的问题：“你对工作感到满意吗？”

是 () 否 ()

(2) 中等结构度的问题：“我觉得自己的工作十分令人满意。”
完全不同意 () 不大同意 ()

有点同意有点不同意 () 比较同意 ()

完全同意 ()

(3) 无结构的问题：“在工作中，你喜欢什么，不喜欢什么？”
(自由回答)

前两种又叫做封闭式问卷，第三种称为开放式问卷。在运用中，我们常常把两种方式结合在一起，以便收集到更多的信息。

问卷法由于采用了量表形式，因此也可以用心理测量的分类方式，即称名型、顺序型、等距型和等比型问卷量表。不同量表类型的记分和统计分析方法是有区别的。例如，如果对称名型问卷表的结果以参数统计方法进行分析，就会得出错误的结论。同时，在统计分析中采用何种统计检验 (χ^2 检验，t 检验，F 检验或者 Mann-Whitney U 检验等)，取决于数据的不同分布，同一种类型的问卷，也许因为数据分布不一样而需要采用不同的统计处理方法。

问卷法一般采用五种项目方式：

(1) 是否式：以是、否或者正、误对问题作出回答，例如，你觉得你工作各个方面的情况使你满意吗？是 ☐ 否 ☐

(2) 选择式：在数个备择答案中选择最符合自己想法的一项，常用的选择式问卷又有称名型和等距型之分。例如：

称名型：你觉得你所在单位的领导在工作中采取了哪一种领导作风？

专制 ☐ 民主 ☐ 放任 ☐

等距型：你对自己的学习常常感到满意吗？

从来没有 ☐ 很少 ☐ 有时 ☐

经常 ☐ 总是如此 ☐

等距型选择问卷实际上是把等距量表转换成选择方式。

(3) 排列式：按照重要性或时间性等标准，对备选答案排出等级或序列。例如：

你对下列活动的兴趣如何？请排出等级顺序：学习（ ）文娱活动（ ）大扫除（ ）社会调查（ ）下工厂（ ）

(4) 填空式：在列出的问题括号中填入自己的情况或看法，例如：姓名 _____ 性别 _____ 年龄 _____ 单位 _____ 收入 _____；我觉得这次学术报告会上最令人感兴趣的论文是 _____，最没有实际价值的论文是 _____。

(5) 量表式：以心理量表方式让受试对问题作出反应，量表有许多类型，主要有 5 点量表、7 点量表和百分量表，例如：

5 点量表的问卷项目：

你有机会参加计算机系统的开发工作吗？

从来没有 难得参加 有时参加 常常参加 一直参加
1 2 3 ④ 5

|-----|-----|-----|-----|-----|

百分量表项目：

你在多大程度上对学校工作感到满意？

0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|

3. 问卷法的优缺点

问卷法的优点很多，它比较客观统一，效率比较高，可以用团体方式进行；结果统计高度数量化、规范化；费用低，不必化很多

力气去训练使用人员；在有的课题的研究中，由于问卷不记名，使得答卷人更加开放，真实地反映自己的各种观点和态度。但是，问卷法也有不足之处，它不够灵活，多数问卷要求以有结构方式回答问题，使人感到不能充分说明自己的态度，有时还会由于不回答许多项目而使问卷失效；问卷法只适合于具有一定文化程度的人，使得应用范围受到限制。研究表明，对文化程度比较低的人使用问卷时，容易出现一些问题，影响研究质量，比如，问卷的回收率低、漏失数据、答卷时作随机反应，等等。这些都是在运用问卷法中应加以注意的。

第二节 访谈法的设计

一、影响访谈效果的因素：

心理学的研究表明，访谈是谈话人之间的一种社会过程和社会交往的产物。谈话双方的心理特征、态度、期望、动机、知觉和行为等相互作用和影响；访谈所处的情境、信息传递的性质等也影响着访谈的效果。图 8-1 表示访谈过程中谈话人之间的交互作用关系。

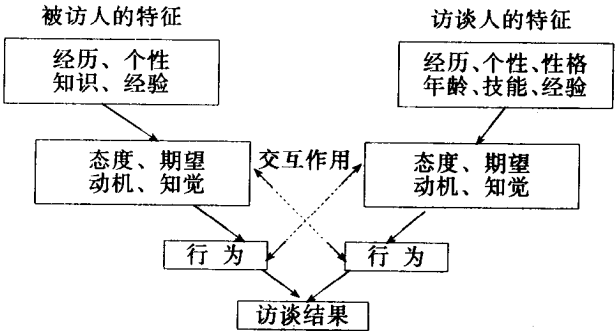


图 8-1 访谈过程中谈话双方的相互作用

1. 成功访谈的条件

访谈要获得成功，不仅要求访谈具备一定的信度、效度和灵敏度，而且还取决于访谈人的技能、被访人的能力与动机以及访谈内容的可行性等因素。概括起来成功的访谈应具备三方面的条件：

(1) 资料的可及性

访谈人需要收集的数据资料是被访人完全能够得到和提供的。被访人的记忆、情绪和拥有的信息等都可能影响资料的可及程度。

(2) 被访人的认知

被访人对于自己在访谈中的角色、访谈要求、问题意义等的理解和认知，直接决定了访谈的实际效果。

(3) 被访人的动机

这是指被访人的谈话动机、价值观、访谈目的与自己需要的一致程度等。心理学研究对被访人动机作出几种解释：内在动机、工具性动机、利他动机或者感情与理智的满足等。

影响被访人动机的因素很多，最近的访谈研究提出一些较普遍的因素，比如竞争、不愿表示出无知、对访谈后果的担心、研究单位的声望、对访谈人的喜爱、社会规范、情绪上的需要，等等。

2. 访谈人的影响

访谈人的特征和行为，对于访谈结果有很大的影响。访谈人不仅需要在访谈过程中观察和记录对方的反应，还必须尽可能取得被访人的合作，在访谈人的提问、记录、激发、探究中，都可能发生影响访谈效果的误差。

(1) 提问方式

访谈问题的提法常常影响对问题的反应，问题的表达如果对双方含义不同，就会造成“提问误差”。因此，了解并善于运用不同提问方式，可以减少这类提问误差。比如，同样是询问对方对自己单位的看法和态度，可以用不同提问办法：

a. “你喜欢这个单位吗？”这种问法较带有感情色彩，而且限制了回答内容；

b. “请你告诉我一些有关这个单位的情况。”这种问法比较中性，可以自由回答，从答话中分析对方对单位的态度。

有的问题比较敏感，容易使被访人产生焦虑甚至敌意，比如，“你上一次什么时候打了人？”这类访谈称为“紧张访谈”或“防御反应访谈”，对访谈人的访谈技巧有较高的要求。

(2) 探究方式

不适当的探究，会造成误差，影响访谈结果。在被访人没有充分时间作出回答，或者没有能充分说明自己的看法时，也会发现这种误差，研究表明，这类误差是与访谈人的经验有关的。

(3) 激发动机

访谈人能不能激发被访人的兴趣和积极性，对于访谈的效果很有影响。许多时候，由于研究需要或者要进一步了解问题，要求进行两次或多次谈话，如果被访人觉得第一次谈话单调乏味、令人不愉快或者并没有意义，就会影响第二次谈话，甚至不愿意再接受访谈。这方面比较重要的一点，是访谈人的真诚与友好，注意建立合作、融洽的关系。

(4) 访谈记录

访谈内容的记录方式也会影响效果。一种方式是在访谈结束后根据记忆作出访谈记录和报告，因而造成比较大的误差。较好的记录办法是在谈话当时作简短的笔记，访谈以后进行详细整理。还有一种方法是录音记录法，有研究证明，运用录音可以减少 25%左右的记忆误差。可是，录音容易引起被访人紧张或注意力分散，而且耗时费钱。实践表明，录音方法比较适合于无结构的访谈。

此外，访谈人的注意、兴趣、倾向、记录能力等也都影响访谈的效果，在访谈过程中应加以注意和控制。

二、访谈的信度、效度和客观性

研究访谈在信度和效度方面有比较高的要求，这也是它区别于一般访谈活动的重要特点。除了前面提到的资料可及性、认知和动机以外，其它许多方法论问题，也会降低访谈数据的效度，例如，问题的设计、编码、被访人的取样，等等。访谈人的技能、个人特征等也都是访谈信度和效度的基础。

访谈信度通常以稳定性来评估，即我们所说的重测信度。还可以用客观性来考察，“客观性”是指两个或多个访谈人（或评分人）在划分反应类别上的一致程度。不同的访谈人或评分人把同样的反应划分为不同类别，则访谈结果就比较主观、可靠性低。因此，客观性也就是评级人的人际信度。严格地讲，除了应从单个访谈人的信度考虑以外，还应该考察多个访谈人的人际信度，又称为复合信度。人们主张以“类别内相关”（intraclass correlation）来估算访谈的信度，这是以方差分析为基础的。著名心理学家吉尔福特（1954）提出单个访谈人的“类别内信度”的计算方式：

$$r_{11} = \frac{V_p - V_e}{V_p + (K-1)V_e}$$

式中 V_p 为被访人或被评人的变异度；

V_e 为差误变异； K 为访谈人的人数。

从公式可以看到，如果 $V_e = 0$ ，则 $r_{11} = 1.00$ 。随着差误变异减少，信度提高。这个公式也可用于多个访谈人的情况。不过，算出的结果只表示了单一访谈人的平均信度水平，复合信度会大于这一信度水平，可以用斯皮尔曼-布朗公式求得信度的增量。

假定由 4 位访谈人分别与 7 位职工一一谈话，访谈人用 10 点量表对被访人作出评定，得到表 8-2 的结果。读者是否能根据量表的统计分析知识，运用吉尔福特的类别内信度公式，估算一下这一访谈

的实际信度 ($SS_{\text{误差}} = SS_{\text{总}} - SS_{\text{被访人}} - SS_{\text{访谈人}}$)。

表 8-2 4 位访谈人对 7 位被访人评定的结果

被 访 人	访 谈 人			
	A	B	C	D
(1)	4	3	4	5
(2)	8	6	9	7
(3)	2	5	3	2
(4)	4	7	9	4
(5)	1	1	2	1
(6)	7	3	9	1
(7)	9	3	6	1

访谈的效度比较难确定，可以根据谈话标准计算预测效度或同时效度，也可以进行构思效度的分析。研究证明，提高访谈信度和效度的最有效办法，是对访谈人进行严格的训练，并且设法控制访谈人的个人特征对于访谈结果（反应）的影响，访谈人个人特征中比较值得注意的是这样一些因素：年龄、性别、专业、文化程度和经验等。例如，年轻的访谈人与中年被访人的交谈，关系容易融洽；同样性别的年轻人交谈，对谈话反应的抑制最少；文化程度较高的访谈人作出的评级高于较低文化程度的人；经验的增长使访谈结果的准确性提高，等等。

三、访谈问题的设计

访谈问题的设计，首先要考虑到成功访谈的三个主要因素：信息资料的可及性、问题的认知因素以及被访人的动机。这就要求在设计过程中注意信息的再认性、语言与用词、概念性质、社会要求与规范以及可能引起防御反应的问题，还应该避免有双重含义的问题和记忆要求高的问题。在设计时，可以根据不同情境、被访人特

点和研究目的，运用不同形式的问题。在访谈设计中经常采用的有以下一些问题形式。

1. 开放式问题与闭合式问题

开放式问题是让被访人自由回答的一类问题；闭合式问题则是限制答话内容或方式的问题形式，包括一些强迫性选择题（在问卷法中常用）。例如：

（1）开放式问题

“当生产任务没有完成时，你们班组里发生一些什么情况？”

“你们班的同学觉得新的奖学金制度怎么样？”

（2）闭合式问题

“当生产任务完不成时，你们班组里的人是不是互相责备和埋怨？”

“你们班里是不是大多数同学都感到新的奖学金制度能够提高学习积极性？”

在访谈设计中究竟应采用开放式问题还是闭合式问题，应该考虑四个方面的因素。

（1）访谈的目的：如果访谈目的不仅要了解被访人的态度或特征，而且还要了解他所持意见的基础、回答问题的参考框架、以及感觉的强度等，则应采用开放式问题，或者用多个闭合式问题了解多方面情况。

（2）被访人的信息水平：开放式问题有助于了解被访人的信息水平。有研究表明，信息水平（对问题情况的了解）不高的被访人希望回答闭合式问题。因此，闭合式问题有助于激发他们的访谈动机。

（3）被访人意见的结构：被访人对于问题是否有明确的态度？是否有足够信息确定自己的意见？意见越明确，就越宜采用闭合式问题。

(4) 对被访人情况的了解程度：在研究初期，由于对被访人情况不够清楚，可以运用开放式问题；而在研究后期，在掌握了相当的资料以后，可以采用闭合式问题进行较大规模的访谈，进行定量分析。

2. 直接问题与间接问题

访谈问题是采用直接方式还是间接提问，这与问题和目的之间的关系有关，即访谈目的是询问一些问题、还是要测量某些概念。直接问题是要求被访人直接表示自己的态度或需要的程度，而间接问题则是以某种理论或推论联系起来的。因此，直接问题可以避免推论的复杂性或风险。但是要求访谈双方有共同的语言。

一般来说，当谈话双方没有很多共同语言、对概念缺乏相同的理解、问题可能产生较大的情绪障碍、被访人的意见会受到社会规范与要求的较大影响时，以间接问题为宜。间接问题有多种方式，可以是意义含糊的图片、故事或句子，也可以是表面直接的问题、但实际上是间接方式，以便取得直接方式得不到的资料。但这样做不容易保证问题的表面效度。

3. 问题的组织与编排

根据访谈研究的结果，在访谈问题的设计时，应注意以下一些组织与编排问题。

(1) 问题的编排使得问题由广泛、一般到具体，由较大的问题到小问题，问题过渡自然。

(2) 访谈初期的问题起说明背景、激发兴趣的作用；重要的问题应放在中部，在被访人积极性已激发起来却还没有感到疲劳时进行关键内容的交谈；容易引起不愉快或疑惑的问题尽可能排在后面。

(3) 对复杂课题，应多用复合性问题或多重问题。在采用多重闭合式问题时，应采用随机化原则，使问题陈述形式有所变化。

近年来，心理学研究中日益主张把定性与定量分析结合在一起，

要求对访谈结果进行量化处理。这就需要设计适当的访谈量表，以便在访谈过程中或者访谈结束后，对被试的反应作出量化评估和编码。这种量表的设计与问卷量表的设计十分相似，我们将在第三节“问卷法的设计”中介绍和讨论其设计原则和方法。

第三节 问卷法的设计

一、影响问卷效果的因素

问卷法是心理学研究的一种十分有效的方法，比较适合于心理学研究的特殊性。但是，问卷法的效果受到许多其它因素的影响，这些因素主要分成三类：人的特征、情境的因素以及测量工具本身的特点。这三类因素还会发生多种交互作用，影响答卷的方式和结果。

1. 受试人的主观倾向

受试人（或被试）变量包含两方面的变异：系统变异和随机变异。在问卷研究中，我们特别注意的是两种系统性的反应偏向或定势，这就是倾向于以符合社会要求的方式答卷和倾向于以默认或接受的方式答题。另外，有的受试人希望表现得合作，显得深思熟虑，造成某种反应偏向。如果受试人在不同测量情境中都有比较相似的反应偏向，则表现出某种个性倾向，需要在问卷设计时注意控制和克服。

2. 工具方面的因素

这是指由测量工具本身所造成的偏差，也有人把它看成被试与测量工具之间的交互作用。例如，过多的问卷项目会使答卷人疲劳、焦虑、注意力下降，使问卷结果（特别是后半部分问卷）的质量大受影响。研究表明，问卷项目的数量一般不要超过 60—70 个，否则，问卷的效果会有明显下降。有的国家规定，在学校与企业中从事任

何问卷调查或研究时，问卷项目不能超过 80 项，这并不是没有道理的。

在问卷设计中采取一些措施，可以克服和控制一部分主观倾向因素，提高问卷质量。例如，运用相反陈述法平衡“默认”定势；用强迫选择法减少社会要求性定势。当然，在问卷中，应该尽量避免采用有较强社会要求性的项目。

对于已造成系统偏差的问卷结果，可以在记分或分析阶段时设法在一定程度上消除这些影响。一种办法是运用“侦察量表”，预测或事后检查被试的反应倾向，根据“侦察”结果来决定校正分数的方法。

3. 情境方面的因素

这方面的因素主要包括问卷实施现场的条件与特点、主持问卷研究的人的行为与态度、实际学习或工作任务的干扰，等等。情境因素中一个比较容易造成问卷“污染”的问题是其它人或同事（同学）的干扰，许多问题调查以团体方式进行，受试人有机会相互讨论或启发，使问卷结果严重污染。对于这个问题，一方面应注意说明问卷答卷要求与规则，另一方面应控制答卷现场。

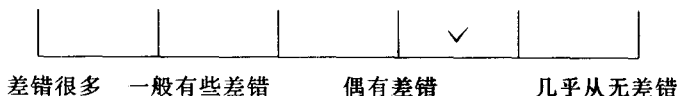
二、问卷设计的一般原则和步骤

1. 问卷量表设计的层次

问卷量表的设计包含着四个层次，这就是问卷的理论构思与目的、问卷的具体形式或格式、问卷项目的语句和问卷的用词。在进行问卷设计时，首先要明确各层次的具体设计要求，以便确定问卷的结构和设计重点。问卷设计中的第一个层次也是最高的层次，是问卷量表的构思与目的。不同的目的和理论依据，决定了问卷项目的总体安排、内容和子量表的构成。第二个层次是问卷的具体形式或格式，例如，应该采用图解式量表还是数字式量表？量表应采取

多少个等级（量表的点数）？是用奇数等级还是偶数？量表上的文字说明该用多少？等等。这些问题都是在第二个设计层次需要考虑的。一般来说，图解式量表比单纯数字式更有利于转达等级意义和评级连续体的心理距离。图 8-2 中，A 是一种图解式量表，B 是数字式。

A：工作成效评定：



B：工作成效评定：1 2 3 4 ⑤ 6

图 8-2 图解式量表与数字式量表

量表等级的点数也是设计中的一个重要因素。根据心理测量理论和研究的结果，多点数一般优于少点数，量表的测量信度随点数增加而提高，辨别力也随之加强。研究表明，当量表点数从 2 点增加到 7 点，信度显著提高；7 点以后，信度提高幅度趋向平稳；当量表点数取到 11 点以上时，信度提高得很少；20 点以上时，信度下降。因此，量表点数在 5 点到 11 点这个范围，是比较理想的。对于具有较强的“趋中”反应定势的被试，或者估计会出现“趋中”定势的项目，应选用偶数等级。

问卷量表设计的第三个层次是问卷的语句，应注意避免复杂语句或带有引导性的问题，还要避免与社会规范有关或有情绪压力的问题，又称为“高负荷问题”。语句层次上特别重要的是使项目用语明确、具体，尽可能避免多重含义或隐含某种假设。问卷的第四个层次是问卷用词，总的要求是避免过于抽象、一般的词语，防止反应定势。

在一些可能有明显社会要求和规范影响的问卷项目上，应设法掩盖研究的目的，采用投射式提问。例如，在有关领导行为的研究中，不直接问被试自己的看法，而是让被试对“周围其他人”的想

法作出评定。被试常常会把自己的看法“投射”到“周围其他人”身上，作出真实反应。

问卷项目的编排应从一般开始，越来越具体，比如先排对学习或工作的整体看法的项目，然后了解有关学习或工作环境、奖励和团体等方面的具体问题。还应该把不影响情绪的问题安排在问卷开始部分，比如，年龄、工作、身体状况等。此外，在项目排列上还应注意顺序效应或位置效应。

2. 问卷设计的一般步骤

问卷设计主要包括下列步骤：

- (1) 根据研究目的与假设，收集所需资料；
- (2) 确定问卷形式，可以从研究者的时间、研究范围、对象、分析方法和解释方法等方面考虑；
- (3) 列出标题和各部分项目；
- (4) 征求意见，修订项目；
- (5) 试测，以 30—50 人为试测样本，求出信度、效度；
- (6) 进行项目分析，重新修订；
- (7) 正式测试。

在问卷设计的各阶段，都需要注意排除和克服影响问卷效果的因素，从严要求。

3. 问卷测量的准确性

问卷设计必须使问卷的测量具有一定的准确性，问卷测量的准确性包括五个方面的特点：效度、测量工具之间的相关、同质性、信度和分数的分布。

(1) 效度

我们在心理测量理论的讨论中介绍过各种效度的特征及其评定方法。常用的指标是问卷测定值与某一标准之间的积矩相关系数。这种方法比较适合于测量项目内容相似的情况，可是多数问卷都包括

几个方面的问题，这就需要分别确定相应的预测项目与标准类别，并求出类别内相关，作为问卷测量的效度评定。常用的公式为：

$$P = 1 - \frac{\sum (X_{i1} - X_{i2})^2}{2n\sigma^2}$$

式中 X_{i1} 、 X_{i2} 分别为问卷量表部分与标准类别的分数； n 为被试人数； σ^2 是总分数的变异值。

(2) 测量工具之间的相关

在心理学研究中，日益主张建立多重效度指标，并检验方法本身（在这里指测量工具）的效应，这就是采用前面介绍过的多特征—多方法分析，检验聚合效度与辨别效度。

(3) 同质性

传统测量理论的模型都假定测量项目之间应比较同质，因此，在问卷设计中应注意其同质性。同质性程度越高，总测量分数的变异一般也越高。常用的检验办法是计算问卷项目的同质比 (HB)，其公式为：

$$HR = \frac{\sigma_t^2 - \sum \sigma_i^2}{(\sum \sigma_i)^2 - \sum \sigma_i^2}$$

式中 σ_t^2 为总分数变异， σ_i^2 是单项目变异， σ_i 为单项目的标准差。

(4) 信度

对于大多数问卷测量来说，不可能获得单一的信度指标，只能在项目水平上保证测量信度。许多人主张，应着重于排除和控制各种影响信度的因素，我们已经讨论过这些因素。

(5) 测量分数的分布

问卷设计研究表明，提高测量准确性的重要途径之一，是控制问卷测量分数的分布。可以通过试测，调整项目结构和项目量表，使分数分布趋向正态分布，提高测量的辨别力。同时，检验项目之间的相关和子量表之间的相关。对于分数分布过分偏态或峰态的问卷，

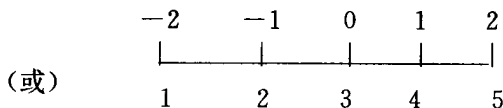
应进行修正，如属于被测行为本身的特点造成偏态分布，需采用非参数统计方法进行数据处理。

三、常用问卷量表的设计

心理学研究中越来越多地需要进行问卷测定，问卷大多运用量表方式。常用的问卷量表有：总加量表法、累积量表法、一致定位量表法和语义区分量表法等。下面就几种主要量表法作一些介绍和讨论。

1. 总加量表法

总加量表法又称利克特量表，是由美国心理学家利克特（Likert, R. A.）于 1932 年创制的。这种量表形式主要有 5 点法、7 点法和 3 点法。典型的 5 点划分为：



它表示了五种反应的连续体。还可以用 7 点法或 3 点法，3 点法常用于有关小学生的研究。

总加量表以重测法和折半法计算信度，量表很灵活，容易设计。所谓“总加”，是指把反应得分加在一起作为测量指标，因此，多数采用单调项目。在设计时，应同时运用积极与消极的陈述，并检验项目分数与总分数之间的相关值。由于这一相关值往往高估，需要进行校正。可以采用校正公式：

$$r = \frac{r_{Ti}\sigma_T - \sigma_i}{\sqrt{\sigma_T^2 + \sigma_i^2 - 2r_{Ti}\sigma_i\sigma_T}}$$

式中 r_{Ti} 为项目与总分数之间的相关；

σ_T 、 σ_i 分别为总分数与项目分数的标准差。

根据校正结果，保留高相关的项目，删除低相关项目，再对问

卷进行构思效度的检验，确定问卷的项目系列。

2. 累积量表法

累积量表法是由古特曼 (Guttman, L.) 于本世纪 40 年代提出的。这种方法强调测量项目的单维性，假定测量同一维度特征的一组项目，可以按难度或数量的连续体排成顺序，表示接受项目所拥有“特征量”的程度。

在项目设计时，把测量项目分解成若干两分小项目，以“是”或“否”方式作答。例如下面一组两分项目：

- | | 是 | 否 |
|-------------------|----|----|
| a. 你身高高于 1.80 米吗？ | __ | __ |
| b. 你身高高于 1.75 米吗？ | __ | __ |
| c. 你身高高于 1.70 米吗？ | __ | __ |
| d. 你身高高于 1.65 米吗？ | __ | __ |
| e. 你身高高于 1.60 米吗？ | __ | __ |

被试的肯定回答都作为 α 反应。量表的逻辑假设是，对 a 答“是”的人总会对其他项也答“是”的，对 b 答“是”的应对 c—e 也答“是”，以此类推。如果在反应中出现不一致（称为反转反应），例如，同意 b 而却不同意 c，那么量表的同质性就可能有问题。这时，需要检验量表的复制指数 (R)。复制指数的公式为：

$$R = 1 - \frac{e}{nK}$$

式中 e 为反转反应的数目， n 是被试数目， K 为项目数目。一般情况下，以 $R = .85 - .90$ 作为可接受的边界， R 低于这个数值时，就应进行修正。例如，某项量表试测得到表 8-3 的数据，可以看到，被试 3、4 分别有一次反转反应，计算复制指数可得：

$$R = 1 - \frac{2}{3 \times 5} = .87$$

由于 $R = .87$ 处于可接受的范围，表明量表的单维性符合要求。当然这是一个假设的研究数据，实际项目分析时，被试与项目都更多，情况也更复杂。

表 8-3 某量表试测结果

项 目	被 试				
	1	2	3	4	5
a	✓	×	✓	✓	×
b	✓	✓	⊗	✓	✓
c	✓	✓	✓	⊗	✓
得 分	3	2	2	2	2

注：✓：是；×：否，⊗：反转反应

累积量表的设计要求比较高，尤其在量表项目的分解时，需要周密的构思，作出顺序结构，在心理学研究中可根据需要采用。目前在能力测量和学业成就测验中运用得比较多。

3. 一致定位量表法

一致定位量表法是由瑟斯顿等在本世纪 20 年代末提出的，是等距量表问卷设计中最常用的方法之一，这方法以心理物理测量为基础，因此比较严密。这种方法的假设是，某一属性的测量，能够按一个量的维度排成等级，这时项目既不是单调、总加性的，也不是累积的，而是表示了不同于其它项目的量值。一致定位量表法要求由多个评判人对项目作出评定，确定量的维度，可以用成对比较法，也可以用连续间距法。然后按照一定规则选定最终量表。一般以 11 点量表作出评级，初选项目应不具有双重含义，并能以赞成或不赞成作出回答。表 8-4 为某项目归类の結果。

这一项目由 80 位评判人以最赞成—最不赞成作出归类评判，并以评判数据分布的中数（该例中为 6.3）作为项目量表值，以分数分

布的第 75 百分位与第 25 百分位之差作为评判的离散度 (Q)。

表 8-4 第 n 个项目的归类评判结果

类 别	把第 n 项目归于每一类别的评判者人数
11 最赞成	3
10	6
9	6
8	8
7	12
6	20
5	14
4	6
3	2
2	2
1 最不赞成	1
合计 80 人	

一致定位量表一般由 20—25 个项目组成,应使所含项目的量表值尽可能分布于 1—11 的范围,如果需要在多个相同量表值项目中选择,应选择评判离散度(以 Q 值表示)较低的项目。选定的最终量表,项目应随机排列。在使用时,要求以同意或不同意对项目作出回答,并用同意项目的量表值的中数作为受试人的问卷分数。

4. 语义区分量表法

语义区分量表是由奥斯古德 (Osgood, C. E.)、苏奇 Suci, G. J.) 和坦嫩鲍姆 Tannenbaum, P. H.) 等在 1957 年根据语义心理学研究提出的,已在问卷设计中被广泛应用。例如,管理心理学研究中有关领导行为的 LPC 量表和 MPC 量表,就是运用语义区分量表法所设计的。语义区分量表的设计采用因素分析方法,确定了语义的三维空间:评价、潜能、活动(评价:好——坏;潜能:强——弱;活动:快——慢),每一维度用 3—15 组双极形容词测定,得出

的分数求出总分，表示人们的态度或看法，还可以比较人们之间的差异。图 8-3 为某语义区分量表实例。

		学校（概念）									
(E) 1.	令人愉快	7	: 6	: 5	: 4	: 3	: 2	: 1			不愉快
(A) 2.	主 动	_	:	_	:	_	:	_	:	_	被 动
(P) 3.	深	_	:	_	:	_	:	_	:	_	浅
(E) 4.	好	_	:	_	:	_	:	_	:	_	差
(A) 5.	快	_	:	_	:	_	:	_	:	_	慢
(P) 6.	强	_	:	_	:	_	:	_	:	_	弱

(E: 评价维度, A: 活动维度, P: 潜能维度)

图 8-3 语义区分量表实例

语义区分量表法的局限是，人们倾向于对自己的喜好作夸大的描述，并倾向于在中区打钩，造成误差。但是，这种方法极为灵活，易于构思，便于使用和记分，因此，虽然历史较短，却已广泛用于心理学各领域。问卷研究证明，这种方法与一致定位量表法有很高的相关。

语义区分量表的数据分析，可以分析分为三个部分分数：概念间分析 (between-concepts)，量表间分析 (between-scales) 和被试间分析 (between-subjects)。奥斯古德等认为 通常的积矩相关系数并不是两个概念之间关系的稳定测定，他们提出 D 统计作为概念之间距离测度：

$$D_{ij} = \sqrt{\sum d_{ij}^2}$$

式中 D 为两个概念之间的线性距离 (i, j)，d 是 i, j 在相同因素 (E、A 或 P) 上的差异量。所以， $\sum d_{ij}^2 = \sum (X_i - X_j)^2$ ； $D_{ij} = \sqrt{\sum (X_i - X_j)^2}$ 。

例如，某被试对五种事物（或概念）在 6 个项目量表上测评的

结果如表 8-5 所示。

表 8-5 某被试对五种概念测评的结果

量表项目	概 念				
	A	B	C	D	E
1	6	2	6	5	3
2	5	2	5	5	2
3	6	1	4	6	2
4	7	1	5	6	3
5	5	3	5	7	1
6	6	2	7	7	2
平均值	5.83	1.83	5.83	6.00	2.17

如果 6 个量表均为“评价”测评，则 A、B 在语义空间内的距离为：

$$D_{AB} =$$

$$\sqrt{(6-2)^2 + (5-2)^2 + (6-1)^2 + (7-1)^2 + (5-3)^2 + (6-2)^2}$$

$$= \sqrt{106} = 10.30$$

依相同方式，也可以求出其余各概念之间的语义距离，并以较近的概念作为同类概念。这种分析方法叫做“距离-聚类分析”。

第四节 观察法的设计

一、观察的类型和特点

科学研究都离不开观察，它和心理学研究的其他方法一样，是收集科学事实和各種心理活动资料的基本途径，也是发展和检验心理学理论的实践基础。观察法是在一定时间内对特定行为表现或活动进行考察，而收集研究资料的一种方法。例如，通过对儿童正常活动的观察，可以了解他们的认知与社会发展的特点，为安排教学

计划和设计儿童活动内容提供证据；通过对操作工人的仔细观察，进行动作分析，为制定技术培训方案提供依据；通过对管理人员日常工作的观察，进行职务分析，作为建立工作责任制和提高管理效率的基础，等等。心理学研究方法的发展趋势中一个引人注目的方面，就是越来越多地运用“参与观察法”，研究者在直接参加被试的学习、工作或活动过程中，系统地收集研究数据。实践证明，把观察法与访谈法及问卷法等结合起来使用，能够比较全面、深入和可靠地了解所研究的心理过程及其特点。

1. 观察法的类型

观察法可以按事先是否确定具体观察项目而分成“有结构观察”和“无结构观察”；还可以按观察者是否直接参加所研究的活动，分成“参与观察”和“非参与观察”。有结构观察中，观察者有比较严密的观察和记录的计划；无结构观察时，则一般没有详细的有关数据收集的观察计划。

参与观察一般比非参与观察效果好，因为观察者参与其中，既有自我体验，又能与被观察人建立融洽的关系，对所观察的活动也有更深刻的了解，并且能更及时地发现新的研究信息。因此，我们主张在有条件时，多采用参与观察。我们（1984，1985）进行的一项有关电脑技术培训的计算机应用心理学的研究，曾经运用参与观察法，直接参加技术培训，从中观察和记录职工在培训过程中认知的、情感的和社会交往方面的特征，取得较好的效果。参与观察的前提条件是观察者能够从事某项工作或活动，因此对观察者的能力要求比较高。

观察还可以按照对行为的不同取样方式，分成“事件取样观察”和“时间取样观察”。事件取样观察只对某种与研究目的直接有关的预先确定了的行为进行观察与记录。时间取样观察则是在一定时间间隔进行观察，对这一时间中发生的各种行为表现作较全面的

记录。时间取样可以随机进行，也可以在可能发生典型行为表现的时间进行，一般应在活动开始、中期和结束阶段都能抽选一定时间作出观察。我们曾在某企业进行一项有关工作责任制的准实验（1981，1982），其中运用了“时间取样观察法”，对职工的工作表现和交往情况进行非参与观察，取得了有关工作表现的比较客观的数据。

2. 观察的特点

在任何观察中，首先要考虑的问题是究竟观察什么，这似乎是显而易见，当然是观察行为特征，其实不然。许多观察之所以失败或者收效甚微，原因之一就是没有十分明确具体的观察对象。例如，要研究儿童独立性与解决问题的能力之间的关系，就希望观察这两个变量。可是，什么是独立性行为呢？需要相当明确而不含糊地作出规定，或者叫做对所观察行为的操作定义。因此，观察者的一项基本任务是需要对行为（观察对象）进行分类和定义。例如，有研究者把课堂行为分成八个行为类别，进而建立课堂行为的全面的观察系统。

观察的第二个特点是观察需要确定“行为单元”，即观察测定中所用的行为成分的大小。我们可以观察与记录很小的行为单元，这样容易取得较高的信度，可是往往使观察效度大受损失。这就涉及到测量程序的“克分子”与“分子”问题，克分子测量方式把较大的行为整体作为观察单元或单位。例如对言语行为的观察，从个体之间的对话或者整个谈话段落或句子作为观察目标。分子测量方式则以较小行为成分作为观察单元。例如，把言语行为的单元细分到短语或者词。采用何种行为单元，这与研究目的、研究的阶段、观察者的经验等都有关系，但对观察的信度与效度影响颇大。观察者应该善于根据不同观察任务，灵活运用克分子测量和分子测量。

观察的第三个特点是观察推论的程度，即观察要求观察者作出

多大程度的推论。分子测量方式的观察要求作出的推论程度很少；克分子测量则要求对观察单元作较大程度的推论。一般来说，观察都要求作出一定推论，大多数研究需要作出较高程度的推论，因而要求对观察者进行严格的训练。在设计观察内容与程序时，应使推论程度保持在适中水平。过大的行为单元还会使不同的观察者对相同行为作出十分不同的解释；但是如果行为单元过于具体，又会使观察太刻板而缺乏灵活性。

观察的第四个特点是观察系统的普遍性或者可应用性。有的观察系统设计可以用于各种不同研究情景，有的则只适合于一定场合，比如课堂、装配线生产车间等。普遍性与可应用性还与行为的取样有关，我们在前面提到过观察的事件取样和时间取样。事件取样使观察比较整体化、效度也比较高；时间取样可以提高行为样本的代表性，但观察内容可能比较零碎，在实际运用时应加以注意。

二、观察的测定

观察法的一个重要方面是观察的测定。在观察时，要求观察者用预先设计好的评级量表对所观察的特征和行为作出评定。许多情况下，观察者当场打分可能会影响被观察者的行为和工作活动，因而不宜边观察边打分，可以在观察结束以后，根据观察笔记，用量表进行有关行为特征的评级。

1. 观察评级中的反应偏向

在观察的评级中，特别需要防止和控制某些反应偏向，其中最常见的有三种。

(1) 晕轮效应

这是以对客体的一般印象而形成恒定的评级倾向。日常生活中常有这种情况：因为某人同意自己的意见，就认为这个人总是聪明的，同时又觉得这个人正直、善良。在测量中则经常有这种晕轮效

应，特别是在评级中，对被观察者的某种一般印象，往往会泛衍到对他的其他各方面特征的评价。研究证明，在所评定的特征的定义不明确、不容易观察或者在伦理道德上比较重要时，晕轮效应特别强烈。在观察评定中，应该十分注意控制这种偏向。

（2）宽大效应

这是指在评级中出现过宽或者过严的倾向，前者称为正宽大效应，后者叫做负宽大效应。在观察评定中也会出现这种倾向，特别应防止对某组或某类型的被观察者评级过宽而对另一组类的人过严的偏向。

（3）趋中效应

这是在评级判断中避免作出极端性判断，而倾向于在量表中段打分的偏向，特别是当观察者对所评定的对象不熟悉时，最容易出现趋中效应。

上述几种偏向都会严重地影响评级的准确性，也使观察测定结果的可靠性下降。为了克服与控制这些偏向造成的差误，研究者提出一些新的观察评级办法，目前比较流行的有混合标准评级量表和行为定位量表。

2. 混合标准评级量表

混合标准评级量表以行为的关键事件作为评级依据。在设计时，先按照观察目的与要求，确定所观察行为的基本维度或方面；然后，为每一维度规定好、中、差三个关键事件，关键事件是指能代表好、中、差行为的典型情况；由此构成混合标准项目，三类事件分别以 G（好）、A（中等）和 P（差）表示，并随机排列。观察时，对照关键事件，对所观察的行为作出“好于”、“一致”或“差于”关键事件的评级，最后把评级结果转换为观察得分。混合标准评级量表的记分方法见表 8-6。

表 8-6 混合标准评级的记分

较好关键事件 (G)	一般事件 (A)	较差事件 (P)	得分
+	+	+	7
0	+	+	6
-	+	+	5
-	0	+	4
-	-	+	3
-	-	0	2
-	-	-	1

表中“+”：好于关键事件；“0”：与关键事件一致；

“-”差于关键事件

3. 行为定位量表

行为定位量表也是以关键事件为基础的。设计时需要由相似的人员分成四个评定小组，最好选择对所观察行为比较熟悉的人员。设计程序包括五个步骤：（1）由第一个评定小组讨论所需观察的行为特征和研究的构想，确定行为表现的所有重要方面；（2）由第二个评定小组就第一组所提出的重要行为方面，分别提出好、中、差的行为事例；（3）由第三个评定小组进行“重译”，这就是给他们提供重要行为方面的表格和随机排列的好、中、差行为事例，要求他们把每一事例重新归类到各个行为方面，如果有不能分回原类的事例，则删除；（4）由第四个评定小组确定经“重译”而保留的项目的量表值，可采用一致定位量表的设计方法（本章第三节）；（5）对保留的项目进行试测和观察评级，求出“评级者间信度”（以评级者间相关为指标），如果符合要求，以 1—9 或 1—11 点的量表轴标出行为事例的量表值位置，制成正式量表。行为定位量表对各行为等级以具体事例定位，定义明确，评级误差小，测量效度较高。

三、观察法的评价

和其它的收集数据的方法一样，观察法有其优点和缺点，应该在心理学研究中扬长避短。

1. 观察法的优点

观察法可以取得被试不愿意或者没有能够报告的行为数据，并且在行为发生的当时作即时记录。因此，观察数据比较客观、全面和准确。同时，观察是在行为发生和心理活动的整个过程中进行的，具有较多的“时间信息”。研究表明，在观察的基础上，比较有把握作出有关心理因果关系的推论，这是问卷或其它方法所难以实现的。

2. 观察法的缺点

观察法也有一些缺点。一是观察结果的质量在很大程度上依赖于观察者的能力和其它特征（疲劳等）；二是在有些情况下，观察活动可能影响被试的正常行为，使观察结果失真；三是观察法的运用，需要化费较大精力和较多时间对观察者进行严格的训练，观察工作的成本也比较高。以上这些方面，都是在使用时应该特别重视的问题。

心理学的研究表明，访谈、问卷和观察等方法各有其长处和特点，因此，在心理学研究中综合地使用这些方法，可以取得更好的效果。我们曾经在有关职工归因及其对工作成效影响的现场研究中，同时运用了这三种方法，测量结果相互补充和验证，提高了研究的效度。有关计算机系统开发的心理学研究中，也先后运用了访谈、问卷和观察等方法。实践证明，这种综合研究模式，比较适合于对较复杂的心理过程和活动作出全面、深入的了解和分析，是心理学应用研究的最有效途径之一。当然，要使这些方法很好地为研究目的服务，还有一些必不可少的前提条件，包括提高研究程序与方法的设计质量、采用严格的实验规范、运用细致周密的分析手段等。

第八章思考题：

1. 研究访谈有哪些特点？访谈法的不同类型有何区别？
2. 访谈法与问卷法有什么优缺点？
3. 影响访谈效果的有哪些因素？如何提高访谈的信度和效度？
4. 在访谈设计和问卷设计中，应分别注意什么问题？
5. 常用问卷量表有哪些类型？在设计中有何具体要求？
6. 观察法有哪些特点？观察测定中应防止什么反应偏向？
7. 如何设计观察评级量表？

第九章 口语报告法、社会测量法和 Q 分类技术

第一节 口语报告法

心理学研究的一个重要趋势，是日益重视认知过程的特点、注意从认知水平上进行深入的分析。在许多领域中，判断、决策和思维过程成为研究的中心课题，包括人们在作出判断和采取行动之前的认知与情绪过程。这种研究注重于对心理活动及其信息加工的整体过程的考察，从而能够详尽、系统地验证心理活动的因果联系，这种研究思路，符合恩格斯所阐明的唯物辩证法的基本思想，“即认为世界不是一成不变的事物的集合体，而是过程的集合体”。进行这类“过程研究”的一个有效的方法，就是口语报告法（verbal protocol），或者叫做“原始报告分析（protocol analysis）”。认知心理学研究中所用的“大声思维（think aloud）”方法，是口语报告分析的一种方式。

一、口语报告法的历史发展

尽管现在所广泛运用的口语报告方法已经具有系统的理论基础和严格的分析程序，它的发展还是应联系到早期内省方法的演变。在心理学早期发展中，对“心理操作”的直接观察，被作为研究心理活动的主要方法。詹姆斯（James, W.）（1890）在其经典著作《心理学原理》中写道：“内省观察是我们必须首先、最重要和总是要依

赖的。”另一位心理学先驱比纳则提出（1912）：“内省是心理学的基础。”因此，在本世纪初，心理学家对于“朴素式”内省（即天真被试作出的内省）的价值已经是很少有疑问了。可是，构造主义学派把内省与特定的实验条件联系起来，又对被试进行专门的理论与实际训练，把内省作为研究所谓“意识经验”的主要途径，用内省方法去描述“感觉成分”，等等，这就使内省方法十分不同于早期的朴素式内省。例如，冯特（Wundt, W.）曾要求被试经过一万次有指导的练习，才能参加正式的实验；铁钦纳（Titchener, E. B.）则主要依赖于实验过程结束后的内省追述报告。与此同时，许多心理物理实验中也采用了内省报告的方法。不久，行为主义心理学家华生（Watson, J. B.）（1913）发动了对意识研究的批评，否定了冯特等人的经典内省方法，特别是分析式内省结果不可重复的问题。同时，他认为，许多心理活动是外显的，思维也可以是外显的活动，只有研究外显活动，心理学才能得到客观的数据。他（1920）写道：“通过让被试对特定问题的出声思维，比起依靠不科学的内省方法来，能够更多地了解和认识思维心理学。”因此，华生（1920）作出了被认为最早的大声思维报告的分析。

早期的口语报告分析，主要是实验者对被试口语报告记录的总结性描述。直到1945年左右出现了录音机，才使得有可能对口语报告进行完整的实时记录，这就显著地改进了口语报告方法。在把记录转换成文字以后，研究者有可能根据课题需要，分析心理过程的任何一个部分。同时，也有可能系统地考察口语报告数据的可靠程度。

口语报告方法的突破性发展，是在50年代后期。当时，纽厄尔（Newell），肖（Shaw）和西蒙（Simon）等（1958）运用计算机，对人类解决问题的认知过程进行了计算机模拟研究，从而提出了有关认知过程的信息加工模型。这一心理学理论上的重大进展，为口语

报告数据的采集、加工、编码和分析，提供了充分的理论依据，使这一方法大大地完善和系统化。对于一些口语报告数据的深入分析和实验表明，利用口语报告方法所获得的信息，可以通过模拟模型再现认知加工过程，这就证明了口语报告方法的可靠性和有效性。70年代以来，这一方法在心理学许多领域的研究中得到运用，使方法本身也得到进一步的发展。近年来，我国心理学研究中也已经采用这种方法，取得了良好的效果。1985—1986年，我们在有关计算机系统开发决策的实验研究中，采用了口语报告法，并对其分析与处理程序作了若干改进，从而系统、深入、全面地考察了管理人员在进行计算机系统开发的决策过程中的信息加工特点，并且取得了有关信息搜索、整合和决策支持信息利用的大量数据，为决策支持系统设计和决策认知技能培养，提供了实验依据。

二、口语报告法的主要程序

口语报告法有各种形式，但是它们都有比较一致的理论假设和基本程序。

1. 口语报告法的理论假设

每一种方法都有其理论依据。口语报告法则是建立在一系列有关人类信息加工过程的基本假设的基础之上的。只有充分了解这些假设，才能正确地运用口语报告方法。

(1) 被试的行为可以看作在一定“问题空间”的搜索，随着这种搜索，被试逐步积累起有关问题情境的知识。

(2) 被试的每一步搜索，都包含着某一种与任务有关的算子(operator)，被试把算子应用于短时记忆(STM)中的知识，从而进入问题空间中新的层次。

(3) 被试的口语报告，相应于其最新获得的信息以及STM中的部分信息。

(4)被试 STM 中所包含的信息（即可报告的信息）主要由算子输入所需要的基本知识、算子所产生的新知识以及目标表征与子目标表征的符号等几方面内容组成。

在以上四项假设的基础上，可以在心理学研究中结合具体课题，运用口语报告法收集有关心理过程的数据，应根据具体目的与要求，选用适当的形式，以便充分发挥这种方法的效用。

2. 口语报告法的基本程序

口语报告法分为问题设计、口语报告、记录与转译、数据编码和结果处理等五个程序。问题设计是这一方法的关键一环，它要求设计出比较容易展开的实验问题。研究表明，提供包含多种因素的问题情境或图解式实验任务，可以使问题设计得更好。

口语报告法要求被试在解决某一问题或者作出某种判断与决策时，根据问题的特点和要求，报告自己的思考与分析过程。要求边考虑边报告，尽可能说明所有的思考细节。与此同时，对口语报告进行实时录音，这种录音应尽量避免干扰被试的思考过程，录音的结果，需要及时地转译成书面材料，供编码和分析用。口语报告结果的编码是这一方法基本程序的重要组成部分，它直接影响着口语报告数据的信度与效度，我们将在下一部分作专门的介绍和讨论。在获得编码数据以后，就可以按照研究的目的和需要，进行统计分析和处理，在对口语报告结果进行处理时，应注意把定量数据和定性材料结合在一起，使之相互补充。除了常见的统计分析以外，还可以进行图解分析或关键词句归类分析等。

3. 口语报告法与传统内省法的区别

现在所广为运用的口语报告法，与传统的内省法有着很大的区别。首先，理论假设与依据完全不同；其次，口语报告法采用天真朴素的被试，不了解有关其行为的理论构思，也没有经过专门训练，而内省法的被试经过专门训练，往往就是研究者自己，明确地了解

理论设想与构思；第三，口语报告法要求在心理活动的当时收集数据，边作业边报告，随着时间和心理活动的发展而测量心理特点，而内省法则在作业完成以后，用问卷、访谈或自我追忆等方式进行。上述区别，也反映出口语报告法的一些特点。正是这种特点，使口语报告法成为心理学研究的有效手段。

三、口语报告的编码和结果处理

对于口语报告记录的材料，需要进行系统的编码。这种编码既有利于对原始数据进行整体、综合的考察，也能够提高结果的信度和效度。进行口语报告的编码，首先要设计相应的编码系统。编码系统的设计，应从两个方面考虑：一方面，编码系统必须能反映任务的特点，适合于特定的内容；另一方面，它必须反映出研究的理论构思，符合理论上的要求。例如，在我们所进行的有关“计算机应用的界面特征和决策策略”（1986）的研究中，设计了“决策过程变量的编码系统”。这个系统包含了认知策略、项目信息加工、记忆加工、信息评估、对策选择和间接陈述等类别，分别对 39 个过程变量进行编码，这个编码系统比较侧重于过程变量，包括搜索、比较、评价、回忆、归因、目标加工和判断（决策）等算子。

在设计出编码系统以后，下一步工作便是实际编码。为了提高并检查编码结果的信度，应由两名以上的人员分别对口语报告录音记录独立地进行编码，从而考察“评级者间信度”（inter-rater reliability），还可以检验口语报告测量的效度。在此基础上，可以对编码数据进行各种统计分析。

口语报告结果的另一种处理方法，是进行“决策树”分析，作为决策判断各步的图解，图 9—1 是“决策树”分析的一个例子。这种分析的一个缺点，是不同的分析者不容易取得一致的决策树，因而难以进行直接的相互比较。西蒙（1975）曾提出以“产生系统”进

行编码，即把口语报告所反映的加工过程，分解成若干“条件-行动”的产生系统，使得编码分析与理论上的考虑更紧密地联系在一起。在采用这种编码方式时，需要具备系统、深入的理论构思，对研究者的分析能力也有更高的要求。

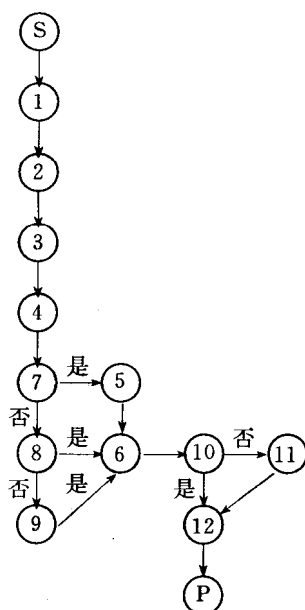


图 9-1 口语报告的决策树分析

四、口语报告数据的运用

口语报告数据有许多用途，按照研究的具体目的与要求不同，主要用于四个方面：一是探索性的研究，作为有关认知过程研究的初期资料，以此寻求有规律性的东西；二是用于对其它数据进行补充，对其它数据作进一步的说明；三是用于检验假设，运用口语报告数据，检验某种理论假设；四是用于建立心理模型，对认知加工过程

进行计算机模拟。总之，口语报告数据是心理学各个领域研究中重要的组成部分，在正确的理论指导下，可以有效地用于不同研究目标。

第二节 社会测量法

一、社会测量的概念和特点

社会测量是由心理学家莫里诺 (Moreno, J. L.) (1934) 提出的一种研究方法，主要用于研究团体内（特别是小团体）成员之间人际关系和人际相互作用的模式，即所谓社会结构——人际知觉方式。这种方法发展至今，已经成为一种比较广义的测量手段，在心理学研究中发挥重要的作用。

社会测量法的基本假设是，在所有团体里都存在相当程度的相互作用，使得各成员在不同程度的积极与消极人际情感的基础上形成一种非正式的组织；这种相互偏爱或疏远的关系会显著地影响团体的士气和效率。在我们运用社会测量方法时，应该注意这种方法的上述基本假设。

社会测量法要求团体成员根据研究者所提供的一种或多种标准，选择一位或多位同伴。目前，社会测量的选择已包含相当广的内容，它已不只是指“人”的选择，还可以选择信息交流渠道、或组织层次；还可以要求被测人对数种选择排出等级。

社会测量法具有一些重要的特点：（1）社会测量法涉及社会性的变量；（2）社会测量容易引起被测人较强的兴趣与动机；（3）测量结果的应用性强，特别适合于团体研究。

二、社会测量法的设计和实施

1. 社会测量的标准

确定社会测量的标准，这是测量设计的首要任务。社会测量标准，是指被测人作出选择的依据。在确定标准时，需要考虑标准的性质、数目以及准许的选择数目。

社会测量标准可以按其重要性分成强标准和弱标准。强标准是指对于工作或生活有重要意义和长期影响的标准，例如一起工作、一起学习等；弱标准往往是临时和短期性的活动或任务，比如，让受测人选择与谁在一起参加某次活动。选择标准时，应注意掌握两条原则：

- (1) 让受测者了解标准的内容；
- (2) 标准尽可能具体，说明一起参加什么工作或活动。

标准可以用积极方式，也可以用消极方式提出，比如，积极方式：“你喜欢与谁一起讨论工作计划？”消极方式：“你不希望与谁一起参加攻关任务”。消极方式容易引起受测者的焦虑和不安，因而要谨慎使用。此外，每次测量一般使用一个标准，作三个选择；如果研究有特殊需要，可以用多个标准和更多选择。

2. 社会测量的实施

在实施社会测量时，应明确指导语，并指出测量目的；让受测者在自愿基础上参加测定，并向他们说明，将不公开测量结果，使他们无拘束地参加测量，也可避免影响正常团体气氛和成员之间关系，并限定团体范围，让受试者在一定范围内选择适当对象。例如，我们曾在某项研究中运用以下的社会测量：

“这儿是你所在小组成员的名单，请按你愿意与谁在一起完成新产品试制任务的程度，对这些成员作出评级，4表示非常愿意与其工作，0表示根本不愿意一起工作，3、2、1则居于其间。”

特别要注意的是，在根据研究目的和任务确定了选择标准和数目以后，应向被测人说明测量意义，熟悉他们的情况，建立融洽的合作关系，使测量顺利进行。

3. 社会测量的几种形式

社会测量有几种常用的方式，上面所举的实例是其中一种，称为等级排列式，对团体成员一一排出等级顺序，然后作出分析。其它的方式有社会测量图、靶式社会图和“猜测”技术等。

(1) 社会测量图：以矩阵方式标出选择对象，按喜爱程度作出等级，然后以等级顺序加权记分。例如，给“最好”的同伴记3分，给“第二好”的同伴记2分，给“第三好”的人记1分。再以这些分数乘以被选次数，得出各人的分数点。

(2) 靶式社会图：这种形式比上一种形式更有用，这是以靶图方式标出被选频次，靶心为频次最高的人，越向外周越是不被人选择的人物所处的区域。

(3) “猜测”技术：这是最简单的方法之一，给受测人呈现一些有关积极或消极特征的简短描述，让他们列出与这一系列描述相匹配的人，然后根据这些选择作出分析。

三、社会测量结果的分析

社会测量结果的分析主要有三种形式：矩阵分析、图解分析和指数分析。

1. 社会测量矩阵分析

这种方法对团体心理特征的研究最有效。通常利用 $n \times n$ 的方形矩阵（ n 等于团体人数）。例如，某研究要求一个五人团体进行社会测量，问：“你愿意与小组内谁一起从事为期两个月的新项目工作？请选择两人。”这一测量取得下列结果（表 9-1）。

这个矩阵表示从 I 选 J 比如 成员 选择 、 ；②选择了 、

等等。分析时考虑两方面的结果：一是选择的性质（单向选择、双向选择还是无选择）；二是选择的数目（合计数）。上述结果中，
、
之间为双向选择（记作 $\leftrightarrow$ ②）；
、
为单向选择（③→④）；
被其他所有成员所选，而
无人选，完全不为人接受。

表 9-1 五人小组的社会测量结果

(成员)	J				
	①	②	③	④	⑤
①	0	1	0	0	1
②	1	0	0	0	1
I ③	0	0	0	1	1
④	0	1	0	0	1
⑤	1	1	0	0	0
合 计	2	3	0	1	4

每一选择以 1 表示。不选到的为 0。

这种分析法主要适合于小团体，当团体增大时，很难从数目差异中纵观整个人际关系，在这种情况下，社会测量图更有效果。

2. 社会测量的图解分析

图解分析又称为“指向图分析”。上例结果如果用指向图表示，可作出图 9-2。

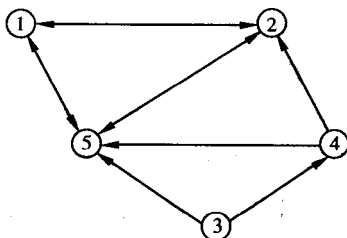


图 9-2 社会测量图解

一看这张图就可发现，成员 处于选择中心； 、 ② ⑤相互选择（双向选择），很可能是团体内的非正式组织（三人以上相互选择）不过用这种图解分析法团体也不宜过大，一般应少于 20 人。否则，较难作出指向图，也不容易解释结果。而且，不同的研究者可以对相同数据作出不同的图解布局，因此没有矩阵分析的结果可比性高。

3. 社会测量指数分析

社会测量指数的分析，也许对心理学研究的实际意义更大，因为它以一个综合指数表示了团体的社会结构关系。指数分析的计算公式如下：

$$\text{个人社会测量指数} = \frac{\Sigma i \text{ 列中被选数目}}{n-1}$$

式中 n 为团体成员人数； Σ 为被选数目之和。

运用这一公式，可以估算出团体中每一成员的社会测量指数，了解每个人在团体结构中所处的位置。上例中（矩阵），成员 和 的指数分别为：

$$\text{成员 的测量指数} = \frac{4}{5-1} = 1.00$$

$$\text{成员 的测量指数} = \frac{2}{5-1} = .50$$

此外，成员之间相互选择，表明了团体内聚力的程度，因此，可以运用团体社会测量指数作为整个团体的内聚力指标，计算公式如下：

$$\text{团体社会测量指数} = \frac{\text{相互选择数目之和}}{\text{所有可能的相互选择总数。}}$$

四、社会测量法的评价

社会测量法是一种简单、经济和自然的采集心理研究数据的方法，可用于各种团体方面的研究。这种方法比较灵活，已被广泛应

用于教育、儿童、管理、人事、消费等心理学领域。

作为一种测量方法，我们还需考虑一下它的心理测量学特征，主要是指它的信度和效度。这方面的研究，采用了重测信度评定和内部一致信度评定方法。结果发现，在使用强标准进行社会测量时，结果比弱标准的测量结果稳定；团体成员年龄对测量信度有影响，随着年龄增长，测量的稳定性提高；当选择数目较少时，测量信度不如多选择时高。研究还证明，测量信息的保密程度和是不是记名对于结果的准确性影响很大，当保密程度较高并且不记名时，测量可靠性比较高。社会测量的效度比较难检查，由于从所问的问题本身并不能理解选择的含义，所以内容效度常常不高。如果社会测量为了选择团体的领导或者组织高效率的新团体，则可以考察一下预测效度或同时效度。实际上，社会测量的选择是与其它许多因素和变量密切有关的，比如团体结构、任务要求、领导作风等，只有把这些因素联系起来，才能真正了解社会测量选择的定义。

社会测量方法也有较大的局限性，了解这一点，可以使我们更好地运用这种方法，适当和积极地弥补这些局限。社会测量的最大局限之一，是选择模式并不能说明选择的理由与原因，选择数据只表明表面性的关系，不能解释其中的因果关系。同样选其人，可以出于十分不同的考虑，选择某成员，并不一定表示偏爱或者吸引。研究证明，团体社会关系是一个动态的复杂过程，除了感情上的好恶，还有认知、动机等各种因素。因此，不能把社会测量的结果作为团体内关系的绝对指标，而必须结合其它手段（个案法、问卷法等）作综合的、深入的、定期的分析。

社会测量结果的分析可以采用等级相关法、 χ^2 检验法、矩阵分析法、因素分析法，各种方法适合于不同研究需要。社会测量数据的应用主要是分组、诊断、治疗和评估四个方面。研究者并不主张完全由“相互选择”的人组成团体，这样会使团体过于同质（感情

上)和封闭化;在诊断方面,主要用于发现“孤立成员”或“学习缓慢者”、“行为问题者”从而进行治疗,包括个别帮助和重建团体等措施;社会测量还可用于对个人情况或学校措施等进行评估。在管理和人事心理学研究中应用也很多,比如,领导才能评价、工作分配、团体结构分析,等等。

社会测量法的发展趋势是广义化、数量化、多元化和系统化。我们在前面已经提到社会测量的广义化,包括了各种社会心理研究。数量化趋势则是与整个心理学研究发展的方向相一致的。现代心理学的研究,不用数量指标和统计方法来进行综合分析的情况已经很少,这已是大势所趋。同时,单纯从感情或“最好的同伴”之类的依据作出的选择,越来越不能满足研究的较高要求,因此,更多采用了多元社会测量,例如,同时收集自我知觉、同伴知觉和教师或上级的评定数据。此外,社会测量法越来越多地与某些理论框架相结合,进行更为系统化的研究。上述趋势,是值得我们运用社会测量法时重视的。

第三节 Q 分类技术

一、Q 技术的性质和意义

我们在前几章讨论的许多分析和研究方法,都是以变量作为分析单元,集中于解释变量之间的关系。还有一种分析方法,这是以“人”作为分析单元,重点在于分析“人际”的关系,这就是 Q 技术。Q 分类技术是由斯蒂夫森(Stephenson, W.) 1953 年在《行为研究》一书中提出的,是用以研究所谓“人间关系”(interpersonal relations)。Q 技术是以 Q 方法论(Q methodology)为基础的,通过 Q 技术基础来说明 Q 方法的具体程序。

Q 技术适宜于小样本或少数被试的研究，特别对个性研究具有重要意义。通常情况下，我们运用测量之间的相关信息进行研究分析，称为 R 方法论。比如用四项测验给一批人测试，然后计算测验所包含的变量之间的关系（相关）。Q 技术则不然，它要计算的是人之间的相关，把人分成几种群类，从而分析不同类型的人员。

二、Q 技术的程序

Q 技术运用等级顺序程序对客体作出分类，这些客体可以是陈述语句、单词、图片等，让受试者根据一定标准把这些客体分类成几堆。例如，让受试者把印有陈述语句的卡片按自己同意与否作出分类，称为 Q 分类。Q 分类通常要求客体数目在 60—120，多数研究采用 100 项，以保证结果的可靠性，而又不使分析过于困难。一般情况下，把卡片按同意的程度不同而分成 11 堆，从“最赞成”到“最不赞成”，并要求按一定比例分类，使分类接近正态分布。表 9-2 为分类的分数和每一堆的卡片数目。

表 9-2 Q 分类的实例

分 数	10	9	9	7	6	5	4	3	2	1	0
每堆卡片数目	2	3	5	7	9	13	9	7	5	3	2 (n=65)
各堆百分比 (%)	3	5	8	11	13	20	13	11	8	5	3

在分类以后，每一陈述获得了相应的分数，例如，某些陈述卡片被分入第 8 堆，则这些陈述均得 8 分。以上是 11 等级的 Q 分类，还可以作 9 等级的 Q 分类。常用的分类数目如表 9-3、9-4 所示。

在数个受试者分类以后（或一个受试者作两次分类）可以求出“人际相关”，例如某 Q 分类研究（四个人）的结果相关（表 9-5），可以说明人与人的关系。

表 9-3 9个等级的 Q分类

	最不赞成 (或最不相同)					最赞成 (或最相同)				
等级	9	8	7	6	5	4	3	2	1	
分数	1	2	3	4	5	6	7	8	9	
卡片数	2	3	6	11	16	11	6	3	2	(n=60)
	4	6	9	13	16	13	9	6	4	(n=80)
	4	6	10	12	16	12	10	6	4	(n=80)
	4	6	10	15	20	15	10	6	4	(n=90)

表 9-4 11个等级的 Q分类

	最赞成					最不赞成					
等级/分数	11	10	9	8	7	6	5	4	3	2	1
卡片数目	2	3	4	7	9	10	9	7	4	3	2 (n=60)
	2	3	5	8	11	12	11	8	5	3	2 (n=70)
	2	4	6	9	12	14	12	9	6	4	2 (n=80)
	2	4	7	10	13	16	13	10	7	4	2 (n=90)
	2	4	8	14	20	24	20	14	8	4	2 (n=120)

表 9-5 Q分类 (四人) 的相关矩阵

人 员	A	B	C	D
A		.92	-.08	-.08
B	.92		-.17	-.17
C	-.08	-.17		.75
D	-.08	-.17	.75	

上面的相关说明了这四个人的相似程度，A、B 两人以十分相似方式分类 ($r=.92$)，C、D 两人也十分相似 ($r=.75$)，其它的相关系数接近零，这说明这里存在两类人。如果这是有关工作态度的结果，则可以说存在 I 型和 II 型不同态度的人员。然后我们以 I、II 两类人所赞成或不赞成的陈述语句的内容，分析这两类人的具体特征。

三、Q 技术的类型

Q 分类有几种方式，主要分成“结构性 Q 分类”和“非结构性 Q 分类”。

1. 非结构性 Q 分类

非结构性 Q 分类是对一组同质项目按不同程度而作出的分类。这种分类方式着重于一种特质的测定，分类项目需要从一项目总体随机选取，预选项目应为实际需要项目数的 1.5—3 倍，让人按相似性评定分类，使之形成近于等距的连续性，再经筛选、修订，形成分类资料。

2. 结构性 Q 分类

非结构性分类就单一领域分类并强调随机抽取项目，结构性分类则不同，它根据费歇尔的实验设计与方差分析原则，就某一理论或假设的变量设计一套项目。例如，有关儿童发展的研究中，可以设计分别反映“内部控制”和“外部影响”两方面内容的项目。因此，这是一个把理论构思结构化的过程。常见的结构性分类有两种：单向分类和双向分类。

(1) 单向结构性分类 (one-way structured Q sorts)

单向结构性分类是运用一种变量分类基础的结构性分类，类似于简单的单向方差分析。例如，以“内倾—外倾”作项目分类设计，或以“开放—保守心理”作另一种项目设计，等等。对分类结果，需

要作方差分析，当然，也可以用相关分析方法，不过，方差分析的结果更为准确。

(2) 双向结构性分类 (two-way structured Q sorts)

双向结构性分类适合于较复杂的理论构思，即进行两个维度的分类。例如，某研究运用 80 项目的分类资料测定社会态度，并检验态度的结构理论，以态度（开放—保守）和抽象性（抽象—具体）两个维度进行分类，得出表 9-6 的数据。我们可以对这一结果进行 2×2 的方差分析（表 9-7）。

表 9-6 双向结构性 Q 分类的平均值

	态 度		平均值
	开 放	保 守	
抽象性 抽象	5.70 ($\sigma=1.59$)	4.15 ($\sigma=1.35$)	4.93
具体	6.70 ($\sigma=1.66$)	3.45 ($\sigma=1.99$)	5.07
平均值	6.20	3.80	5.00

(n=80)

这一结果说明，态度与抽象性交互作用并不显著，而该分类人的态度趋向于开放。

表 9-7 2×2 方差分析结果

方差来源	平方和	自由度	均方	F
态度	115.20	1	115.20	41.74 *
抽象性	.45	1	.45	.16
态度×抽象性	14.45	1	14.45	5.24
误差	209.90	76	2.76	
总和	340.00	79		

(* $P < .01$)

四、Q 技术的评价

对 Q 方法论和 Q 技术的评价，离不开因素分析，即对某项目或人进行因素分析的检验。对 Q 技术的效用，还是颇有争论的，这里讨论一下它的优缺点。

1. Q 技术的优点

Q 技术灵活易用，很受研究者的欢迎，它的最大的优点是与伦理密切结合。结构性 Q 分类完全是理论定向的，因此，项目设计的逻辑性和实用性较强。它的另一个优点是适用于单一被试或很少的样本的研究情境。此外，Q 技术可以用于检验自变量对于复杂因变量的效应，例如，可以使用方差分析和因素分析，有效地评估在各种因素影响下的态度改变过程。Q 技术还特别适宜于进行探索型的研究，有助于提出和发现新的思想和假设。最后，Q 技术所收集的资料，可以用方差分析、相关分析和因素分析等多种方法进行整理和分析，使这一方法更为实用。

2. Q 技术的缺点

Q 技术也有一些缺点和局限。首先是人员的取样问题和代表性，Q 技术作出的有关人员类型分类的普遍性往往不高，因此，应该用横切法研究加以补充；Q 技术的另一个局限，是项目不易满足某些统计处理方法的假设，比如，项目反应的独立性、资料项目数据的等距性，等等。当然，这种影响并不严重。此外，Q 技术具有强迫选择和分类的性质，使得限制了受试者的自由反应，并可能满足重要的信息。可以通过提高受试者的积极性和合作态度弥补这一局限，同时，用访谈法收集补充信息，作进一步的分析和解释。

第九章思考题：

1. 口语报告法的理论假设是什么？口语报告法有哪些基本程序？
2. 如何对口语报告结果进行编码？口语报告法对于心理学研究有何实际意义？
3. 如何设计实施社会测量法？社会测量法的结果有哪几种分析方法？
4. 在进行社会测量时，应注意哪些问题？
5. 什么是 Q 技术？Q 技术有哪几种类型？
6. Q 技术对于心理学研究有何意义？

第四编 研究数据的分析和结果的报告

本书第四编着重于讨论研究数据的统计分析方法和研究结果的总结与报告。第十章“研究数据的基本统计方法”，从数据的特征出发，分别讨论了相关分析、回归分析和方差分析等最常用的统计分析方法。由于心理学和行为科学其它领域研究的大部分数据都是非参数性质的，因而非参数统计日益为人重视。第十一章介绍了若干非参数统计方法及其适用范围，并且讨论了贝叶斯统计和统计决策理论，指出了这些方法在心理学研究中的应用前景。心理活动是由多方面因素所决定的，因此，多元分析方法成为研究复杂情境中心理活动特点的有效手段。第十二章“多元分析方法”是本编的重要内容。在介绍了多元分析的基本概念和主要方法以后，详细讨论了因素分析和多元回归分析两种方法。第十三章“计算机统计分析和研究结果的报告”，介绍了运用计算机统计软件包进行统计处理的方法，并且说明了研究结果解释与总结的方式，最后，讨论了研究报告与论文撰写的规范、各部分的特点及其相互关系。

第十章 研究数据的基本统计方法

第一节 统计分析基础

一、数据的特征

研究数据的统计分析，是心理学研究方法的重要组成部分。要对数据进行有效的统计分析，首先要了解数据的特征。

1. 数据的分布特征

当我们对一组人的特点或一个人多次反应的特征进行测定时，就得到一组数据，这组数据实际上形成一定的频次分布。心理学研究中常见的分布有正态分布 $\left(Y = \frac{N}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \right)$ 、二项分布、t 分布、F 分布和 χ^2 分布等等。数据分布具有一些重要的特征，可以从图 10-1 说明这些特点。

(1) 分布 A 的分数低于分布 B 的分数，分别表示出分布的一般水平，即分布的集中趋势；

(2) 两种分布的变化程度不同：分布 A 中的分数比较紧密地聚集在一起，而分布 B 中的分数则比较分散，这表明数据分布具有不同的离散程度；

(3) 从图中可以看到，B 的分数分布是对称的，而 A 的分数分布却是偏斜的，因而数据分布又具有不同的偏态度；

(4) 分布 A 非常陡峭，而分布 B 相对扁平，这种陡峭或扁平的

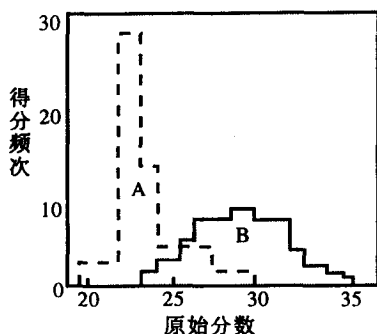


图 10-1 两种不同特征分数的分布

特征称为峰态度。

2. 集中趋势

数据分布的集中趋势有不同的指标，常用的有算术平均数、众数和中数。

算术平均数是数据分布的平衡点，它的计算公式为：

$$\bar{X} = \frac{X_1 + X_2 + \cdots + X_n}{N} = \frac{\sum X_i}{N} \quad (\text{其中 } \bar{X} \text{ 为算术平均数, } X_1, X_2, \cdots, X_n$$

为各次测定值， N 为测定个数)。算术平均数具有两个重要特征：

(1) 各测定值与其平均数之差的总和为零： $\sum (\bar{X}_i - \bar{X}) = 0$ ；

(2) 各测定值与其平均数之差的平方和 ($\sum X^2$) 小于测定值与任何其它点的差数的平方和。

众数是指分布中频次最高的分数，它最少受极端分数的影响，在作粗略估算时，计算简便。

中数是分数的频次分布上 50% 处的分数值，在这个值的两边各有相同个数的分数。当分数数目为偶数时，取分布 50% (最中心处两个分数的平均数作为中数)。中数也有两个重要特点：

(1) 中数对于非平衡性的极端分数没有算术平均数那样敏感，因此，在分布非常偏斜的情况下，最好以中数代替算术平均数，作为

集中趋势的指标；

(2) 中数的取样稳定性不如算术平均数，多样本的中数变化度比较大，常常不够稳定。

3. 标准差和方差（变异）

标准差和方差（变异）是分数离散度的指标。

(1) 标准差

标准差常常用 σ 和 SD 表示(SD: Standard Deviation)。一个原分数总体的标准差计算公式为：

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}} = \sqrt{\frac{\sum x^2}{N}}$$

(其中： x 为原分数， μ 为总体均值， N 为分数数目)

由于研究中常常涉及的是某一总体的样本，如果用上一总体标准差公式计算样本标准差，会出现低估的问题。因此，我们以 $N-1$ 代替 N 来计算标准差。并用 S 表示样本标准差：

$$S = \sqrt{\frac{\sum (X - \bar{X})^2}{N-1}} = \sqrt{\frac{\sum x^2}{N-1}}$$

(其中： X 为原分数， $\bar{X}$ 为样本均值， N 为分数数目)

标准差主要有两个特点：

- 标准差对分布中的每个分数都很敏感；
- 标准差在取样波动的情况下比较稳定。

(2) 方差

方差也称为变异，一组分数的方差正好是其标准差的平方，对于样本分数而言，方差 S^2 为：

$$S^2 = \frac{\sum x^2}{N-1}$$

总体分数的方差 σ^2 为： $\sigma^2 = \frac{\sum x^2}{N}$

由于方差是测量的乘方单位，因而在描述性统计学中用得并不

多。但是，方差经常用于推论性统计学（关系与预测）。在计算时，还可以采用另一更实用的公式：

$$\sigma^2 = \frac{\sum X^2}{N} - \bar{X}^2$$

4. 偏态度 (skewness)

一组数据分布偏斜的程度，可以用偏态度 (Sk) 来表示。按偏斜的方向不同，又可以分为正偏态（向左偏）和负偏态（向右偏），见图 10-2。另一个决定正负偏态的方式是以算术平均数 ($\bar{X}$) 与众数 (M_0) 的距离为指标，当 $\bar{X} > M_0$ 时，分布为正偏态；当 $\bar{X} < M_0$ 时，分布为负偏态。

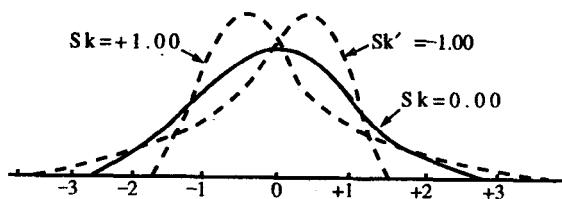


图 10-2 正态分布与两种不同的偏态分布

在实际计算中，Sk 的公式为：

$$Sk = \frac{\sum X^3 / N}{\sigma_x^3}$$

5. 峰态度 (kurtosis)

数据分布的峰态度和原分数与均值之差的四次方有关，常用的指标为 Ku：

$$Ku = \frac{\sum X^4 / N}{\sigma_x^4} - 3$$

峰态分布又可分为尖峰态和低峰态（或负峰态），如图 10-3 所示。

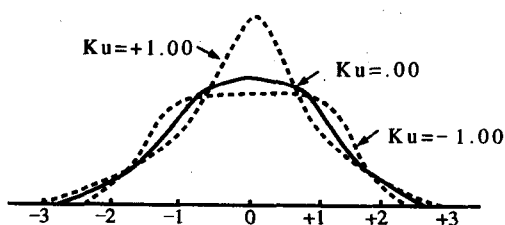
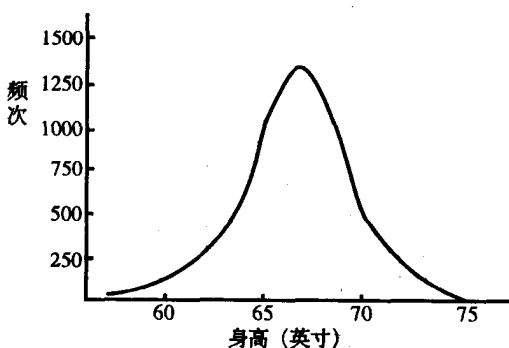


图 10-3 正态分布和两种不同的峰态分布

二、正态分布和标准分数

1. 正态分布

正态分布是统计学中最基本、最重要的分布，也是心理学研究中最常用的数据分布。正态分布的研究至少可以追溯到 17 世纪。当时人们发现，多次观察的差误分布呈有规则的形式，即现在所说的正态形式，因此，这种分布最初被称为“差误的正态曲线”（normal curve of errors）。不久，人们又发现，不仅测量差误，而且许多其它观察数据都呈正态或近似正态的曲线。19 世纪末，高尔顿（Galton, F.）曾对 8000 多人的身体、心理和心理运动测定所得的结果，数据



($M = 67.02$ $\sigma = 2.56$)

图 10-4 8585 名男子身高的频次分布

分布也十分接近正态分布。图 10-4 为身高的测定结果。

在帕斯卡尔 (Pascal) 和贝诺利 (Bernolli) 等人研究的基础上, 德莫弗 (MeMoivre, A.) 提出了正态分布的数学公式

$$Y = \frac{N}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

其中, $Y=X$ 值的频次, N =观察数目, σ =分布的标准差,

π =常数 3.1416, e =自然对数之底 2.7183,

X =分布中的任何分数, μ =分布的均值。

虽然不少人对于这种常见的钟形曲线并不以为然, 而且在心理学研究中也常有误用正态分布的现象, 但是正态分布的许多数学特征, 使得它在统计学中处于某种支配地位。统计学的许多技术问题, 只有在假定总体观察值为正态分布时才得以解决。当然, 有许多变量的分布并非正态分布。例如, 许多社会学变量 (社会经济地位、收入、文化程度、家庭规模等) 呈偏态分布; 某些社会态度变量呈双峰分布 (bimodal distribution) 还有许多“制度性行为”变量呈 J 形分布。

在正态分布中, 平均值和标准差之间具有特定的关系, 即均值与不同标准差之间在曲线下所包含的面积成一定的比例, 表示了分数的百分数, 图 10-5 为正态分布中的分数百分数分配。

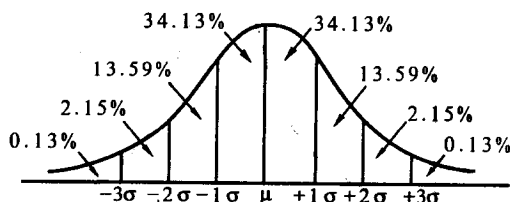


图 10-5 正态曲线下的不同 σ 中的面积比例

2. 标准分数

为了了解个别原分数与总体分布的关系, 应同时考虑总体分布

的均值和标准差，这就需要计算标准分数。不同测量结果的相互比较，也需要先使测量分数标准化。

这里主要讨论心理学研究中两种常用的标准分数： z 分数和标准九分数。

(1) z 分数

z 分数的计算公式为：

$$z = \frac{X - \mu}{\sigma}$$

(其中： X 为原分数， μ 为平均值， σ 为标准差)

z 分数有三个特征： z 分数与原分数的分数分布的形状相同； z 分数的平均值总是等于零 ($\mu_z = 0$)； z 分数的标准差总是等于 1 ($\sigma_z = 1$)。

(2) 标准九分数 (stanine)

另一种标准化方式是采用标准九分数，在计算时，将原分数按其频次分布的积累百分数转换成 1—9 的标准分数。表 10-1 为积累百分数与标准九分数的转换表。

表 10-1 积累百分数与标准九分数的转换

标准九	1	2	3	4	5	6	7	8	9
百分数	4	11	23	40	60	77	89	96	100

标准九分数的平均值为 5.0，其标准差为 1.96。采用标准九分数的好处之一，是便于计算机数据处理，因为计算机许多穿孔卡片为每行九位。标准九分数比 z 分数运算简便，并且更易于比较，在心理学研究中运用比较广泛。

数据分布的其它几种主要类型，如二项分布、 t 分布、 F 分布和 χ^2 分布等，各有其特点和用途，我们不在这里详细讨论，读者可以参阅有关的心理统计学著作。

三、统计功效和统计检验

我们已经知道，在统计检验中主要有两类错误：Ⅰ类错误和Ⅱ类错误。统计学原理指出， α 水平限制了犯Ⅰ类错误的概率，常常把 α 定在 .05 或 .01 从而把错误拒绝 H_0 的可能性限制在较低的水平。同时，把 β 定义为犯Ⅱ类错误的概率。所谓统计功效，则是在统计检验中不犯或避免犯Ⅱ类错误的概率，即：

$$\text{统计功效} = 1 - \beta$$

统计功效 ($1 - \beta$) 与 α 和 β 之间在统计检验中的关系，可以表示于图 10-6 图 10-6 说明，随着 α 的降低， β 相应地增长，而 $1 - \beta$ 相应减小。图中 H_0 为虚无假设， H_1 为备择假设，可以看到，犯Ⅱ类错误的可能性比Ⅰ类错误更大，即研究者较少有可能在 H_0 为真时接受 H_1 ，而更有可能在 H_1 为真时接受 H_0 。

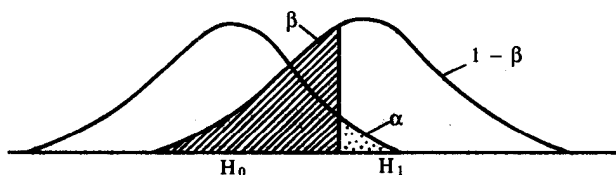


图 10-6 统计功效 ($1 - \beta$) 和 α 、 β 的关系

在一项研究或者实验中，统计功效是发现自变量真实效应的敏感测度。在下列两种情况下，考虑统计功效问题是特别重要和有用的：

(1) 在实验设计之初，设法在研究中提高检验自变量效应的能力，避免犯统计检验错误的可能性；

(2) 当需要对实验结果没有发现任何真实效应这一情况作出解释时，根据实际统计功效的高低，作出适当的分析和推论。

统计检验的功效一般取决于四方面的因素，这四方面因素是我

们在设计和完成一项研究或实验时应该着重考虑的。这四个因素是：

(1) 研究所确立的备样假设 H_1 ，即当虚无假设 H_0 为伪时，假定为真的那个假设；

(2) 研究者所选择的 α 值和相应于 α 的拒绝区域，一般来说，当 α 值较小时，统计功效降低；

(3) 样本的大小，当 $1-\beta > \alpha$ 时，随样本增大，统计功效提高；

(4) 所研究总体的变异程度，在其它条件均等的情况下，总体离散度减小，统计功效将增大。

一些常用的统计检验方法，例如， t 检验、 F 检验、 χ^2 检验以及各种非参数检验，都有特定的假设和适用于不同的数据分布，在运用时，应注意尽可能满足这些假设，防止误用。具体的检验方法，在各种心理统计著作中均有详细介绍。在对多组变量同时作统计检验时，不宜采用 t 检验。这时，应运用方差分析的方法。

第二节 相关分析与回归分析

一、相关分析

1. 相关的概念

心理学研究中，常常需要考虑变量之间的关系，例如，学校班级的规模是否与学习成绩有关系？职工的缺勤率与工作积极性有何联系？等等。要回答这些问题，就需要进行相关分析。相关是用于描述两个变量之间关系的测度，也是许多高等统计技术的基础。

相关的思想是由高尔顿首先提出的。他于 1877 年提出了回归线和回归的标准误差等概念。皮尔逊 (Pearson, K.) 继而推导出相关的统计学度量，称为积矩相关系数。本世纪以来，皮尔逊相关系数已广泛应用于社会科学与自然科学的许多研究领域。在心理学研究

中，布赖恩 (Bryan) (1892) 在他的经典研究 “随意运动能力的发展” 中，第一次把相关方法应用于心理学研究课题。1904 年，斯皮尔曼提出的著名的 “衰减校正公式” (correction for attenuation) 和桑戴克 (Thorndike, R. L.) 提出的 “心理和社会测量的理论” 都显著地促进了相关方法在心理学研究中的应用。

相关与正态分布和双变量正态分布有密切的关系。双变量正态分布呈一个三维的表面，如图 10-7 所示。

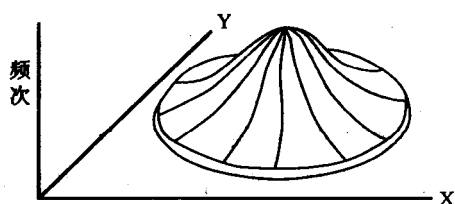


图 10-7 双变量正态分布图

所有的双变量分布都有以下特征：

- (1) 对于 X 的每一个值，与之关联的 Y 值的分布都是正态的，反之亦然；
- (2) X 与 Y 为线性关系；
- (3) 散点分布具有同方差性。

相关系数的取值范围为 $-1.00 \leq r \leq +1.00$ 用以表示双变量数列间关系的方向和强弱。 $r = +1.00$ 为完全正相关， $r = -1.00$ 为完全负相关， $r = 0$ 为无任何相关。相关系数 -0.50 与 $+0.50$ 所表示的关系强度相等，只是方向不同。皮尔逊相关是心理学中运用最广泛的，它的计算公式为：

$$r = \frac{\sum xy}{\sqrt{\sum x^2 \sum y^2}}$$

其中 x 、 y 分别为 X、Y 两个变量的测定值与其平均数之差 ($X - \bar{X}$)、($Y - \bar{Y}$)。

相关系数可以表示出不同的相关程度。如图 10-8 所示。

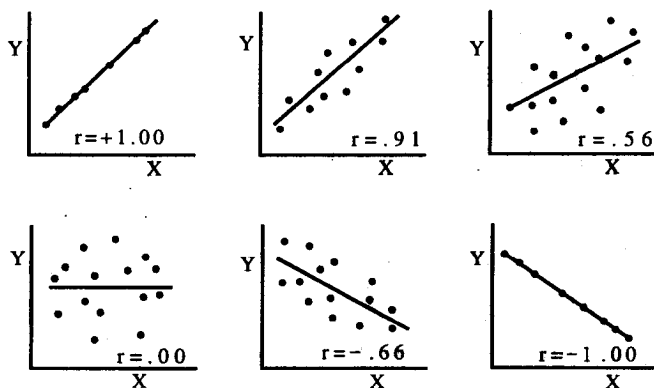


图 10-8 不同的相关程度

2. 相关类型和变量性质

相关有各种不同类型，分别适合于不同性质的变量。皮尔逊相关适用于两个等距或等比测定的连续变量；当所研究的两个变量中一个是两分变量或类别变量，而另一个是连续变量时，应采用点双列相关系数。如果两个变量都是两分变量，则可运用 ϕ 系数。当我们考虑的是两组等级数之间的相关时，需要计算斯皮尔曼的等级相关系数，它的公式为：

$$\rho = 1 - \frac{6 \sum d^2}{N(N^2 - 1)}$$

(d ：两列成对等级的差数)

由于心理测量或其它测定中的差误会降低或衰减相关的程度，因此需要对所求得的相关数作出衰减校正。为此，斯皮尔曼提出一个著名的衰减校正公式：

$$r_t = \frac{r_{12}}{\sqrt{r_{11}r_{22}}}$$

式中 r_i ——真相关, r_{12} ——原相关, r_{22} 、 r_{11} ——两列分数的信度系数。从式中可见, 测量信度越高, 则相关衰减越少, 校正量也越小。

3. 偏相关和复相关

(1) 偏相关 (partial correlation)

在研究包含多个变量的关系时, 我们常常要用到偏相关。两个变量之间偏相关的系数, 是在一个或多个其他变量的所有线性影响被排除的情况下, 这两个变量之间仍然存在线性关系的指标。在三个变量时, 可以得到两两相关系数三个 (r_{12} 、 r_{13} 、 r_{23}) 这样, 在我们使变量 z 恒定的条件下, 变量 x 与 y 的偏相关为:

$$r_{12 \cdot 3} = \frac{r_{12} - r_{13}r_{23}}{\sqrt{(1 - r_{13}^2)(1 - r_{23}^2)}}$$

这一偏相关又称为一阶偏相关, 因为只恒定了一个变量。运用偏相关方法, 可以在研究多个变量的相互关系时, 分离出单一变量之间的关系。例如, 某项研究发现, 文化程度 A、工作经历 B 和工作成效 C 等三个变量之间的关系分别为: $r_{AB} = .50$, $r_{AC} = .50$, $r_{BC} = .30$, 为了明确变量 A、B 在对方恒定条件下与变量 C 的关系, 就需要计算偏相关系数 $r_{AC \cdot B}$ 和 $r_{BC \cdot A}$

$$r_{AC \cdot B} = \frac{.50 - .50 \times .30}{\sqrt{(1 - .50^2)(1 - .30^2)}} = \frac{.35}{.83} = .42$$

$$r_{BC \cdot A} = \frac{.30 - .50 \times .50}{\sqrt{(1 - .50^2)(1 - .50^2)}} = \frac{.05}{.75} = .07$$

可见, 在这项研究中, 文化程度与工作成效实际相关比较高, 而工作经历与工作成效却关系很少。

当我们研究四个变量之间关系时, 可以恒定其中两个变量, 得到二阶偏相关:

$$r_{12.34} = \frac{r_{12.3} - r_{14.3}r_{24.3}}{\sqrt{(1-r_{14.3}^2)(1-r_{24.3}^2)}}$$

(2) 复相关 (multiple correlation)

在心理学研究中，需要同时用多个指标（变量）预测某种心理特征或过程，这就要运用复相关的方法。复相关是数个变量的联合效应与某一变量之间的相关，以大写字母 R 表示。例如， $R_{1.23}$ 表示变量 X_2 和 X_3 的联合效应和变量 X_1 之间的相关，可以通过 X_1 、 X_2 、 X_3 之间的零阶相关系数求得：

$$R_{1.23}^2 = \frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}$$

研究表明，当我们采用两个互不相关的预测指标时，将获得最高的复相关。在心理预测研究方面，还可以用复相关方法发现变量的“抑制”效应。具有“抑制”效应的变量称为抑制变量，它与高效度的指标结合在一起，可以提高和改进整个预测效度。例如，变量 P_1 的效度为 .40，变量 P_2 的效度为 .00，而 P_1 和 P_2 之间的相关为 .80。在这种情况下，把 P_1 和 P_2 结合在一起运用，可以提高预测效度。如图 10-9 所示， $r_{p1c} = .40$ ， $r_{p2c} = .00$ ， $r_{p1p2} = .80$ 。这里， P_2 是抑制变量，它能够解释 P_1 中的一部分误差变异。

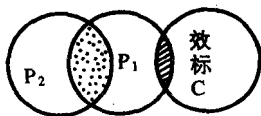


图 10-9 运用抑制变量提高效率

$$\begin{aligned} R_{c.p1p2}^2 &= \frac{r_{cp1}^2 + r_{cp2}^2 - 2r_{cp1}r_{cp2}r_{p1p2}}{1 - r_{p1p2}^2} \\ &= \frac{.40^2 + .00^2 - 2 \times .40 \times .00 \times .80}{1 - .80^2} \\ &= \frac{.16}{.36} \approx .44 \end{aligned}$$

所以 $R_{c.p1p2} = \sqrt{.44} \approx .67$

因此，尽管 P_2 本身的效度为零，但由于 P_2 与 P_1 相结合，提高了原来单独用 P_1 时的效度，即从原效度系数 .40 提高到 .67。复相关一般以 R^2 直接作出解释，表示共同的变异量。计算复相关时，应注意样本的大小和变量的数目。小样本时，复相关 R 值的误差将比较大，因而，一般以 200 人的样本较为合适，此外，由于相关的误差量会随着变量数目的增加而增大，不宜在复相关设计中包含太多变量。

4. 决定性系数

相关系数表明了两个变量之间的关系。可是却很少能说明这种关系的性质，只能说存在某种相对密切的关系或关系很小。如果要进一步说明相互关系的性质，就需要运用一个新的概念，叫做“决定性系数”(coefficient of determination) 表示为 $r^2 \times 100\%$ 。决定性系数决定了两个变量共同变异量的百分数。下面是三个不同的决定性系数所表示的含义。

- (a) 如果变量 X 和 Y 之间的相关为零，则决定性系数为 $0^2 \times 100\% = 0\%$ [图 10-10 (a)];
- (b) 如果变量 X 和 Y 之间的相关为 .80，则决定性系数为 $.80^2 \times 100 = 64\%$ ，这说明两个变量之间有 64% 的共同变异，或者说有 64% 的说明变异原因的因素对于 X 、 Y 变量是共同的。[图 10-10 (b)];

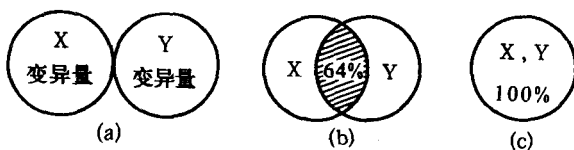


图 10-10 三种决定性系数所表示的关系

(c) 如果变量 X 和 Y 之间的相关为 1.00 则决定性系数为 $1^2 \times 100\% = 100\%$ 。[图 10-10 (c)]。

由于相关分析是心理学研究中最常用的统计分析方法之一，正确地运用决定性系数，可以在更高层次上对相关分析结果作出明确的解释，提高相关分析的功效。

二、回归分析

1. 回归效应和回归分析的假设

(1) 回归效应 (regression effect)

统计方法的主要用途之一是预测将要发生的事件，人们可以运用统计方法，根据学生在中学期间的成绩和某些能力倾向测验的分数，预测他们在大学期间的学习成就。通过统计分析，能够最大限度地提高自变量 X (预测指标) 对因变量 Y (标准或结果变量) 的预测精度。一般来说，相关越高，预测越准确；相关越小，预测误差就越大。统计学的原理告诉我们，除非存在完全相关 ($r=1.00$ 或 $r=-1.00$)，从 X 对 Y 的所有预测都包含着向平均数的回归。

高尔顿在研究了父母及其子女的特征之间的关系以后，首先提出了“回归效应”，例如，父亲高或父亲矮的孩子都倾向于朝父亲身高的平均数“回归”，他把这种回归趋势称为“回归定律”。

在回归分析中，我们并不能肯定因果关系，因此，可以从变量 X 预测变量 Y，也可以从变量 Y 预测变量 X，这取决于我们如何定义这些变量和它们之间的关系。

(2) 回归分析的假设

回归分析都是在一定的基本假设上进行的。回归分析主要有三个假设：

回归线上各点的 Y 分数呈正态分布，即剩余值 (residuals) 为正态分布，但对自变量没有正态分布的假设；

- b. 回归线上各点的 X 和 Y 之间为线性关系，所有剩余值的平均数为零；

回归线上各点的剩余值的方差为同质的，即同方差值。

在大多数回归模型中，还假设自变量都具有固定的值，而不是随机的，例如性别和年龄。

2. 回归方程

图 10-11 为智力测验和学校成绩之间关系的回归线。

回归线可以用一方程式表示：

$$\hat{Y}_i = a_Y + b_{Y \cdot X} X_i$$

其中： a_Y 为 X 为零时的 Y 值或 Y 截距； $b_{Y \cdot X}$ 为回归系数。图 10-11 中所表示的回归方程为： $Y_i = 10 + 2.5X_i$ 。这是从 X 预测 Y ，也可以表示为从 Y 预测 X ：

$$X_i = a_X + b_{X \cdot Y} Y_i$$

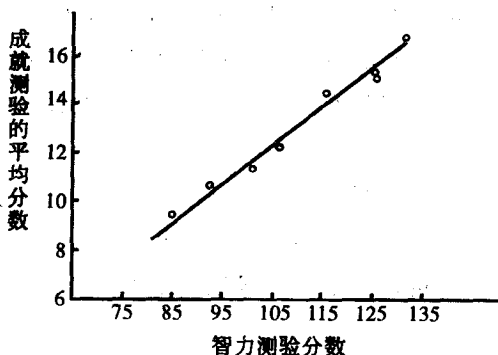


图 10-11 智力测验分数与成就测验平均成绩的回归关系

回归方程中回归系数的计算公式为：

$$b_{Y \cdot X} = \frac{\sum xy}{n\sigma_x^2} \quad \text{或} \quad b_{X \cdot Y} = \frac{\sum xy}{n\sigma_y^2}$$

回归方程中截距值的计算公式为：

$$a_y = \frac{\sum Y_i - b_{y \cdot x} \sum X_i}{n} \quad \text{或} \quad a_x = \frac{\sum X_i - b_{x \cdot y} \sum Y_i}{n}$$

3. 最佳直线拟合和预测误差

(1) 最佳直线拟合

最佳直线拟合是指回归线与散点图中所有各点达到最佳的配合。例如，在图 10-12 中，九个散点表示出两个变量之间的关系，

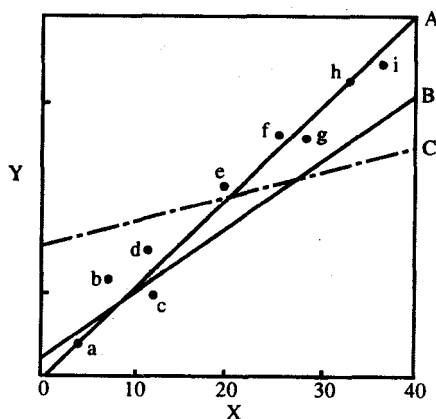


图 10-12 a-i 九个点的三条不同的拟合线

从图中看到直线 A 可以达到最佳的拟合度或配合度，实际上，这条直线与各点的差距是最小的，即这些差异值的平均值最小。在拟合过程中，绝对差异数的平均值并不能完全区分拟合度（例如直线 c），而差异数平方的平均值更能表示这种拟合度。由于这一原因，加上某些数学运算上的考虑，人们以差异数乘方的平均值最小作为最佳拟合的标准，这就是以最小乘方法求出最佳拟合的回归线，其基础是运用了平均数的一个重要特征：平均值是最小乘方点。因此，我们现在可以从两重意义上看回归系数：它可以看作由分数的行列平均值构成的回归线的斜率；也可以是在最小乘方意义上最佳拟合线

的斜率。

(2) 预测误差

回归方程的价值不仅在于它能够描述双变量分布，而且在于它可以利用来作为预测的工具，即运用从一个样本数据得到的回归方程对相似样本的被试作出预测。如果我们得到回归方程 $Y=2+3X$ ，则可以预计， X 得分为 4 的被试，其 Y 值将是 14。回归方程的预测功能对于教育和管理领域十分重要，正在各方面得到广泛的应用。

我们知道，回归方程使我们可以根据个体在某一变量上的分数来预测他在另一变量上的分数，最好的预测为 y_i 或 x_i 。在预测中，我们还需要知道预测误差的大小，对任一个体来说，预测误差为 $y - y_i$ ，即个体的实际分数 y 和预测值 y_i 的差异程度。所有个体的误差值的分布是从 x 预测 y 的准确性的很好指标。这一分布越宽，误差越大，即分布的离散程度 $\sigma_{y \cdot x}$ ，这实际上是偏标准差 (partial standard deviation) 也可以是 $\sigma_{x \cdot y}$ ，即在 x 恒定条件下， y 的离散程度。

在考虑回归线的离散程度时，方差为：

$$\frac{\sum v^2}{n} = \sigma_y^2 - b_{y \cdot x}^2 \sigma_x^2$$

由此可得，
$$\frac{\sum v^2}{n} = \sigma_y^2 (1 - r_{xy}^2)$$

可写作，
$$\sigma_{y \cdot x}^2 = \sigma_y^2 (1 - r_{xy}^2)$$

或
$$\sigma_{x \cdot y}^2 = \sigma_x^2 (1 - r_{xy}^2)$$

即，
$$\sigma_{y \cdot x} = \sigma_y \sqrt{1 - r_{xy}^2}, \text{ 或 } \sigma_{x \cdot y} = \sigma_x \sqrt{1 - r_{xy}^2}$$

这就是“预测的标准误差”的公式。预测标准误差的解释与标准差相似，当分数分布呈正态分布的条件下，可以得到非常明确的解释。

4. 回归和相关的比较

尽管回归和相关都涉及变量之间的关系，并且在计算机都出现

r 和 r^2 这一指标，但是它们却是包含着不同心理学含义的不同概念。在回归分析中，往往区分出自变量和因变量；而在相关分析中，并不区分这两种变量。在回归分析中常常固定 x 值，并假定 y 为正态分布的随机变量；在相关分析中， x 和 y 均为随机变量并且呈双变量正态分布。在回归模型中，不能把 r 解释为 x 和 y 变量之间的线性关系， r 实际上决定于回归线周围散点和 x 的离散度，这两个因素。因此，回归和相关建立在不同的统计假设基础之上，适合于不同的研究目的和要求。当研究的重点在于解释和预测因变量特征时，最好运用回归分析方法。有的心理学家建议，在心理学研究中，只要有条件，就应使用回归分析，只有当不可能运用回归分析时，才考虑使用相关分析。

第三节 方差分析

方差分析（ANOVA: analysis of variance）不只是一种统计方法，它实际上是一种研究方法和研究思路。有人把 ANOVA 与多元回归和因素分析一起，作为现代统计和研究方法的精华，这是一点也不过分的。因此，我们在这里并不把 ANOVA 作为单纯的统计手段，而更侧重于这种方法在心理学研究中的用途和意义。

一、方差分析和它的基本成分

ANOVA 作为一种统计技术，是要对 J 组均值之间的系统性（非随机的）差异作出评估。ANOVA 由统计学家费歇尔（Fisher, R. A.）在他 1925 年的经典著作《研究工作者的统计方法》中首先提出的。它与 t 检验方法相比，主要有以下优点：它可以在进行多组比较时准确地确定犯 I 类错误的概率；它的统计功效更高；它还可以同时评估两个以上自变量的效应。因此，ANOVA 在心理学以及整个行为

科学研究中成为十分有效的工具。

我们在前几章已经谈到，心理学研究所涉及的方差或变异主要有两类：一类是系统变异或实验变异，即显示自变量实际效应的变异；另一类称为差误变异，即由机遇或随机因素而引起的变异。通常，在心理学研究中，把出现在各组之间的由受控制的实验因素所引起的变异（即“组间变异 V_B ”）作为实验变异而把发生在一组内部的、由被试个别差异或实验误差所造成的变异（即“组内变异 V_w ”）作为差误变异，这两方面构成了总变异（ V_t ）：

$$V_t = V_B + V_w$$

组间变异（ V_B ）的计算公式为 $V_B = \frac{\sum x^2}{N}$ （在小样本时为 $\frac{\sum x^2}{N-1}$ ）。例如，在以三种教学方法对三组学生的教学效果的评定中，三组学生的成绩平均值分别为 20，15 和 10，则组间变异（实验效应）为：

	X	x^2	
第一组：20	5	25	
第二组：15	0	0	$V_B = \frac{\sum x^2}{N} = \frac{50}{3} = 16.67$
第三组：10	-5	25	

M: 15

$\sum x^2$ 50

组内变异（ V_w ）的计算方法与 V_B 相似，不过它是就一组内部的数据计算变异或方差量，还可以以同样方式求得总变异（ V_t ）。

在 ANOVA 中需要检验 V_B 所反映变异的统计显著性，这是用 F 比率来运算的： $F = \frac{V_B}{V_w}$ ，并以 F 表评价所求得的 F 比率。在 ANOVA 时，需要分别计算 $\sum x$ ， $(\sum x)^2$ 和 $\sum x^2$ 并以 $C = \frac{(\sum x)^2}{N}$ 作为校正

项，从而分别求得组间和组内平方和（SS）以及均方 $\frac{SS}{df}$ 并以组间和组内均方（MS）之比计算 F 比率： $F = \frac{MS_b}{MS_w}$ ，最后，检验其统计显著性。

ANOVA 分析总是与研究的设计结合在一起的。ANOVA 中最简单的设计称为单向方差分析，又叫“简单随机组设计”。这种设计把随机选取的被试随机地分配到不同实验条件中去，使每种条件人数相等（新的 ANOVA 变式已可处理被试人数不相等的条件），并在不同条件下收集各组的数据。例如，有一项研究试图寻求最佳的授课方式。研究选取了 27 名被试，随机分配到三种授课条件（每种条件 9 人）：（1）一般讲授；（2）一般讲授加补充阅读材料；（3）阅读材料加课堂讨论。在采用上述授课方式以后，对被试进行了课程内容的测试，得到表 10-2 的数据。对这一结果进行了如下单向方差分析。

表 10-2 三种授课方式效果测试得分

条件 1：一般讲授		条件 2：讲授+阅读		条件 3：阅读+讨论	
x_1	x_1^2	x_2	x_2^2	x_3	x_3^2
92	8464	86	7396	81	6561
86	7396	93	8649	80	6400
87	7569	97	9409	72	5184
76	5776	81	6561	82	6724
80	6400	94	8836	83	6889
87	7569	89	7921	89	7921
92	8464	98	9604	76	5776
83	6889	90	8100	88	7744
84	7056	91	8281	83	6889
$\Sigma: 767$		819	74757	734	60088
$N: 9$		9		9	
$X: 85.222$		91		81.556	

从表 10-2 数据, $N=27$, 总分 $=2320$, $\Sigma x^2=200428$,

$\bar{X}_{\text{总}} = \frac{\Sigma x}{N} = 85.926$ 。研究的虚无假设为 $\mu_1 = \mu_2 = \mu_3$ 三种方式效果相同。以 $\alpha = .50$ 检验 F 值。

$$\begin{aligned} SS_B &= \left[\frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} \right] - \frac{(\Sigma x)^2}{N} \\ &= \left[\frac{(767)^2}{9} + \frac{(819)^2}{9} + \frac{(734)^2}{9} \right] - \frac{(2320)^2}{27} \\ &= 408.074 \end{aligned}$$

$$\begin{aligned} SS_W &= \Sigma x^2 - \left[\frac{(\Sigma x_1)^2}{n_1} + \frac{(\Sigma x_2)^2}{n_2} + \frac{(\Sigma x_3)^2}{n_3} \right] \\ &= 200.428 - \left[\frac{(767)^2}{9} + \frac{(819)^2}{9} + \frac{(734)^2}{9} \right] \\ &= 671.778 \end{aligned}$$

$$SS_t = \Sigma x^2 - \frac{(\Sigma x)^2}{N} = 200428 - \frac{(2320)^2}{27} = 1079.852$$

其中, $df_B = K - 1 = 2$, $df_W = N - K = 27 - 3 = 24$,

$df_t = N - 1 = 26$ 。

$$S_B^2 = \frac{SS_B}{df_B} = 204.037, S_W^2 = \frac{SS_W}{df_W} = 27.991,$$

所以, $F = \frac{S_B^2}{S_W^2} = \frac{204.037}{27.991} = 7.29$, 查表, F 值显著。

表 10-3 三种授课方式效果的 ANOVA 结果

方差来源	SS	df	S ² (MS)	F
组 间	408.074	2	204.037	7.29
组 内	671.778	24	27.991	
总 计	1079.852	26		

(当 $\alpha = .05$ 时, F 临界值 $= 3.40$, 因此, 拒绝虚无假设)

表 10-3 表示出 ANOVA 的结果, 表明三种方式有显著的差异。

二、因子方差分析 (factorial analysis of variance)

心理学研究中经常要涉及到多个自变量独立地或共同地对于一个因变量的效应。因子方差分析就是一种分析两个或多个自变量对于一个因变量的独立或交互效应的统计方法。这种分析方法是现代研究设计和统计学中最显著的发展之一，它使得我们得以对两个以上变量的同时性操作和交互作用作出计划和分析。图 10-13 表示了教学方式 A 和学习动机 B 对于学习成绩作用的因子方差分析图，从而考虑 A、B 在不同水平上的组合对于学习成绩发生的效应。

		教学方式 A	
		A ₁	A ₂
学习动机 B	B ₁	A ₁ B ₁	A ₂ B ₁
	B ₂	A ₁ B ₂	A ₂ B ₂

图 10-13 对两种自变量 A、B 效应的因子方差分析

因子方差分析的运算与一般方差分析方法相似，需要分别计算所有组合的组间平方和 (SS_B) 以及所有分数的总平方和 (SS_t) 并求得组内平方和：

$$SS_w = SS_t - SS_B$$

然后，我们还要分别求出各因子的组间平方和。在因子方差分析中，还可能存在着另一种变异或方差，即由于多个自变量的交互作用所引起的变异，当这种变异量达到显著水平时，就需要对这种交互性变异作进一步的单效应分析 (simple effect analysis)。由于现实生活中的各种因素都是以交互方式起作用的，因此在心理学研究中，人们对自变量的这种交互性变异影响越来越感兴趣。交互作用的变异量可以用下式计算：

$$\text{交互作用变异} = \text{所有组间变异} - \text{各因子组间变异之和}$$

在因子方差分析中，因子组间变异称为主效应（main effect），各因子之间都有可能发生交互作用或联合性交互作用。

这里用另一个研究实例，说明双因子的 ANOVA 的运算过程和结果表达方式。教育心理学家一直对教材设计和编排方式十分感兴趣，认为不同的教材编排会直接影响学习的效果。有一项研究对思考题在教材内容中的位置安排和思考题的性质进行了研究，考察不同思考题（知识题与解题能力题）放在教材内容前、后不同位置的效果。研究者设计了 2×2 的因子式实验，即问题位置（前与后） $\times$ 问题性质（知识与能力），每个因子有两种水平，并把 24 名被试（每组 6 名）随机分配到上述四种条件，在学习十个小时以后，对 24 名被试分别给予 50 项目的测试，获得了表 10-4 的结果（测试得分）。

表 10-4 四种实验条件所得的测试分数
问题位置（因子 B）

		内容前 (B_1)		内容后 (B_2)	
问题性质 (因子 A)	知识性 (A_1)	19	23	31	28
		29	26	26	27
		30	17	35	32
	能力性 (A_2)	27	21	36	29
		20	26	39	31
		15	24	41	35

用单向 ANOVA 相似的计算方法，可以求得，

$$SS_A = 18.17 \quad MS_A = 18.17;$$

$$SS_B = 532.42 \quad MS_B = 532.42;$$

$$SS_{A+B+AB} = 627.45;$$

$$SS_{AB} = 627.45 - 18.17 - 532.42 = 76.86 \quad MS_{AB} = 76.86;$$

$$MS_w = 20.13.$$

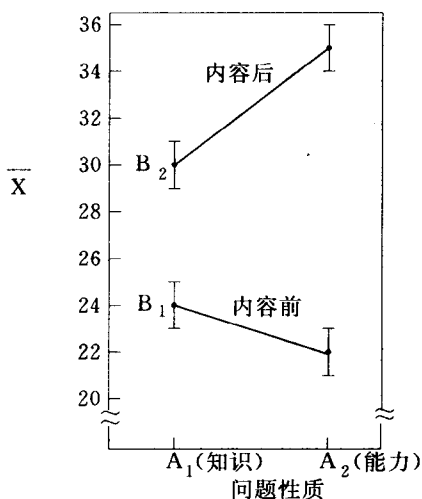
双因子 ANOVA 的结果以表 10-5 表示，可以看到，问题性质因子（A）对学习效果没有显著的主效应，问题位置因子（B）却表现

出十分显著的主效应 ($p < .001$)。同时, A、B 两个因子之间 ($A \times B$) 发现有一定的交互作用。图 10-14 表示出这种交互作用的具体效应。

表 10-5 2×2 的 ANOVA 分析结果

方差来源	df	MS	F
问题性质 (A)	1	18.17	.90
问题位置 (B)	1	532.42	26.45**
$A \times B$	1	76.86	3.82*
因子内	20	20.13	

** $p < .001$, * $p < .10$



(注: 图中 I 符号为置信区间)

图 10-13 问题性质与位置的交互作用

图 10-13 表明, 一方面, 问题安排在内容后比编制在内容前的效果要好得多, B₂ 条件 (在 A 因子的两种水平) 的得分显著高于 B₁ 条件; 另一方面, A、B 两个因子存在交互作用。问题性质不同, 问题

位置的效应也不一样，对于知识性问题，位置的影响要小一些；而能力性问题，影响大得多，在内容前位置条件下不如知识性问题，在内容后位置却显著优于知识性问题。这些研究结果为教材设计和思考题及练习题的安排，提供了具体的心理学依据。

因子方差分析有不少优点，它可以同时操纵和控制两个或两个以上的变量；它能够控制没有操纵的那些变量；它比简单 ANOVA 更精确；尤其重要的是，我们可以用它研究自变量对于变量的交互性效应。正因为如此，因子方差分析在心理学各领域的研究中得到日益广泛的应用。

第十章思考题：

1. 在研究结果的统计分析中，应注意哪些数据特征？不同的统计检验方法分别适合于什么数据？
2. 为什么要使用标准分数？如何运用标准九分数？
3. 什么叫统计功效？如何提高统计功效？
4. 在心理学研究中如何运用偏相关、复相关和抑制变量？决定性系数表示什么含义？
5. 回归分析对于心理学研究的意义是什么？在回归分析中如何决定预测误差？
6. 方差分析有哪些基本成分？因子方差分析有哪些特点？

第十一章 非参数统计、贝叶斯统计 和统计决策理论

第一节 非参数统计方法

一、非参数统计的基本问题

1. 什么叫非参数统计

在心理学研究中，常常需要对所研究的对象作出各种统计判断。统计判断主要涉及两类问题：估算总体值和检验假设。总体值即参数，因此这种统计判断或检验方法称为参数统计方法，它的统计学基础是建立有关总体分布形态的基本假设，一般都是在正态分布的假设下检验虚无假设和作出总体参数推断的。但是，在许多情况下，我们并不知道或不可能肯定总体的分布形态，因此统计学家发展了一种不限定总体分布形态的统计手段，即非参数统计（nonparametric statistics），它主要用于检验假设的方面，同时也已经应用到心理学研究设计中。非参数统计主要适用于称名型和顺序型的数据资料，也可以运用于等距或等比型的数据；而参数统计则仅适用于等距或等比型资料，常用的等级相关就是非参数统计的例子。西格尔（Siegel, S.）（1956）的经典著作《行为科学的非参数统计》，对非参数统计方法及其应用作了系统的介绍，使非参数统计方法日益流行。

2. 非参数统计检验的假设

大多数非参数统计检验方法都有一定的基本假设，主要有以下方面：（1）各观测之间是相互独立的；（2）变量具有基本的连续性和一致性；（3）数据呈顺序型或者称名型。

在采用非参数检验法时，应该注意满足上述基本假设。实际上，非参数统计检验的假设远远没有参数统计检验那样严格，因此它的应用范围比参数统计方法更广。同时，由于非参数统计检验的功效（power）是随 N 大小增大而提高的，并且在心理学和行为科学其它领域的研究中很少能获得真正适合于参数统计的测量，因此，非参数统计检验法在心理学研究中越来越受人重视。

然而，在参数统计方法的所有假设能够得到满足、数据测量也能达到所要求水平的情况下，还是应采用参数统计方法。这时，参数统计方法比非参数统计方法更为有效。

二、常用非参数统计检验方法和样本的特点

大部分有关心理与教育统计的书籍，都对非参数统计检验方法有专门介绍。这里，我们在介绍非参数统计方法的基本概念的基础上，着重讨论如何在心理学研究中根据研究目的和样本特点，选用适当的非参数统计检验方法。

目前可供运用的非参数统计方法很多，在统计检验中首先要考虑数据的某些特征。例如，某些数据具有奇偶的两分特征（例如投钱币），这种数据应采用二项式统计（binomial statistics）。数据的另一特征是其范围（range）在小样本情况下范围是数据变化度的良好指标；数据还具有周期性（periodicity）的特征，等等。

选用何种非参数统计方法，在很大程度上与数据样本的情况有关，取决于所检验的数据来自单样本还是多样本、相关性样本还是独立样本。

1. 单样本检验

在单一样本情况下，常用的非参数统计检验法有二项式检验、 χ^2 单样本检验和 Kolmogorov—Smirnov 单样本检验等等。其中，Kolmogorov—Smirnov 检验法较适合于检验各样本分布之间或样本分布与理论分布之间配合的良好性（goodness of fit）这种方法对于小样本的检验特别有效。

kolmogorov—Smirnov 方法以虚无假设 H_0 条件下的理论分布的累积分布值，与观察值的累积频次分布作比较，并以两者的“最大离差” D 作为检验的指标。例如，某研究考察操作者对于信号颜色明亮程度的偏爱，以便选出适当的设计参数，由 10 名被试对五种明暗等级作出选择，得到表 11-1 的数据。分析可得， $D = \text{最大值} [F_0(x) - S_{10}(x)] = \frac{5}{10} = .500$ ，查表，当 $N=10$ 时， $D \geq .500$ ，达到显著水平，被试更喜欢较亮的颜色。

表 11-1 对信号颜色的偏爱研究结果

颜色等级：暗	1	2	3	4	5	明
被试选择的频次	0	1	0	5	4	
理论累积分布 $F_0(x)$	$\frac{1}{5}$	$\frac{2}{5}$	$\frac{3}{5}$	$\frac{4}{5}$	$\frac{5}{5}$	
观察值累积分布 $S_{10}(x)$	$\frac{0}{10}$	$\frac{1}{10}$	$\frac{1}{10}$	$\frac{6}{10}$	$\frac{10}{10}$	
$ F_0(x) - S_{10}(x) $	$\frac{2}{10}$	$\frac{3}{10}$	$\frac{5}{10}^*$	$\frac{2}{10}$	0	

* $p < .01$

Kolmogorov—Smirnov 检验法可以用于很小的样本，并分别地处理各个观察值，因此在不少方面优于 χ^2 检验方法，既简便，又具有较高功效。

2. 两个相关样本的检验

所谓相关样本，是指运用“匹配”等程序对外源性差异加以控

制的两个样本。在参数统计中，常常用 t 检验法对这类样本作出差异检验；而在非参数条件下，常用的方法有：**McNemar** 的变化显著性检验法、符号检验法和 **Wilcoxon** 的配对符号等级检验法等等。**McNemar** 检验法常常用于“事前—事后”研究设计，而且被试同时作为对照组的情况；符号检验法适合于用正负号作为数据的两个样本的差异检验，但它只利用了各对差异的方向信息；**Wilcoxon** 配对符号等级 (matched—pairs signed-ranks) 检验方法则不但考虑到成对数据的方向差异信息，而且考虑到差异值的相对大小，因而能更充分利用数据信息，提高检验精度。这种方法对成对被试测量得分差异的等级进行分析，使等级与差异值相同符号，选择其中符号频次较少的等级，以这些等级之和的绝对值 (T) 作为检验指标，然后查表看其显著性水平。**Wilcoxon** 配对符号等级检验法已成为心理学研究中最有用的统计检验方法之一。

3. 两个独立样本的检验

在许多研究中，难以获得相关样本或者不适宜采用这种样本，这时，可运用两个独立样本（以随机方法获得）。对于这类样本的非参数统计检验，主要有 **Fisher** 准确概率检验法、 χ^2 独立样本检验法、中数检验法和 **Mann—Whitney U** 检验法等。当两个独立样本大小相同时，**Fisher** 准确概率检验法是分析称名型或顺序型离散数据的极其有用的非参数技术，中数检验法用于检验两个独立样本集中趋势的差异，它适合于顺序测量水平以上的各种数据。对于这类数据，还可以用 **Mann—Whitney U** 检验法来检验两个独立样本是否来自相同总体，这种检验法比中数检验法的功效更高。

4. 多样本检验和非参数相关测定

在包含多个相关样本或独立样本的研究中，除了采用 **Cochran Q** 检验法或 χ^2 检验法以外，还常常运用非参数方差分析。有关非参数方差分析的方法将在后面作详细介绍。

在非参数统计中，非参数相关测定及其显著性检验占有重要的地位。非参数相关测定是对称名型和顺序型数据的相关运算，其方法比参数统计中的皮尔逊相关计算容易一些。常用的非参数相关指标有：列联系数 C 、斯皮尔曼等级相关系数 ρ 、Kendall 等级相关系数 r 、Kendall 等级偏相关系数 $r_{xy \cdot z}$ 和 Kendall 一致性系数 W 等等。它们的计算公式分别为：

$$(1) \text{ 列联系数: } C = \sqrt{\frac{X^2}{N + X^2}};$$

$$(2) \text{ 斯皮尔曼等级相关系数: } \rho = 1 - \frac{6 \sum d^2}{N(N^2 - 1)};$$

$$(3) \text{ Kendall 等级相关系数: } r = \frac{S}{\frac{1}{2}N(N-1)},$$

$$(4) \text{ Kendall 等级偏相关系数: } r_{xy \cdot z} = \frac{r_{xy} - r_{zy}r_{xz}}{\sqrt{(1 - r_{zy}^2)(1 - r_{xz}^2)}};$$

$$(5) \text{ Kendall 一致性系数: } W = \frac{S}{\frac{1}{12}K^2(N^3 - N)}。$$

三、非参数方差分析

1. 单向方差分析：Kruskal-Wallis 检验

这是一种运用等级数据进行方差分析的非参数检验方法。它的基本公式为：

$$H = \frac{12}{N(N+1)} \sum_{j=1}^k \frac{R_j^2}{n_j} - 3(N+1)$$

其中 N 是等级的总数， n 为一组中等级的数目

R 为任一系列中的等级总数。

下列为三组人员成绩评定等级的列表（表 11-2）要求检验三个小组的等级是否存在差异。

表 11-2 三组成绩评定的等级

	小 组		
	I	II	III
	12	11	4
	14	16	3
	10	5	8
	17	7	1
	15	6	9
	18	2	
Σ 等级:	81	47	25
M:	13.50	7.83	5.00

先计算

$$\sum_{j=1}^k \frac{R_j^2}{n_j} = \frac{(81)^2}{6} + \frac{(47)^2}{6} + \frac{(25)^2}{5}$$

$$= 1093.50 + 368.17 + 125.0$$

$$= 1586.67$$

再运用 Kruskal-Wallis 公式：

$$H = \frac{12}{17(17+1)} \times 1586.67 - 54$$

$$= 62.22 - 54 = 8.22$$

H 的分布近似于 χ^2 分布 自由度为 $K-1$, 即 $3-1=2$ 。从 χ^2 表, 我们可以发现 H 值处于 .02 显著性水平。

Kruskal-Wallis 方法类似于单因子方差分析, 简单有效, 可用于实验性和非实验性数据的分析。在一些心理学、社会学和教育研究中, 如果对测量结果是否可以作参数分析有疑问, 都可以通过对原分数评定等级顺序而运用 Kruskal-Wallis 检验方法。即使对于十分不规则但与等级有些相似的数据, 也可采用这一方法。

2. 双向方差分析: Friedman 检验法

在心理学研究中，当对被试运用了匹配程序，或者对相同被试进行了一次以上的观察和测定的情况下，可以运用弗里德曼（Friedman, M.）（1937）首先提出的等级顺序方差分析法，也可采用普通的双向等级方差分析。

例如，在一项有关角色和对教学能力的知觉之间关系的研究中，要求教师作出相互评定，并且让行政管理人员和学生评定这些教师，参加评定的各类人员人数不等，需要检验的是这些评定有没有显著差异。评定结果如表 11-3。对于这些数据，可以采用普通的双向 ANOVA，但要求这些数据基本满足参数分析的假设。Friedman 方法用以检验行或列之间的等级差异，他提出的公式为：

$$\chi_r^2 = \frac{12}{kn(n+1)} \sum_{j=1}^k R_j^2 - 3K(n+1)$$

其中 χ_r^2 为弗里德曼检验统计值；K 为评级的数目；n 为所评等级的客体数目； R_j 为每一列中等级的总和； $\sum_{j=1}^k R_j^2$ 为总和平方之和。

表 11-3 对六位教师的评定数据

教 师	同事之间	行政管理者	学 生	ΣR
A	(3)28(3)	(1)19(1)	(2)22(1)	5
B	(1)22(1)	(2)23(2)	(3)36(3)	6
C	(2)26(2)	(1)24(3)	(3)29(2)	7
D	(2)44(6)	(1)34(4)	(3)48(6)	16
E	(1)35(4)	(2)39(6)	(3)40(4)	14
F	(2)40(5)	(1)38(5)	(3)45(5)	15
ΣR	11	8	17	

（表中右括号内为列等级 左括号内为行等级 数目越大 所知觉的能力越高。）

首先计算 $\Sigma R_j^2, \Sigma R_j^2 = (11)^2 + (8)^2 + (17)^2 = 474$ 因为 $K=6, n=3$ 所以

$$X_r^2 = \frac{12}{(6)(3)(4)} \times 474 - (3)(6)(4) = 79 - 72 = 7$$

$df = n - 1 = 3 - 1 = 2$,查 χ^2 表 , 差异显著水平为 .05。可是 , 当 K 和 n 都比较小的情况下 , 显著性水平是值得怀疑的。

上述结果表明 , 三组评定之间有显著差异。我们还可以检验被评级的教师之间有无差异 , 这时 $K = 3$, $n = 6$, 从公式可得 ,

$$X_r^2 = \frac{12}{(3)(6)(7)} \times 787 - (3)(3)(7) = 11.95$$

由于 $df = n - 1 = 6 - 1 = 5$,查 χ^2 表 , 教师之间在 .05 水平上有显著差异。有人对同样数据 , 以普通 ANOVA 作了检验 , 得到相同结果 , 这说明 , 在许多情况下 , 两种方法相当一致 , 而等级 ANOVA 要简便得多。

3. 一致性系数 : W

测定等级之间的联系程度 , 是检验假设的更直接方法。著名计学家肯德尔 (kendall) 提出的一致性系数是这种等级相关的重要指标。大家知道 , 常用的等级相关系数 (如斯皮尔曼等级相关系数) 表示了两个变量等级之间的关系 , 而肯德尔的 W 系数可以表示 K 个变量之间的关系。因此 , 肯德尔的 W 系数特别适合于多变量的研究或者有关多次评定之间信度的研究。例如 , 我们希望了解上例中各列等级的一致性程度 , 如果两组评定人的评级之间没有联系 , 则等级相关系数应接近于零。 W 可以表示等级之间的平均一致性程度 (从 .00—1.00) 。我们可以用两种方式确定 W 。

一种方法是以完整的等级方差分析的组间 (或等级间) 平方和与总平方和之比来表示 W , 这一比率为等级数据的相关比率的平方 E^2 。

当有 n 个客体和 k 个等级时 , Kendall 的一致性系数为 :

$$W = \frac{12S}{K^2 (n^2 - n)}$$

其中 S 为 n 个等级与其平均等级差异值的平方和，即等级的组间平方和 (SS_b)。

从上例中可以简化为表 11-4 的数据：

表 11-4 三类人员对教师的评定结果

教师	同事之间	行政管理者	学生	Σ	Σ^2
A	3	1	1	5	25
B	1	2	3	6	36
C	2	3	2	7	49
D	6	4	6	16	256
E	4	6	4	14	196
F	5	5	5	15	225
Σx	21	21	21	63	$\Sigma x_i = 63$ $(\Sigma x_i)^2 = 3969$ $\Sigma x^2 = 273$ $\Sigma (\Sigma)^2 = 787$

我们可以用公式 $S = \Sigma X^2 - (\Sigma X_i)^2/n$ ：

$$S = (5^2 + 6^2 + \dots + 15^2) - (63)^2/6 = 787 - 661.5 = 125.5。$$

Kendall 用另一种方法也可以得到相同结果，即以每一等级与总和平均值之差的平方和作为 S 值，因为 $63/6 = 10.5$ ，所以

$$\begin{aligned} S &= (5 - 10.5)^2 + (6 - 10.5)^2 + \dots + (15 - 10.5)^2 \\ &= (-5.5)^2 + (-4.5)^2 + \dots + (4.5)^2 = 125.5。 \end{aligned}$$

在该例中 $K=3$ ， $n=6$ ，所以：

$$W = \frac{12 \times 125.5}{(3^2)(6^2 - 6)} = \frac{1506}{(9)(216 - 6)} = \frac{1506}{1890} = .797 = .80，$$

运用等级方差分析，也可得到同样结果（表 11-5）， $E^2 = .80$ 。运算过程如下：

$$\text{以上例数据得到：} C = \frac{3969}{18} = 220.50，$$

总方差 = $273 - 220.50 = 52.50$,

评级者之间方差 = 0,

教师之间方差 = $\frac{787}{3} - 220.50 = 41.83$,

表 11-5 等级方差分析

方差来源	df	SS	ms	F
评级者之间	2	0	0	
教师之间	5	41.83	8.366	7.84 (.01)
剩余方差	10	10.67	1.067	
总 和	17	52.50 (SS _t)		

$$E^2 = \frac{SS_b}{SS_t} = \frac{41.83}{52.50} = .797 = .80$$

由此可见,三组等级之间的关系很密切。W系数的显著性水平可以用F比率进行检验,这一例中 $W = .80$,查F表,W达到.01显著水平。

从上述两种方法的比较来看,一致性系数的运算更简便,适用范围也更广,在心理学许多领域的研究中,特别是在心理测量的结果分析中,都可以运用W系数考察多个变量之间的关系。

第二节 贝叶斯统计

一、贝叶斯统计的一般问题:条件概率和贝叶斯定理

我们知道,传统或经典的统计方法和程序,都是以样本信息作为基础的,即以样本均值、样本方差等样本统计数据去研究和考贝有关总体特征的假设。这种方式又称为“取样理论”式的统计推论。从广义上讲,经典统计学包括统计推论和统计决策程序,它们都只

是以样本资料为基础的。

第二种统计推论和决策方式称为贝叶斯统计。它与传统或经典统计学的主要差别在于，它正式地利用了样本信息以外的信息，并使各种信息结合在一起作为统计推论和决策程序的基础。用以结合信息的机制就是贝叶斯定理（Bayes theorem）。贝叶斯定理是 19 世纪由英国牧师贝叶斯（T. Bayes）发现的。随着主观概率理论的发展，贝叶斯统计学在 50 年代后期发展起来，特别在商学和社会科学中越来越受人重视。

之所以要采用贝叶斯方法（Bayesian methods），是由于人们需要运用任何或所有可取的信息（无论是样本信息还是其它信息）进行统计推论和决策。这一点对于决策理论特别重要，因为单凭样本信息不容易作出非常准确的决定。在心理学其它研究中，贝叶斯方法的意义也是显而易见的，它可以充分利用研究者有关课题的知识，对各种假设确定先验概率，并运用贝叶斯定理作出结论和推断。因此，许多心理学家主张逐步采用这种统计方法。

贝叶斯统计的基础是条件概率的概念，即事件 A 在事件 B 已经发生的条件下可能发生的概率 $[P(A/B)]$ ，这一条件概率为：

$$P(A/B) = \frac{P(AB)}{P(B)}$$

例如，一个坛子里有 100 个球，其中 40 个大球，50 个蓝色球，10 个既是 大球又是蓝色的。问题是要求出在已抽取的球为蓝色这一条件下，随机抽取大球的概率 $P(\text{大} / \text{蓝色}) = \frac{P(\text{大而蓝色})}{P(\text{蓝色})}$

根据题意，

$$P(\text{大} / \text{蓝色}) = \frac{P(\text{大而蓝色})}{P(\text{蓝色})} = \frac{.10}{.50} = .20,$$

这就是在蓝色球条件下抽取大球的条件概率。同样可得：

$$P(\text{蓝色} / \text{大}) = \frac{P(\text{大而蓝色})}{P(\text{大})} = \frac{.10}{.40} = .25,$$

即 $P(L/B) = P(LB) / P(B)$ $P(B/L) = P(LB) / P(L)$;

由此可得 $P(LB) = P(L/B) P(B)$ 和 $P(LB) = P(B/L) P(L)$

$P(L)$

所以, $P(L/B) P(B) = P(B/L) P(L)$ 。

这样, 我们可以得到贝叶斯定理:

$$P(B/L) = \frac{P(L/B) P(B)}{P(L)},$$

在上例中, $P(B/L) = \frac{.17 \times .60}{.50} = .20$ 。

二、贝叶斯统计的运用

在心理学实验研究中, 往往会遇到两个相互排斥的假设 H_1 和 H_2 , 在实验之前, 研究者预计, H_1 为真的概率为 .50, 即 H_1 有 50 : 50 的机会为真。然后做了实验, 获得数据 D, 那么 D 是如何影响 H_1 为真的概率呢? 这就可以运用贝叶斯定理:

$$P(H_1/D) = P(H_1) \frac{P(D/H_1)}{P(D)}$$

其中 $P(H_1)$ 是 H_1 为真的先验概率, $P(H_1/D)$ 是 H_1 为真的后验概率, $P(D/H_1) / P(D)$ 为或然比 (Likelihood ratio)。

对于两个假设 (相互排斥而又只有两个假设) 的情况, 可采用贝叶斯定理的另一种形式:

$$P(H_1/D) = \frac{P(D/H_1) P(H_1)}{P(D/H_1) P(H_1) + P(D/H_2) P(H_2)}$$

例如, 某种诊断技术, 可以在某严重疾病的症状出现以前正确地确认 70% 的患此疾病的人, 但有 20% 的可能性作出不正确诊断。假定有 1000 人作了检查, 其中 300 人具有阳性征兆。我们可以用贝叶斯定理的上述形式, 决定被诊断为有病而实际上确有病的可能性。

从例中数据可得:

有该病的概率 $P(A_1) = .30$

没有病的概率 $P(A_2) = .70$

$P(D/A_1) = .70$ $P(D/A_2) = .20$

$$\begin{aligned}\text{所以 } P(A_1/D) &= \frac{.70 \times .30}{.70 \times .30 + .20 \times .70} = \frac{.21}{.21 + .14} \\ &= \frac{.24}{.36} = .60\end{aligned}$$

这就是说，被诊断为有病而实际上有病的概率为 .60。

贝叶斯统计与传统统计学最大的差异，是它正式地把先验意见 (prior opinion) 结合进统计程序。在传统统计分析中，尽管实验者对某些实验结果总有预先设想，但并不把这种设想综合到统计分析中去；而贝叶斯统计则把有关实验结果的设想结合到分析中去。

在实际生活中，我们经常碰到各种不确定的事件，或者需要在不确定条件下作决策，这就需要运用贝叶斯统计分析方法。例如，一位四年级学生转学到一个新的地区去学习，由于两个地区教学水平不同，需要决定是进三年级还是进四年级学习。为此，学校对这位学生进行了一系列测验，各测验项目的难度相等。过去，学校测验的结果是 60% 的四年级学生能正确地回答这些测验项目，而只有 30% 的三年级学生也能回答。这一问题的两个假设就是：

假设 H_1 ：“这位学生适合进三年级”；

假设 H_2 ：“这位学生适合进四年级”。

我们以 R 表示这位学生做对测验项目，以 W 表示做错项目，并假定一系列测验的结果为：

WWRWRWWWRRWWW。

根据这一结果，应如何作出决定呢？这里需要用“先验证据” (prior evidence) 的概念 即在所了解的有关问题的所有情况 X 的条件下， H_1 的证据 $ev(H_1/X)$ 。另外，对于数据进行序贯分析 (sequential analysis)，应运用贝叶斯方程的证据形式：

$$\underset{\text{后验证据}}{\text{ev}}(H_1/DX) = \underset{\text{先验证据}}{\text{ev}}(H_1/X) + 10 \log_{10} \frac{P(D/H_1X)}{P(D/H_2X)}$$

后验证据 先验证据 证据的增加和减少

其中最后一项说明在考虑新的数据 D 时，有利于 H_1 的证据是增加还是减少。

在本例中，学校对 H_1 还是 H_2 并无确定的信息，即

$$P(H_1/X) = P(H_2/X) = .50$$

这就是说，

$$\text{ev}(H_1/X) = 10 \log_{10} \left(\frac{0.5}{0.5} \right) = 0$$

即赞成“这位学生适合进三年级”的假设 H_1 的先验证据为零。

现在，看一看证据随数据变化是增加还是减少。我们有两种数据 R、W，需要计算每种数据对证据的变化：

(1) 如果一个项目的答案正确 (R)，则

$$10 \log_{10} \frac{P(R/H_1X)}{P(R/H_2X)} = 10 \log_{10} \frac{.30}{.60} = -3.01\text{dB}$$

这就是说，这位学生每一次正确地回答一个项目 (R) 对假设 H_1 的证据会减少 3.01dB。

(2) 如果一个项目的答案错误 (W)，那么

$$10 \log_{10} \frac{P(W/H_1X)}{P(W/H_2X)} = 10 \log_{10} \frac{.70}{.40} = +2.43\text{dB}$$

即这位学生每一次答错，对假设 H_2 的证据增加了 2.43dB。

这样，我们可以逐个地考虑“后验证据”：

做完项目 1， $\text{ev}(H_1/WX) = 0 + 2.43 = 2.43\text{dB}$ ；

做完项目 2， $\text{ev}(H_1/WWX) = 2.43 + 2.43 = 4.86\text{dB}$ ；

做完项目 3， $\text{ev}(H_1/RWWX) = 4.86 - 3.01 = 1.85\text{dB}$

以此计算做完 12 个测验项目时的证据情况，然后以概率术语或

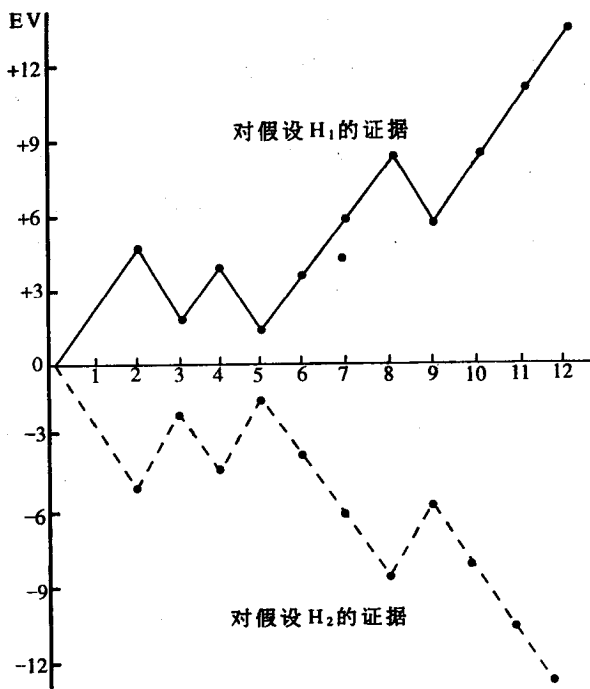


图 11-1 对假设 H_1 和 H_2 证据的序贯分析图

优势比率 (odds ratio) 作出解释。图 11-1 是对两个假设 (H_1 和 H_2) 证据的序贯分析。

从图 11-1 中可以看到 当学生做完 12 个项目以后 对假设 H_1 的证据已经达到 12.84dB 查表可知, H_1 成立的概率已高于 .95 ,即可以在 95% 的置信度上作出结论, 应该把这位学生安排到三年级学习。

第三节 统计决策理论

近年来, 统计分析中的一个新领域发展很快, 这就是统计决策

理论 (statistical decision theory)。这个理论着重于研究和分析不确定情景下的决策问题，其主要课题是决策规则的选择，即如何根据有关证据决定“怎样决策”的问题。这个理论实际上补充了一般统计理论，通过获得样本结果概率的信息，提供了以最佳方式达到决策目标的信息。因此，这个理论对心理学以及经济学、社会学等其它学科的研究和发展具有重要的意义。在这里，我们就统计决策理论的主要概念作一些初步的介绍。

一、决策的损失和效用

决策理论主要是从经济决策问题的研究中发展起来的。在大多数情况下，人们所面临的都是不大确定的决策情景，决策者并不确切地知道事件将会出现什么结果。这时，存在“决策风险”，也存在多种“外界状态”(states of the world)，即统计推论所建立的假设。作为“最佳的决策”，其目标总是要使预期的损失最少。这里涉及两个重要概念：决策的损失和效用。决策的损失是指在作出某项决策时预期的损失 (expected loss)。在决策理论中，人们往往以损失 (负的收益) 来描述各种决策结果。表 11-6 为某决策问题的损失函数 $L(a, \theta)$ 。

表 11-6 某决策问题的损失函数 $L(a, \theta)$

状态 θ 行 动 a	天雨 (θ_1)	天晴 (θ_2)
a_1	-20	10
a_2	5	5
a_3	25	-7

表 11-6 所示 a_1 、 a_2 、 a_3 分别为某个体户在足球赛时的销售决策行动：

a_1 : 只销售雨伞 ;

a_2 : 既售雨伞 , 又售汽水 ;

a_3 : 只出售汽水。

当采用 a_1 时 如果天雨 这位个体户将盈利 20 元 但要是天晴 , 则损失 10 元。

此外 , 假定天气变化的概率分布 $P(\theta)$ 为表 11-7 所示 :

表 11-7 天气变化 θ 的概率分布

状 态	天雨 (θ_1)	天晴 (θ_2)
概率 $P(\theta)$	20	80

在这种情况下 , 最佳决策是什么呢 ?

统计决策理论告诉我们 , 采用行动 a 时 , 预期的损失 $L(a)$ 为 :

$$L(a) = \sum L(a, \theta) P(\theta)$$

所以上例中行动 a_1 、 a_2 、 a_3 的预期损失分别为 :

$$L(a_1) = -20(.20) + 10(.80) = 4$$

$$L(a_2) = 5(.20) + 5(.80) = 5$$

$$L(a_3) = 5(.20) - 7(.80) = -6$$

可见 , 由于 a_3 使预期损失最小 , 所以 a_3 是最佳行动。

从 “ 预期损失 ” 的金额去考虑上述决策问题 , 这在重复决策中都是有效的。但是 , 当某些决策只作一次时 , 以预期损失的金额作为标准就可能犯决策错误。请看下例 :

a_1 : 采取行动 a_1 后 , 肯定能获得 10,000 元 ;

a_2 : 采取行动 a_2 后 , 有 50% 的可能获得 21,000 元。

研究表明 , 大多数人宁愿选择行动 a_1 , 而不是 a_2 , 尽管 a_2 的预期金额高于 a_1 :

$$(10,000 \text{ 元} \times 1.00) < (21,000 \text{ 元} \times .50)$$

这就说明，决策并不是建立在金额本身的基础之上，而是以对钱的主观评价。即钱的“效用”为基础的。决策效用是指对于决策后果的主观评价。研究证明，在许多情况下，效用是比预期损失更适当的决策指标，决策应该建立在“期望效用”的基础上（expected utility）。图 11-2 为上例两种选择的期望效用 $u(M)$ ：

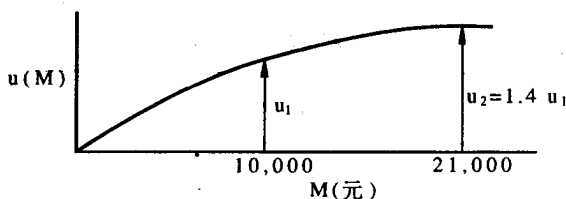


图 11-2 对钱的主观评价

如图 11-2 所示， $u_1(1) = u_1$ ， $u_2(.5) = 1.4u(.5) = .7u_1$ ，因此，显然应该选择 a_1 行动。效用理论提出一条重要的决策规则或称 EU 标准：在决策中使期望效用值（EU）最大。

应该指出的是，影响和决定效用高低的既有经济因素，也有各种非经济因素，例如，声誉、商业道德、顾客满意感和其它社会因素。不同的人会有不同的效用函数，因而会在相同情景中作出不同决策。统计决策论的重要方面之一就是确定不同的效用函数。

二、对策论

对策论是对不确定情景中一些“冲突”，或“竞争”情景的决策分析，旨在寻求最好的对策策略。对策论主要研究两类对策问题：严格规定的对策和混合策略的对策。较复杂的对策问题需要运用线性规划等高等数学方法。

对策论目前在心理学研究中常用的方法是“两难对策法”。这是利用对策论的原理，创设一定的任务情景和条件，研究奖励、动机、沟通和认知加工能力等因素对于作出“竞争”还是“合作”对策的

影响。图 11-3 是一项两难对策研究的图解，表明了 A、B 两方在采

		A	
		X	Y
B	X	+10 +10	+15 -15
	Y	-15 +15	-5 -5

图 11-3 两难对策图式

取合作（X）或竞争（Y）对策（行动）时可能的得益或损失。图 11-3 的图式往往形成竞争行动。例如，A 选竞争（Y）而 B 选合作（X）的话，A 将得 15 分，而 B 将失去 15 分。这类两难对策法在社会心理、管理心理和教育与儿童心理的研究中应用得比较广泛。

许多对策论研究运用了概率决策规则，通过主观概率的考察，把对策论问题看作不确定情景下的决策问题。这时，决策者必须估算其对手所有可能行动的概率分布，并评价行动概率 $P(\theta_1)$ 、 $P(\theta_2)$ 、 $\dots P(\theta_n)$ ，从而选择应采取的对策。

我们看到，传统统计推论建立于样本信息基础之上，贝叶斯推论则进一步包括了其它可取信息，而在决策论中，又考虑了第三种信息输入，即与多种决定有关的损失或代价的信息。因此，贝叶斯统计和决策理论为深入、准确地认识和分析心理活动的动态过程提供了新的思路和方法，在心理学研究中越来越受人重视。

第十一章思考题：

1. 非参数统计有哪些主要方法？在心理学研究中如何选用这些

方法？

2. 一致性系数有何优点？如何运用一致性系数？
3. 什么是贝叶斯统计方法？它和传统统计方法的主要区别是什么？
4. 在心理学研究中如何运用贝叶斯定理？
5. 统计决策理论有何特点？它对心理学研究有何意义？

第十二章 多元分析方法

第一节 多元分析的基本方法

多元分析方法又称作多变量分析 (multivariate analysis)。它是一组用于分析较复杂数据的统计技术，用于研究涉及多个自变量 (IV) 和因变量 (DV) 之间关系的问题。现代心理学的发展，使人们日益认识到，人类行为和心理过程是由多种变量和因素所共同决定的。因此，心理学研究离不开多元分析方法。事实上，多元分析方法已成为心理学、教育学、管理学乃至社会科学其它各学科研究方法论中不可缺少的组成部分。

一、多元分析方法的发展

多元分析方法是由高尔顿首创的。高尔顿把双变量的正态分布方法运用于传统的统计学研究，发现了相关系数和线性回归线，为多元分析奠定了基础。以后，皮尔逊运用多种方法，发展了相关分析方法；费希尔 (Fisher, D.) 提出了方差分析和判别分析法；霍特林 (Hotelling H.) 确定了主成分分析法和典型相关分析法；威尔克斯 (Wilks, S.) 发展了用于方差检验的多元分析方法。多元分析方法发展中特别有意义的事件，是斯皮尔曼根据有关智力的研究所提出的二因素理论。这一理论以及有关的研究，大大促进了因素分析方法的发展和应用，人们因此把斯皮尔曼称为“因素分析之父”。

电子计算机的发展和运用，使各种多元分析方法得到迅速的发

展，许多现代多元分析方法已经与最初的方式大不相同。用于多元分析的统计软件包的开发，使我们在心理学研究中得以广泛地应用多元分析方法，大大提高了研究的水平。其中最常用的软件包有：SPSS (Statistical Package for the Social Sciences) 、SAS (Statistical Analysis System) 和 BMDP (Biomedical Programs) 等。

需要指出的是，多元分析方法的运用，是比单变量分析复杂得多、困难得多的任务。要求我们进行更全面、深入的理论思维，具备更多的数学知识，承担更繁重的数据与资料处理工作。更重要的是，应该更为积极地从事各种心理学实际研究和应用工作，用多元分析方法解决社会生活各领域的实际问题，并且促进心理学科学理论的完善和发展。

二、多元分析的基本概念

在学习多元分析的具体方法以前，我们介绍和讨论一些在多元分析中常常遇到的基本概念。除了我们已经熟悉的自变量与因变量、实验与非实验等概念以外，数据的连续性、变量间的正交性 (orthogonality) 标准分析与层次分析等，也是比较重要的基本概念。

1. 数据的连续性

在运用多元分析方法时，需要考虑数据的数目和所表示的特征或行为之间关系的性质，从而明确地区分连续数据、离散数据和两分数据。

(1) 连续数据

这是用某些量表值平滑变化的量表测量的数据，可以取量表范围内任何值。比如，时间、收入、年龄、温度、距离和某些测验分数，当连续数据的变量呈正态分布时，特别适合于运用于作多元分析。

(2) 离散数据

这类数据常常用称名量表测定，比如，不同类型的领导作风、不同组织层次的职工、按不同的年龄组划分成人和儿童等。在多元分析中，往往把离散变量再分类为一系列两分变量，这一程序称为“亚变量编码”程序。一般来讲，两分变量的分析。比多个离散变量的数据分析灵活得多。

2. 正交性

正交就是指完全没有联系。如果两个变量的向量成直角、相关为零时，它们呈正交。这时，从一个变量无法了解另一个变量的线索。

在进行统计分析时，常常要求有这种正交性。在单变量统计分析时，我们假定因子相互独立，这样，在正交实验设计中可以明确地把某种因果关系归之于各种主效应和交互作用。在多元分析中正是如此，可以通过正交性为假设，并以实际数据加以验证，使得能够以多个预测变量（或自变量）以简单的可加方式预测标准变量（或因变量）。

3. 标准分析与层次分析

在大多数研究中，多数变量并不是正交的，而是相互关联的。例如，行为表现常常与个性特征、动机等许多因素相关。在这种情况下，变量之间就共同分担或者重叠了一部分变异量，多元分析的一项主要任务就是决定如何来处理这种重叠的变异量。这就是选择两种不同的策略：标准分析和层次分析。

（1）标准分析

标准分析是在评估每一变量对于结果的贡献时忽略重叠部分，而只把这部分变异作为对总结果（如 R^2 ）的贡献。图 12-1 为标准分析图示。图中， X_1 、 X_2 分别是预测指标，在考虑 X_1 、 X_2 与 Y 的关系时，忽略带点的重叠部分，而只以斜线部分评估 X_1 、 X_2 对于 Y 的效应。

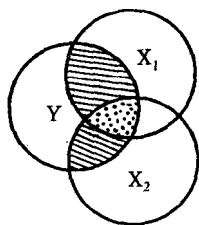


图 12-1 标准分析图解

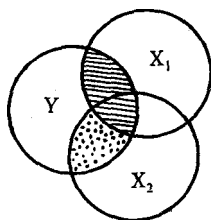


图 12-2 层次分析图解

(2) 层次分析

这种分析不同于标准分析。在层次分析中，研究者把优先顺序分配给各个变量，第一个变量分配到其它变量的任何重叠变异，随后的变量则取其独特变异与任何剩下的重叠变异。以此类推，图 12-2 为层次分析图示。

在图 12-2 中，对变量 X_1 给予了“优先权”。虽然总的关系与上例相同，但变量 X_1 和 X_2 对于变量 Y 的相对贡献已经改变。我们把斜线部分的变异量归于 X_1 ，而只把散点部分的变异量归于 X_2 。

在多元分析中，选用标准分析还是采用层次分析，这对于研究结果很有影响。应该根据不同研究目的和数据特点，选用适当的分析策略。

4. 多元分析技术的分类

多元分析技术或方法主要分成两大类：一类方法用于分析自变量与因变量之间的关系，在多个自变量的基础上预测某一因变量的情况；另一类方法则用来解释一组变量之间的相互关系或者对多个变量进行分类，并不明确地划分自变量与因变量或者作出有关因果关系的假设。属于前一种方法的有多元回归、多元判别函数分析、多变量方差分析和典型相关分析等。第二类方法主要包括因素分析、聚类分析和多维量表分析等。

在多元分析中，具体分析方法的选择，将取决于研究的目的和

数据的特点，尤其是测量量表的性质。因此，在对多元分析技术分类时，还需要考虑量表的性质，我们把名称与等级量表作为非参数量表，而把等距与等比量表作为参数量表。图 12-3 是主要多元分析方法的分类程序，说明了如何在心理学研究中选用适当的多元分析方法。

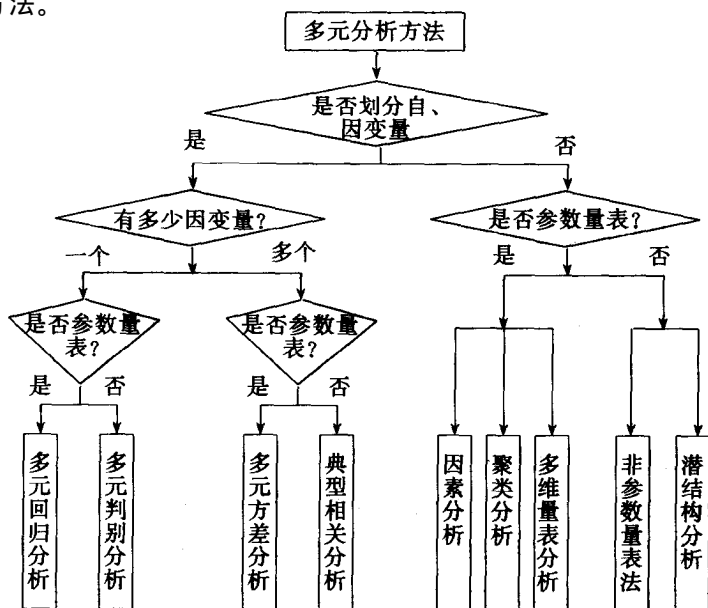


图 12-3 多元分析方法的分类程序

三、多元分析的基本方法

多元分析有许多方法，心理学研究中常用的有以下一些方法。

1. 多元回归 (multiple regression)

多元回归是一种十分有力的统计分析技术，用以评估和分析一个因变量和多个自变量之间的关系，例如，从性别、学前感知运动发展水平等自变量，预测小学生的阅读能力，先决定因变量和自变

量的关系程度，然后确定各个自变量对于这种关系的重要性。多元回归的优点在于，它既可以分析因变量与几个相关自变量的关系，也可以分析自变量相互独立的问题；多元回归还可以用于连续变量或两分变量的统计分析。它的主要用途，除了前面所讲的决定关系及自变量的重要性以外，还可进行自变量预测效力的比较，只根据自变量数据来预测某一样本在因变量上的分数。多元回归的一个特殊用途，是进行因径分析（path analysis）或因果分析（causal analysis），这也是多元回归的最重要应用之一。

多元回归有“标准多元回归”和“层次多元回归”等几种类型。前者同时对所有自变量作出回归方程；后者则逐步评估每一自变量加入方程后的作用。标准多元回归分析用于简单地评价变量之间的关系；层次多元回归则使得研究者控制整个回归的过程，比较适合于建立和检验假设或者理论。我们在本章第三节中详细介绍和讨论多元回归方法。

多元回归分析也存在一些局限性，其中最重要的局限，是因果关系的推论问题。因果关系的证明，需要进行逻辑的、实验的研究并不是统计方法所能解决的。因此，在运用多元回归（包括一般回归分析）方法时，必须对所假定的因果关系提供充分的研究证据，建立严格的理论构思。另一个问题是变量的选择：应该如何测定因变量和自变量？应增加什么自变量才能最大程度地提高预测能力？这些都是在多元回归分析中应该慎重考虑的。

2. 典型相关（canonical correlation）

典型相关是一种很复杂的统计技术，它在心理学及其它学科的研究中都运用得很少。典型相关分析主要用来了解两个或多个因变量（如，效标）与多个自变量（预测指标）之间的关系。我们知道，多元回归分析用于研究单个因变量与多个自变量之间的线性关系。典型相关则是多元回归的某种发展。它把两组变量组成成对的线性

结合，称为典型变数（canonical variate）。第一对典型变数（或变量）使一组线性结合与另一组线性结合之间的相关最大；第二对典型变数则使排除第一对变数的方差以后，变量线性结合之间的相关最大。这两对典型变数假定为相互独立，这样成对地分析，直到剩余相关矩阵中多组变量之间没有显著联系为止。

假如某项研究希望确定若干个性变量（预测指标组）与某些购买行为（效标组）之间的关系。可以作出若干购买行为（采用购买单、使用商店减价票、访问商店等）的线性组合为：

$$Z = a_1x_1 + a_2x_2 \cdots + a_nx_n$$

同时，个性变量的线性组合为：

$$W = b_1y_1 + b_2y_2 \cdots + b_ny_n$$

然后，可以计算确定相应的典型相关系数或加权量，以此作出解释。典型相关分析是运用得最少的一种多元分析方法。原因之一，是典型相关使变量的线性组合之间相关最大，却很难对这种结果作出解释。虽然由于计算机软件包的运用，典型相关分析的应用也有所增加，但是结果解释的问题仍然没有能解决。目前，这种方法的使用主要作为描述性分析方法，并不是良好的假设检验方法。

3. 多元方差分析（MANOVA: multivariate analysis of variance）

多元方差分析适合于多个参数量表式因变量的情况。这时，自变量可以是一个或多个非参数量表式的变量。这种方法比通常的ANOVA更能防止犯I类错误，还能发现ANOVA没有能发现的差异。与此有关的是协方差分析（analysis of covariance）主要用于提高对自变量的检验功效，它可以从差误项中分离出可预测的变异量，起到减少“噪音”的作用；它也可以用来作为统计匹配程序，用于被试不可能随机分配于实验处理的情况。此外，协方差分析可以在多个因变量的情况下，隔离自变量差异的来源。因此，通过协方差

分析，可以检验多个自变量的主效应和交互作用，并且比较单一自变量的多种水平。

多元方差分析的局限是因果关系假设的问题。进行这类分析，应该有明确的理论构思和假设，对变量作认真的定义。同时，这类方法运算程序十分复杂，结果的含义也比较含糊，因此，有人建议，在能够不用这些方法时，就避免使用它们。

4. 判别函数分析 (discriminant function analysis)

判别函数分析 (DISCRIM) 的主要目的，是根据各种预测变量的情况来预测它们的组合效应。例如，应该对预测变量作出何种最佳的组合，才能使组间差异最大？判别分析需要解决三个方面的问题：(1) 决定两组以上分数差异的统计显著性；(2) 对变量进行一定组合，使组间判别程度最大；(3) 决定各组成员的特征。其中第(2)项是判别分析主要解决的问题。图 12-4 表示了判别函数 (Y) 的

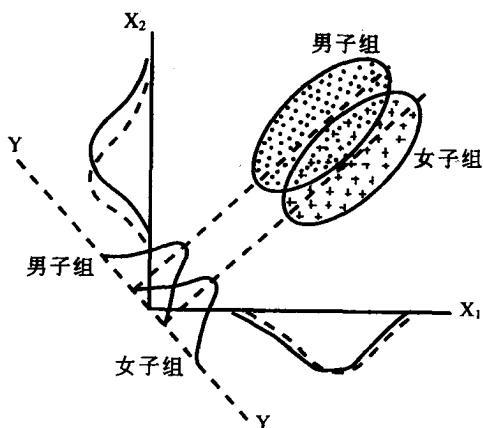


图 12-4 判别函数 (Y) 的判别作用

意义，男、女两组被试在 X_1 和 X_2 两个变量上的分数分布重叠在一起，无法根据 X_1 和 X_2 上的分数来区分两组被试（比如，作出某种

临床诊断)，通过判别分析，可以找出判别函数 Y ：

$$Y = a_1 x_1 + a_2 x_2$$

式中 a_1 、 a_2 为两个自变量的加权量； x_1 、 x_2 为两个预测变量。这样，可以根据函数 Y 来很好地判别和区分两组被试， a_1 、 a_2 这两个加权量是使得下列方差比率最大时而获得的：

$$\frac{Y \text{ 均值之间的方差}}{Y \text{ 的组内方差}}$$

这就是费希尔（1936）所提出的 F 比率，从那时至今，这一规则一直是判别分析的基础。

在大多研究中，都运用两个以上的变量，因而需要采用“多重判别函数”（multiple discriminant function）：

$$Y_1 = a_1 x_1 + a_2 x_2 + \cdots + a_K x_K$$

$$Y_2 = b_1 x_1 + b_2 x_2 + \cdots + b_K x_K$$

$$\vdots \quad \vdots \quad \vdots \quad \vdots$$

$$Y_h = h_1 x_1 + h_2 x_2 + \cdots + h_K x_K$$

推导判别函数的主要运算步骤是进行因素分析。不过，并不是在变量之间相关的基础上进行分析，而是在组内变异平方和的量表上作出分析。

判别分析也分为标准分析和层次分析，前者在所有预测变量的基础上同时求出判别函数；后者使变量以一定顺序进入判别函数方程，从而评价各变量对判别函数的贡献。

由于判别分析主要用于预测自然存在的小组的特征，而不是实验者随机分配的小组，因而没有解决因果关系问题，即解决组间差异的原因。判别结果不容易一般化到其它有关的情境中去。

5. 主成分分析和因素分析（principal components analysis and factor analysis）

这种分析主要应用于没有规定自变量和因变量的多变量问题。它们对整个相关矩阵进行分析，从而发现变量中的子集。因素分析又分成探索型和验证型。探索型因素分析旨在通过变量组合而总结数据，往往在研究的初期进行，为提出假设提供基础；验证型因素分析则用于检验有关潜在结构的假设，常在研究后期运用。主成分分析获得若干“成分”，因素分析则得到“因素”，两者的程序基本相同，只是相关矩阵和变异（方差）的分析不同。在主成分分析中，变量的所有方差都包含在结果中，而在因素分析中，只注重共同变异并设法排除误差变异和独特变异。我们在本章第二节中详细介绍和讨论这些方法。

6. 聚类分析（cluster analysis）

聚类分析的目的与因素分析很相似，它用于把许多变量分成小组或群类，这类的变量之间有高相关而与其它群类的变量相关比较低。作聚类分析时，需要把负相关转换为正相关。然后，找到最高相关的一对变量，组成第一群类，并把其它与这群类（两个变量）有高相关的变量也包括进去。在这以后，找出另外有高相关而与第一群类中的变量相关很低的一对变量，组成第二群类。如此分析下去，直到找不到显著相关的剩余变量。

当群类结构十分明确时，使用聚类分析比因素分析来得简便，但实际研究数据往往不是这种情况。只有当变量数目较少时，才能准确地找出群类，可是，当研究的变量较多时（例如，20个以上）则很难从相关矩阵中辨别群类的位置。

聚类分析也可以分成“非层次法”和“层次法”，（也有人称为“等级法”）上面介绍的分析方法属于层次法。研究表明，各种聚类方法可能得出不同结果，应以一种以上的方法对相同数据作出分析，如果得出相似结果，则聚类结论更加可靠；还可以运用相同测量工具对不同样本进行重复分析，这是与“多特征—多方法”的分析思

路相一致的。通过聚类分析，指出指标或样本中的群类，发现亲疏关系。

上述几种多元分析方法各有特点，需要根据研究目的和要求，决定采用何种方法。多元分析方法有很多优越性，却都不是万灵药，每一种方法都有它特定的假设、条件和数据要求，例如，正态性、线性和同方差性等。在运用多元分析方法时，十分重要的一条，是在研究计划的阶段，确定理论框架和构思，决定收集何种数据、怎样收集和如何分析数据资料。

第二节 因素分析

因素分析方法是多元分析的主要方法之一，至今已有 80 多年历史。因素分析是目前运用最广泛的心理学研究手段，并在经济学、教育学、社会学、生物学、分类学、气象学、医学和考古学等学科广为应用，在理论上和方法上都取得很大的进展。可以说，因素分析是心理学对于其它学科研究方法的最重要贡献。每一个心理学工作者，都应该学习和掌握这种分析方法及其有关的理论。

一、因素分析的发展和应用

因素分析属于统计科学，它是从心理学研究中发展起来的，也是 20 世纪以来在心理测量学中的一个突出成就，对心理学各方面的理论和实践起着重要的作用。

1. 因素分析的发展

人们一般都把因素分析归于斯皮尔曼的创造，称他为“因素分析之父”。这主要由于他早期所进行的有关智力结构的研究，他所提出的二因素论，即一般能力和特殊能力之分，成为因素分析最重要的基本理论之一（1904）。

科学的发展，实际上是在一定社会历史条件下一批人的实践成果。在谈到因素分析的发展历史时，至少还应该提到两个人，这就是高尔顿和皮尔逊。高尔顿在 19 世纪末、20 世纪初提出两条重要的思路，从而奠定了因素研究的基础。第一，他认为“一般智力”是以连续的方式形成从极笨到极聪明的分布，并提出“特殊能力”问题，冲击了当时所流行的认为由离散的“心理官能”（mental faculties）支配行为的观念；第二，高尔顿提出相关的思想，创造性地用定量方法来说明变量之间的相互依存性。皮尔逊（1901）第一个作出因素研究的明确程序，它运用多维空间几何，推导出他的运算公式，提出了重要的“主轴法”，进而为因素分析提供了方法论基础。

到了 30 年代，人们已经比较清楚，斯皮尔曼的二因素论并不总是能准确地描述心理学研究的结果，因此，需要一种新的思路，多因素或群因素的思想便流行起来。首先提出多因素分析概念的是加尼特（Garnett）（1919）；以后，瑟斯顿（Thurstone）总结了这方面的研究。瑟斯顿的主要贡献是，把斯皮尔曼的“四差标准”（tetrad difference criterion）运用于相关矩阵的等级分析，用以决定公共因素的数目，从而把矩阵方法紧密地结合到因素分析之中，大大促进了因素分析方法的发展。瑟斯顿还提出了“简单结构”的概念和方法，使得能够比较客观地对公共因素作出定义，并为计算机数据处理中的因素识别提供了标准。因而，瑟斯顿的理论和方法很快得到广泛运用。50 年代后期，电子计算机开始用于因素分析的运算，成为因素分析发展的重要里程碑。但是这一时期的因素分析是比较盲目的、探索式的，引起的争论很多，这种方式可以很好地运用于分类目的，却并不能促进结构理论。60 年代后期，人们日益关心因素分析中假设检验的方法论。例如，吉尔福特运用普罗克拉斯提斯变换，证明了他的三维智力模式所预测的一些因素。随着计算机应用的深入，出现了验证型因素

分析，从而为心理学理论的发展和研究的精细化、数学化提供了强大的武器，成为因素分析方法发展的最新成就。

2. 因素分析的应用

因素分析的主要用途，是从众多变量的交互相关中找出起决定作用的基本因素；其次，为建立科学理论提供明确的证据。因此，因素分析可以定义为“从大量测量数据（ n 组）中决定基本变量 K 个， $K < n$ ）数目和性质的一种方法”，因素分析正在广泛地应用于心理学各领域的研究，包括教育、儿童、临床、管理、工程、社会和医学心理等各方面。

因素分析应用历史最长、最广泛的领域是智力和能力结构等的研究。例如，斯皮尔曼的二因素论（1904）、瑟斯顿的群因素论（1938）、吉尔福特的智力结构论（1956）、卡特尔（Cattell）的个性研究（1970）和哈曼（Harman）等的认知与气质研究（1975），都是因素分析应用的典范。在心理测验的设计和检验中，因素分析的应用也很多。例如，斯坦福一比纳量表和韦克斯勒智力量表等，都运用了因素分析方法进行设计和修订。

在其他一些应用研究领域，管理心理学中奥尔德弗（Alderfer）的“存在—关系—成长”理论（ERG）、赫兹伯格的双因素论以及有关领导行为的许多研究，也都采用了因素分析方法。此外，沃伊斯（Voiers, W. D.）（1964）的声学通讯研究、格雷格（Gregg, L. W.）（1961）的飞行事故研究、萨克曼（Sackman, H.）等（1964）的人机系统研究等，也都把因素分析的思想和方法应用于各自的研究。在教育心理、临床心理和儿童心理等领域，因素分析的应用更是层出不穷。近年来，我国心理学家也在研究中日益重视因素分析方法，在越来越多的课题中采用因素分析方法，取得了很有意义的研究成果。因素分析的广泛应用，极大地推动了心理学的发展，丰富了人类对于心理过程及其发展规律的认识。

现代科学研究是多学科协同作战的，因素分析方法已经应用于其它各学科的研究。例如，经济学中有关系系统实效的评价（1972）和不确定条件下的投资决策（1962）；医学领域的心血管病研究（1965）、心电图与脑电图分析（1960）。其它一些学科有有关重矿物数据的分析、地质图的研究、气象预报的研究、城市区域结构与经济发展的分析、生物分类研究，考古分析技术的评价，等等。

3. 因素分析时应注意的问题

在讨论因素分析的概念和具体方法以前，应该指出它的某些局限性，以便在看到它的强大生命力的同时，谨慎地运用这种方法。

（1）在因素分析时，不要过分地解释相关系数的意义。在讨论因果关系问题时，必须考虑到“相关不能说明因果关系”的道理，努力从实际活动中找出因果关系的证据。

（2）防止测量差误所造成的分析错误。比如，避免低信度的测量数据，注意相关显著性的检验。

（3）重视样本的选择。从不同总体选取的混合样本，往往会使特定于某一总体的因素含糊不清。

（4）数据分布应近于正态分布，相关应是线性的。过于偏态或峰态的数据不能得出有意义的基本因素，曲线关系也会使结果十分不可靠。

二、因素分析的基本概念

1. 因素和因素荷重

因素是一种构思，是存在于大量变量之中的某种共同变异。因素是由因素荷重作出操作定义的，因素荷重是变量与因素之间的相关系数，或者说是单位矢量在 x 轴上的投影，即 $\cos\theta$ （图 12-5）。‘

因素分析的基本模型为：

$$Z_{ji} = a_{j1}F_{1i} + a_{j2}F_{2i} + \dots + a_{jm}F_{mi} + d_{ji}U_{ji}$$

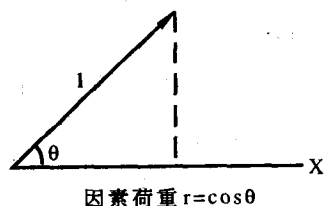


图 12-5 因素荷重

式中, Z_{ji} : 个体 i 在变量 j 上的标准分数;

F_{1i} : 个体 i 在第一个共同因素上的标准分数;

F_{2i} : 个体 i 在第二个共同因素上的标准分数;

F_{mi} : 个体 i 在第 m 个共同因素上的标准分数;

U_{ji} : 个体 i 在独特因素上的标准分数。

因素分析的这一模型说明, 个体 (i) 在变量 (j) 上的分数, 可以构想为少量的、推导得到的共同因素上分数的加权和。其中 a_{j1} 、 a_{j2} ... a_{jm} 为因素荷重。

与因素荷重有关的一个重要概念, 叫做公共因素方差 (communality), 也称“共同性”表示为 h^2 , 公共因素方差是所有共同因素荷重的平方和。例如, 某变量矢量在 A、B 两个共同因素上的荷重分别为 .60 和 .40, 则这个变量的公共因素方差为 $h^2 = .60^2 + .40^2 = .52$ 。如图 12-6。

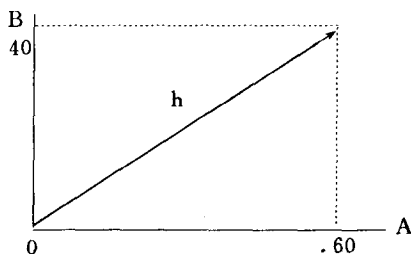


图 12-6 公共因素方差 (h^2)

公共因素方差的值一开始并不知道，必须从数据进行估算，把估算值插入相关矩阵的对角线部分。估算的方法很多，常用的有：

- (1) 取相关矩阵每一行或列中最大的相关值；
- (2) 复相关的平方，即每一变量与 $n-1$ 个剩余变量的复相关；
- (3) 利用重测信度 r_{jj} 实际上 r_{jj} 是公共因素方差的上限估计值， R_j 是下限值，即 $R_j \leq h^2 \leq r_{jj}$ ；

(4) 采用迭代方法，在插入 h^2 估计值后，进行因素分析，再从共同因素荷重的平方和作为 h^2 的新估计值，这一程序重复到两次估算值十分接近为止。

2. 变异 (variance)

变异的概念是因素分析的重要基础，它在统计学上是分数离散性的指标的平方值（即标准差）。因此，因素荷重的平方表示了变量在共同因素上所说明的变异量。一项测定的总变异量可以表示为 $1 = h^2 + b^2 + e^2$ 其中 b^2 为特殊变异量， e^2 为误差变异量。我们所说的信度就是 $r_{tt} = h^2 + b^2 = 1 - e^2$ 、误差 $e^2 = 1 - r_{tt}$ 。因素分析的基本目的之一，是通过第一个共同因素，抽取出尽可能多的公共变异量，然后在第二个因素中抽取剩余变异的最大量共同变异，直至没有什么共同变异。

当我们测量 n 个变量时，可以用变异量表示因素分析的数学模型：

$$Z_1 = P_{11}F_1 + P_{12}F_2 + \cdots + P_{1m}F_m$$

$$Z_2 = P_{21}F_1 + P_{22}F_2 + \cdots + P_{2m}F_m$$

$$\vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots$$

$$Z_n = P_{n1}F_1 + P_{n2}F_2 + \cdots + P_{nm}F_m$$

式中 P_{ij} 等为因素荷重，也是公共因素变异。

两个变量之间的相关 r_{ij} 为：

$$r_{ij} = h_i h_j \cos \theta_{ij}$$

式中 i, j : 两个变量 (如, 测验) ;

h_i, h_j : 两个变量 i, j 矢量的长度 ;

θ_{ij} : i, j 两矢量的夹角。

我们已经知道 h 也是公共因素方差 (共同性) 的平方根, 因此, 以 $h_i h_j \cos \theta_{ij}$ 表示两个变量的相关, 在因素分析中可以起重要作用。

3. 因素的数目

这是指在因素分析时应抽取多少因素为宜, 这既与理论考虑有关, 也与数据特点有关。人们在实际应用中根据不同要求提出不同的决定标准, 常用的有以下几种:

(1) 凯泽的标准

这个标准是, 当因素的潜根 (latent root) 或 “特征值” (eigenvalue) 小于 1 时 停止抽取因素。潜根是每一因素的总荷重的平方和, 它是矩阵代数的表达方式, 其值从第一因素向后减少。表 12-1 为常见的因素矩阵的形式。凯泽 (Kaiser, H. F.) 的标准特别适合于主成分设计, 并且在变量数为 20—50 之间时最可靠。

表 12-1 因素矩阵 A

变 量	因 素			h^2
	I	II	III	
1	r_{11}	r_{12}	r_{13}	h_1^2
2	r_{21}	r_{22}	r_{23}	h_2^2
3	r_{31}	r_{32}	r_{33}	h_3^2
4	r_{41}	r_{42}	r_{43}	h_4^2
5	r_{51}	r_{52}	r_{53}	h_5^2
特 征 值	$\sum r_{i1}^2$	$\sum r_{i2}^2$	$\sum r_{i3}^2$	$\sum h^2$

(2) 卡特尔的陡阶检验法:

卡特尔 (Cattell, R. B.) 检验法用以弥补凯泽标准的缺点, 办法是将每一因素所能解释的变异量作成图, 并连接成曲线, 以曲线

呈台阶（不平滑处）以后的那个因素为基准，决定应抽取的因素数目。图 12-7 为 25 个因素的“陡阶检验”的图解。可以看到，曲线在第 8 个因素处出现台阶，因此，以第 9 个因素为基准，取 9 个因素。这种方法在这一例子中比凯泽标准多取 4 个因素。这种方法的解释是，陡阶之前曲线陡降，所包含的因素可以解释相当大比例的变异量，而陡阶以后曲线趋于平坦，仅能解释极少部分变异量，可将这些因素舍弃。

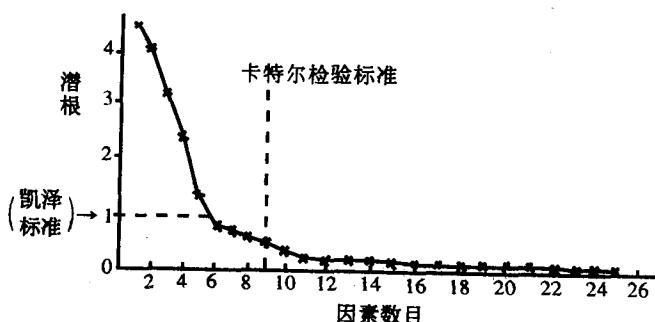


图 12-7 对 25 个因素的“陡阶检验”图

(3) 其他标准

除了上面的两种标准，还有一些别的标准，也为研究人员所采用。例如，有人认为凯泽标准缺乏统计学基础、太简单，因此提出以变量数目与信度平方值的乘积作为选取变异量的标准，认为超过这一数值就是差误变异量了；还有人提出以残余相关量的比例作为标准；与此相似的一种标准是当因素荷重低于 ± 0.20 时，停止抽取因素。

4. 因素分析的基本类型

因素分析可以分成三种基本类型：R 型、Q 型和 P 型

(1) R 型因素分析

R 型因素分析是在所有被试的基础上计算成对变量的相关，着

重于研究变量之间的相关关系。

(2) Q 型因素分析

这种分析是在所有变量的基础上计算被试之间的相关，对被试样本作出分类和解释。

(3) P 型因素分析

P 型因素分析主要分析单一个体，计算个体在多次测定后变量之间的相关，找出个体所特有的基本因素。

目前最常用的是 R 技术，Q、P 技术用得还不多。

三、因素参照轴的旋转和简单结构

在因素分析中需要对因素的参照轴作出旋转，进行这种旋转主要有两条原因：(1) 因素参照轴是因素在向量空间的参考框架，作出转轴，可以使结果更容易解释；(2) 转轴后得到的因素结构的可靠性更高，可以重复，便于验证。

1. 转轴

转轴是指因素坐标轴绕其原点旋转至一定的位置，使得因素结构更为明确。因素分析有两类方法：第一类称为“直接法”，即运用数学模型从相关矩阵直接得到因素矩阵；第二类是在直接法基础上进行参照轴的旋转，取得新的因素矩阵，称为“推导法”。在心理学的大多数研究中，因素结构都比较复杂，需要对直接法取得的因素矩阵进行转轴，才能对基本因素作出明确的有意义的解释，因此，推导法应用得更加广泛。

因素分析中采用两种转轴方法。一种方法是“直转法”，这种方法假定基本因素之间相互独立，从而作出正交旋转（坐标轴之间呈 90° ）；另一种方法称为“斜转法”，这种方法允许因素之间有相关。两种方法相比，前者比较简便，因为它不必考虑因素之间的关系。图 12-8 为某直转法的图解，参照轴由 F_1 、 F_2 按顺时针方向转动了 30° ，

得到 F'_1 、 F'_2 的位置，可以用“分析法”求出新的因素荷重，旋转后因素的共同性不变。

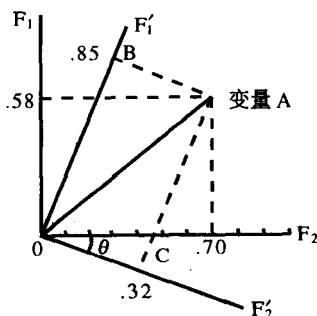


图 12-8 直转式旋转参照轴

转轴以后，变量 A 的新的因素荷重由以下公式计算：

$$OB = F_1 \cos \theta + F_2 \sin \theta = .85$$

$$OC = F_2 \cos \theta - F_1 \sin \theta = .32。$$

2. 简单结构

瑟斯顿在 30 年代提出因素的“简单结构”理论，并首先进行了参照轴旋转的尝试。简单结构的原则使得因素分析有可能达到结果的“不变性”。瑟斯顿的名著《群因素分析》中，把简单结构定义为一种参考构架，其中的每一测验变量包含于一个或多个座标超平面 (hyperplane)。他还提出形成简单结构的五条标准：

- (1) 因素矩阵的每一行至少包含一个零荷重；
- (2) 如果有 n 个共同因素，则每一因素中至少有 n 个零荷重；
- (3) 每对因素，应有几个变量在一个因素中荷重值为零，或同时在另一因素中有显著荷重；
- (4) 对于每对因素，有大部分变量在两因素中的荷重近于零；
- (5) 每对因素应有少量变量在两因素中都有显著荷重。

瑟斯顿的上述简单结构标准在实际应用中被研究者广泛接受，

并已编入各种统计软件包，成为确定因素分析结果和转轴的标准。当然，实际研究所涉及的各种变量之间的关系十分复杂，加上测量误差所造成的影响，使因素分析的结果往往只能基本上满足简单结构的要求。

3. 因素荷重与原变量之间相关的关系

因素分析从变量之间的相关系数出发，求出因素荷重，那么，因素荷重与原变量间的相关之间的关系如何呢？可以用图 12-9 来表示。

我们已经知道，变量间的相关可以表示为它们夹角的余弦，即 $r_{12} = \cos\alpha$

根据数学推导，我们可以证明，变量 1、2 之间的相关系数，可以用这两个变量在基本因素参照轴上的因素荷重来表示。

$$r_{12} = \cos\alpha = x_1x_2 + y_1y_2$$

这个公式适用于基本因素之间不相关的情况。当存在多个基本因素时，公式为：

$$r_{jk} = a_{j1}a_{k1} + a_{j2}a_{k2} + \dots + a_{jm}a_{kn}$$

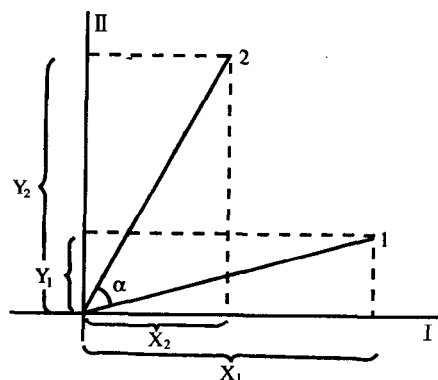


图 12-9 两个变量在基本因素上的荷重

式中 $a_{jm}a_{km}$ 分别为 j 、 k 两个变量在 m 个基本因素上的荷重。在实际计算时，由于取样误差及其它误差，从因素荷重复制的变量相关系数会略不同于原相关，它们之间的差异称为“残余相关”。

四、因素分析的主要方法

因素分析的方法很多，心理学研究的统计软件包常常包括多种可选用的方法，例如，SAS 软件包中的因素分析方法就有这样几种：主因素分析、映象分析、Alpha 因素分析、迭代主轴因素分析和最大似然因素分析等，其中主因素分析 (principal factor analysis) 使用得最广泛，它可以运用每一变量与其它所有变量的复相关平方 (SMC) 作为公共因素方差的估计值。近年来，人们更多地倾向于采用最大似然因素分析 (maximum likelihood factor analysis) 的方法，这个方法以迭代方式估算公共因素方差值，分析精度更高，但花的时间要多得多。因此，人们主张以主因素分析确定基本因素数目，然后以最大似然因素分析求出因素荷重的准确值。此外，参照轴旋转的方法也很多，SAS 软件包中就包含了四种旋轴方法。

在心理学实验研究中，也可以不用计算机而通过手工运算来进行因素分析，这就是传统因素分析方法中目前仍常常采用的重心法。当没有条件用计算机软件包进行因素分析时，这种方法可以给我们一幅因素结构的粗略图解，并作出有意义的心理学解释。

1. 因素分析的基本步骤

因素分析包括一些基本步骤，从数据的采集到因素的解释和总结，每一步都需要我们认真细致地进行，以便排除各种可能造成误估的因素。

(1) 数据的采集：因素分析一般要求连续的、近于正态的数据资料。应从同一总体抽样，运用等距或等比量表测定，获得原始数据；应力求数据测量的高效度，防止取样误差和测量误差。

(2) 求出相关矩阵：因素分析的第二步是求出相关矩阵，并在相关系数的基础上进行因素分析，一般应计算皮尔逊相关系数。

(3) 因素运算：求出基本因素矩阵，根据变量在各因素上的因素荷重，算出各因素的特征值，决定因素数目。

(4) 参照轴旋转：对取得的因素矩阵进行转轴，可以用几何方法也可以用分析法，并按照简单结构的原则确定最佳旋转角度。

(5) 因素的解释和命名：对转轴后的因素结构作出解释。因素的解释应以研究的理论构思和实际因素荷重为基础，从最大荷重的变量得出因素的主要含义。因素的解释从高荷重变量往低荷重变量进行，对于同时包含正、负荷重的因素（称为双极因素），应先作出荷重符号的整理。

因素分析是与有关变量基本结构的理论构思分不开的，因此，在研究设计中就必须十分慎重地作出有关研究变量的基本理论构思。

2. 重心法因素分析

重心法(centroid method)是以多个变量(或测验)的矢量的重心作为基本因素参照轴位置的因素分析法。用这种方法抽取出的因素又叫做重心因素(centroid factor)。

下面以某研究得到的相关矩阵为例，说明重心法的基本步骤和方法。

表 12-2 为某研究的相关矩阵，我们以每一列相关系数中最大的系数作为公共因素方差的估计值。

表 12-2 相关矩阵

变 量	1	2	3	4	5	6
1	(.63)	.63	.00	.00	.27	.00
2	.63	(.63)	.32	.36	.21	.00
3	.00	.32	(.72)	.72	.27	.24
4	.00	.36	.72	(.72)	.00	.00
5	.27	.21	.27	.00	(.72)	.72
6	.00	.00	.24	.00	.72	(.72)

(1) 抽取第一个重心因素

求出各行各列相关系数之和（包括对角线中的公共因素方差估计值），并用以下公式求出第一个重心因素的荷重：

$$a_1 = \frac{E_j}{\sqrt{T}} = mE_j$$

式中 a_1 ：第一因素荷重；

E_j ：各列相关和；

T ：所有相关系数之和， $m = \frac{1}{\sqrt{T}}$ 。

可以得到下列结果（表 12-3）。

表 12-3 抽取第一重心因素

变 量	1	2	3	4	5	6	
1	.63	.63	.00	.00	.27	.00	1.53
2	.63	.63	.32	.36	.21	.00	2.15
3	.00	.32	.72	.72	.27	.24	2.27
4	.00	.36	.72	.72	.00	.00	1.80
5	.27	.21	.27	.00	.72	.72	2.19
6	.00	.00	.24	.00	.72	.72	1.68
E_j	1.53	2.15	2.27	1.80	2.19	1.68	11.62 = T
$a_1(mE_j)$	.449	.631	.666	.528	.642	.493	$1/\sqrt{T} = .293$ = m

(2) 计算第一因素残余

在抽取了第一因素以后，需要计算相关矩阵的相关残余量，以便进行第二个因素的抽取运算。计算残余量 (ρ_{ij}) 的公式为：

$$\rho_{ij} = r_{ij} + (a_i)(-a_j)$$

式中 r_{ij} 为原相关， a_i 、 a_j 分别为 i 、 j 两个变量的第一因素荷重。

这样，表 12-3 例中变量 1、2 的相关残余为：

$\rho_{12} = .63 + (.45)(-.63) = .35$ 。以此类推，可以得到新的残

余相关矩阵，准备作第二因素的抽取运算。第二因素的抽取比较复杂，可按表 12-4 的格式进行，先求出第一因素荷重之和 $\Sigma a = 3.41 = V$ ，并求出原相关各列之和 $\Sigma r = S_1$ ；然后，以 $(S_1 - Va_1 + a_1^2 = S_2)$ 和 $(-S_2/2 = A)$ ，求出 A 值；对 A 以后每行最大的正相关系数作出反射，直至所有行值中括号外的均为负值或者括号内的均为正值。“反射”是指使每一变量矢量保持相同长度但反转到相反方向。最后，计算出第二因素荷重 a_2 ，表 12-4 中分别为 .156、.444、.230、.444、-.598 和 -.665，通常我们只取小数点以后两位。在此基础上，我们还可以继续求出第二因素的相关残余量，准备第二个基本因素的分析，直至满足所需要的因素数目为止。

表 12-4 中， S_1 为第一因素分析时各列相关系数之和（排除对角线的系数）； $\Sigma |E| = T = 6.430$ ， $m = 1/\sqrt{T} = .3944$ 。

采用同样方式，可以求得第三个因素的因素荷重。表 12-5 为重心因素矩阵，我们需要求出特征值或潜根，并算出各因素对于公共因素方差的贡献量，表中三个因素的贡献量分别为 45.2%、29.1% 和 25.7% $\left(\frac{\Sigma a^2}{\Sigma h^2} \times 100\% \right)$ 。

表 12-4 第一因素残余和第二因素的抽取

第一因素荷重 a_1	.449	.631	.666	.528	.642	.493
变量	1	2	3	4	5	6
1		.346	-.299	-.237	-.018	-.221
2	.346		-.100	.027	-.195	-.311
3	-.299	-.100		.368	-.158	-.088
4	-.237	.027	.368		-.339	-.260
5	-.018	-.195	-.158	-.339		.403
6	-.221	-.311	-.088	-.260	.403	

$V = 3.409$

续表

S_1	.900	1.520	1.550	1.080	1.470	.960
$S_2 = S_1 - Va_1 + a_1^2$	-.429	-.233	-.276	-.441	-.307	-.478
$-S_2/2 = A$	.2145	.1165	.1385	.2205	.1535	(.2385)
+第6行	-.0065	+.1945	.0505	-.0395	.5565	(.2385)
B: +第5行	-.0245	-.3895	-.1075	-.3785	(.5565)	(.6415)
$-2B = C$	.049	.779	.215	.757	-1.113	-1.283
D	.346	.346	.368	.368	-.403	-.403
$C + D = E$	.395	1.125	.583	1.125	-1.516	-1.686
$mE = a_2$	.156	.444	.230	.444	-.598	-.665

表 12-5 重心因素矩阵

变 量	I	I	II	h^2	h^2 估计值
1	.449	.156	-.617	.607	.63
2	.631	.444	-.432	.782	.63
3	.666	.230	.522	.769	.72
4	.528	.444	.498	.719	.72
5	.642	-.598	-.132	.787	.72
6	.493	-.665	.147	.707	.72
Σa^2	1.978	1.271	1.122	4.371	

对于以上因素矩阵，还需要作出参照旋转，然后作出因素的解释，读者可以运用“分析法”试作转轴，并以简单结构原则决定参照轴旋转的最佳位置，手工转轴有时要进行多次，取得近似的简单结构。在转轴时，还可以通过作图，找出近似的旋转角度，然后计算新荷重。

五、验证型因素分析——LISREL

因素分析最新的发展是进行验证型因素分析。验证型因素分析是基于以下的考虑而提出的，即社会与行为研究中的许多变量都是

不可观察的或者是潜变量（例如，焦虑、动机、智力、态度等），因此，不应该以单一的指标来表示特定的构思，而需要用多重指标，才能保证有效、可靠地说明变量的特征。通过“结构方程模型”（structural equation model），可以运用多重指标来表达潜在的变量。70年代末，Jöreskog 等研制了称为 LISREL 的专用软件，使得能够用计算机实现结构关系的运算。目前，验证型因素分析已经在心理学各个领域成为最强有力的研究手段之一（参阅本书第二章），它既是心理学理论发展方面的重要工具，又促使研究者细致、认真地考虑研究的理论构思与变量结构，使研究更为严密并富有理论与实际意义。

LISREL 是以最大似然（ML）统计理论为基础的，而不是象其他许多传统统计方法那样以最小乘方统计理论为依据。LISREL 包括结构模型和测量模型，结构模型表示了外源变量和内源变量之间的关系，这些变量大多数是潜变量。这种模型表示为：

$$B_{\eta} = \Gamma \xi + \zeta$$

式中 η 为 $m \times 1$ 的内源潜变量；

ξ 为 $n \times 1$ 的外源潜变量； ζ 为残余或误差；

B 为内源变量对内源变量（ η ）效应系数的 $m \times m$ 矩阵；

Γ 为外源变量（ ξ ）对内源变量（ η ）效应的 $m \times n$ 矩阵。

测量模型用来说明潜变量与测量变量之间的关系，并用两个方程表示：

$$y = \Lambda_y \eta + \epsilon$$

$$x = \Lambda_x \xi + \delta$$

其中 y 为因变量测量， x 为自变量测量， Λ 为它们的因素荷重， ϵ 、 δ 为 y 、 x 的测量误差。测量模型可以用来检验测量的信度和效度。

图 12-10 为一个简化的因果关系结构模型，表示了以上所介绍的变量之间的关系。验证型因素分析在运算以后，需要用 χ^2 检验来检查实际数据与理论模型之间的配合度（goodness of fit）。

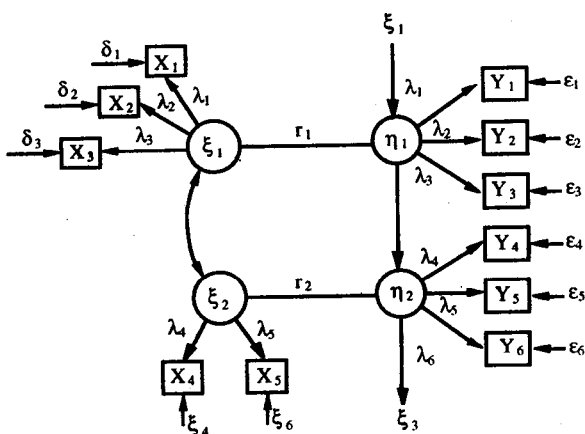


图 12-10 简化的因果关系模型

验证型因素分析比探索型复杂得多，但对于心理学研究具有十分深远的影响，它的重要前提条件是符合实际的理论构思和严格的测量数据。实际应用表明，在这两个条件不满足或者比较差时，模型的数学逼近将会中断或者得到配合度极低的因素结构。研究还表明，验证型因素分析特别适合于纵向研究的数据。我们曾于 1985 年在有关计算机应用的界面特征和决策策略的研究中运用验证型因素分析方法（1984 年第五版 LISREL 软件），以现场研究数据为基础，分别建立了“人与计算机界面特征的 LISREL 结构模型”和“界面特征高价因素的 LISREL 结构模型”，深入研究了人与计算机界面各特征之间的系统结构和较深层的心理过程，并据此提出了“人与计算机界面层次理论”（1986）。实际研究表明，验证型因素分析是一种更严密的验证因果关系特征的方法，它所建立的心理模型具有“参数不变性”的优点，并且较好地解决了测量误差、公共因素方差估算和参照轴旋转不变性等问题。

第三节 多元回归分析

我们在本章第一节中介绍了多元回归的基本特征和局限，在这一节里，将具体介绍和讨论多元回归的方法。

表 12-6 阅读成绩 (Y)，言语能力 (X_1) 和成就动机 (X_2) 的多元回归数据

Y	X_1	X_2	Y'	$Y - Y' = e$
2	2	4	3.031	-1.031
1	2	4	3.031	-2.031
1	1	4	2.353	-1.353
1	1	3	1.960	-.960
5	3	6	4.494	.506
4	4	6	5.172	-1.172
7	5	3	4.668	2.332
6	5	4	5.062	.938
7	7	3	6.023	.977
8	6	3	5.346	2.655
3	4	5	4.778	-1.778
3	3	5	4.101	-1.101
6	6	9	7.706	-1.706
6	6	8	7.313	-1.313
10	8	6	7.880	2.120
9	9	7	8.950	.050
6	10	5	8.841	-2.841
6	9	5	8.164	-2.164
9	4	7	5.565	3.435
10	4	7	5.565	4.435
Σ : 110	99	104	0	
M: 5.50	4.95	5.02		
SS: 770	625	600	$\Sigma e^2 = 81.609$	

SS: 原分数的平方和; e: 回归残余量

一、基本统计计算

假定我们进行了一项教育心理学研究，测得了 20 位学生的阅读成绩、言语能力和成就动机的分数，需要计算阅读成绩（Y）对言语能力（ X_1 ）和成就动机（ X_2 ）的回归，研究数据为表 12-6 所示。

在表 12-6 的数据结果的基础上，可以分别求出：

$$\Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{N} = 165.00;$$

$$\Sigma x_1^2 = \Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N} = 134.95;$$

$$\Sigma x_2^2 = \Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N} = 59.20;$$

$$\Sigma x_1 y = \Sigma X_1 Y - \frac{(\Sigma X_1)(\Sigma Y)}{N} = 100.50;$$

$$\Sigma x_2 y = 39.00; \quad \Sigma x_1 x_2 = 23.20;$$

$$S_y = \frac{\Sigma y^2}{N-1} = \frac{165}{20-1} = \sqrt{8.6842} = 2.9469;$$

$$S_{x_1} = \sqrt{\frac{\Sigma x_1^2}{N-1}} = 2.6651; \quad S_{x_2} = 1.7652。$$

此外，三个变量之间的相关分别为： $r_{x_1 y} = .6735$ 、 $r_{x_2 y} = .3946$ 、 $r_{x_1 x_2} = .2596$ 。

二、多元回归运算与分析

上一数据例子中，我们准备计算多元回归方程： $Y' = a + b_1 x_1 + b_2 x_2$ 以便从个体的 x 值预测 Y' 值，即根据个体的言语能力和成就动机分数来预测阅读成绩分数。多元回归运算和分析应包括以下步骤：

1. 我们需要求出回归方程中的常数（ a 、 b_1 和 b_2 ）从而可以从

X_1 、 X_2 的值插入方程获得 Y' 值。

2. 求出所“说明”的变异量，即复相关系数 $R^2_{Y.12}$ ，以了解 Y 的总变异中有多少份量是由于 Y 对 X 言语能力和成就动机的回归。

3. 对回归结果进行统计显著性的检验，考察 Y 对 X 的回归系数是否具有显著性。

4. 需要决定不同的自变量 (X_1 、 X_2) 在解释 Y 时的相对重要性，这一点对于心理学研究具有很大的意义，也是回归分析中比较困难的一步。

1. 回归系数的计算

回归方程中 b 的计算可运用以下公式（对两个自变量的情况）：

$$b_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2};$$

$$b_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}。$$

在上例中，可求得 b_1 、 b_2 为：

$$b_1 = \frac{(59.20)(100.50) - (23.20)(39.00)}{(134.95)(59.20) - (23.20)^2} = .6771;$$

$$b_2 = \frac{(134.95)(29) - (23.20)(100.50)}{(134.95)(59.20) - (23.20)^2} = 3.934。$$

a 的公式为： $a = Y - b_1 x_1 - b_2 x_2$

$$\begin{aligned} \text{因此, } a &= 5.50 - (.6771)(4.95) - (.3934)(5.20) \\ &= .1027 \end{aligned}$$

所以，得出多元回归方程：

$$Y' = .1027 + .6771x_1 + .3934x_2$$

在这一回归方程中， b_1 和 b_2 称为“非标准回归系数”。目前在心理学研究中更多采用的是标准回归系数 (β)。因此回归方程可以表示为标准分数的形式：

$$z' y = \beta_1 z_1 + \beta_2 z_2$$

在两个自变量时， β 的计算，可用以下公式：

$$\beta_1 = \frac{ry_1 - ry_2 r_{12}}{1 - r_{12}^2}, \quad \beta_2 = \frac{ry_2 - ry_1 r_{12}}{1 - r_{12}^2}$$

从上例数据可以求得：

$$\beta_1 = .6123, \quad \beta_2 = .2356$$

因此，标准回归方程为：

$$x'y = .6123z_1 + .2356z_2$$

2. 复相关系数

关于复相关系数的运算，我们已在第十章中作了介绍和讨论，可用公式：

$$R_{y \cdot 12}^2 = \frac{ry_1^2 + ry_2^2 - 2ry_1ry_2r_{12}}{1 - r_{12}^2}$$

从表 12-6 数据可得：

$$R_{y \cdot 12}^2 = \frac{.6735^2 + .3946^2 - 2(.6735)(.3946)(.2596)}{1 - .2596^2}$$

$$= .5054$$

3. 显著性检验

多元回归的显著性检验主要包括三个方面：(1) 对 R^2 的检验；
(2) 对回归系数的检验；(3) 对变量所说明变异量的检验。

R^2 的检验采取 F 检验的方法：

$$F = \frac{R^2/K}{(1 - R^2) / (N - K - 1)}$$

式中 K 为自变量的数目， N 为样本大小， K 和 $N - K - 1$ 为自由度 (df)。

我们已经知道表 12-6 的数据的 $R_{y \cdot 12}^2 = .5054$ ， $N = 20$ ，可求得：

$$F = \frac{.5054/2}{(1 - .5054) / (20 - 2 - 1)} = 8.684$$

因此，可查得这一 F 值已达到 $P < .01$ 的显著性水平。

也可以用相同公式检验回归系数的显著性，但这里， R^2 为 $R_{y1}^2 = \frac{SS}{\sum y^2}$ ，（式中 $SS=b\sum xy$ ），可以用表 12-6 的数据求得， $F(1, 18) = 14.92$ ， $P < 0.1$ 。

4. 变量的相对重要性

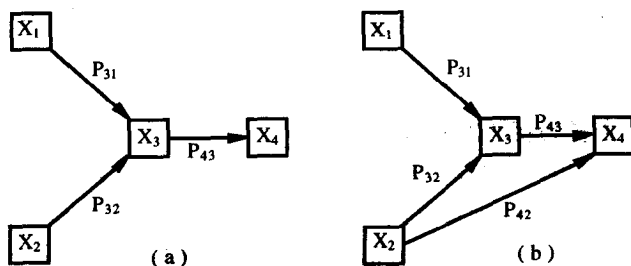
变量的相对重要性，涉及到对多元回归结果的解释，因此是比较复杂的任务。这既要考虑回归系数的相对大小，又要考虑其统计显著性，当统计显著性相似时，系数大小的影响比较小一些，更重要的则是考虑研究变量的含义和研究的理论构思，并重视多变量所说明变异量的比重。

多元回归分析的目的在于解释和预测对象。当以预测为主时，我们着重于选择最佳的预测指标，以获得更好的 R^2 ；当解释作为主要目的时，重点就放在对回归方程作出解释，特别是回归系数的大小。多元回归分析的运用，使心理学研究提高到一个新的水平，它与因素分析一样，已经成为心理学研究和分析的最重要工具之一。

三、因径分析

因径分析 (path analysis) 是多元回归分析的一种形式，它运用“因径图”帮助构建研究思路或者检验复杂的假设。通过因径分析，可以估算出自变量对于因变量的直接或间接的效应，并以因径系数表示。因径系数实际上是标准化了的回归系数 (β)。

因径图以变量之间的结构关系作为框架，表示出不同层次的因果效应。图 12-11 中有两个模型，都表示出社会经济地位 (x_1)、智力 (x_2) 对于成绩 (x_4) 的影响。在 a 模型中，社会经济地位和智力都直接影响人的成就动机或需要 (x_3)，而 x_3 又影响成绩；b 模型则在 a 模型基础上增加了智力 (x_2) 对于成绩的直接效应。究竟哪一模型更符合实际情况呢？可以用因径分析来进行检验。



x_1 : 社会经济地位; x_2 : 智力; x_3 : 成就需要; x_4 : 成绩

图 12-11 两个因径模型

因径系数的计算并不复杂。可以先计算出 x_3 在 x_1 、 x_2 上的回归系数, 求出 β 权重, 然后计算 x_4 在 x_3 上的系数以及 x_4 对 x_2 的回归系数, 看 x_2 与 x_3 、 x_4 的关系。佩德沙 (Pedhazur, E.) 的重要著作《行为研究的多元回归分析》(1982 年第二版) 详细介绍和讨论了因径分析的理论依据和具体方法。我们曾 (1981, 1982) 就不同工作责任制条件下职工的归因及其作用, 进行了因径分析, 发现了不同归因倾向对于随后情感、期望和行为意向及工作绩效的积极效应或影响, 取得了较好研究成果。在运用多元回归分析和因径分析时都应注意, 它们都需要较大的样本, 并且, 对于测量信度和测量程序等都有更严格的要求, 主要原因是, 这些统计分析容易受到测量误差的影响。

第十二章思考题:

1. 为什么说全面、深入的理论思维在心理学研究中是运用多元分析的前提条件?
2. 多元分析有哪些主要方法? 它们各有何特点?

3. 因素分析的发展中有哪些重要研究和理论？
因素分析有哪些应用？它对心理学研究有何意义？
4. 在因素分析中应注意哪些问题？为什么？
5. 因素数目的确定有哪几种方法？如何进行因素参照轴的旋转？
6. “简单结构”的标准有哪些？因素分析包括哪些基本步骤？
7. 如何运用重心法进行因素分析？验证型因素分析有何特点？它对心理学研究的意义是什么？
8. 多元回归主要有哪几个步骤？这些步骤各有何特点？

第十三章 计算机统计分析和研究报告的撰写

第一节 计算机统计分析

计算机统计分析是心理学研究结果分析的最重要手段。特别是统计软件包的开发和运用，使计算机统计分析在心理学研究中的应用出现了新的突破。

计算机统计分析是运用计算机进行研究数据和资料的统计、处理、分析、贮存和打印输出。除了可以用 BASIC、FORTRAN、PASCAL 等计算机程序语言自编统计分析程序以外，还可以使用计算机统计软件包，心理学研究中最常用的统计软件包有 SPSS，SAS，BMDP 和 Minitab 等。

一、计算机与统计理论

计算机通常可分为巨型、大型、中型、小型和微型。由于微型计算机的迅速发展，已经可以在微机上配备 SPSS 等专用软件包进行统计运算，不过，运算的时间比中、小型机长得多。国外心理学系或研究机构，大都采用中型计算机形成终端网络，可以随时在办公室接通数据库和计算机，进行各种统计分析。同时，配备多种个人计算机，进行专项统计分析和计算。

实践证明，计算机对于统计分析和检验是极为有用的，既快又准确。不过，从现在的情况来看，它们还只是一个计算工具。计算

机并不能告诉研究者特定的检验为何是适宜的，或者，如何解释统计结果；计算机也不知道特定的研究设计是不是有方法论上的缺陷。对于这类问题，研究者需要根据统计理论和研究方法的知识来加以解决，需要以这些理论和知识来运用计算机进行统计计算和检验，正如许多计算机用户所说：“垃圾进、垃圾出”，计算机统计分析的数据本身的质量是至关重要的。这就是说，有效的计算机分析的前提，是严格的研究设计、准确的数据测量和充分的理论知识，这是在进行计算机统计分析时特别需要重视的。

二、统计软件包的使用

统计软件包在计算机统计分析中应用日益广泛。它由一系列统计程序组成，并有一定的操作规程，每一种软件包都有专门的起用指令，在计算结束后，需要用特定指令退出程序。

计算机的数据处理可以分为“联机”（on-line）和“整批”（batch）处理两个类型。联机式是一种交互式处理，用户给计算机的指令立即被执行并随即给予新的指令；整批处理时，所有数据打入穿孔卡系统，成批加工，没有随即反馈。在使用统计软件包时，我们多半作联机运算，而只有在计算机加工负荷较大，统计任务又比较复杂时，才作整批处理。例如，在运算 LISREL 时，往往采用整批处理方式。

心理学的研究涉及各种各样的变量，需要收集被试有关某些变量的数据，这些数据是所有统计软件包的输入，数据形式是每一列表示特定的变量。常见的数据结构如图 13-1。

在输入所有数据以后，我们用一定的指令让软件包完成所要求的数据分析。统计软件包具有一定的数据分析能力，包括描述性统计、相关、列联表、统计显著性检验，数据修正和选择以及各种多元统计方法。

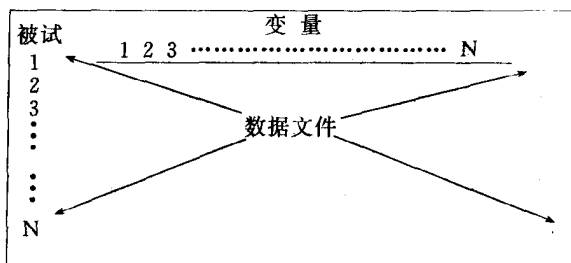


图 13-1 数据文件结构

所有统计软件包都有描述性统计的功能，主要有平均值、中数、众数、方差和标准差等，还包括数据的偏态度和峰态度、频次分布等。统计软件包还有各种数据修正或选择功能，比如，对变量分数进行转换（乘方、开方和取绝对值）；对多变量的分数进行加、减、乘、除的合并或划分；按一定的选择标准部分被试进行某种分析；对变量值作出重新编码，等等。

三、SPSS 和 SAS

SPSS 和 SAS 是心理学研究中最常用的统计软件包，在这里作一些专门介绍。

1. SPSS 软件包

SPSS 是“社会科学软件包”的英文缩写，它有联机系统和整批系统两种。SPSS 的使用有两个步骤：建立数据文件或数据库；写出 SPSS 程序，执行数据分析。

(1) 数据文件

数据文件的结构（图 13-1）需要对所有数据作出数字编码，即使是“是”与“否”的反应，也应以“1”和“2”作出编码；对于某些背景变量（如性别、职业等），也可以用数字编码。

在许多研究中，被试会对一个或多个变量不填写答案或不作出反应，这种情况称为漏失数据。数据漏失的原因多种多样，人们常常简单地排除任何包含了漏失数据的被试，索性不对这些被试的测

量结果作出分析。但是大多数研究者宁愿保留这些被试，而对漏失数据作特殊的编码，并分析其余的数据，SPSS 就可以以这种特殊方式处理漏失数据。最简单的办法是留出空格，以空格作为漏失数据而以数字为“真”数据。较多采用的是另一种方法，即用一个不可能为真实数据的数字来表示漏失数据，比如，用“99”作为漏失数据的编码。另外，还可以同时用不同方式进行漏失数据编码，以区别漏失的不同原因。

在建立数据文件时，比较重要的一点，是要求在相同数列中对每一被试作出每一变量的编码。表 13-1 为若干数据的实例。

表 13-1 变量与被试的数据格式

被试	变量→								
001	1	528	12	2	7	3	6	5	2
002	2	528	13	1	4	5	3	4	3
003	1	510	16		3	4	5	1	1
004	1		8	1	6	2	1	2	
005	2	500	14	3	5	6	4	7	5

从表 13-1 可以看到 前三列数据成为一组 表示被试的编码 然后空出一列；第五列数字是性别的编码，1 表示男职工、2 表示女职工，接着又空出一列；第七——九列为经济收入的编码；第十——十二列是文化程度的编码，等等。空一列的目的是便于检查数据，当然也可以不空。在数据全部输入以后，需要对数据作出具体规定和定义，告诉计算机，哪几列属于哪一变量。在不少情况下，每个被试都有大量数据，这时，可以对每一被试提供多行数据，但仍需特别注意各数列变量的一致性。

(2) 程序与指令

SPSS 程序由一系列执行某种功能的指令所组成，每一指令跟随着特定的操作说明。SPSS 指令又分为数据定义指令和数据修正

与选择指令。统计处理指令等几种类型。例如，用于变量命名的指令：VARIABLE LIST；用于从被试中选择一定子样本的指令：SELECT IF；用于对连续变量进行描述性统计的指令：CONDESCRIP-TIVE；用于计算两个或多个变量之间的皮尔逊相关系数的指令：PEARSON CORR，等等。

在使用 SPSS 进行统计分析时，一般的程序格式为：

100 统计指令；规定变量

110 功能选择；表

120 统计；表

可以根据具体需要，选用具体的指令和说明，下面是 SPSS 的一个方差分析（ANOVA）程序的实例：

10 VARIABLE LIST

15 SUBNUM NOISE CROWD, ERRORS

20 INPUT FORMAT

25 (F3.0 1X F1.0 1X F2.0)

30 N OF CASES

35 30

40 ANOVA

45 ERRORS BY NOISE (1 3) CROWD (1 2)

50 STATISTICS

55 ALL

2. SAS 软件包

SAS 是“统计分析系统”的英文缩写，是比较新的软件包。SAS 包含许多有用的功能，例如，数据处理、图像输出、图表以及某些商业应用等，它包括了其它统计软件包所拥有的各种数据定义、修正和统计分析功能。SAS 也分为联机交互式 and 穿孔卡成批系统。

SAS 程序所用的语句都以分号（；）结尾，第一行程序为数据语

句：DATA；然后是输入语句，用以表示每一变量名称（不超过 8 个字符，并以字母开头），例如：

```
INPUT SUBNUM1-3 SEX5;
```

这一语句表示输入变量；被试号 1-3 列，性别第 5 列。

SAS 自动地把空白作为漏失数值，还把句号作为漏失值。SAS 的统计程序采用 PROC（或 PROCEDURE）语句，而且必须包括统计程序的名称，即“PROC 程序名称”；例如，对“收入”、“文化程度”变量进行描述统计分析时，用以下程序：

```
10 PROC MEANS;
```

```
20 VAR INCOME EDUC;
```

如果对一组数据作 t 检验，程序为：

```
DATA TTEST;
```

```
INPUT SUBNUM1-3 NO1SE 5 ERRORS 7-8;
```

```
CARDS;
```

```
001 1 1
```

```
002 1 2
```

```
003 1 0
```

```
004 1 1
```

```
005 1 1
```

```
006 1 2
```

```
007 1 1
```

```
008 1 2
```

```
009 1 0
```

```
010 1 2
```

```
PROC TTEST;
```

```
CLASS NOISE;
```

```
VAR ERRORS;
```

我们还可以用 SAS 进行相关分析 (PROC CORR)、方差分析 (PROC ANOVA), 因素分析 PROC FACT 以及其它各种多元统计分析。

SPSS 和 SAS 都有各自的使用和操作手册。在运用这些统计软件包之前, 必须认真阅读手册, 掌握操作要领, 记住注意事项。这些统计软件包还有不同的版本, 新版本往往包含一些新的功能或在程序上作了新的改进。新版 SPSS⁺ 可以在 IBM 微机上运行, 使用程序也更为简便。不论运用何种计算机统计分析程序, 都需要首先明确研究目的和统计分析要求, 使计算机统计软件包更好地为心理学研究服务。

第二节 研究结果的解释和总结

一、结果解释的意义和原则

在对研究数据进行分析以后, 就应该从研究的理论框架和目的出发, 根据数据分析结果, 作出合理的解释, 使结果的心理学意义充分地展现出来。

结果解释的工作是一种创造性的活动, 不但需要理论的指导和足够的知识, 而且需要丰富的想象力, 解释也是对研究程序和资料的实际评估, 又是一种沟通心理学含义的艺术。

在对结果的解释中, 首先需要考虑所研究的问题, 分清主次, 并对名词作出解释; 其次, 应该对研究方法的科学性作出分析, 包括取样的代表性、研究设计的效度、研究者的训练、测量工具的信度和效度、数据的误差和特征、统计分析方法的合理性等等。在此基础上, 对结果的含义恰如其分地作出解释和分析, 特别应重视研究假设中没有考虑到的关系或没有预期的研究发现。对于无结构访谈

和观察等的结果，必须先进行专门的编码，编码可以由研究者（访谈或观察者）自己担任，也可以训练专业编码员进行编码。前者常常可以结合现场体验，使编码更有依据，但可能会受到某些无关因素的干扰，比如，被访人的外表、态度和双方关系等；后者判断比较客观、可靠性较高，但必须对编码员作专门训练和培养。

在对结果作出解释之前，还需要考虑具体的数据分类原则，其中最重要的是根据研究课题性质和目的来确定类别；第二，各类别相对独立，并且不重复，这就要求变量的定义十分明确；第三，各个类别必须是处于同一维度或层次，这些都是在进行数据分类时必须注意遵循的。

二、结果的解释

我们知道，心理学研究的目的，是解释、预测和控制心理过程以及影响心理活动的各种因素，尤其是需要找出变量之间的关系和发展趋势。在这方面，研究的效度对结果解释的影响很大，例如，内部效度会影响变量之间具体关系的明确性，关系到对研究本身含义的解释；外部效度则影响研究结果的普遍意义，关系到在解释中如何作出普遍性的结论。

结果解释是在一定的理论框架下进行的，具体的方法主要有四种。

1. 推论法

这是指从结果统计证据，作出逻辑推论，即从已知的事实，推出未知的原理。

2. 演绎法

这是由一般的原则推论出局部的道理。常用的方法是三段论，从多个已知命题推论出某一未知命题，演绎是作出科学预见和发展理论的重要阶段和环节。

3. 归纳法

归纳法与演绎法正好相反，它以许多特殊的研究结果为基础，归纳出普遍的、一般的原则，在目前的心理学研究中，归纳法用得比较多。60年代末，格拉斯（Glass, B. G.）和斯特劳斯（Strauss, A. L.）在他们的名著《发现基础理论：定性研究的策略》中，提出在心理学和社会学等领域的研究中以演绎方式发展理论的新途径。这在心理学研究方法的发展中产生了很大的影响。

归纳法的具体方法很多，例如枚举归纳法、不完全归纳法。归纳法的推理方式大致有五种：契合法（求同法）、差异法、同异共用法、共变法、剩余法，在运用这些方法进行心理学研究结果的解释与分析时，应记住恩格斯的话：“归纳和演绎，正如分析和综合一样，是必然相互联系着的。”（恩格斯：《自然辩证法》）

第三节 研究报告和论文的撰写

一、论文报告的意义和要求

1. 研究报告的意义

研究报告是对于研究的成果、事件、情况、进展或者解释结果的陈述。撰写研究报告，是心理学研究的最后阶段，撰写明确的有说服力的研究报告，也是一种十分重要的研究技能。如果不能通过有效的、科学的研究报告，明确地与其他人交流研究成果，那么，研究本身的质量再高，也是没有价值的。心理学的许多研究，都是因为良好的研究报告而为人们所接受和运用，从而在实际生活与进一步研究中产生影响。例如，著名心理学家米勒（1956）的“魔数7，加減2：我们信息加工能力的某些局限”和费斯汀格等（1959）的“强迫顺从的认知后果”等经典论文，都通过相当的论文报告技巧，

把他们的研究及思想传递给人们，成为认知心理、社会心理等各心理学领域发展的重要标志。也有一些研究，由于论文报告令人费解，而使研究结果失去意义。当然，如果研究本身漏洞百出，那就不可能通过论文报告来弥补了。正因为如此，研究报告被作为研究各环节中不可缺少的一环，它又是与研究的构思、方法和结果紧密联系在一起，研究报告和论文的撰写，成为心理研究方法中重要的内容。

2. 研究报告和论文的要求

研究报告形式较多，常见的有单一研究的报告，篇幅较短，针对某一具体实验或调查作出报告。学位论文的要求就高一些，例如，学士论文详细报告某一研究的结果；硕士论文包括对同一课题的一、二项研究的综合结果，篇幅长一些，讨论深入一些；博士论文则一般要求围绕同一课题报告三-五项研究，作出比较全面、透彻和系统的分析。其范围较之硕士论文更广泛，分析、解释或评估某一主题思想。

研究报告和论文的目的不同于其它类型的写作，它是要尽可能明白、准确地报告研究的内容和结果。好的研究报告除了明白、准确以外，还应经济、简炼，这就要求研究者进行充分的思考和总结，研究报告一般都经过几稿的修改和征求意见。

1974年出版的《美国心理学会出版手册》（第二版）详细地提出和规定了心理学研究论文的格式和写作方式。这一手册已为世界各国心理学家所接受，成为各心理学杂志发表论文报告的标准。

在论文报告中，应该设法简单和直截了当地说明研究的思想，避免不必要的复杂词语和句子，文章的展开应该平稳和富有逻辑性，并注意上下文的连贯性，真正“面向读者”，使人看得懂。

在撰写心理学研究报告和论文时，应该准备写几稿，在第一稿写好后，应让同行阅读和评价一下。《美国心理学会手册》就建议，

应该由与自己意见不同的人（“敌人”）来阅读和评价，而不应只让“朋友”来评定，这样就更容易发现问题或者错误了。然后，进行认真的补充和修改。论文报告还应注意书写整洁和段落分明，有研究表明，整洁的论文阅读效果更好。

心理学研究报告与论文的评价有八个标准，可以供我们在撰写论文报告时参考：

- （1）具有一个或多个具有理论意义的新结果；
- （2）结果具有重要的实际意义和可应用性；
- （3）论文思想新异，或者以新方式说明老问题；
- （4）研究结果的解释明确合理；
- （5）论文以简明方式综合了新的研究数据；
- （6）论文指出了先前所持的某种思想的问题或局限；
- （7）研究有明确的研究图式或者实验操作；
- （8）所报告的结果或理论具有普遍性。

二、研究报告和论文的组织

心理学研究的报告和论文包括四个主要部分：引言、方法、结果和讨论，它还应该有标题、摘要和参考文献，大多数报告还有注释和图表。

1. 标题

论文报告的标题应该简短地表示出研究的主要问题，一般可以表明所研究的主要变量。例如，一项研究的标题为“个性与职业信息对于工作评定的影响”，这就比较简明地说明了研究的问题和主要变量，如果再加上“有关……的实验研究”或“有关……的调查”就多余了，应该尽量避免。

2. 摘要

论文报告的摘要是研究的主要内容与结果的简短总结，一般应

在 500—800 字左右（英文摘要 100—150 字）它并不是整个论文的段落大意，而是简短而直接地说明研究的问题、所用的方法、结果和主要结论，从而使读者在看了摘要以后能决定是否值得读全文。因此，摘要是论文的重要部分。

3. 引言

在引言部分，应该说明研究的理论框架以及同一课题上先前的文献，但并不是“文献综述”。“引言”使读者了解研究的目的和实际原因，并且表明先前研究结果对于从事进一步研究的动力，指出将在先前研究的基础上探讨的新问题和假设。

4. 方法

“方法”部分详细地说明用以检验假设的设计和程序，应使读者获得足够的信息，有可能重复这项研究。《APA 出版手册》规定这一部分“应包括理解和重复该研究所需要的基本信息”。

方法部分又分成几个小节，分别说明研究的被试、测量和实验仪器、研究的具体程序等。

（1）被试：被试的数目、取样方法、特征（年龄、性别、来源等），动物被试还应报告其种类等；

（2）仪器：说明研究所用的设备、仪器、测量工具、问卷等评定量表的设计和参数等；

（3）程序：具体说明研究的过程、指导语、变量的水平及其测量、被试的分配和实验设计、对无关变量的处理方法、实验处理的内容与性质、主试的情况、研究的步骤，等等。

5. 结果

总结研究的数据和有关的统计分析，应简短地说明每一结果与研究假设的关系，而且，应报告所有有关结果，包括与自己假设不一致的结果。图表往往可以明确、简洁、经济地表示出主要结果，特别是当需要报告大量的研究数据；在作出图表时，应加上简短的解

释，使人能看懂内容。

6. 讨论

这部分内容是对研究结果的含义和意义的评价，应说明结果是否支持了研究的假设，并讨论其效度和理论意义及实际意义。讨论中，还应指出研究的局限或进一步需要研究的问题。

7. 参考文献

在论文报告的末尾，应该列出研究报告中所直接提到的或引用的资料来源，包括资料的时间、内容、发表刊物及页码等。

参考文献的排列可以按照作者的姓氏笔划为序，也可以按研究报告所提到或引用的顺序排列。

8. 附录

研究所用的主要问卷或测评表，以及研究所涉及的重要统计推导或公式等，应该在“附录”中列出，以便读者在重复研究时使用，也便于对研究的思路、方法和统计分析作较深入的了解和评价。

第十三章思考题：

1. 在进行计算机统计分析时，应重视哪些问题？如何使用计算机统计软件包？
2. SPSS 与 SAS 各有何特点？如何处理漏失数据和编写执行程序？
3. 结果的解释有何意义？结果解释的有哪些主要原则和方法？
4. 研究报告和论文的撰写应注意哪些要求？论文评价的主要标准是什么？
5. 论文报告包括哪些部分？各部分的特点及其相互关系如何？

参 考 文 献

1. 恩格斯：《自然辩证法》，人民出版社，1971年。
2. 陈立：《工业心理学简述》，浙江教育出版社，1983年。
3. 陈立：“我对测验的看法”，《心理科学通讯》1982年第3期。
4. 陈立：“见习统计方法中的误用与滥用”，《心理科学通讯》1985年第3期。
5. 陈立：“管理心理学新动向”，《应用心理学》1986年第1期。
6. 叶佩华等：《教育统计学》，人民教育出版社，1982年。
7. 赫葆源等：《实验心理学》，北京大学出版社，1983年。
8. 王重鸣：“决策研究的若干理论和方法”，《应用心理学》1986年第4期。
9. 王重鸣：“心理学研究方法发展的几个新特点”，《杭州大学学报》哲学社会科学版，第16卷，第3期，1986年。
10. 王重鸣：“心理学的研究方法”，汪安圣主编《心理学及其在工业中的应用》中第15章，机械工业出版社，1987年。
11. 王重鸣：《劳动人事心理学》，浙江教育出版社，1988年。
12. Anastasi, A. Psychological Testing, 5th. ed. 1982.
13. Cook, T. D. & Campbell, D. T. ; Quasi-Experimentation; Design & Analysis Issues for Field Settings, 1979.
14. Cozby, P. C. : Using Computers in the Behavioral Sciences, 1984.

15. Cronbach, L. J. ; Essentials of Psychological Testing, 4th. ed. 1984.
16. Dunnette, M. D. (ed). Handbook of Industrial and Organizational Psychology. 1983.
17. Chisell, E. E. , Campbell J. P. , and Zedeck, S: Measurement Theory for the Behavioral Sciences, 1981.
18. Glass, G. V. , McGaw B. and Smith, M. L. ; Meta-Analysis in Social Research, 1981.
19. Guilford, J. P. ; Psychometric Methods, 2nd. ed. 1954. \=
20. Harman, H. H. ; Modern Factor Analysis, 1976.
21. Hays, W. L. ; Statistics, 3rd ed. 1981.
22. Hopkins, K. D. & Stanley, J. C. ; Educational and Psychological Measurement and Evaluation, 6th. ed 1981.
23. Kerlinger, F. N. ; Foundations of Behavioral Research 3rd ed. 1986.
24. Landy, F. J. ; Psychology of Work Behavior, 3rd. ed. 1985.
25. Lord, F. M. ; Applications of Item Response Theory to. Practical Testing Problems, 1980.
26. Marken, R. ; Methods in Experimental Psychology, 1981.
27. Pedhazur, E. J. ; Multiple Regression in Behavioral Research; Explanation and Prediction, 2nd. ed. 1982.
28. Rubenowitz, S. ; Utrednings-och forsknings-metodik, 1980.
29. Rubenowitz, S. ; Test och deras anvandning vid urval, 1959.
30. Saz, G. ; Foundatons of Educational Research.
31. Siegel, S. ; Nonparametric Statistics for the Behavioral Sciences, 1956.
32. Stone, E. ; Research Methods in Organizational Behavior 1978.

33. Tabachnick, B. G. & Fidell L. S. : Using Multivariate Statistics
1983.
34. Zikmund, W. G. : Exploring Marketing Research, 2nd Ed.
1985.

前 言

王重鸣博士的新著《心理研究方法》一书，是根据他四年来讲授该课的讲义精心修改而成的。我很赞赏费歇尔（R. A. Fisher）在其晚年的著作《统计方法与科学论断》的扉页上所引的18世纪阿诺德（A. Arnauld）的一句名言：“我们是用推理来改进科学，然而我们应该用科学来改进推理。”我的业师斯皮尔曼教授，不能说是统计学专家，但因为智力理论的需要而创建了因素分析的方法，成为今日统计方法中的一个重要分支。他又因为心理学有些数据只能应用序数，因而提出了等级相关公式，开创今日非参数统计的先河。我对心理研究方法便存着同样的期待。本书第一编对各种方法的叙述，我认为已提纲挈领地体现了心理学理论发展与研究方法繁殖的血缘关系。

本书第二、三两编，可以说已经深入到许多关于方法论的问题。如此，它就不能局限于数据的处理，而必涉及到实验设计（广义的）和实验方法各方面的问题。我认为这是非常值得重视的课题。数据的科学性，其准确程度，都是在数据是如何取得的过程中决定的。现代的科学方法论，即所谓科学学的，从维也纳的逻辑学派开始，曾经主张科学的统一论（unity of science），好像只有一种科学方法。我觉得这只是哲学的思辩。具体问题要具体分析。所以从实际研究工作来讲，我们不能不顾及研究对象的多样性而采取不同的方法与技术。清人龚自珍曾有“不拘一格降人才”的俊语，我觉得研究方法也应该采取这种宽容的态度，不能一律奉数学模型为圭臬。内省法和访谈法等，也都应该取得恰当的地位。

第四编可以说是统计方法，似乎是全书的梁柱。统计方法在应

用上，已经遍及结果处理的各个领域。这里所提出的，大都是现代计算机《社会科学统计软件包》（SPSS）所大体包涵了的。所以熟悉这些内容，不仅是心理学工作者，其它诸如社会科学（包括管理科学等）、教育科学、甚至许多工程技术也都用得着的。因此本编标题没有局限于心理学。这样，本书就可以作为许多社会科学、各种应用科学工作者的参考用书。本书可资利用范围之广，实在可观。

统计方法通常分成两大内涵，描述的与推断的。本书的重点在推断，例如对因果的联系，并且深入到多因果的实际。但我不得不指出，对描述性统计，特别是对图解法的不足，我希望在再版时能加以补充。1986年，英国的洛维（A. D. Lovie）主编的《心理学与社会科学统计学的新发展》一书，即强调图解法和图画在统计方法中的重要性。他指出图解法不仅使数据表达得更加紧凑，一目了然，而且它能积极地帮助提出结果的假设并加以证实。我也许过份强调了描述与推断统计方法的对立，实际上，描述与推断是有联系的。就如平均数与差异值，在描述与推断统计方法中，作用是同等重要的。正因为如此，在现代统计方法中，图解就变得越来越重要。

本书最后一章对计算机统计分析的介绍，我认为是很及时的。读者会感到实惠。最后，连论文报告的撰写都有所指引，这都不能视为画蛇添足。为补充我在上面提出的有关图解法的意见，本书所引用的软件包，MINITAB即可应付裕如。所以对读者来说，本书如果利用得当，也可以说是应有的了。我对本书的出版是抱有很高期望的。希望广泛的热情读者，能从中摄取应有的帮助与启发，推陈出新，这无疑会给作者以更大的鼓舞，会在方法学的研究与应用中，得到进一步的发展。

陈 立

1989年 7月于杭州大学