心理统计学

Psychological Statistics

中国心理学会心理学教学工作委员会推荐教材
高等院校·应用心理学专业教材

胡竹菁　主编

高等教育出版社
HIGHER EDUCATION PRESS

目　　录

第一版前言

从 2008 年 12 月接受中国心理学会心理学教学工作委员会副主任苏彦捷教授和高等教育出版社单玲编辑的邀请，承担《心理统计学》一书的编写工作至今，已是近一年的时间了。

接受任务后，本书副主编之一董圣鸿博士很快就组织起写作班子，他们是来自不同高校的八位具有博士学位的中青年教师，一直从事心理统计学的研究和教学工作。董圣鸿博士与各位作者按照心理统计学的学科体系确定了本书初稿的提纲和编写分工。经过一年的努力，我们终于完成了编写工作。

与国内同类教材相比，本书在以下三个方面作了特别的努力：

一是试图较全面地论述各种假设检验相对应的统计检验力（power of test）和效果大小（effect size）的计算方法，这与当前心理统计学的发展趋势是一致的。这部分内容包括统计检验力和效果大小的基本含义、评估原理，以及选择平均数差异检验、方差分析、卡方检验或相关系数等方法对研究数据进行假设检验后，其结果的统计检验力与效果大小的具体计算方法。这也许是本书最重要的特色。

二是调整了全书内容的编排。我国目前常见的心理统计学教材中，对“参数估计”和“假设检验”两部分内容是分开介绍的，但又都涉及平均数、方差、比率及相关系数等统计量。本书作者认为，参数估计和假设检验是推断统计的主要内容，对平均数、方差、比率及相关系数等统计量的推断都包括这两方面的内容，都是基于抽样分布进行推断，从本质上说是相通的。为了强调这点，本书在统稿时将这一部分内容编排为五章：第五章阐述参数估计与假设检验的数学原理，第六章至第九章分别对四种主要样本统计量的参数估计和假设检验进行论述，即平均数的参数估计与假设检验、方差的参数估计与假设检验、比率的参数估计与假设检验以及相关系数的参数估计与假设检验。这样似乎在逻辑上更为合理，读者也更容易接受。

三是试图通过“一般线性模型的方差与协方差分析”这一章，将方差分析和回归分析这两方面的内容在协方差分析模型上统一起来。

在近两个月的统稿工作中，江西师范大学戴海琦教授提出了许多宝贵意见和建议，在此，向戴海琦教授致以最诚挚的感谢！我们还要对苏彦捷教授、单玲编辑在本书编写过程中所给予的支持和指导表示衷心的感谢。本书主编师从张厚粲教授学习和研究心理统计学，张厚粲教授在心理统计领域所做的开创性工作，对于主编及其他作者都具有较大的影响；本书编写过程中也参考了国内外许多心理统计学方面的教材和资料，在此，对张厚粲教授和这些文献的作者表示衷心的感谢。此外，全书统稿期间，江西师范大学心理学院刘建平书记携同所有院领导为本书的统稿工作创造了一个良好的工作环境，在此一并表示感谢。

全书的编写分工如下：第一章、第四章、第十一章，胡竹菁；第二章，康春花；第三章，周骏；第五章、第六章、第七章，张阔、康春花；第八章、第九章，张阔、康春花、刘声涛；第十章、第十二章，董圣鸿；第十三章，涂冬波；第十四章，刘声涛；第十五章，文剑冰。

由于水平有限，书中疏漏之处，敬请读者批评指正。

本书出版正逢江西师范大学建校 70 周年，谨以此表达我们真诚的祝贺。并希望本书的出版也能为我国心理统计学的发展作出贡献。

编者

2009 年 11 月 15 日

第二版前言

本书自 2010 年出版至今已近五年时间了。去年三月底，本书三位正副主编根据使用本教材的教师所提供的意见，参考国外著名心理统计学家所撰写的教材每隔 3-5 年都会进行修订以使该教材更加完善的经验，在与高等教育出版社责任编辑单玲女士协商后，决定对全书修改出版第二版。经过近一年的努力，终于完成了修改任务。与第一版相比较，主要修改内容有如下述：

（一）三位主编分别对本书的三个编写特色进行了重点修改

第一版前言中曾提到与国内同类教材相比较，本书包含以下三个特色：

1.“较全面地论述与各种假设检验相对应的统计检验力和效果大小的计算方法”（第 11 章）

我国著名心理统计学家张厚粲先生早在 1986 年就在她主编的《心理与教育统计学》一书中多次指出“统计检验力（$1-\beta$）是个相当重要的概念”（北京师范大学出版社，1986，P233-234，P268），但由于各种原因，国内心理统计学教材似乎都没有对如何求取“$1-\beta$”的值进行过系统的介绍，以至在 2007 年暑假召开的“第四届全国心理学学术期刊联席会议（福建-武夷山）”上，当有专家提议国内实证性心理学学术论文是否应该像美国心理学学术期刊那样要求提供统计检验力和效果大小指标时，因为国内心理统计学教材没有相应的教学内容而作罢。2011 年，当本书主编向张厚粲先生单独汇报如何求取“统计检验力和效果大小”的内容时，张先生听完汇报后的第一句话就是“你不能只对我一个人讲这些内容，应该让更多的心理学研究者知道”，并亲自组织北京师范大学心理统计学方向的一些教师和研究生与本书主编一起进行专题讨论……。随后，本书主编分别在“中国心理学学会学术年会”、“华人心理学家学术研讨会”、“心理学与中国发展高端论坛”、“首届中国心理学博士后学术论坛”等学术会议上，以及在北京师范大学、北京大学、复旦大学、南开大学、华南师范大学等十多所高校做“统计检验力和效果大小的计算原理及方法”的学术报告。我国著名专业学术期刊《心理科学》

在李其维主编的推动下，从 2013 年第 3 期开始在国内率先要求实证性心理学学术论文必须提供统计检验力和效果大小指标，江西师范大学主办的《心理学探新》杂志随后跟进，决定从 2014 年第 1 期开始也有此要求……。

根据这几年的教学和学术期刊对这两个指标使用过程中的经验，虽然“统计检验力”和“效果大小”是密切相关的两个概念，但在报告研究结果时“效果大小”指标更为重要，因此，，胡竹菁博士对该章按新的编写思路进行重写过程中，将第一版重点论述“统计检验力”改为重点论述“效果大小”的求取方法。此外，他还负责对第 1，4，13，15 章内容进行了修订。

2. 较全面地介绍了“一般线性模型的方差与协方差分析”（第 14 章）

董圣鸿博士在确定本书第一版的写作提纲时已经注意到了“一般线性模型的方差与协方差分析”在现代心理统计学发展过程中的重要地位，并安排刘声涛博士撰写这一章内容（第 14 章）。统计分析有多种方法，人们很容易认为各种方法之间没什么关系，但事实上并不是这样，很多方法从某个角度看是相同的。读者在本书第 14 章将发现，*t* 检验、方差分析和多元回归原来可以统一到一个模型下，这个模型就是一般线性模型。根据这几年的教学经验和使用本教材师生的反馈意见，董圣鸿博士对该章按新的思路进行了重写。此外，他还对第 2，3，10，12 章内容进行了修订。

3. “在内容编排体系上有所创新”

国内心理统计学教材通常把推断统计中有关“参数估计”和“假设检验”这两部分内容分别安排一章进行论述。但是，正如本书第一版前言中所指出，我们在进行数据处理时，对平均数、方差、比率及相关系数等各统计量的统计推断实质上都包括“参数估计”和“假设检验”两方面的内容，其基础都是基于抽样分布而进行的推断，从本质上说是相通的。因此，我们可以通过一章的篇幅（第 5 章）专题阐述参数估计与假设检验的数学原理，而后分别用专章（第 6-9 章）分别对上述四种主要样本统计量中每一种的参数估计和假设检验统一进行论述。经过四年的教学实践，现在我们仍然认为这样编排在逻辑上更为合理，读者也更容

易接受。由于在第一版的写作过程中，这部分内容主要由张阔博士负责，因此第二版仍然由他对这部分的五章内容进行修改。

（二）对本书涉及到使用 SPSS 软件包的例题中，统一用 SPSS 中文版（18.0）运算后的结果进行编写。

本书第一版的编写过程中，由于不同编者所拥有的 SPSS 版本不一样，因此全书不同章节中的 SPSS 软件包界面，以及对其运算结果的介绍过程都有差异。这次修改过程中我们在版本上统一使用正版 SPSS 中文版（18.0）进行运算，对运算过程也统一使用第一版第十章通过“流程框”的介绍方式。

（三）对各章的关键术语与正文中的论述进行了相应的修改，并统一在正文中附上了相应的英文词组。

有如本书第一稿的统稿过程那样，江西师范大学戴海琦教授在本书第二版的修改工作过程中仍然给予了全程的关注和支持，提出了许多宝贵意见和建议，在此，向戴海琦教授致以最诚挚的感谢！同时也对单玲女士在本书修改全程中给予的许多帮助和指导表示感谢。最后需要指出的是，虽然本书第二版主要由三位正副主编负责修订，但本书仍然是参与第一版撰写工作的八位博士的共同成果。

谨希望本书第二版能继续为我国心理统计学的发展做出贡献。

由于水平有限，书中疏漏之处，敬请读者批评指正。

编者

2015 年 3 月 30 日

第一章 绪 论

学习本章内容，将有助于你对以下问题的理解与思考：

1 心理统计学的研究对象是什么？

2 心理统计学研究的数据有何特点与类型？可分为哪几种类型？

3 我们为什么要通过统计推断工作来研究心理现象的有关规律？

4 心理统计分析主要包括哪些内容？

第一节 心理统计学的研究对象

一、心理统计学的定义

心理统计学（psychological statistics）是研究如何运用概率论与数理统计学的原理和方法，对搜集到的心理现象的有关数据资料进行科学推论，以找出心理活动规律的一门科学。心理统计学是数理统计学与心理学的一门交叉学科，也是应用统计学的一个分支，同时，它也是进行心理学研究的一种科学方法。

概率论 (probability theory) 是研究随机现象数量规律的数学分支。它有悠久的历史，它的起源与博弈问题有关。16 世纪，意大利的一些学者开始研究掷骰子等赌博中的一些简单问题，例如比较掷两个骰子出现总点数为 9 或 10 的可能性大小。使概率论成为数学的一个分支的真正奠基人是瑞士数学家雅各布·伯努利 (Jakob Bernoulli) ,他建立了概率论中第一个极限定理 —— 伯努利大数定律。而后，在许多数学家的研究基础上，柯尔莫哥洛夫于 1933 年出版了他的《概率论基础》一书，书中第一次给出了概率的测度论式的定义和一套严密的公理体系。这一公理体系的提出是概率论发展史上的一个里程碑，为现代概率论的蓬勃发展打下了坚实的基础，并迅速应用到物理学、化学、生物学、经济学等众多研究领域，成为数学的一个分支学科。

数理统计学 (mathematical statistics) 是数学的一个分支学科，是伴随着概率论的发展而发展起来的。它研究怎样去有效地收集、整理和分析带有随机性的数据，以对所考察的问题作出推断或预测，为采取一定的决策和行动提供依据和建议。数理统计学可用于种种专门领域 (物理、化学、工程、生物、经济、社会等) ，但只涉及其中带随机性的数据的分析问题，而不是以任何一种专门的知识领域为研究对象。在用数理统计方法分析带随机性的数据时，从统计模型的选择、实验方案的制定、统计方法的正确应用以至所得结论的恰当解释，都离不开所研究领域的专门知识、经验以至良好的组织。

心理学在诞生之日起，就一直希望成为像其它自然科学一样的实验性学科，因此，人们也把概率论与数理统计学应用到心理学研究领域，使心理统计学成为指导心理学实验研究的最重要的分支之一。

二、数据的含义与类型

如前所述心理统计学是研究如何运用概率论与数理统计学的原理和方法，将搜集到的心理现象的有关数据资料进行科学推论，以找出心理活动规律的一门

科学。

数据 (data) 在统计学中又称为资料，有时也称为数据资料。它是关于事件的一组离散的数字描述，是构成信息和知识的原始材料。

心理现象的有关数据资料与其他资料一样，按不同的分类标准可分为不同的类型。

1．从数据是否具有连续性来划分，可把数据区分为离散数据和连续数据两大类

离散数据 (discrete data) 是指在任何两个数据点之间所取数值的个数是有限的。例如，某所大学学生党员的数量，教授的人数等。连续数据 (continuous data) 则是指在任何两个数据点之间都可以插入无限多个大小不同的数值。例如，身高、体重等。

2．从数据的观测方法和来源来划分，研究数据可区分为计数数据和测量数据两大类

计数数据 (count data) 是指在心理研究中，对研究对象通过计算个体数目而获取的数据。例如，在学校一次考试中，有多少个学生得分为优秀，多少个学生得分为不及格等用整数形式表达的数据形式。在两个相连的整数数字之间没有其他的数字。例如，在 2 个人和 3 个人之间，不存在 2.5 个人或 2.67 个人这样的数据。

心理研究中除了可以通过计算个体数目而获取的数据之外，有很多数据是需要借助于一定的测量工具或一定的测量标准来获取的，例如考试成绩的分数，对人类个体感觉阈限的测量，身高的测量、体重的测量等，这些需要借助于一定的测量工具或一定的测量标准来获取的数据称为测量数据 (measurement data) 。

3．从数据反映的测量水平来划分，可把数据区分为称名数据、顺序数据、等距数据和比率数据四种

上述计数数据作为研究变量时，一般只能对其进行不同性质或不同类别的定性分类，例如，学生性别可以分为男性、女性两类，人们对某件事的态度可以分为赞成、无所谓、反对等。因此将这类数据称为称名数据 (nominal data) ，有时也称为分类数据。这种数据既无相等单位也无绝对零点，不能进行四则运算。

上述测量数据，可以按其是否有相等单位和绝对零点而分为三种测量水平，分别称为顺序数据、等距数据和比率数据。

顺序数据 (ordinal data) 是指既没有相等单位，也没有绝对零点的数据，即按照研究对象某种属性的多少或大小，按次序将其加以排列后获得的数据资料。如将学生的成绩区分为优、良、中、差，或者将人们对商品喜爱程度进行排序等。由于没有相等单位和绝对零点，顺序数据也不能进行四则运算。

等距数据 (interval data) 是指有相等单位但没有绝对零点的数据，例如温度、离差智商等。这类数据因为有相等单位，因而可以进行加减运算，但因没有绝对零点，不能进行乘除运算。

比率数据 (ratio data) 是指既有相等单位也有绝对零点的数据。例如人们的身高、体重等数据。心理学实验中获取的有关人们各种感觉阈限的物理量、某种认知过程中的反应时等数据都属于这类数据。这类数据可以进行四则运算。

三、数据处理过程中几组常用的概念

在实际工作中，经常需要对在不同条件下实验或是观察 (包括调查) 所得到数据进行分析，以判断不同条件对结果是否有影响，或是不同条件下的结果是否存在显著差异。在这种情况下，常涉及几组常用的概念：

(一) 自变量与因变量

在实验或观察中，研究者操纵或改变某变量的状态以考察其对因变量的影响，这个变量**因素**或**因子**（factor）也称为独立变量或**自变量** (independent variable) ，用英文字母 A，B，C 等表示。

在实验或观察中，因子所取的状态，称为因子的**水平** (level) ，因子 A 的水平常记为 A_1，A_1，…。

实验的结果或是观察指标，称为反应变量或**因变量** (dependent variable) ，常用英文字母 Y 表示。

在实验或观察中，不同水平的组合而形成的特定的实验条件，这被称为**处理** (treatment) 。如果实验只有一个因素，那么一个水平就是一个处理。假设实验有两个因素，A 因素有 2 个水平，B 因素有 3 个水平，那么就有 2×3 共 6 个处理。

（二）总体、样本和抽样

如果我们对人 (或动物) 的某种心理现象感兴趣，例如，高校教师的工作压力状况，或者某市小学一年级新生的智力水平等，想通过研究得到有关该心理现象的特征，首先想到的方法是对所有高效教师进行工作压力感的调查，或对某市小学一年级新生进行智力测验。但这样进行研究，有时会因为成本过高而无法实现，有时要“找到所有成员”本身也是一件不可能的事。因此，如果有一种方法成本较低，对人数不多的个体进行研究后就能得到一般性的结论，无疑会对我们的研究有很大帮助。数理统计学家为我们提供了这样的方法。要了解这种方法，需要先了解总体、样本和抽样等基本概念。

总体 (population) 又称为全域，在统计上是指所要研究的具有某种特征的一类事物的全体。总体中的每个基本单元叫做**个体** (individual)。总体是由一些具有共同性质或特征的个体所组成的。

总体的大小往往随着研究问题的不同而改变，一般可分为有限总体和无限总体两类。如果总体所包含的个体是有限的，则称该总体为有限总体。例如，要考察“某市小学一年级新生的智力水平”，可以对该市小学一年级新生进行智力测验，而任何城市某一年新招收的小学一年级学生总数都是有限的，所以该研究的总体是有限总体。如果总体所包含的个体是无限的，就称为无限总体。例如，要通过试验考察掷硬币时正面朝上的概率，理论上投币的次数是无限的。

在心理学、教育学及其他领域的研究中，直接对总体进行考察大多是比较困难的，有时甚至是不可能的。因此在实际研究中，往往从总体中抽取一定数量的个体加以考察，这些个体就组成了总体的一个**样本**(sample)。从总体中抽取样本的过程就叫做**抽样**（sampling）。所抽取的样本中包含的个体的数目，称为**样本容量**（sample size），通常用n表示。

由于样本中包含的个体较少，所以研究样本比研究总体容易得多。对样本进行研究的最终目的是对总体进行统计推断。例如，要考察“某市小学一年级新生的智力水平”，我们可以从该市小学一年级新生的总体中，随机抽取 5%（或其他百分比例）的学生进行智力测验，并根据测验结果对该市全体小学一年级新生的智力水平进行推断。通过考察样本来推测总体，可以节省对总体调查所需要的大量人才、物力、财力和时间，而且如果抽样方法合理、科学，所得样本对总体有较强的代表性，样本信息能够充分反映总体的特征。

根据样本的数字特征来推断总体的数字特征的过程，称为**统计推断**（statistical inference）。通常认为，统计推断主要包括参数估计和假设检验两大部分内容。无论是进行哪一种统计推断，推论的准确性都会受到样本代表总体的程度的影响，即受到样本代表性的影响。通常样本容量越大，对总体的代表性也越强。此外，样本的代表性也受抽样方法等其它因素的影响。例如，随机抽样的代表性一般较强，而方便取样的代表性则要差一些。

（三）总体参数和样本统计量

在统计研究中，需要对随机变量分布的数字特征进行描述，常用的统计指标有平均数、比率以及方差和标准差等。如果这些统计指标描述的是总体的特征，称为总体参数（population parameter，简称为参数）；如果这些指标描述的是样本的特征，即如果这些指标是通过抽取总体中的部分个体而计算得到的，一般称这些指标为样本统计量(sample statistic)或样本特征值。

(四）函数关系和相关关系

心理学研究的主要任务是分析各种变量之间的相互关系。事物之间的关系包括两种类型，一是函数关系，二是相关关系。所谓**函数关系**（functional relationship）是指事物之间存在着严格的依存关系，其特征是现象与现象之间的数量关系是一一对应的，即对某一个变量的每一个数值，另一个变量都有唯一确定的值与之对应。它可以用数学表达式十分准确地表达出来。所谓**相关关系**（correlativity）指两个事物或现象之间存在一定的数量关系，但又不能确定何者为因何者为果的数量之间的关系。

第二节 心理统计学的主要研究内容与本书的框架

心理统计学主要分为描述统计与推断统计两大部分。

一、描述统计

描述统计（descriptive statistics）主要是阐述在对一组或几组数据进行整理后，如何用统计图、统计表或诸如平均数、标准差、比率和相关系数等特征量数来表达数据特征的统计方法。

本书第二章讲述如何通过绘制统计图和统计表来汇总数据，通过表格或图示，能够直观、形象地呈现数据的概貌，包括数据的散落范围，数据的变化情况，各组数据的结构等，可以使我们加深对原本杂乱无章的数据的认识。

但是，只用统计图和统计表来汇总数据，对于一批数据资料的描述和利用是不够充分、不够全面的，我们还需要用更为凝练、更为概括的数量指标来描述这批数据的统计特征，这样才能更简洁明了、更经济并且更精确地了解这批数据，以便在后续研究中能更进一步地使用它。因此，本书第三章将讲述描述统计学中的一些数值方法，主要包括集中量数、差异量数、相关程度的数值量度等，另外一个较为重要的特征量数即比率的数值计算方法放在第九章介绍。

二、推断统计

通过样本的数字特征来推断总体的数字特征的过程，称为统计推断。利用对样本调查的结果去推论总体是统计学的核心内容，这方面的各种统计分析方法通常被统称为**推断统计**（inferential statistics）。推断统计的数学基础是概率论和数理统计原理，本书第四章将简要介绍概率论的有关知识。抽样的概率分布是各种推断统计方法的基础，其中最主要的四种抽样分布即 Z 分布、t 分布、χ^2 分布和 F 分布，本书将分别在第四、六、七等章它们首次出现时加以介绍。

通常认为，推断统计主要包括参数估计和假设检验两大部分内容。

(一)参数估计与假设检验

参数估计（parameter estimation）是指从总体中抽取的随机样本，来估计总体分布中未知参数的过程。从估计形式上看，可以分为点估计和区间估计。**点估计**（point estimation）是用样本统计量的单一数值估计未知的总体参数。**区间估计**（interval estimation）是依据样本统计量，根据一定的精确度的要求，推断总体参数所在的区间。本书第五章第二节将以平均数的区间估计为例介绍根据样本统计量进行参数估计的基本原理。

假设检验(hypothesis testing)是指根据抽样结果，在一定可靠程度上，对一个或多个有关总体分布的原假设做出拒绝或接受结论的统计程序。当使用这种方法的目的在于检验某个样本统计量的数值与其假设的总体参数之间是否存在显著差异，或者检验两个样本统计量之间的差异是否显著进而推断各自总体参数之间的差异是否显著时，称为差异显著性检验（significance test of difference）。在假设检验过程中，用于判断接受还是拒绝原假设的统计量被称为“检验统计量（statistic of test）”。本书第五章第 3-4 节将以平均数的差异显著性检验为例介绍假设检验的一般原理。

由于样本统计量包括样本平均数，样本方差 (或样本标准差)，样本比率和样本相关系数等不同形式，每种样本统计量又都包含参数估计与假设检验两种统计推断内容，因此，本书第六、七、八、九等章分别介绍样本平均数、样本方差 (或样本标准差)、样本比率和样本相关系数等四种样本统计量的参数估计和假设检验方法。

从某种意义上说，方差分析是平均数差异显著性检验的延伸和方法上的扩展，本书第十章将论述如何用方差分析的方法来解决多于两个平均数和多于一个自变量的假设检验问题。

上述四种样本统计量的假设检验和方差分析，都是在总体分布情况已知，并且仅由少数几个参数便可确定的条件下进行的、其检验方法通常是基于总体呈正态分布的假定，称为参数检验（parametric test）。在统计推断中，还存在总体分布为非正态甚至不知总体分布形态的情况。这种情况下的假设检验称为非参数检验（nonparametric test）。本书第十二章将介绍如何进行非参数的统计推断。

此外，我们还用专章对线性回归 (第十三章)、一般线性模型的方差与协方差分析 (第十四章) 和某些常见的多元统计方法 (第十五章) 作了简要介绍。

(二) 单总体与两总体的假设检验

在心理研究中，在不知道某种现象的总体特征如何的情况下，如果是想知道该总体的某种参数值可能的具体值 (即点估计) 或可能的范围内 (即区间估计)，可以通过抽取包括 n 个个体的一个样本，采用参数估计的方法进行统计推断。一般情况下，参数估计是针对单总体而言的。

在进行假设检验时，要注意存在两种不同的情况：一种情况是在已经知道某种心理现象总体分布情况下，对所抽取的一个样本的样本统计量与总体参数之间的差异情况进行差异显著性检验，这样的统计推断过程称为单总体的假设检验(one-sample hypothesis test)。如果在研究中根据需要在总体一中抽取了包括 n_1 个个体的一个样本，在总体二中抽取了 n_2 个个体的另外一个样本，试图通过检验两个样本的样本统计量之间差异，来推断其各自代表的两个总体之间的差异，这样的统计推断过程称为两总体的假设检验 (two-sample hypothesis test)。

由于样本统计值主要包括样本平均数，样本方差 (或样本标准差)，样本比率和样本相关系数等形式，本书第六、七、八、九、十等章分别论述这些统计量的假设检验时，都包括单总体的假设检验和两总体的假设检验两种情况。

(三) 独立样本与相关样本的假设检验

进行两总体假设检验时，我们还需要注意来自这两个样本的数据之间的关联程度。如果两个样本的数据之间没有关联，即不存在一一对应的关系，则称这两个样本为独立样本（independent sample）。例如，从来自两所不同学校的被试身上获得的某种心理测试

的数据。依据独立样本数据所进行的假设检验称为独立样本的假设检验 (independent sample hypothesis test) 。如果两个样本的数据之间有关联即存在一一对应的关系，则称为相关样本(correlated sample)，例如通过配对 (match) 获得的两组数据，或者是同一组被试在两种不同实验条件下作出的反应数据。依据相关样本数据所进行的假设检验称为相关样本的假设检验 (correlated sample hypothesis test) 。

(四) 本书其他内容

当今国内外心理统计学最重要的发展趋势之一，是有关统计检验力 (power of test) 和效果大小 (effect size) 的计算方法问题。本书第十一章对有关统计检验力和效果大小的基本含义、评估原理以及与各种检验方法相应的统计检验力与效果大小的具体计算方法作了较为详尽的介绍。

本书第十四章则对心理统计学另外一个发展趋势即“一般线性模型”的统计方法做了详尽的介绍。

最后，本书用了一章的篇幅（第十五章）简要介绍了一些多元统计方法在心理学研究中的使用。

第二章　数据的整理：表格法和图形法

学习本章内容，将有助于你对以下问题的理解与思考：

1. 数据的整理的有何意义？主要通过哪些方法进行整理？
2. 统计表的基本结构包括哪些内容？主要有哪些类别？
3. 统计图的基本结构包括哪些内容？主要有哪些类别？
4. 次数分布表与次数分布图主要有哪些类别？
5. 在什么情况下应该用探索性数据分析来进行数据整理工作？

1.有研究者对某市113位男性市民和105位女性市民对城市绿化分布格局的满意情况作了调查，得到的主要结果是：

（1）对该市绿化分布格局满意者：男性15人，女性8人；

（2）对该市绿化分布格局比较满意者：男性11人，女性9人；

（3）认为该市绿化分布格局一般者：男性47人，女性43人；

（4）对该市绿化分布不满意者：男性40人，女性45人。

我们是否能用一张表格或者一幅图的形式，将上述结果直观地呈现给读者？这样的表和图应遵循什么要求？

2.下面是某班40名学生在某项测验中的成绩：

94 89 88 84 89 97 84 99 90 85 87 86 81 84 80 77 75 83

80 77 89 82 73 62 62 58 72 77 92 84 73 79 77 67 83 70

86 88 77 79

请问该班学生此项测验成绩分布有何特征？

第一节 数据的初步整理与常用的统计表和图

在心理学研究中，我们将直接收集得到的数据称为原始数据或观测数据，它们常常是“杂乱无章”的，很难直接提供我们想要的信息。例如，本章开头的问题2，仅看原始数据，我们无法回答提出的问题。只有通过对原始数据进行整理、分析、概括，我们才可以发现数据的特征和规律，从而获得大量规律性知识和有用的信息。对数据进行整理、分析、概括的常用方法包括编制统计表、绘制统计图和计算统计指标等方法，在此先介绍编制统计表与绘制统计图两种方法。

一、数据的初步整理

在对数据进行分析、概括其特征之前，常常需要对数据进行初步的整理，以便制定出简单明了的统计图表。数据初步整理的基本方法有排序和统计分组两种。

（一）数据的排序

数据的排序（sort 或 order）是按照某种标准，对事先收集到的杂乱无章的数据按照一定顺序标准（升序或降序）进行排列的过程。数据排序后，就可以为数据的进一步处理做好准备。对测量数据，如得分、成绩、身高、体重等，可以按数值大小或等级次序排序；对属性（称名）数据，如性别、职业类别等，可以按一般的表达习惯排列，如果没有约定俗成的表达习惯，则可以按照汉语拼音、笔画数或字母顺序排序。数据排序后，对于数值型数据，可进一步划分组别。

（二）统计分组

数据的统计分类又称统计分组，它是对研究中所获得的大量数据进行整理的重要步骤。所谓统计分组(data grouped)，就是根据被研究对象的特征，将所得数据划分到各个类别中去的数据整理过程。整理数据时的分类工作，就是对观测中的分类再次核对、加工，使

分类更趋合理、正确，这样才能使对数据的进一步分析研究建立在比较坚实的基础上。

统计分组一定要以与被研究对象的本质有关的特性为依据，才能确保分组的正确。对数据进行分组时，所依据的特性称为分组标志。分组标志要明确且在整理数据过程中要前后一致。同时，分组标志必须能够将全部数据包括进去，不能有所遗漏。

分组标志按形式大致可分为性质类别和数量类别两种。

1.性质类别

主要是根据事物的属性不同将被观测的事物加以划分，仅反映事物在组别、种类上的不同，不说明事物之间的数量差异。例如将一组被试按性别分为男生与女生，这不能说明数量上的差别；又如按成绩优劣，将这组被试分为优、良、中、差四组，虽然该分类标志本身有好坏之分，但不能反映出各个组别之间的数量差异，因此该分类标志也属于性质类别。性质类别的数目及分类的详细程度，要视研究对象的性质和研究任务而定。

2.数量类别

主要是以数据的取值大小为分类标志，将数据按数值大小排出一个顺序来。主要有两种情况，第一种情况是直接按数值大小，由小到大排序；第二种情况是等级排序，将数据按等级大小排列成序。在这种排序中，项目本身就显示了分类的数量信息，这一点是与性质类别明显不同之处的。

二、统计表

统计表（statistical table）是用表格的形式来表达数据或事物分类等内容的一种有效表达方式。由于表格具有简明、清晰、准确的特点，同时表格的逻辑性和对比性又很强，因而在对数据进行统计分类以后，一般都用统计表的方式加以表达。如果表格设计合理规范，不仅会使文章论述清楚，还可收到美化与节省版面的效果。下面我们将介绍统计表的基本结构与种类。

（一）统计表的基本结构

心理学研究中的统计表普遍采用三线表的形式（中国心理学会，2002）。三线表是表格的一种，三线表通常只有 3 条横线，即顶线、底线和栏目线，没有竖线（见表 2-1）。其中顶线和底线为粗线，栏目线为细线。当然，三线表并不一定只有三条线，必要时可加辅助线。三线表的组成要素包括：表序、表题、项目栏、表体、表注。

项目栏 表序 表题 顶线 栏目线

表 2-1　儿童同伴关系的聚类分析

	自我知觉型	良好型	同伴拒绝型	友谊质量型	同伴忽视型
Z_p	0.029	2.13	−0.63	−0.19	−0.4
Z_n	−0.25	−0.22	2.82	−0.16	−0.3
$Z_{F总}=Z_{Fp}-Z_{Fn}$	−0.5	0.68	0.51	1.89	−1.1
Z_S	0.68	0.56	−0.016	−0.47	−0.78
Z_L(孤独感)	−0.27	−0.84	0.[illegible]3	−0.13	0.68

注：表中数据为各类型在各指标上得分的平均Z分。

表体 底线

表注

1.表序和表题

表序即表的序号。根据表格在文章中被提及的顺序，用阿拉伯数字进行排序，如“表 1”、“表 1-1”等。表题即表格的标题。与文章的标题类似，表题应当准确得体，简短精练。要避免使用泛指性的词语做表题，如“数据表”、“对比表”、“计算结果”等。表序和表题之间空 1 个字距，不用任何点号。表序和表题放置于表的上面。

2.项目栏

项目栏指表格顶线与栏目线之间的部分。项目栏中一般要放置多个“栏目”。所谓栏目，就是该栏的名称，它反映了该栏信息的特征或属性。一个表格应当能把相关的数据归类，以使读者能够进行比较。项目栏确立了数据组织的逻辑，并确定了栏目下数据栏的性质。与标题一样，栏目也应当简单明了，不要过于繁琐、冗长。

3.表体

底线以上、栏目线以下的部分叫做表体，它容纳了表格内的大部分或绝大部分信息，是表格的主体。表体内的数字一般不带单位，百分数也不带百分号（%），应把单位符号和百分号等归并在栏目中。表体中的数字以个位数（或小数点）对准上下对齐。缺失数据项要划“-”。

4.表注

表注是对统计表或者表内的某些内容进行补充说明和解释，般置于表格下方（张厚粲，徐建平，2004）。有 3 种类型：(1)一般注解；先对表格进行整体的描述，说明并提供相关的信息。再对表中的缩写符号等进行解释，一般注解的格式为“注：……(文字说明)”。(2)特殊注解；用于对某一列、行或个别项目进行说明。有特殊注解的部分，用标注于左上角的小写字母来注明。(3)概率注解。概率注解是用来注明检验结果的统计学显著性。应用星号标出拒绝原假设的 p 值。如 $*p<0.05, **p<0.01, ***p<0.001$ 等。

（二）统计表的种类

统计表可按形式及内容不同，划分成不同的类型。不同类型统计表的具体功能不同。

1.简单表

这是指只列出调查名称、地点时序或统计指标名称的统计表，如表 2-2。

表 2-2 儿童同伴关系研究调查样本结构

年级	四年级	五年级	六年级
人数	83	85	81

2.分组表

这是指只有一个分类标志分组的统计表，如表 2-3。

表 2-3 同伴关系不同类型的学生孤独感水平

	自我知觉型	良好型	同伴拒绝型	友谊质量型	同伴忽视型
Z_L(孤独感)	-0.27	-0.84	0.3	-0.13	0.68

3.复合表

这是指有两个或两个以上统计分组标志的表，如表 2-4。只有两个分组标志的称为两项表；有三个分组标志的称为三项表，如此类推。

表 2-4 不同年龄儿童在两维空间方位下推理的成绩（$M \pm SD$）

年龄（岁）	推理成绩	
	单模型	双模型
7	0.65±0.89	0.58±0.76
9	1.21±0.99	1.25±0.98
11	1.79±0.94	1.79±1.02
全体	1.22±1.04	1.21±1.04

资料来源：心理学报, 2004，36（2）：174-178.

（三）SPSS 制作统计表简介

使用 SPSS 制作统计表，首先要准备好数据。SPSS 的数据一般而言，一行即为一位被试的原始数据，例如本章开头的问题一的数据就应该有 218 行，具体参见本书所附光盘中例 02-1.sav 的数据。如果使用本章开头问题一所呈现出来的整理后的数据（参见本书所附光盘中例 02-1-2.sav 的数据），那么在 SPSS 分析过程中要把人数作为加权变量（见图 2-1）。

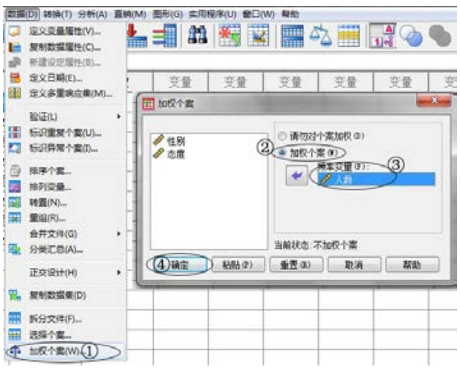


图 2-1 SPSS 数据加权过程示意图

使用 SPSS 制作统计表，即可通过“分析\表\设定表”来交互式地制作统计表，也可采用“分析\描述统计\频率（或交叉表）”制作统计表（如图 2-2 所示）。以本章开头问题一的原始数据为例（所附光盘中例 02-1.sav 的数据），简要说明制作交叉表的过程。按图 2-3 所示，把“性别”作为行分类变量，把“满意程度”作为列分类变量，制作出来的交叉表如表 2-5 所示。制作统计表的详细过程请参阅有关 SPSS 书籍。

图 2-2 SPSS 制作统计表常用菜单命令示意图

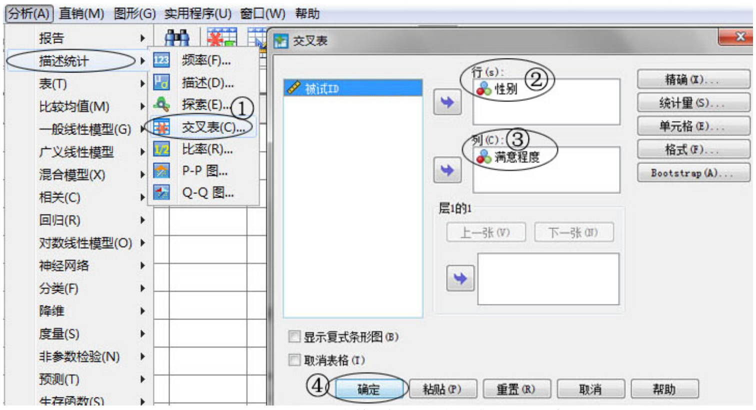


图 2-3 SPSS 制作交叉表过程示意图

表 2-5 不同性别对城市绿化分布不同满意程度的人数分布表

		满意程度				合计
		满意	比较满意	一般	不满意	
性别	男	15	11	47	40	113
	女	8	9	43	45	105
合计		23	20	90	85	218

三、统计图

经过汇总整理与加工取得的统计资料，不仅可用统计表来表现，还可用统计图来表现。统计图是统计资料整理结果的基本表现形式，也是统计分析的重要工具。所谓统计图（statistical graph(or diagram)）就是依据数字资料，应用点、线、面、体、色彩等制成的整齐而又有规律，简明而又知其数量的图形。统计图在数据整理中占有很重要的地位。一图知万言，一张简单的图形，就可以把一大堆数据中的有用信息概括地表现出来。图形比数字更为具体，能把事实或现象的全貌形象化地呈现出来，给人以清晰、深刻的印象，因而便于理解和记忆。统计图还有一定的艺术性，它可以表现得生动、有

趣。现在统计软件的使用十分普遍、便利，画统计图变得十分简单，可以充分发挥统计图在呈现数据时的直观、形象的作用，而且使用统计软件制作统计图，可有效避免传统手工制作统计图存在的不足或易出现的错误。为有效地反映数据的特征，关键在于根据数据的类型选择恰当的图形。

（一）统计图的基本结构

1.图序和图题

图序即图的序号。根据图在文章中被提及的顺序，用阿拉伯数字进行排序，如“图 1”、“图 1-1”等。图题即图的标题，要求简明准确，要有较好的说明性和专指性，不要为追求形式上的简洁而选用过于泛指的图题。图序和图题之间空 1 个字距，不用任何点号。放置于图的下面。

2.标目

标目是用来说明坐标轴的含义的，应该与被标注的坐标轴平行，居中排印在坐标轴和标值的外侧，文字方向与坐标轴方向相垂直。

3.标值

坐标值上的标值范围是根据图形的数据确定的，但坐标轴上标值的间距则是任选的，不同的间距使图形表现出不同的效果。例如男女生的平均成绩相差 5 分，如果以 5 分或 10 分作间距则觉得相差不大，如果 1 分作间距则觉得相差很大。因此标值的选择要慎重，以准确表达数据的特点为根本。

4.图例

当图中用某一符号来代表某一变量时，就需要用图例来说明该符号的意义。在许多类型的图中，图例是不可缺少的组成部分。

5.图注

图中未能表达又必须要表达的信息应在图注中说明，如需要解释的度量单位、符号以及缩略语等。它可以帮助读者理解统计图中所示资料的意义，提高统计图的使用价值。

统计图的基本组成要素如图 2-4 所示。

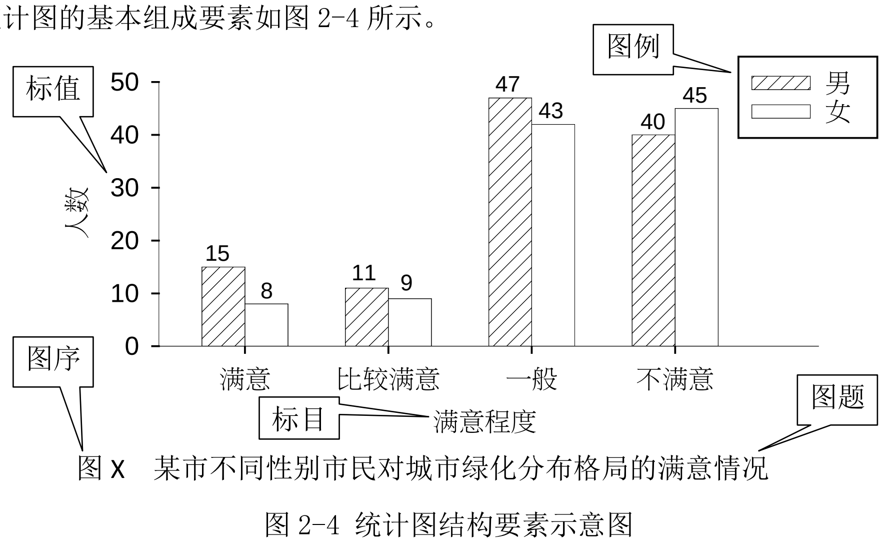


图 2-4 统计图结构要素示意图

（二）几种常用的统计图

1.条形图

条形图（bar charts）是用宽度相同的长条的高度来表示各个统计事项之间的数量大小及差异情况。条形图通常适用于描述离散型变量（如类别变量）的统计事项。常见的条形图有简单条形图、复合条形图和分段条形图。

简单条形图（Simple Bar Chart）是用同类的直方长条来比较若干统计事项之间的数量关系的一种图示方法。它适用于统计事项仅按一种特征进行分类的情况。如图 2-5 是根据本章开头问题一关于某市市民对城市绿化分布格局的满意情况的题目绘制的一个简单条形图。

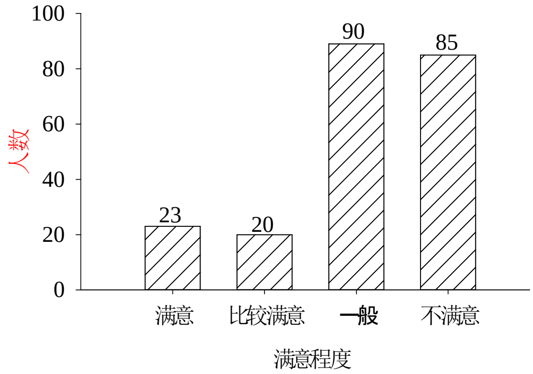


图 2-5　某市市民对城市绿化分布格局的满意情况示意图

复合条形图（Clustered Bar Chart）是指由两个或两个以上直条组成条组的条形图，各条组之间有间隙，组内直条之间无间隙。它常用于呈现多特征分类下的统计事项之间数量关系。例如我们要了解不同性别市民对城市绿化分布格局的满意程度之间的差别，可按照被调查对象的性别和对城市绿化的满意程度两个特征来划分，采用复合条形图呈现调查结果。图 2-4 所示的就是复合条形图。

分段条形图（Stacked Bar Chart）和复式条形图一样，也需要多考察一个分组因素，是一种以直条全长代表某个变量的总量，而用其中的分段长度表示不同来群对总量的贡献（构成比或数量大小）的一种条形图。图 2-6 就是分段条形图。与图 2-4 的复合条形图相比，图 2-6 主要描述了不同性别内部不同态度人数的构成情况，而图 2-4 主要描述了持相同态度人员中不同性别人数的比较。

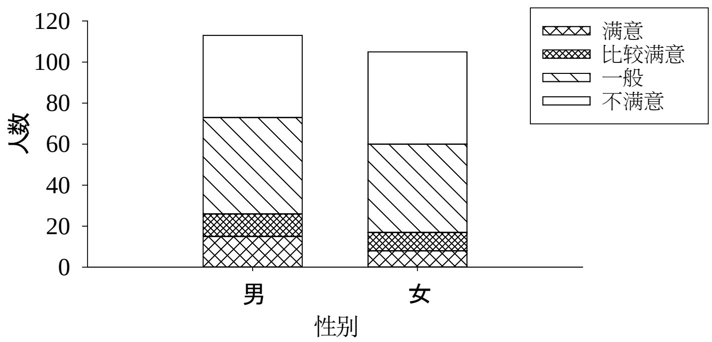


图 2-6　某市不同性别市民对城市绿化分布格局的满意情况

使用 SPSS 制作条形图的过程见图 2-7 至图 2-8。在图 2-7 的第二步可选择条形图的类型，一旦选定哪个类型，外面会有一个黑色边框，图中显示选择的简单条形图。选择不同类型图后点击“定义”，即可进行图 2-8 中所示的 a、b、c 图之一，根据图中椭圆区的设置，即可分别得到图 2-5、图 2-4 和图 2-6 的结果。

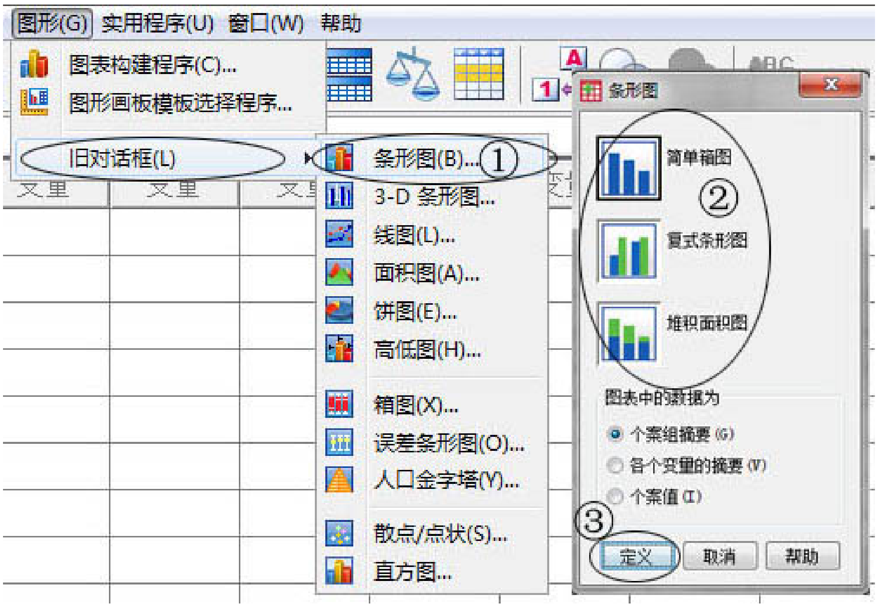


图 2-7　SPSS 制作条形图示意图一

图 2-8 SPSS 制作简单条形图、复合条形图和分段条形图示意图二

2.圆形图和百分条图

圆形图（circle graph），又称饼图（pie），是以单位圆内各扇形面积所占整个圆形面积的百分比来表示各统计事项在其总体中所占相应比例的一种图示方法。它主要用于描述离散型变量，特别是具有百分比结构的分类数据。它可以直观地显示各统计事项在总体中所占的比例大小，也可以反映各部分之间的数量差异。参照图 2-9 呈现的 SPSS 制作圆形图的过程并对图形作适当编辑，可制作出图 2-10 的圆形图，从中可以看出市民对城市绿化分布格局满意程度的比例情况。

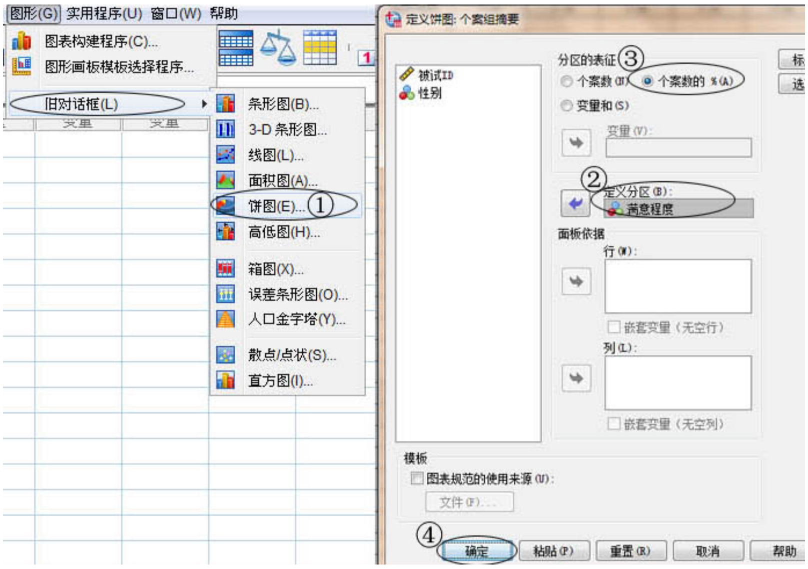


图 2-9　SPSS 制作圆形图过程示意图

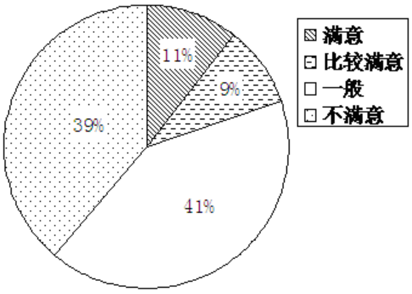


图 2-10　某市市民对城市绿化分布格局的满意情况示意图

百分条图（Percent Bar Chart），是以直条内部各部分面积的大小表示事物内部各组成部分所占的百分构成比的条形图，如图 2-11 所示。百分条图与圆形图的作用相似，但在比较多组群体内部构成比例时，百分条图比圆形图要更加直观。百分条图是在分段条形图的基础上，把各组的总人数都统一为 100%而成，使用 SPSS 制作百分条图的过程如图 2-12 所示。

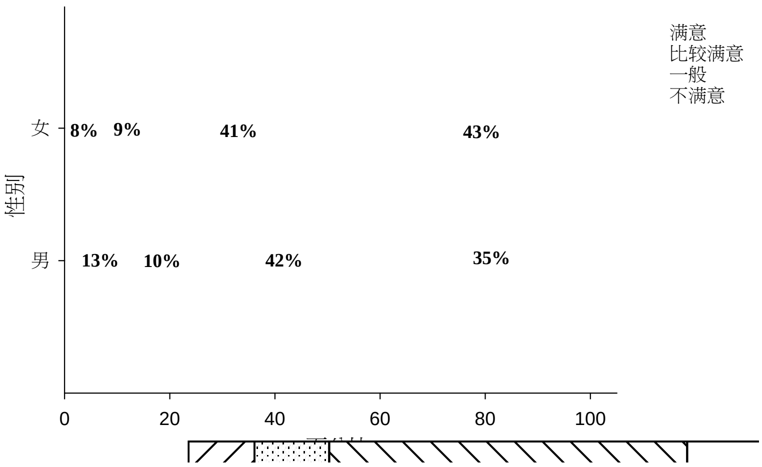
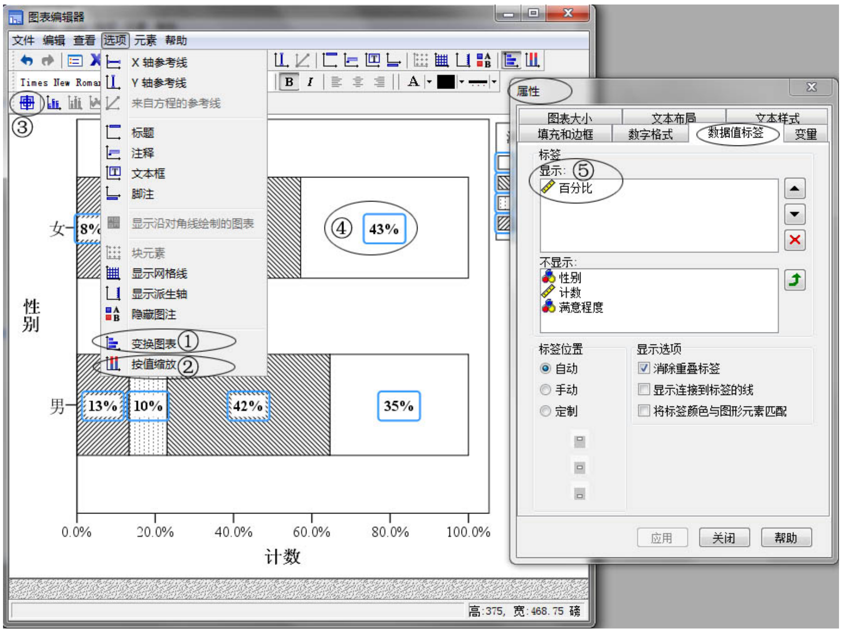


图 2-11　某市两男女市民对城市绿化分布格局的满意度构成比例示意图



（①在图 2-6 的基础上选择“变换图表”把分段条图变换为水平状；②选择“按值缩放”把两组数据全部变为“100%”；③点击“数据标签模式”，然后在每个类别上“盖章”，显示每个类别的“数值”；④如果类别“数值”不是百分比，则双击任一个类别“数值”，弹出“属性”窗口；⑤在“属性”窗口选择“数据值标签”，把百分比放入“显示”栏。）

图 2-12　SPSS 制作百分条图过程示意图

3.线形图

线形图（line graph）是以起伏的折线来表示某种事物的发展变化及演变趋势的统计图。线形图适用于描述某种事物在时间序列上的变化趋势，也适用于描述一种事物随另一事物发展变化的趋势模式，还可适用于比较两组或多组数据资料之间的变化特征及其相互联系。它能形象生动地刻画各个统计事项之间的关系，既可对有关统计事项进行数量比较，又可看出其变化趋势。例如，从图 2-13 可以看出：从 2000-2008 年，某区新增教师人数总体呈逐年递增趋势；2004-2006 年，该区新增男教师人数超过新增女教师人数。根据图 2-14 所示 SPSS 制作线形图的过程（数据为本书所附光盘中例 02-2.sav），再作必要编辑即可得到图 2-13 的线形图。

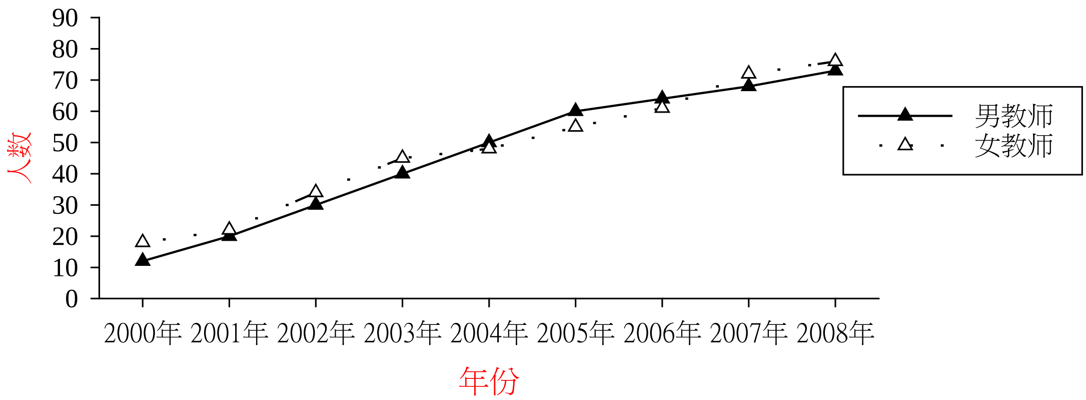


图 2-13 某区 2000-2008 年新增老师人数变化趋势图

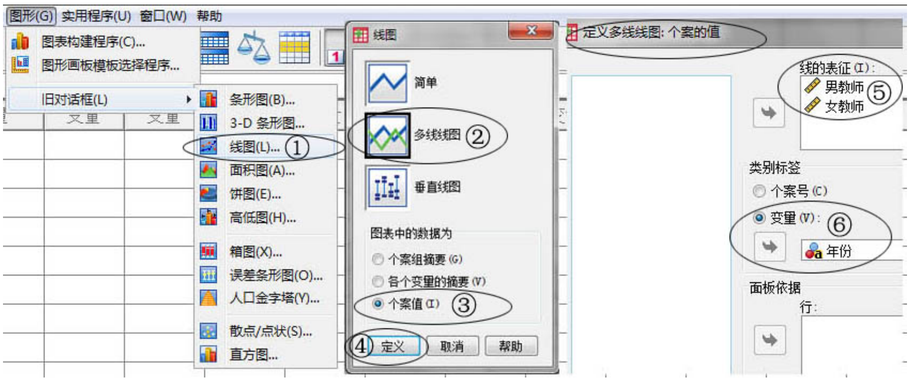


图 2-14　SPSS 制作线形过程示意图

4.散点图

散点图（scatter plot）是用平面直角坐标系上点的散布图形来表示两种事物之间的相互关系及联系模式。散点图适用于描述两个变量的观测数据，在心理学研究中应用广泛。它对于探究两种事物、两种现象之间的关系起着重要作用。我们可以根据散点图中点群的分布情况，结合自己的专业知识和统计学素养，推测两种事物或现象之间的相关程度与联系模式，并进一步采用有关统计方法进行深化研究。这方面的具体内容将在本书第三章第三节相关分析中进一步介绍。

绘制散点图的注意事项：（1）在平面直角坐标系中，横轴一般代表自变量，纵轴一般代表因变量。横轴既可是连续变量的量尺也可是离散变量的量尺，但纵轴一般代表连续变量的量尺；（2）点的描绘依两个变量的观测数据而定，但在具体描绘时应注意用细线画坐标轴，用稍粗黑点描绘各个坐标点，点的位置的确定按平面解析几何学中的方法；（3）根据数值的大小合理划分坐标标值，为了保持标值刻度的均匀，从坐标原点到最小值之间可用“//”断开（参见图 2-15）。

根据图 2-16 所示 SPSS 制作散点图的过程（数据为本书所附光盘中例 02-3.sav），再作必要编辑即可得到图 2-15 的散点图。

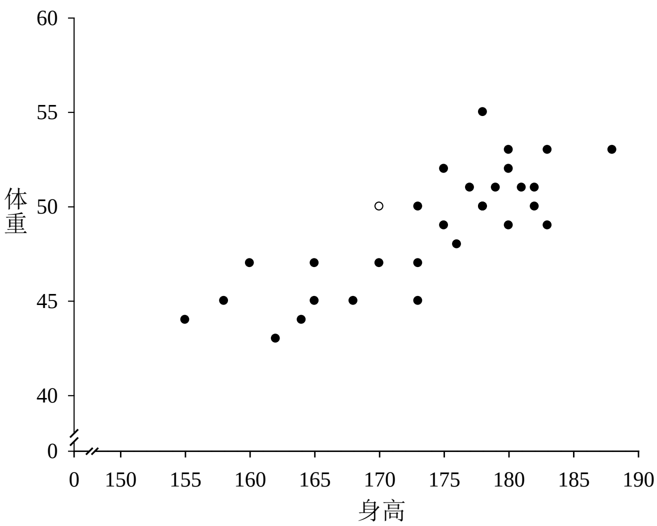


图 2-15　28 名被试的身高与体重散点图

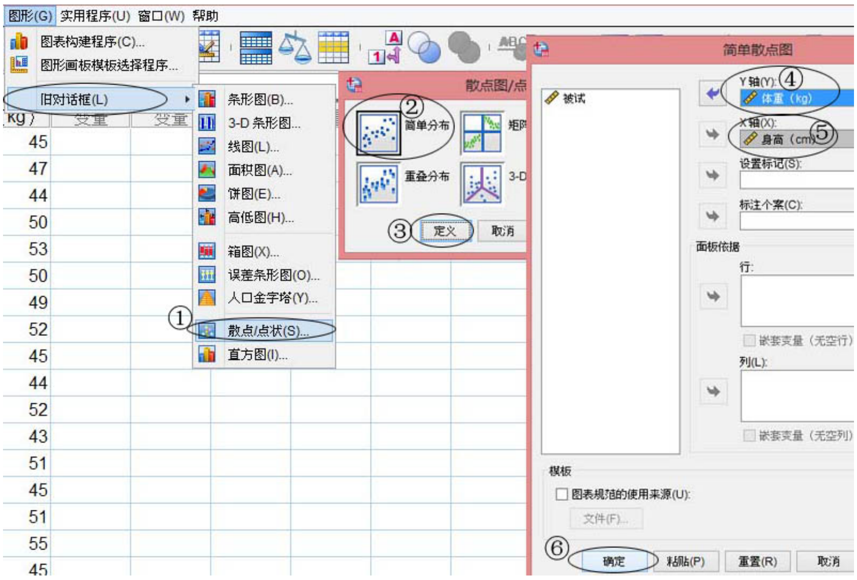


图 2-16　SPSS 制作散点过程示意图

以上几种统计图形，在运用时究竟选择哪种图形，要针对表现的对象，充分发挥各种图的优势，根据需要选用。一般而言，条形图、圆形图和百分条图可用于表现内容；线形图多用于表现变化；条形图、圆形图、百分条图和线性图都可用于比较；散点图一般用于表现变量之间的相互关系。

第二节　次数分布表与次数分布图

次数分布指的是一批数据中各个不同数值所出现的次数情况，或者是指一批数据在量尺上各等距区组内所出现的次数情况。根据它次数的产生方式，次数分布一般可区分为简单次数分布、相对次数分布、累积次数分布、累积相对次数分布等。将一组数据的次数分布情况用规范的表格形式加以体现，这就是次数分布表，若用图形来表达，那就叫次数分布图。

前面我们讲到的条形图、线形图、圆形图主要用于表示离散性变量的观测数据，散点图主要用于两个变量的观测数据之间相关性的探索，而次数分布主要适用于描述单一连续变量的观测数据。据此编制的次数分布表或图，是对数据进行初步整理的结果，有助于了解大量数据的分布情况，不仅是用少量数字有效地概括了大量原始数据，还可以节约呈现数据的时间。编制适用的次数分布表或图，可以为统计计算奠定重要的基础。

一、次数分布表

（一）简单次数分布表

简单次数分布表，通常简称为次数分布表，其实质是反映一批数据在各等距区组内的次数分布情况。下面我们以本章开头的第 2 个问题为例，具体介绍简单次数分布表的编制步骤（数据见本书所附光盘中例 02-4.sav）。

1.简单次数分布表的编制步骤

（1）求全距

全距，也称两极差，用符号 R 表示，是指一批数据中最大值（Max）与最小值（Min）之间的差距。其计算公式为：

$$R = Max - Min \qquad \text{（公式 2-1）}$$

求全距首先找出其中的最大值和最小值，根据公式计算出全距 R。显然在本例中，$R = 99 - 58 = 41$。

（2）定组数

定组数就是要确定把整批数据划分为多少个等距的区组。组数用符号 K 表示。组数的多少要根据数据的多少来确定。组数多，则要求总的数据量大，否则会出现有的组距内无次数分布的现象，使整个数据的分布规律显示不明显，也就会削弱对数据分组整理的功用了；组数少，则会引起较大的计算误差，会埋没数据内含的重要信息。

一般来说，当一批数据的个数在 200 个以内时，组数可取 8-18 组。如果数据来自一个正态总体，则可利用下述经验公式来确定组数，即：

$$K = 1.87(N-1)^{\frac{2}{5}} \qquad \text{（公式 2-2）}$$

公式中，N 为数据个数，在本例中，$N = 40$，根据公式 2-2 计算后取整，得 $K = 8$。

需要指出的是，根据公式计算或事先计划得到的组数可能与实际分组时因考虑组距取整及最低一组的起点位置不同而略有差异，这种差异是正常的，最终结果应以实际划归的组数为准。

（3）定组距

组距是指各组的起点值和终点值之间的距离，用符号 i 表示，确定组距的一般原则是取奇数或 5 的倍数，如 1，3，5，7……等等。若已知全距和组数，则具体的组距可以通过全距 R 与组数 K 的比值来取整确定。在例 2-2 中，$R/K = 41/8 = 5.125$，故可以把组距 i 定为整数 5。

（4）列出组限（分组区间）

组限是每个组的起点值和终点值之间的界限，又叫分组区间。起点值称为组下限，终点值称组上限。在以往的心理统计学文献中，组限的表述方法不尽一致，易引起混淆与误解。表 2-6 中列出的是关于组限的几种不同表达方式。

表 2-6　组限的三种表述方法（$i = 5$）

方法 组中值	（一）	（二）	（三）
22	20～24	[19.5, 24.5)	20～
17	15～19	[14.5, 19.5)	15～
12	10～14	[9.5, 14.5)	10～

对于连续变量来讲，尽管表 2-6 中的三种表述方法形式不同，但它们所包含的意义却是一致的。

组限有表述组限和实际组限两种，这里要注意区别。如表 2-6 中，“15～19”这一表述组限，其真正的实际组限为一个左闭右开的区间：[14.5,19.5)。其中左端点 14.5 称为该组的实下限，右端点 19.5，成为该组的实上限。在“15～19”这一组限中，包含了“实下限”这个点，但不包含“实上限”这一点。

在呈现表格时，各分组区间使用表述组限，并通常采用整数形式，即表 2-6 中“一”这一列的形式；为了书写方便，看起来更简明，也可以采用表 2-6 中“三)”这一列的形式。但在据此对数据进行归类划记时，一定要按照实际组限进行归类。例如，数据 19.5 应划归到 20～24 这组，而不是 15～19 这组。

（5）求组中值

组中值是各组的组中点在量尺上的数值，其计算公式为：

组中值=（组实上限+组实下限）/2　　　　　　（公式 2-3）

例如，表 2-6 中，方法（一）中“15～19”这一组限，其实上限为 19.5，实下限为 14.5，故该组的组中值为 17[（19.5+14.5）/2]。

（6）归类划记

完成上述步骤后，我们就可以设计一个表格，将原始数据准确地归到其所属的组限。一般采用画线计数或写“正”字的方法。为确保登记准确，~~第一次~~登记后需进行核实~~一遍~~。

（7）登记次数

根据归类画记的结果，计算各组的次数 f 。并核对各组次数总和与数据的总个数是否相等。

根据上述步骤，我们将制成表 2-7，表中包含了较多内容，目的是方便编制后面一些其他形式的次数分布表。表 2-7 中的第一栏、第二栏和第四栏拼在一起，就构成本例中~~所指的~~“40 名学生某项测验成绩的简单次数分布表”。

表 2-7 40 名学生某项测验成绩的简单次数分布表

组别	组中值	划记	次数（f）
（1）	（2）	（3）	（4）
95～99	97	II	2
90～94	92	III	3
85～89	87	卌 IIII	9
80～84	82	卌 卌	10
75～79	77	卌 III	8
70～74	72	IIII	4
65～69	67	I	1
60～64	62	II	2
55～59	57	I	1
$\sum$			$N=40$

2.简单次数分布表的意义与缺点

编制后的简单次数分布表，可将一堆杂乱无序的数据排列成序。从表中可以直接发现各个数据的出现次数是多少，其分布的状况如何。由表 2-7 可知，处在区间[79.5,84.5)的人数最多，90 分以上及 70 分以下的人数较少。

简单次数分布表也有缺点。从表 2-7 看，原始数据不见了，只有各分组区间及各组的次数。根据这样的统计表计算得到的平均值，会与用原始数据计算的值有一定的出入。因为简单次数分布表假设各区间的数据分布是均匀的，采用各组的组中值代表原始数据进行计算。但实际上，原始数据在各区间的分布并不一定是均匀的。因此两者计算得到的平均值就存在误差。我们将这个误差称作"归组效应"。

同一组数据，分组的组距越大，分组数目越少，引起的误差越大，反之则越小。不过根据上述我们提到的组距及组数的确定步骤可知，在编制次数分布表时，组距和组数的确定不是没有任何约束的。因此，就一组数据而言，组距的变化引起的误差不会太大，对进一步的统计分析，不会带来需要注意的影响。一般而言，任一组观测数据被分为 10～15 组，全部信息就基本被保留下来了。

（二）相对次数分布表

相对次数就是各组的次数 f 与总次数 N 之间的比值，若以 R_f 表示相对次数，则 $R_f = f/N$ 。表 2-8 中第五栏的数值就是各组的相对次数。表 2-8 第一栏、第二栏和第四栏拼在一起就构成了一个相对次数分布表。相对次数较大的组，说明落入该组内的数据个数占全部数据个数的比例越高，反之，则越低。

表 2-8　40 名学生某项测验成绩的分布统计结果

组别	组中值	次数（f）	相对次数	累积次数	累积相对次数	累积百分数
（1）	（2）	（3）	（4）	（5）	（6）	（7）
95～99	97	2	0.05	40	1.00	100
90～94	92	3	0.075	38	0.95	95
85～89	87	9	0.225	35	0.875	87.5
80～84	82	10	0.25	26	0.65	65
75～79	77	8	0.2	16	0.40	40
70～74	72	4	0.1	8	0.20	20
65～69	67	1	0.025	4	0.10	10
60～64	62	2	0.05	3	0.075	7.5
55～59	57	1	0.025	1	0.025	2.5
$\sum$		$N=40$	1.00			

相对次数分布表主要是反映各组数据的百分比结构。相对次数分布表和简单次数分布表各有不同的用途，它们既可联合使用，也可单独使用。当我们主要对各组的绝对次数感兴趣时，可编制简单次数分布表。当我们侧重各组次数的相对比例结构时，通常要编制相对次数分布表。

（三）累积次数分布表

若我们想知道位于某个数值以下或以上的数据个数有多少时，则要在简单次数分布的基础上，计算累积次数，制作累积次数分布表。累积次数是把各组的次数由下往上或由上往下累加在一起得到的，最后一组的累加次数应等于数据的总次数。相应的，累积次数分布表可分为“以下（向上)”累积次数分布表与“以上（向下)”累积次数分布表。

“以下（向上)”累积次数分布表，其目的在于反映位于某个分数“以下”的累积次数共有多少，在编制时，是从表 2-8 中第三列下面最低组往最高组方向依次向上累加次数。“以上（向下）”累积次数分布表，其目的在于反映位于某个分数“以上”的累积次数共有多少，在编制时，是从表 2-8 中第三列上面最高组往最低组方向依次向下累加次数。例如，将表 2-8 的第一栏、第二栏、第五栏拼在一起，就构成一个以下（向上）累积次数分布表。由此可以看出，90 分以下的有 38 人，75 分以下的 16 人等等。

（四）累积相对次数分布表和累积百分数分布表

累积次数分布表是对简单次数进行累积的结果，若我们对相对次数进行累积，得到的结果则是累积相对次数分布表。累积相对次数仍是小数，但累加到最后一组的结果必然为 1。将表 2-8 中第一栏、第二栏和第六栏拼在一起，就构成一个累积相对次数分布表。

将得到的累积相对次数乘上 100，则得到“百分数”，从而可将“累积相对次数分布表”转化为“累积百分数分布表”。把表 2-8 中的第一栏、第二栏和第七栏拼在一起，就构成一个累积百分数分布表。

应指出的是，累积相对次数分布表和累积百分数分布表同样有“以上”和“以下”两种。在具体应用时，应根据情况选择其中一种即可。

二、次数分布图

为了更直观、形象地表达一个次数分布的结构形态及特征，我们可进一步地从次数分布表出发，绘制相应的次数分布图。

（一）简单次数分布图

1. 次数直方图

次数直方图（frequency histogram）是由若干宽度相等、高度不一的直方长条紧密排列在同一基线上构成的图形。一般用直角坐标系中的纵轴表示数据的次数 f，横轴表示数据的等距分组点，即各分组区间的上下限，有时用组中值表示。纵轴的刻度通常从零开始，横轴的刻度可以从任何合适的数字开始，但应与数据的分布范围和组距有关。直方长条的宽度直接受组距的大小影响，底边的两端点分别为精确上下限，直方长条的高度由每组的次数决定。同时，为了标明各组次

数的大小，可在各矩形顶端标上具体的次数数值。

次数直方图是以矩形的面积表示连续性随机变量次数的分布图。直方图下的面积与总次数相等，所以一个矩形的面积大小与每组的次数分布大小是等价的。如果将总面积定为 1，那么，直方图中每一部分矩形的面积就是该直方长条表示的分组内的次数与总次数的比值。图 2-17 便是根据表 2-8 中的简单次数分布，按上述规则绘制成的次数直方图。

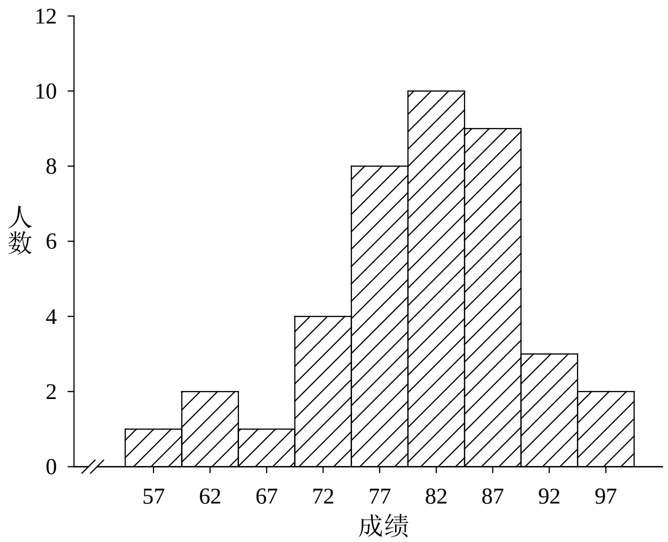


图 2-17 40 名学生某测验成绩的次数直方图

前面介绍的条形图似乎与次数直方图在形式上有相似之处，但实际上，两者是有很大区别的：（1）描述的变量类型不同：条形图用于描述离散变量，如称名数据或计数数据，因此不同类别的直方长条之间是有间隔；次数直方图主要用来描述分组的连续变量，故各个直条是紧密相接的。（2）表示数据多少的方式不同：条形图用直条的长短表示数据的多少或大小，直条长条的宽度没有意义；直方图用矩形的面积表示数据的多少，宽度表示的是组距，因此，某矩形面积与直方图总面积之比就等于该分组区间内的次数与总次数的比值。

2. 次数多边图

次数多边图（frequency polygon）是利用闭合的折线构成多边形以反映次数变化情况的一种图形。直角坐标系中的纵坐标表示数据的次数 f，横坐标表示每个组的组中值。以每个组的次数为纵坐标标点，以相应的组中值为横坐标标点，画出各个点并连接起来，就成为一条折线。为构成闭合多边形和计算面积，可将折线两端分别增加前一组和后一组的组中值点。

次数多边图是以闭合多边形的面积表示连续性随机变量次数的分布。与次数直方图相比，其轮廓显得更好，组与组之间的次数过渡是连续而直接的。从理论

上讲，当一批数据个数足够大时，随着分组间距的不断变小，绘制成的次数多边图就会越来越连续和光滑。若分为无数组时，就形成一条极其光滑的曲线，这种曲线在统计学上称为次数分布曲线。在后面的学习内容中，我们会接触到诸如正态分布、t 分布、F 分布、χ^2 分布等概率分布曲线。另外，次数多边图还可用于多个同质的次数分布的比较，这一点我们将在相对次数分布图中说明。图 2-18 便是根据表 2-8 中的简单次数分布，按上述规则绘制成的次数多边图。

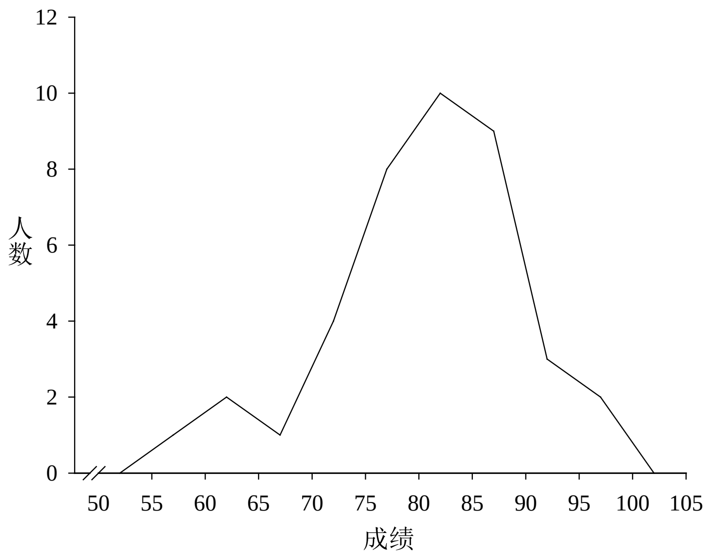


图 2-18　40 名学生某测验成绩的次数多边图

（二）相对次数分布图

相对次数分布图可分为相对次数直方图和相对次数多边图两种。相对次数分布图与简单次数分布图类绘制方法大致相同。不同之处是，简单次数分布图的纵轴表示的是次数 f，相对次数分布图的纵轴表示的是相对次数 R_f。

根据相对次数多边图的特点，可用其表示多批不同数据的次数分布。虽然各次数分布的总次数不等，但可将其用相对次数表示，注意要保证各次数分布的组距相同。这样，就可在同一个图中，用不同的图注来表示各个分布。

（三）累积次数分布图

累积次数分布图也可分为累积次数直方图和累积次数曲线图两种。比较常用的是累积次数曲线图，它是根据累积次数分布制作而成的。其绘制方法同次数多边图基本相同，横轴意义不变，纵轴表示累积次数。

“以下（向上）累积次数曲线图”呈上升趋势，“以上（向下）累积次数曲线图”呈下降趋势。这里称它为曲线图而不是多边图，主要是因为累积次数曲线图的形状不会由于组距的不同而发生较大的变化，因而从抽样数据制成的累积次数曲线图是比较稳定的。图 2-19 便是根据表 2-8 中的累积次数分布，按上述规则绘制成的累积次数直方图和曲线图。

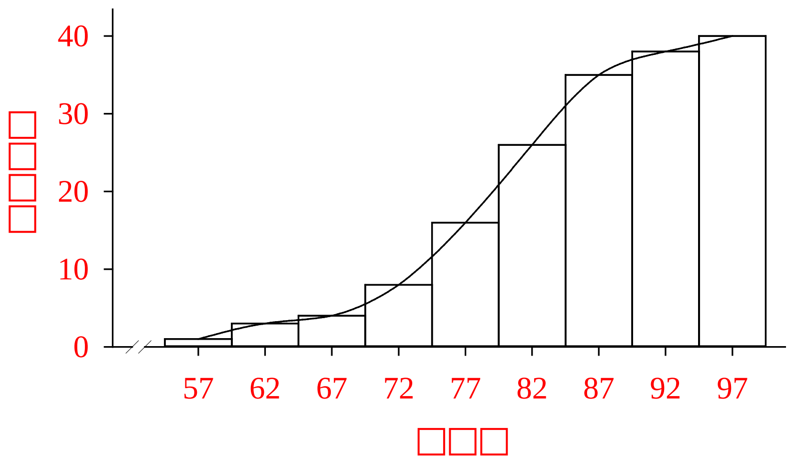


图 2-19　40 名学生某测验成绩的累积次数直方图与曲线图

（四）累积相对次数曲线图和累积百分数曲线图

根据累积相对次数分布以及累积百分数分布，可相应地绘制累积相对次数曲线图和累积百分数曲线图。它们的制作方法大体上与上述的累积次数曲线图相同，不同之处是纵坐标表示的是累积相对次数或累积百分数。

三、用 SPSS 制作次数分布表与次数分布图

（一）用 SPSS 制作次数分布表

SPSS 制作次数分布表，首先需要把测验成绩分组（参见图 2-20、图 2-21），然后通过描述统计的频率分析（参见图 2-22）即可得到简单次数、相对次数和累积百分数分布表（见表），在此结果基础上稍加计算即可得到累积次数分布表。其中，SPSS 输出结果中的“相对次数”包括“百分比”和“有效百分比”两列，二者的区别在于是否剔除了缺失数据，“有效百分比”是剔除了缺失数据后得出的各组的相对次数百分比，而“百分比”把“缺失数据”也作为一个类别进行统计。

图 2-20　SPSS 数据分组过程（可视离散化）示意图

（①输入对成绩进行分组后的分组变量的变量名，如“成绩分组”；②输入变量的名称标签，不要变量名称标签也可以；③选择上端点“包含”，即例如 59 属于“55-59”这一组；如选择“排除”，59 则属于“60-64”这一组了；④点击“生成分割点”弹出图中右侧对话框；⑤输入第一个分割点的位置，即输入第一组的上限；⑥输入“宽度”，即输入组限；⑦点击“分割点数量”的空白处，程序自动生成分割点数量；⑧点击返回主对话框；⑨点击“生成标签”，生成诸如“55-59”的标签。）

图 2-21　SPSS 数据分组过程（可视离散化）示意图续

（说明：第④步点击“格式”，在弹出对话框中选择“按值的降序排序”是为了次数分布表中各组排序由下到上的顺序是分数由低到高，反之选择则是由高到低了。）

图 2-22　SPSS 制作次数分布表示意图

表 2-9　SPSS 频次分析结果

组别	频率	百分比	有效百分比	累积百分比
95～99	2	5.0	5.0	5.0
90～94	3	7.5	7.5	12.5
85～89	9	22.5	22.5	35.0
80～84	10	25.0	25.0	60.0
75～79	8	20.0	20.0	80.0
70～74	4	10.0	10.0	90.0
65～69	1	2.5	2.5	92.5
60～64	2	5.0	5.0	97.5
55～59	1	2.5	2.5	100.0
合计	40	100.0	100.0	

（二）用 SPSS 制作次数分布图

SPSS 制作次数分图，即可在原始数据基础上制作（数据见本书所附光盘中例 02-4.sav），也可用整理后的数据（已编制成次数分布表的数据，见本书所附近光盘中例 02-4-2. sav），二者的过程基本相同。在此，以原始数据制作累积次数直方图与曲线图为例，来说明如何用 SPSS 制作次数分布图。

根据图 2-23 的过程可制作出组中值对应的累各次数“条形图”，如要得到图 2-19 一样的图形，还必须在 SPSS 输出窗口双击“条形图”调出“图形编辑器”窗口，作必要设置：（1）根据图 2-24 步骤把“条形图”设置为“直方图”；（2）根据图 2-25 步骤把“曲线”添加到图中。

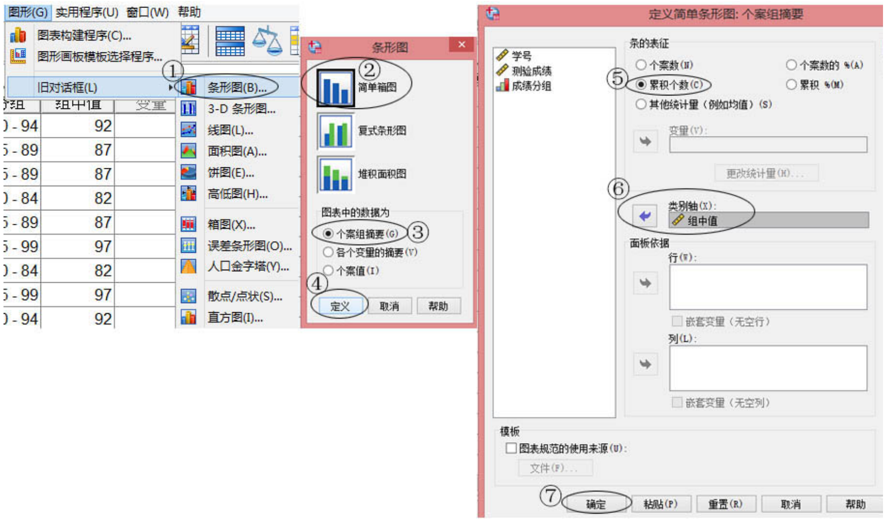


图 2-23　SPSS 制作累积次数直方图和曲线图示意图

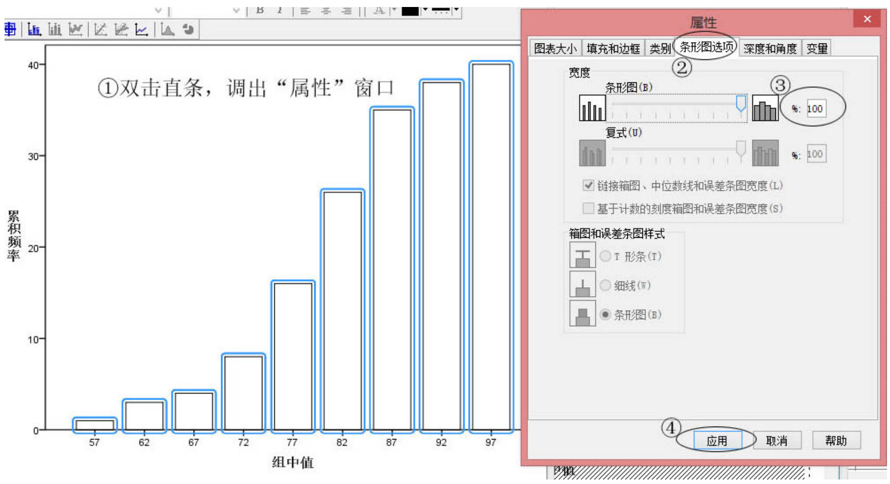


图 2-24　SPSS 制作累积次数直方图和曲线图示意图续一

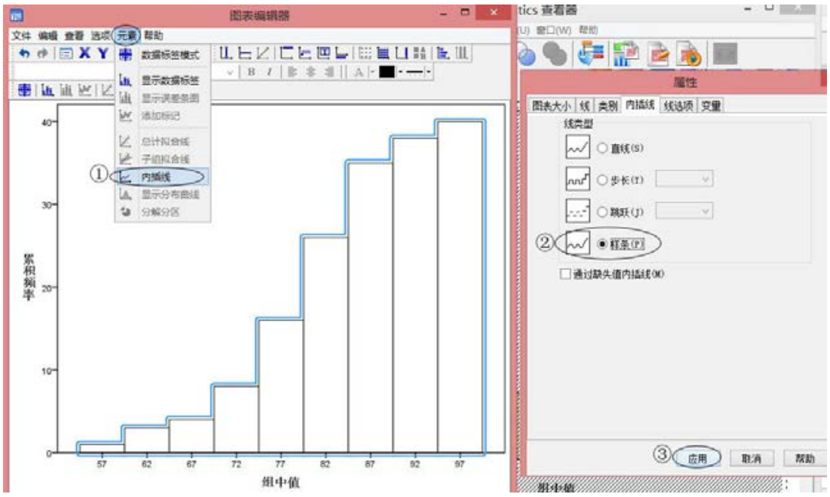


图 2-25　SPSS 制作累积次数直方图和曲线图示意图续二

第三节　探索性数据分析与茎叶图

一、探索性数据分析的意义

1977 年，美国统计学家普林斯顿大学的约翰·图基(John W. Tukey)出版了《探索性数据分析》一书，引起了统计学界的关注，成为关于探索性数据分析(Exploratory data analysis ,EDA)的第一个正式出版物，他通过箱形图（Box plot）等简单方法，指出了统计建模应该结合数据的真实分布情况，对数据的分析，不应该从理论分布假定出发去构建模型，而应该从数据的特征出发去研究和发现数据中有用的信息。探索性数据重新提出了描述统计在数据分析中的重要性，这种简单而直观的方法在理解数据方面是极为有用的。探索性数据更深刻的意义则在于，它为统计学指明了新的发展方向—和数据相结合。

实际上早在图基的《探索性数据分析》问世之前，西方几乎所有从事统计分析和研究的人，在分析某一组具体数据或资料前，都常常对数据作一些预先考察，这些考察实际就是探索分析。与之相对应的证实性分析(如假设检验等)则是后来才产生的。图基将探索分析方法具体化、系统化、规范化，并明确赋予它们以“探索性数据分析”这一名称。具体讲，探索性数据分析是对调查、观察所得到的一些初步的杂乱无章的数据，在尽量少的先验假定下进行处理，通过作图、制表等形式和方程拟合、计算某些特征量等手段，探索数据的结构和规律的一种数据分析方法。它连同高速计算机的广泛应用，带来了数据审查、分析和压缩方法上的新潮流(包括新的制图和制表方法)。探索性数据分析，被认为对 20 世纪统计学的发展产生了重要的影响。

从目前的使用来看，探索性数据分析只是正式建立统计分析模型之前的铺垫，是统计分析的一个环节，其至少有如下三方面的作用：(1) 发现数据中的一些错误，确保正式分析时是排除错误后的数据；(2) 以考察数据是否满足正式分析的条件；(3) 发

现一些新的信息。一般来说，仅仅使用探索方法并不能完成独立全面的统计分析，尤其是在统计方法和技术如此发达的今天。但探索分析作为完整、科学的统计分析过程中的一个环节是不可或缺的。将探索分析的思想运用于统计分析，将使统计分析更加深刻，并引导统计分析走向科学化。

二、茎叶图

适用于探索性数据分析的统计图主要有茎叶图和箱形图，在本节仅介绍茎叶图，对于条形图将在第三章第二节差异量数那一部分进行介绍。

茎叶图（**Stem-and-leaf diagrams**）是由图基提出的一种初步整理数据的方法。它的作用与次数分布表大致相同。与直方图相比，当样本数载 100 以下时，选择茎叶图来描述数据的分布或结构是一种非常有用和更有效的方法。前面我们已经知道，当原始数据转换成次数时，在次数分布表或直方图中原始数据被完全丢失，而茎叶图则避免了这种情况，它不仅能够完整地展示数据的全貌，同时保留了原始数据。我们以表 2-10 所示某班学生心理统计学的考试成绩为例，来看看如何用茎叶图表示。采用 SPSS 制作茎叶图如图 2-26 所示（数据见本书所附近光盘中例 02-5.sav）。

表 2-10 心理统计学期末考试成绩（N=70）

95	57	76	93	86	80	89	76	76	63	74	94	96	77	65	79	60	56	72	
82	70	67	79	71	77	52	76	68	72	88	84	70	83	93	76	82	96	87	69
89	77	81	87	65	77	72	56	78	78	58	54	82	82	66	73	79	86	81	63
46	62	99	93	82	92	75	76	90	74	67									

```
统计学成绩 Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     1.00 Extremes    (=<46)
     2.00        5 .  24
     4.00        5 .  6678
     4.00        6 .  0233
     7.00        6 .  5567789
     9.00        7 .  001222344
    16.00        7 .  5666666777788999
    10.00        8 .  0112222234
     7.00        8 .  6677899
     6.00        9 .  023334
     4.00        9 .  5669

 Stem width:    10
 Each leaf:       1 case(s)
```

图 2-26 心理统计学期末考试成绩茎叶图（N=70，茎宽=10）

我们看到图 2-26 是一个结构简单清楚地图形。茎叶图的结构由三个部分组成：左边一列是次数(Frequency)，中间一列称为茎（stem)，右边一列为叶（leaf)。数据表中的每个数据都可分为两个部分两表示，“茎”+“叶”。例如，数据表中第一个数 95 在茎叶图中被表示为：

茎	叶
9	5

因为茎的宽度（Stem width）等于 10，茎等于 9，实际表示的是 90，因此 95 这个数据分两个部分表示，就是“9×10+5”。图中中间一列的茎用来表示高位数部分，根据数据的大小，可以是 1 位，2 位，等等；叶表示数据的最后一位。如果是一个 3 位数，则茎表示前 2 位，叶表示个位。~~还可以引入……=茎 x 茎宽+叶~~在茎叶图中，茎可以重复出现。如图 2－26 中，中间列的茎 5，6，7，8，9 都分别出现两次。如果数据较多，一个茎可出现 3 次，4 次或更多。茎的多次重复将改变茎叶图的现状，使其向垂直方向延长，有较长的茎。如果每个茎只出现一次，则茎较短，叶较宽，如果我们将图 2-26 改为每个茎只出现一次，其茎叶图如图 2-27 所示。

4	6
5	2 4 6 6 7 8
6	0 2 3 3 5 5 6 7 7 8 9
7	0 0 1 2 2 2 3 4 4 5 6 6 6 6 6 6 7 7 7 7 8 8 9 9 9
8	0 1 1 2 2 2 2 2 3 4 6 6 7 7 8 9 9
9	0 2 3 3 3 4 5 6 6 9

图 2-27　心理统计学期末考试成绩茎叶图（N=70，茎宽=10）

茎叶图中，每个叶代表了一个数据（Each leaf:　1 case(s)），在实际工作中，数据量较大的情况下，可以以每个叶代表多个数据。每个茎之后叶的多少乘上第个叶代表的数据个数，即为每个茎中数据的资料。例如本例中“9”这个茎，第一个“9”茎的次数为“6×1=6”，第一个“9”茎的次数为“4×1=4”。另外，从茎叶图中还可以看出数据中是否存在“极端值”，例如在本例中，图 2-26 显示有一个极端数据（Extremes　(=<46)）。

如果我们将茎叶图逆时针旋转 90°，将茎作为横坐标，我们会发现它的分布于直方图非常相似。只是直方图丢失了原始数据，而茎叶图将它们保留下来。

三、用 SPSS 制作茎叶图

用 SPSS 制作茎叶图过程十分简单（如图 2-28 所示），按图示步骤即可制作出图 2-26 一样的茎叶图。

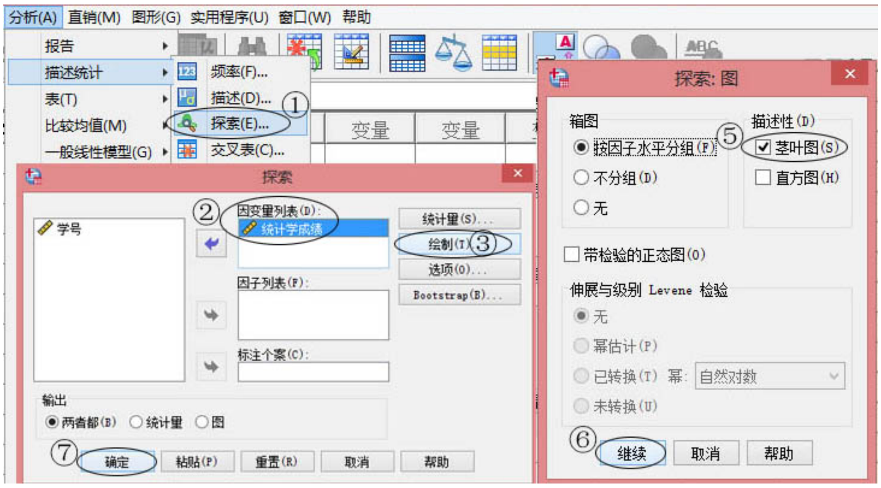


图 2-28　SPSS 制作茎叶图过程示意图

小 结

在分析心理学研究数据之前，首先需要对数据进行初步整理，常用的整理数据的方法有排序、分组、制作统计图和统计表以及计算描述统计量数等。希望读者掌握各种数据整理的方法，尤其是各种统计表与统计图的含义、其适用条件以及具体的画法，要能读懂各种统计表与图所传达的信息。

关键术语

数据的排序：是按照某种标准，对事先收集到的杂乱无章的数据按照一定顺序标准（升序或降序）进行排列的过程。

数据的统计分组：是根据被研究对象的特征，将所得数据划分到各个类别中去的数据整理过程。

统计表：是用表格的形式来表达数据或事物分类等内容的一种有效表达方式。

三线表：通常是指只有三条横线，即顶线、底线和栏目线，没有竖线的一种表格。

统计图：就是依据数字资料，应用点、线、面、体、色彩等制成的整齐而有规律，简明且知其数量的图形。

条形图：是用宽度相同的直方长条的高度来表示各个统计事项之间的数量大小及差异情况的统计图。

圆形图：是以单位圆内各扇形面积所占整个圆形面积的百分比来表示各统计事项在其总体中所占相应比例的一种图示方法。

百分条图：是以直条内部各部分面积的大小表示事物内部各组成部分所占的百分构成比的条形图。

线形图：是以起伏的折线来表示统计事项的发展变化及演变趋势的统计图。

次数分布：是指一批数据中各个不同数值出现的次数情况，或者是一批数据在量尺上各等距区组内出现的次数情况。

探索性数据分析：是对调查、观察所得到的一些初步的数据，在尽量少的先验假定下进行处理，通过作图、制表等形式和方程拟合、计算某些特征量等手段，探索数据的结构和规律的一种数据分析方法。

茎叶图：一种同时排列数量数据顺序并提供有关分布形态的深入信息的探索性数据分析技术。

思考与练习

1.简述三线表的基本组成要素。

2.简述条形图、线形图、饼图和散点图的含义、适用情况和绘制方法。

3.简述次数分布表的意义及缺点。

4.根据例 2-2 制作“以上”累积次数分布表。

5.根据例 2-2 绘制“以上”累积次数分布图。

6.用 SPSS 软件，模拟数据，绘制条形图、线形图、饼图和散点图。

第三章　数据特征的描述

学习本章内容，将有助于你对以下问题的理解与思考：

1. 在一组数据中，如何寻求一个最能代表集中趋势的数据？
2. 在一组数据中，如何寻求一个最能代表离中趋势的数据？
3. 箱形图在描述统计中有何作用？
4. 如何利用标准分数来确定每个数据在该组数据中的相对地位？
5. 如何对两列变量之间的关系进行度量和分析？

1. 在第二章中介绍了用统计表或统计图的方法来描述 40 名学生测验成绩的特点，除统计表、图之外，能否通过报告一些统计指标来描述 40 名学生测验成绩的特点？一组数据的特点常常可以从哪些方面来进行描述？

2. 已知某班期末考试语文平均分为 80，标准差为 10；数学平均分为 70，标准差为 15；英语平均分为 85，标准差为 12。甲生的语文成绩为 85，数学成绩为 81，英语成绩为 90 分。乙生的语文成绩为 88，数学成绩为 85，英语成绩为 83 分。问：这三门功课的总成绩，哪个学生较好？

第一节　集中趋势的度量：集中量数

前面我们谈到了次数分布表与次数分布图的制作。观测数据的次数分布，一般至少有两个方面的特征--一是一组数据的大量数据向某个点集中的程度即集中趋势，二是一组数据彼此离散的程度即离中趋势~~如何~~，这两个特征决定了一批数据的分布形态。用来描述数据的集中趋势的特征量数称为集中量数（meature of central tendency）；用来描述数据的离中趋势的特征量数称为差异量数（meature of dispresion tendency）。本节我们先介绍集中量数。集中量数包括算术平均数、中数、众数等。集中量数反映一批数据的典型水平，是一批数据的典型代表值。

一、算术平均数

（一）算术平均数的计算方法

n 个观测数据 $x_1, x_2 \ldots, x_n$ 的算术平均数（mean），就是所有观测数据的总和除以数据个数所得的商。常见的算术平均数计算公式之一是：

$$\overline{X} = \frac{(x_1 + x_2 + \cdots + x_n)}{n} \qquad \text{（公式 3-1）}$$

简记为：

$$\overline{X} = \frac{\sum_{i=1}^{N} x_i}{n} \qquad \text{（公式 3-2）}$$

在上式中，Σ 为求和符号，i 为观测值的下标，表示观测值的序号，$i = 1,2,\cdots$ N，$\sum_{i=1}^{n} x_i$ 表示从第一个观测值 x_1 开始，一直到第 N 个观测值 x_n 为止的所有数据的总和。在使用连加和符号 Σ 时，如果没有写下标和起迄点，是表示所有数据的连加和。如，$\sum X$ 即指所有数据的连加和。

【例 3-1】 第二章开头的第 2 个问题中 40 名学生的平均测验成绩是多少？

解：将这些数据代入公式(3－1)得：

$$\overline{X}=\frac{(x_1+x_2+\cdots+x_n)}{n}=\frac{(94+89+\cdots+79)}{40}=80.98$$

答：这 40 名学生的平均测验成绩是 80.98。

（二）算术平均数的性质

算术平均数具有如下性质：

（1）数据组全部观测值的和等于观测值的个数乘以观测值的平均数，即 $n\overline{X}=\sum X$

（2）对于一组数据来说，各数据的离差（每个数据与平均数之差，又称离均差）之和等于零，即 $\sum(X-\overline{X})=0$。

（3）对于一组数据来说，各数据与该组数据平均数的离差平方和为最小，即 $\sum(X-\overline{X})^2=$ 最小。

（4）在一组数据中，每个数据都加上或减去、乘以或除以一个常数 C，所得平均数等于原来的平均数加上或减去、乘以或除以这个常数 C。

（三）算术平均数的优缺点

算术平均数是应用最普遍的一种集中量数，具有以下优点：

（1）能很灵敏地反应数据的变化。在一批数据中，只要有一个数据大小发生变化，算术平均数也会发生相应的变化。

（2）同一批数据的平均数可以唯一地严密地确定下来。

（3）计算简单，意义清晰明了。

（4）适合作进一步的代数运算。

正因为算术平均数具备这些优点，当数据呈正态分布或单峰对称分布时，一般都会采用算术平均数作为一组数据的集中量数。但是，算术平均数也有一些缺点：

（1）易受极端数值的影响。极端数据是一批数据中比绝大多数数据大很多或小很多的数据。数据中如果有极端数据，那么计算出来的算术平均数就会偏离数据的典型水平，不足以代表这批数据的集中趋势。所以，当数据呈偏态分布时，采用算术平均数也是不恰当的。

（2）要求数据完整而确切。因为计算算术平均数时，每个数据都参与运算，所以出现了大小模糊不清或不够确切的数据时，就无法计算算术平均数。

二、中数

中数（Median），又称中点数、中位数、中值，指的是在按大小顺序排列的一组数据中，位置居中的那个数，记为 M_d 或 M_{dn}。

（一）中数的计算方法

要先将数据由小到大或由大到小排序，排序后处于中间位置的数就是中数。

如果项数是奇数，中位数是数列中第（$n+1$）/2 项（即正中央的那一项）；如果项数是偶数，中位数是数列中第 $n/2$ 项和第（$n+1$）/2 项的平均数（正中央的两项的平均数）。

【例 3-2】 五名学生的语文成绩为 60，68，71，80，83，求这些成绩的中数。

解：由于位于中间的第 3 个数 71 两边各有两个数据，因此中数为 71。

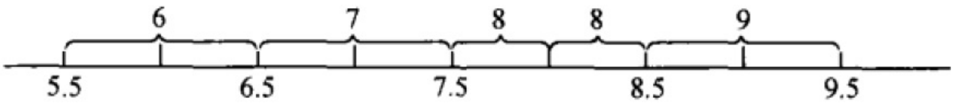
【例 3-3】 六名学生的语文成绩为 68，60，72，80，83，90，求这些成绩的中数。

解：先把成绩排序 60，68，72，80，83，90

$$M_{dn}=\frac{72+80}{2}=76$$

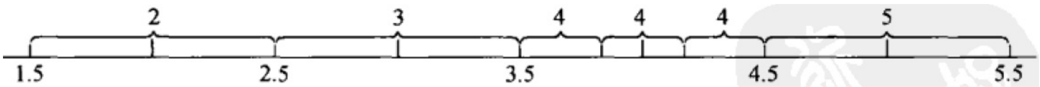
但是当数列的中间位置前后有相同数值时，求中数的方法就比较复杂了，这同样也分为数据个数 n 是奇数或偶数两种情况。

（1）数据个数是奇数，中数附件有相同的数值，例如：6，7，8，8，9。假设这些数据为连续变量，任何数据都代表一个区间（数的实限），则这五个数在数轴上就可表示为：



如上图所示，8 代表的数值区间为 7.5~8.5，但在这一区间有两个 8，而且数列的中间位置在第一个 8，所以中数就是 7.5~8.0 这个区间的代表值（即这一区间的中点）(7.5+8）/2=7.75。

（2）数据个数是偶数，中数附件有相同的数值，例如：2，3，4，4，4，5。假设这些数据为连续变量，任何数据都代表一个区间（数的实限），则这六个数在数轴上就可表示为：



如上图所示，4 代表的数值区间为 3.5 ~ 4.5，但在这一区间有三个 4，而且数列的中间位置在第一个 4 和第二个 4 的交界点，所以中数就是 3.83（将 3.5 ~ 4.5 三等分的第一个点）。

（二）中数的优缺点

由于中数是由数据按小大排序后位于中间位置的数据所决定，不是由每个数据都参与运算求得，因此不易受极端数据的影响。这也是中数比算术平均数更优良的地方。但是，也正因为不是每个数据都参与运算，因此，中数反应不灵敏。另外，中数也不适宜于作进一步的代数运算，例如，我们不能将几个中数综合求

出一个总的中数。由于以上原因，中数的运用没有算术平均数广泛。当一批数据中有极端数据，或者一批数据中有不确切的数据，亦或数据呈偏态分布时，用中数是合适的。

三、众数

（一）众数的计算方法

众数有粗略众数和理论众数两种定义方式。粗略众数是指一组数据中出现次数最多的那个数，可以通过观察法直接求得。理论众数是指与次数分布曲线最高点相对应的横坐标上的一点，因为计算复杂，因此在实际工作中较少使用。众数用符号 M_O 表示。

对于一批原始数据来说，出现次数最多的那个数就是这批数据的粗略众数。对于分组数据（次数分布表），次数最多组的组中值就是这批数据的粗略众数。

【例 3-4】 10 名学生教育学的成绩为 61，63，76，80，80，80，80，92，93，96，求其众数。

解：在这批数据中，80 出现的次数最多，因此众数为 80。

（二）众数的优缺点

众数具有简明易懂，较少受极端数据的影响的优点。但是，众数的求取不准确。在次数分布表中，如果组距发生变化，那么众数值也可能发生变化。众数由数据的次数决定，因此反应不灵敏。众数也不适合进一步的代数运算，不能将几个众数综合求一个总的众数。所以，一般情况下很少采用众数作为一组数据的集中量数。但在两种情况下，我们需采用众数：（1）当我们需要快速而且只要粗略地找出一批数据的代表值时，用众数是合适的；（2）当数据呈多峰分布时，可以同时报告多个峰顶对应的横坐标值。因此，在计算统计量描述数据特征的时候，常常需要先采用统计图表了解数据的分布特征，再决定计算哪个统计量。

第二节 变异程度的度量：差异量数

一、计算差异量数的意义

在很多情况下，我们关心的重点在集中量数，但数据的离散程度也是非常重要的。如果数据离散程度很大，测量值的大小参差不齐，集中趋势（集中量数）的意义就十分有限。另一方面，数据离散趋势的指标（差异量数）的指标本身也具有描述意义，它反映数据间差异的程度。我们先来看表 3-1 所列的几组数据。

表 3-1　分组数据表

第一组	第二组	第三组
7	6	4
7	6	4
4	4	4
4	4	4
1	2	4
1	2	4

经过简单的计算，可知这三组数据的平均数都是 4，也就是说这三组数据的中心位置是一样的，集中量数是一样的。但很明显，三组数据是不同的。也就是说仅有集中量数不能说明数据的全部特征。观察一下，三组数据的差异在于，第一组中，数字之间的距离很大；而第三组中，数字均集中在一个点上。因此，描述数据的特征除了考察数据的中心位置之外，还有必要考察数据的离散程度。

二、平均差

将一批数据中每个个体值与总体平均数之差即$(X-\bar{X})$称为离均差（简称离差）。由于离均差有正也有负，一组数据离均差的总和等于零，不便直接用离均差来反映该组数据的差异状态。因此数学家们用离均差的绝对值来反映各个数据的变异性。平均差（Average Deviation）就是各数据与平均数的离差绝对值的平均数，用符号 AD 表示。对于原始数据，平均差的计算公式为：

$$AD=\frac{\sum\left|x_i-\bar{X}\right|}{n} \qquad \text{（公式 3-3）}$$

【例 3-5】 有 8 名被试的反应时（单位：毫秒）实验数据分别为：18, 16, 20, 22, 17, 17, 20, 21，求其平均差。

解：按公式 3-3 分步计算

1. 求平均数：$\bar{X}=\frac{(x_1+x_2+\cdots+x_n)}{n}=\frac{(18+16+\cdots+21)}{8}=\frac{151}{8}=18.9$

2. 求平均差：$AD=\frac{\sum\left|x_i-\bar{X}\right|}{n}=\frac{\left|18-18.9\right|+\left|16-18.9\right|+\cdots+\left|21-18.9\right|}{8}=1.88$

答：这 8 名被试在实验数据的平均差为 1.88 毫秒。

平均差充分考虑了每一个数值偏离平均值的情况，完整地反映了全部数据的分散程度，在反映离散趋势方面比较灵敏，计算方法也比较简单。它的缺陷在于，由于它的敏感性，使得它易受极端值影响；另外，绝对值运算给数学处理带来很多不便。

三、方差和标准差

为了数学上计算的方便，人们用离均差平方的平均值反映个体的差异情况，称为方差（variance），记为S^2，其计算方法如公式 3-4 所示，其含义可表述为“方差是一组数据所有离均差的平方和的算术平均数”。

$$S^2=\frac{\sum(X-\bar{X})^2}{n} \qquad \text{（公式 3-4）}$$

在计算方差的过程中，经过平方后，各数据的单位都平方了，这样就改变了原来数据的内涵，使人不容易理解。为了与原始数据以及其他统计量在单位上保持一致，再把方差开方，得到另外一个差异量数，即标准差（standard deviation），记为S，其计算方法如公式 3-5 所示，其含义可表述为“标准差是一组数据所有离均差的平方和的算术平均数的算术平方根”。

$$S^2 = \sqrt{\frac{\sum (X-\bar{X})^2}{n}} \qquad \text{（公式 3-5）}$$

【例 3-6】 10 名学生的心理学考试分数为 73，87，83，80，77，79，75，78，72，86，求其标准差。

解：

1 求这组数据的平均数：

$$\bar{X} = \frac{(x_1+x_2+\cdots+x_n)}{n} = \frac{73+87+\cdots+86}{10} = \frac{790}{10} = 79$$

2 求这组数据的标准差：$S = \sqrt{\frac{\sum (X-X)^2}{n}} = \sqrt{\frac{236}{10}} = \sqrt{23.6} = 4.86$

答：这 10 名学生的心理学考试分数的标准差是 4.86 分。

四、其他差异量数

（一）全距

全距（Range）又称两极差，用符号 R 表示。全距是一组数据中最大值与最小值之差。若在分组数据中，全距则是最高组的精确上限与最低组的精确下限之差。其计算公式如下：

$$R = X_{\max} - X_{\min} \qquad \text{（公式 3-6）}$$

全距是描述数据离散趋势最简单的指标。由于全距只由数据中最大值和最小值所决定，所以非常不稳定，如果某一方向的极值发生变化，全距也会随之改变，因此在统计分析中很少使用全距作为离散程度的指标。全距一般只在非常简单地报告数据离散程度时才使用。例如，教师常常这样报告测验成绩的离散程度：“这次测验的最高分为 86 分，最低分为 42 分，最高分与最低分之间相差 44 分”。44 分就是这次测验分数的全距。

（二）百分位数与百分位差

百分位数（percentile，有时译为百分位点，记为P_p）是指量尺上的一个点，对于一组按大小升序排列的数据，位于此点以下的数据的个数占数据总个数的一定百分比。它是用于衡量数据的位置的量度，但它所衡量的不一定是中心位置。所谓第p百分位数（记为P_p）是这样一个值，它使得至少有p%的数据项小于或等于这个值，且至少有（100-p）%的数据项大于或等于这个值。百分位数提供了有关各数据项如何在最小值与最大值之间分布的信息。对于无大量重复的数据，第p百分位数将它分为两个部分。心理测验成绩经常以百分位数的形式报告。比如，假设某个考生语言能力测试成绩的原始分数为104分，相对于参加同一测试的其他学生来说，他的成绩如何并不容易知道。但是如果原始分数104分恰好对应的是第70百分位数，我们就能知道大约70%的学生考分比他低，而约30%的学生考分比他高。

由于用全距表示一组数据的离散程度易受两端极端数据的影响，因此有人提出先取数据两端各10%的数据之后，用$P_{90}-P_{10}$之间的距离这一差异量数，这一差异量数称为百分位差（percentile deviation)。计算百分位差关键在于计算两个百分位数：P_{90}和P_{10}。

原始数据计算第 p 百分位数的步骤如下：

（1）以递增顺序排列原数据（即从小到大排列)。

（2）计算指数i：$i=(\frac{p}{100})n$

式中，p是所求的百分位数的位置，n是数据的个数。

（3）若i不是整数，将i向上取整。大于i的毗邻整数指示第p百分位数的位置；若i是整数，则第p百分位数是第i项与第（i+ 1）项数据的平均值。

下面我们用例3-7来说明上述计算过程。

【例3-7】 有12个数据分别为：2350、2420、2210、2825、2225 2380、2450、2380、2390、2630、2440、2550，试计算该列数据的第85百分位数。

解：

（1）将这12个数据以递增顺序进行排列。

2210 2225 2350 2380 2380 2390 2420 2440 2450 2550

2630 2825

（2）$i=(\frac{p}{100})n=(\frac{85}{100})12=10.2$

（3）由于i不是整数，将其向上取整。则第85百分位数的位置是大于10.2的相邻整数，即第11项。

答：该列数据第85百分位数是第11项所对应的数值，即2630。

现在我们来计算第50百分位数。运用第2步，可以得到：由于i是整数，由百分位数计算步骤中第3步~~的（2）~~可知，第50百分位数即为第6项与第7项的平均值。所以第50百分位数就是（2390+2420）/2=2405 。注意第50百分位数即为中位数。

（三）四分位数与四分位差与箱形图

1.四分位数与四分位差

四分位数（quartile）也叫“四分位点”、“四分点”，是指将一组数据从小到大排列，然后把它们划分为个数相等的四部分，与此对应出现的三个分位点。其中最小的一个分位点记作 Q_1，称为第一四分位数，数据组中有 25%（1/4）的数据小于或等于 Q_1。第二个分位点记作 Q_2，称为第二四分位数，数据组中有 50%（1/2）的数据小于或等于 Q_2。第三个分位点记作 Q_3，称为第三四分位数，数据组中有 75%（3/4）的数据小于或等于 Q_3。可以看出，Q_1 就是第 25 百分位数；Q_2 就是第 50 百分位数，也就是中数；Q_3 就是第 75 百分位数。第一四分位数与第三四分位数之间的距离，称为四分位数间距，记为 QR，即 $QR = Q_3 - Q_1$。第一四分位数与第三四分位数之间距离的一半，称为四分位差（quartile deviation），记为QD，计算公式如下：

$$QD = \frac{Q_3 - Q_1}{2} \qquad \text{（公式 3-7）}$$

【例 3－8】以【例 3－7】数据为例。已知 Q2 即第二四分位数（中位数）为2405。试求该数据的四分位差。

解：

（1）计算四分位差首先要计算第25百分位数和第75百分位数，计算过程如下：

$$Q_1 : i = (\frac{p}{100})n = (\frac{25}{100}) \times 12 = 3$$

由于 i 为整数，由百分位数计算步骤中第 3 步的（2）可知，第一四分位数即第 25 百分位数为第 3 项与第 4 项的平均值。所以 Q1=（2350+2380）/2=2365。

$$Q_3 : i = (\frac{p}{100})n = (\frac{75}{100}) \times 12 = 9$$

同样i为整数，由百分位数计算步骤中第3步的（2）可知，第三四分位数即第75百分位数即为第9项与第10项的平均值。所以Q3=（2450+2550）/2 = 2500。

（2）计算四分位差

如下所示，四分位数将12个数据分为了4个部分，每个部分含有25％的数据项。

2210 2255 2350 ｜ 2380 2380 2390 ｜ 2420 2440 2450 ｜ 2550 2630 2825

Q1=2365　　　　Q2=2405　　　　Q3= 2500

（中数）

四分位差QD的计算如下：

$$QD = \frac{Q_3 - Q_1}{2} = \frac{2500 - 2365}{2} = 67.5$$

2.箱形图

箱形图(**Box plot**)也称箱须图（Box-whisker Plot），是利用数据中的五个统计量：最小值、第一四分位数、中位数、第三、第四分位数与最大值来描述数据的一种方法，它可以粗略地看出数据是否具有对称性，分布的分散程度等信息，特别可以用于对几个样本的比较。

我们将箱形图延至这一节才讲是因为它的关键是计算中位数和四分位数 Q_1 和 Q_3。此外还将用到四分位数间距。

我们使用【例3-8】中的数据来介绍箱形图的绘制步骤：

（1）画一个方盒，其边界恰好是第 1 和第 3 四分位数。【例 3-8】中的月工资数据，我们已经算出 Q_1＝2365，Q_3＝2500。这个方盒包含了数据组中间的50%的数据（如图3-1）。

（2）在方盒上中位数的位置画一条横线（对于【例 3-8】中的月工资数据，中数为2405）（如图3-1）。中数将数据分为相等的两个部分。

（3）利用四分位数间距来设定界限。箱形图的界限定于低于 Q_1 以下 1.5 个 QR 和高于 Q_3 以上 1.5 个 QR 的位置。上述月薪数据中，$QR＝Q_3－Q_1$＝2500－2365=135。因此，上、下限分别为：2500＋1.5×135＝2702.5 和 2365-1.5×135=2162.5，在上限值和下限值对应处分别画一条短横线（图3-1）。上、下限以外的数值作为异常值。

（4）在图 3-1 中的竖线叫做须线（whisker），须线从方盒的边线出发，分别至上、下限之内的最大值和最小值。对于月薪数据，其须线止于2210和2630。

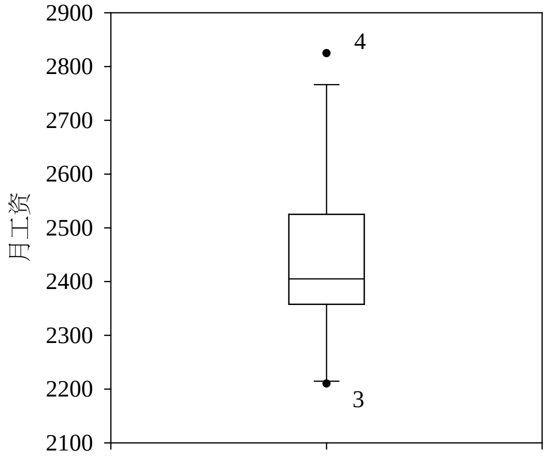


图3-1　月工资的箱形图

（5）最后，任一异常值的位置以符号“.”标出。在图 3-1 中可以看到一个异常值——2825（图 3-1 中标记的 4 表示这个界外值是原始数据的第 4 个数据）。

五、集中量数和差异量数的应用

（一）标准分数的含义

本章前面所说的集中量数和差异量数谈到的都是对一批数据的认识，有时候我们也要了解某个数据在它所在的数据组中所处的地位。反映次数分布中各数据所处地位的量，就叫地位量数。如张三的智商在班里是第一的，这个第一就是一个地位量数。从某种意义上说，前面所说的百分位差和四分位差都属于地位量数，而标准分数则是一种最为重要的地位量数。

标准分数又称 Z 分数，是将原始分数与其平均数之差除以标准差所得之商，计算公式为：

$$Z=\frac{X-\mu}{\sigma} \qquad \text{（公式 3-8）}$$

从标准分数的计算公式中可以看出，标准分数既与平均数有关，也与标准差有关，它是用原始数据离开其平均数多少个标准差来衡量原始数据的相对地位的。

例如，假定有三个分数分别为 8，4，2，这三个分数所在的那批分数的平均数是 4，标准差是 2。根据公式（3－8），我们很容易知道三个分数的标准分数分别为 2，0，－1，其中 2 表示原始分数高于平均数 2 个标准差，0 表示原始分数等于平均数，－1 表示原始分数低于平均数 1 个标准差。

（二）标准分数的性质

标准分数具有如下性质：

（1）标准分数的平均数为 0，标准差为 1。

（2）标准分数是原始分数的线性转换，标准分数的分布形状和原始分数的分布形状相似。

（3）一组原始分数的平均数的标准分数为 0，大于平均数的原始分数的标准分数为正值，小于平均数的原始分数的标准分数为负值。

把原始分数转换成标准分数的实质，就是把单位不等距和缺乏明确参照点的分数转换成以标准差为单位，以平均数为参照点的量表分数。上面提到了，标准分数的平均数为 0，标准差为 1。因此，将原始分数转换成 Z 分数，就是转换成了以 1（1 个标准差）为单位，以 0 为参照点的量表分数，从而明确每个原始分数的相对地位，并且分数间也有了相互比较的基础。

（三）标准分数的应用

由于所有的标准分数的平均数都为 0，标准差都为 1。可以把所有的标准分数都刻画在以 1（1 个标准差）为单位，以 0 为参照点的量表分数上，在这个量表上，同一个体在不同课程的成绩，或不同个体在同一课程的成绩，以及不同个体各门课程的总成绩就可以相互比较了。

标准分数在心理与教育研究中具有很重要的地位，我们可以利用标准分数来明确每个原始分数在分布中的相对地位；也可以利用标准分数来比较两种测验成绩的优劣；还可以利用标准分数来计算不同测验的总分和平均分。

【例 3-9】 以本章开头问题二的数据为例，问甲乙两生谁的三门功课总成绩更好一些？

解：

由于原始分数不具备相同的参照点和相等的单位，不是等距变量，不可以相加，所以用原始总分的和作为三门功课总成绩来比较是不恰当的。根据前文关于标准分数的描述可知，采用标准分数的总分来比较总成绩的优劣是可以。

（1）求甲生标准分数成绩总分：

①求甲生各科标准分数：

$$Z_{语}=\frac{85-80}{10}=0.5\text{；}\quad Z_{数}=\frac{81-70}{15}=0.73\text{；}\quad Z_{英}=\frac{90-85}{12}=0.42$$

②计算甲生三门功课的标准分总成绩：$Z_{甲}=Z_{语}+Z_{数}+Z_{英}=0.5+0.73+0.42=1.65$

（2）求乙生标准分数成绩总分：

①求甲生各科标准分数：

$$Z_{语}=\frac{88-80}{10}=0.8\text{；}\quad Z_{数}=\frac{85-70}{15}=1.0\text{；}\quad Z_{英}=\frac{83-85}{12}=-0.17$$

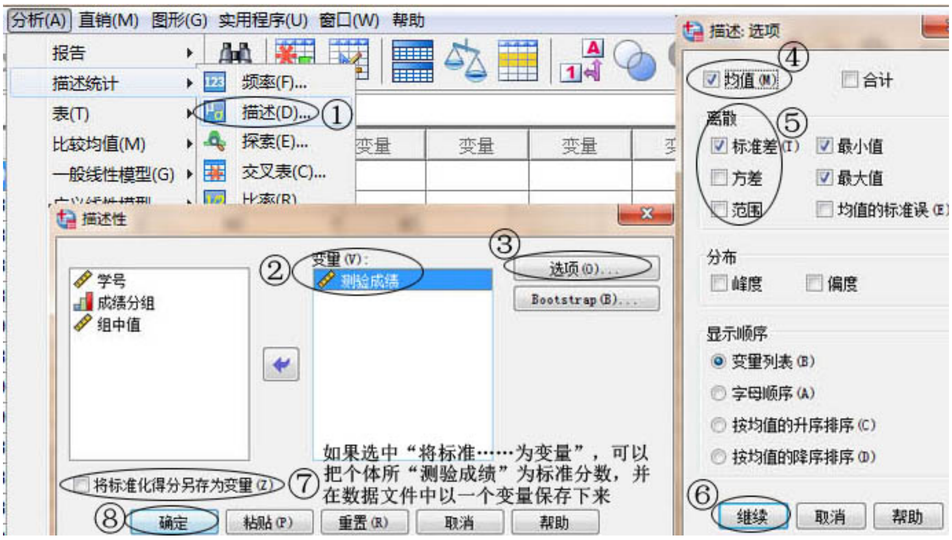
②计算乙生三门功课的标准分总成绩：$Z_{乙}=Z_{语}+Z_{数}+Z_{英}=0.8+1.0-0.17=1.63$

答：虽然甲乙二人在这三门课上的总成绩都是 256 分，但由这两人在这三门课上的标准分数的总和可知，$Z_{甲}>Z_{乙}$，因此，甲生的总成绩较好。

六、用 SPSS 计算集中量数、差异量数、标准分数和制作箱形图

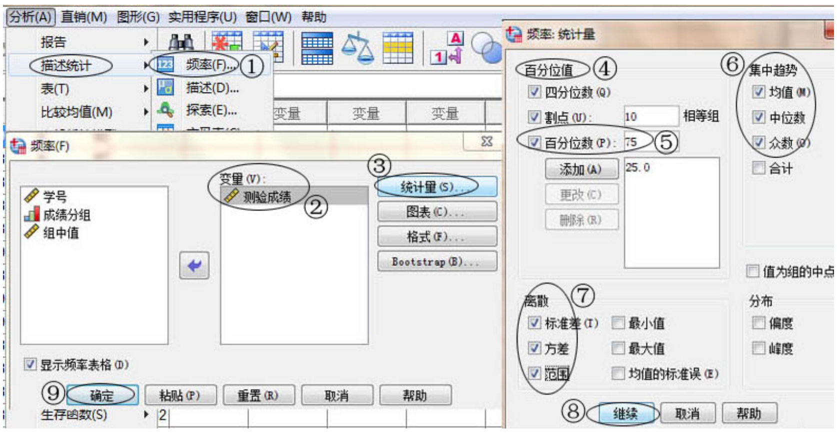
（一）用 SPSS 计算集中量数、差异量数和标准分数

用 SPSS 计算集中量数和差异量数可以采用两个不同的过程：（1）“分析→描述统计→描述”过程（见图 3-2）；（2）“分析→描述统计→频率”过程（见图 3-3）。这两个过程能给出的统计量存在一些差别：第一个过程可以计算均值、标准差和方差等统计量，以及把原始数据转换为标准分数，但无法给出百分位分数；第二个过程除了不能给出标准分数之外，能给出第一个过程计算出来的各个统计量，还可以给出任意指定的百分位数。



（如果在第⑦步选中“将标准化得分另存为变量”，则会在数据文件中把“测验成绩”的标准分作为一个变量保存下来，默认的变量名为“Z 测验成绩”。）

图 3-2　用 SPSS 计算集中量数、差异量数和标准分数的过程示意图一

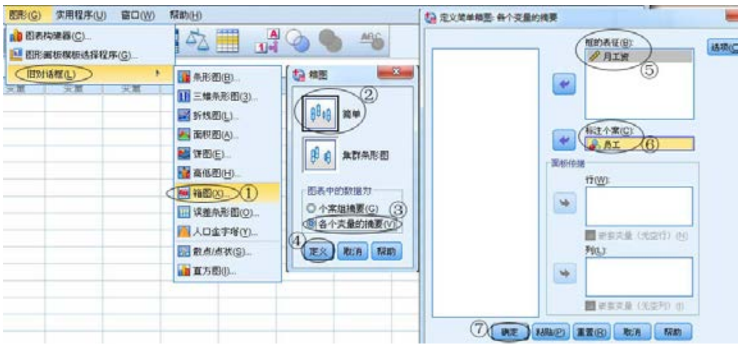


（在第⑤步可以输入任意一个百分数，并点击“添加”，即可计算出相应的“百分位数”。）

图 3-3　用 SPSS 计算集中量数、差异量数的过程示意图二

（二）用 SPSS 制作箱形图

按照图 3-4 的步骤（数据见本书所附光盘中例 03-7.sav）即可制作出简单的箱形图（如图 3-1 所示）。



（在第⑥把“员工”加入到“标注个案”框，则会在生成的图中使用员工号码标注界外值。）

图 3-4　用 SPSS 制作简单箱形图过程示意图

另外，正如前文所言，箱形图特别可以用于对几个样本的比较，例如从图 3-5 中就可明显看出来女生的平均水平高于男生，且离散程度小于男生，另外男、女生成绩都有一些界外值。制作多组数据分布比较箱形图的过程见图 3-6（数据参见本书所附光盘中例 02-5.sav）。

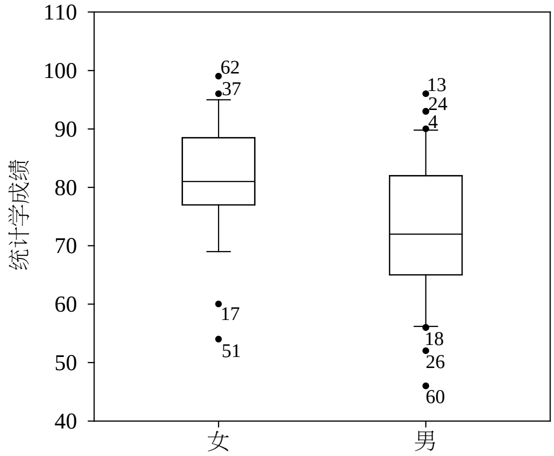


图 3-5　不同性别学生统计学成绩分布的比较

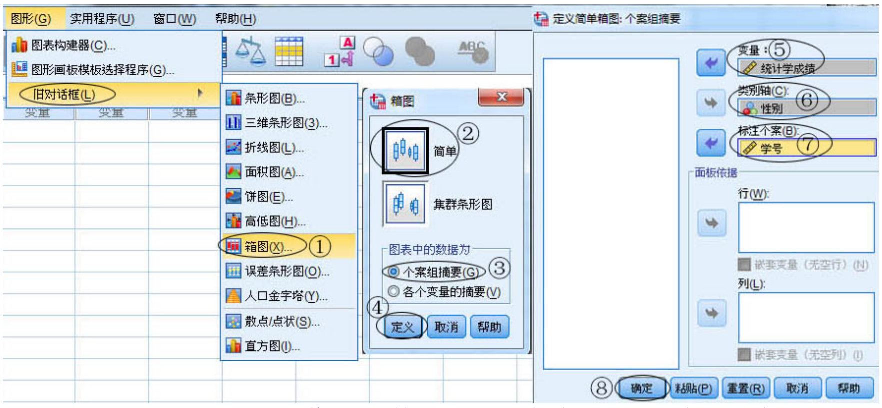


图 3-6　用 SPSS 制作多组数据分布比较箱形图的过程示意图

第三节　两变量关系的度量：相关分析

一、相关关系、相关分析与相关系数的含义

集中量数和差异量数主要描述单变量的数据资料的分布特征。在心理研究中，经常会涉及两个或多个研究对象，需要了解研究对象之间的相互关系。研究两个乃至多个变量之间关系的统计技术有多种，这里主要介绍相关分析。

客观事物是普遍联系着的，科学研究的任务之一是分析各种变量之间的相互关系。事物之间的关系包括两种类型，一是函数关系，二是相关关系。

函数关系（functional relationship）是指事物之间存在着严格的依存关系，其特征是现象与现象之间的数量关系是一一对应的，即对某一个变量的每一个数值，另一个变量都有唯一确定的值与之对应。函数关系可以用数学表达式十分准确地表达出来。如圆的半径和周长之间的关系就是函数关系，对于圆的某一半径（r）值就有一特定的周长（c）值与它相对应，两者的关系的数学表达式为：$c = 2\pi r$。

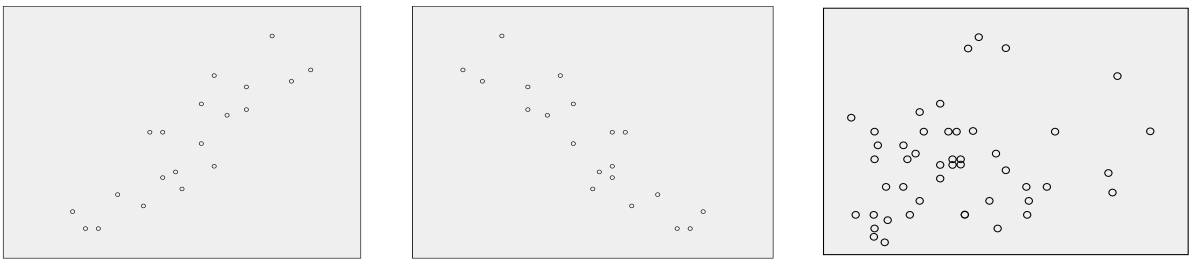
两个事物或现象之间存在一定的数量关系，但又不能确定何者为因何者为果

时，就称这种关系为相关关系（correlativity），简称相关。也就是说，事物之间存在相关关系时，并不能说明就一定存在因果关系。例如学生的智商和学习成绩的关系。我们知道智商和学习成绩有密切的关系，却不能根据智商的值来准确确定学生的学习成绩。因为学生学习成绩不仅受智力因素的影响，还受到非智力因素以及考试时的生理和环境因素的影响。

相关分析（correlation analysis）就是对变量之间是否存在相关、存在什么样的相关以及相关的程度如何等问题进行分析的统计方法，常用的方法有绘制散点图和计算相关系数两种。

将具有相关关系的两种现象的成对观测值标在平面直角坐标系中，这就是相关散点图。相关散点图可以非常直观地呈现出变量间的相关关系，阅读相关散点图就可以知道两个变量的成对观测值之间是属于什么种类的相关。

按变量关系的表现形态，相关关系可分为直线相关和曲线相关。直线相关是指两变量之间呈线性趋势（图 3-7（1）和（2））。曲线相关是指两变量之间呈



（1）　　（2）　　（3）

图 3-7　线性相关散点图

现某种曲线趋势，也称非线性相关（图 3-8）。本书中，所涉及到的都是线性相关，而不涉及曲线相关。

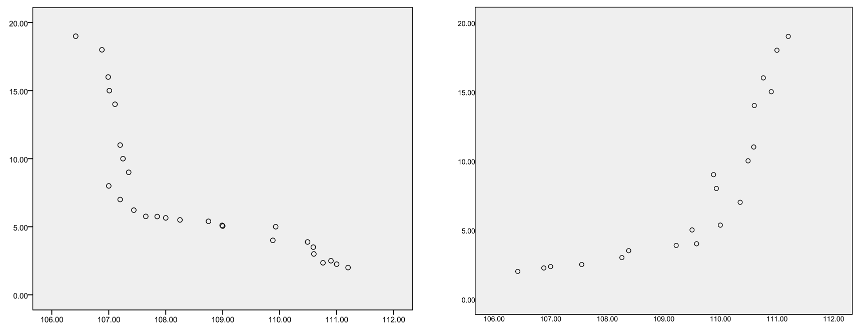


图 3-8　曲线相关散点图

相关系数是指描述变量间相关的方向和相关程度的统计量，通常用符号 r 表示。相关系数的取值范围在＋1 和－1 之间。相关系数的正负号表示相关关系的方向，相关系数的绝对值表示相关的程度，绝对值越大表示相关越强，绝对值越小表示相关越弱。

按相关系数数值变化的总趋势，相关关系可分为正相关和负相关。正相关表现为当一个变量增大或减小时，另一个变量在总趋势上也增大或减小，两个变量变化方向的趋势相同（图 3-7（1）)，此时相关系数的符号为+。负相关表现为当一个变量增大或减小时，另一个变量在总趋势上则减小或增大，两个变量的变化方向相反（图3-7（2)），此时相关系数的符号为 -。

按两个变量联系的紧密程度，相关系数可分为完全相关、高度相关、低度相关和零相关。完全相关指两个变量之间是一一对应的、完全确定的相关关系。这实际上就是函数关系。完全相关表现在散点图上就是两变量的各个点都在一条直线上。在心理学的研究中，完全相关的变量是极少见的。高度相关又称强相关，是指在变量 X、Y 中，当 X 值变化时，与之对应的 Y 值增大（或减小）的可能性非常大，表现在相关散点图上就是散点比较集中地分布在某条直线的周围。低度相关又称弱相关，是指双变量 X、Y 中，当 X 值变化时，与之对应的 Y 值增大（或减小）的可能性较小，表现在相关散点图上就是散点比较离散地分布在某条直线的周围。零相关是指双变量 X、Y 中，当 X 值变动时，与之对应的 Y 值可能有变动，也可能无变动，而且毫无规律，也就是变量之间不具线性关系（图3-7（3))。

心理学研究中常用的相关系数有积差相关系数、等级相关系数、质量相关系数等。

二、积差相关系数的求法与应用

积差相关又称积矩相关，是 20 世纪初英国统计学家皮尔逊（Karl Pearson）提出的一种计算直线相关的方法，因而又称为皮尔逊相关。

积差相关的适用条件是：（1）两列变量都是等距或等比的测量数据；（2）两列变量所来自的总体必须是正态的或近似正态的对称单峰分布；（3）两列变量必须具备一一对应的关系。

积差相关系数的计算公式为：

$$r=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{\sum(Y-\overline{Y})^2}}$$（公式 3-9）

在上式中，r 指积差相关系数，$\overline{X}$ 表示 X 变量的平均数，$\overline{Y}$ 表示 Y 变量的平均数。

若记 $x=x_i-\overline{X}$，$y=y_i-\overline{Y}$，而将 $\sqrt{\sum(X-\overline{X})^2}$、$\sqrt{\sum(Y-\overline{Y})^2}$ 分别用标准差的形式表示，则上式成为

$$r=\frac{\sum xy}{nS_XS_Y}$$（公式 3-10）

公式 3-10 中，$\frac{\sum xy}{N}$ 称为协方差，即当 x 大 y 也大时，xy 也大；当 x 小 y 也小时，xy 也小。N 表示成对的 X 变量和 Y 变量的对数。也就是说， X、Y 二列变量的一致性程度决定了 $\sum xy$ 的绝对值大小。因此，协方差的绝对值大小直观地反

映了两列变量的一致性程度。但是，在研究实践中，往往 X 变量与 Y 变量具有不同的测量单位，因此，通常不用协方差直接表示变量间的一致性，而是将各变量的离均差分别除以各自的标准差，使之成为没有实际测量单位的标准分数，然后再求其协方差，这样就可以对两列具有不同测量单位的变量进行一致性的计算了。即

$$r=\frac{\sum xy}{N\sigma_X\sigma_Y}=\frac{1}{N}\sum\left[\frac{x}{\sigma_X}\right]\left[\frac{y}{\sigma_Y}\right]=\frac{1}{N}\sum Z_X Z_Y=\frac{\sum Z_X Z_Y}{N} \qquad \text{（公式 3-11）}$$

我们可以看到，积差相关系数是两列成对观测值中各对观测值的标准分数的乘积和的算术平均数。

我们也可以象公式 3-17 那样用原始数据的方式来计算积差相关系数。

$$r=\frac{N\sum XY-\sum X\sum Y}{\sqrt{N\sum X^2-(\sum X)^2}\sqrt{N\sum Y^2-(\sum Y)^2}} \qquad \text{（公式3-12）}$$

公式 3-9 至公式 3-12 都可以用来计算积差相关系数，其中公式 3-9 和公式 3-10 需要计算原始数据的离差，公式 3-11 需要根据数据的标准分数来计算，而公式 3-12 则是在原始数据的基础上进行计算。读者可以根据实际情况选择合适的计算公式。

【例 3-10】 为了解记忆能力（记为 X）和学业成绩(记为 Y)的关系，收集了 15 名被试在这两种测验上的成绩（表 3-2），试求记忆能力和学业成绩的积差相关系数。

解：将 15 名被试的成绩列入表中，并将所需统计量及其结果也列入表中。用公式 3-9 计算需要下表中第 1 至第 7 列的信息。用公式 3-12 计算需要下表中第 1、2、8、9、10 列的信息。

表 3-2 15 名被试在记忆能力和学业成绩及这两个变量相关系数计算表

	第 1 列	第 2 列	第 3 列	第 4 列	第 5 列	第 6 列	第 7 列	第 8 列	第 9 列	第 10 列
	X_i	Y_i	$X_i-\bar{X}$ (x)	$Y_i-\bar{Y}$ (y)	$(X_i-\bar{X})^2$	$(Y_i-\bar{Y})^2$	$x\cdot y$	X^2	Y^2	$X\cdot Y$
	82	83	0.8	2.4	0.64	5.76	1.92	6724	6889	6806
	75	79	-6.2	-1.6	38.44	2.56	9.92	5625	6241	5925
	69	76	-12.2	-4.6	148.84	21.16	56.12	4761	5776	5244
	92	83	10.8	2.4	116.64	5.76	25.92	8464	6889	7636
	86	82	4.8	1.4	23.04	1.96	6.72	7396	6724	7052
	76	79	-5.2	-1.6	27.04	2.56	8.32	5776	6241	6004
	85	76	3.8	-4.6	14.44	21.16	-17.48	7225	5776	6460
	72	79	-9.2	-1.6	84.64	2.56	14.72	5184	6241	5688
	79	85	-2.2	4.4	4.84	19.36	-9.68	6241	7225	6715
	86	80	4.8	-0.6	23.04	0.36	-2.88	7396	6400	6880
	83	78	1.8	-2.6	3.24	6.76	-4.68	6889	6084	6474
	90	82	8.8	1.4	77.44	1.96	12.32	8100	6724	7380
	86	78	4.8	-2.6	23.04	6.76	-12.48	7396	6084	6708
	81	86	-0.2	5.4	0.04	29.16	-1.08	6561	7396	6966
	76	83	-5.2	2.4	27.04	5.76	-12.48	5776	6889	6308
Σ	1218	1209			612.4	133.6	75.2	99514	97579	98246

将表中有关结果代入公式 3-9 中，就得到：

$$r=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{\sum(Y-\overline{Y})^2}}=\frac{75.2}{\sqrt{612.4}\sqrt{133.6}}=0.263$$

或者将表中有关结果代入公式 3－12 中，

$$r=\frac{N\sum XY-\sum X\sum Y}{\sqrt{N\sum X^2-(\sum X)^2}\sqrt{N\sum Y^2-(\sum Y)^2}}$$

$$=\frac{15\times98246-1218\times1209}{\sqrt{15\times99514-(1218)^2}\times\sqrt{15\times97579-(1209)^2}}=0.263$$

三、斯皮尔曼（Spearman）等级相关

等级相关适用于下列情形：

（1）对具有等级顺序的测量数据求等级相关。

（2）搜集到的数据是等距或等比的，但其总体分布不是正态，不满足求积差相关的要求，因此，把等距或等比的数据转换成等级顺序数据，再求等级相关。

需要计算的两列数据的等级相关时，可以用斯皮尔曼等级相关系数。斯皮尔曼等级相关系数用 r_R 表示，基本公式为：

$$\gamma_R=1-\frac{6\sum D^2}{N(N^2-1)}$$ （公式 3-13）

在上式中，D 为各对偶等级之差，n 为对偶数据个数。

【例 3-11】根据表 3-3 两种儿童智力测验成绩的等级资料，求取等级相关系数。

表 3-3　两种儿童智力测验成绩的等级资料表

被试编号	智力测验一（R_X）	智力测验二(R_Y)	$D=R_X-R_Y$	D^2
1	4	1	3	9
2	5	7	-2	4
3	1	3	-2	4
4	7	5	2	4
5	2	2	0	0
6	8	6	2	4
7	3	4	-1	1
8	6	8	-2	4
$\sum$	36	36	0	30

解：根据表中数据可得，

$$\gamma_R = 1 - \frac{6\sum D^2}{N(N^2-1)} = 1 - \frac{6\times 30}{8(8^2-1)} = 0.6429$$

答：两种儿童智力测验成绩的相关系数是 0.643。

数据中如果出现相同等级，那么用上式计算相关就要产生误差，需要进行矫正。校正系数 C

$$c = \frac{t\left(t^2-1\right)}{12} \qquad \text{（公式 3-14）}$$

上式中 t 为某一等级的相同数。比如，有一批被试在某变量上的等级数据为：1.5， 1.5， 3， 4， 5， 6， 7 在这批等级数据中，1.5 等级的相同数有两个，因此 t＝2。

当一列变量中有多个相同等级出现时，就要用公式 3-20 来计算多个校正系数的连加和。

$$\sum c = \sum \frac{t\left(t^2-1\right)}{12} \qquad \text{（公式 3-15）}$$

比如，有下列一批被试在某变量上的等级数据，

1.5， 5.5， 5.5， 1.5， 8， 5.5， 3， 9， 5.5， 10

在这批等级数据中 1.5 等级的相同数有 2 个，5.5 等级的相同数有 4 个，

因此，$\sum c = \sum \frac{t\left(t^2-1\right)}{12} = \frac{2\left(2^2-1\right)}{12} + \frac{4\left(4^2-1\right)}{12}$

有了校正系数的信息，就可以计算相同等级下的斯皮尔曼相关系数了。

有相同等级时计算斯皮尔曼相关系数的公式为:

$$r_{RC} = \frac{\sum x^2 + \sum y^2 - \sum D^2}{2\sqrt{\sum x^2 \sum y^2}} \qquad \text{（公式 3-16）}$$

式中：$\sum x^2 = \frac{N(N^2-1)}{12} - \sum \frac{t(t^2-1)}{12}$

$$\sum y^2 = \frac{N(N^2-1)}{12} - \sum \frac{t(t^2-1)}{12}$$

D 为对偶等级差数。

【例 3-12】 10 位被试在两次测验中的成绩排名顺序情况有如表 3-4，试求其相关系数。

表 3-4 10 位被试在两次测验中的成绩排名顺序表

被试编号	测验一(x)	测验二（y）	D	D^2
1	1.5	5.5	-4	16
2	5.5	3	2.5	6.25
3	5.5	5.5	0	0
4	1.5	3	-1.5	2.25
5	8	9.5	-1.5	2.25
6	5.5	3	2.5	6.25
7	3	1	2	4
8	9	7.5	1.5	2.25
9	5.5	7.5	-2	4
10	10	9.5	0.5	0.25
$\sum$	55	55	0	43.5

解：在本例中出现了相同等级，需要使用校正公式。

$$\sum x^2=\frac{10\times(10-1)}{12}-\left(\frac{2\times(2-1)}{12}+\frac{4\times(4-1)}{12}\right)=77$$

$$\sum y^2=\frac{10\times(10-1)}{12}-\left(\frac{2\times(2-1)}{12}+\frac{3\times(3-1)}{12}+\frac{2\times(2-1)}{12}+\frac{2\times(2-1)}{12}\right)=79$$

$$\sum r_{RC}=\frac{\sum x^2+\sum y^2-\sum D^2}{2\sqrt{\sum x^2\cdot\sum y^2}}=\frac{77+79-43.5}{2\times\sqrt{77\times79}}\approx0.72$$

答：其相关系数为 0.72

四、质量相关

质量相关适用于一列变量为等比或等距的测量数据，另一列变量是按性质划分的类别的数据。主要包括点二列相关和双列相关。

（一）点二列相关

点二列相关适用于两列变量中有一列为来自正态总体的等距或等比的测量数据，另一列变量为二分称名变量。所谓二分称名变量是指分为两类的称名变量，

如男与女，对与错，接受与拒绝等。

点双列相关的计算公式为：

$$r_{pb} = \frac{\overline{X}_p - \overline{X}_q}{S_t}\sqrt{pq} \qquad \text{（公式 3-17）}$$

在上式中，p 为二分称名变量中取某一值的数据个数占二发变量总个数的比率，

q 为二分称名变量中取另一值的数据个数占二分变量总个数的比率

$\overline{X}_p$ 为等距或等比变量中与 p 对应的那部分数据的平均值

$\overline{X}_q$ 为等距或等比变量中与 q 对应的那部分数据的平均值

S_t 为全部等距或等比变量的标准差

【例 3-13】 已知 10 名被试在某测验中的得分，这些分数的总体分布为正态。用 1 表示女性，0 表示男性（表 3-5）。试求测验得分和性别的相关。

表 3-5　10 位被试在某测验中的得分表

考生	1	2	3	4	5	6	7	8	9	10
性别	1	1	1	1	0	0	1	0	0	1
测验分数	63	55	65	47	46	57	57	55	51	53

解：性别为二分称名变量，因此计算点双列相关系数。

$n = 10 \qquad S_t = 5.80$

$p = 0.6 \qquad q = 0.4 \qquad \overline{X}p = 56.67 \qquad \overline{X}q = 52.25$

$$r_{pb} = \frac{\overline{X}_p - \overline{X}_q}{S_t}\sqrt{pq} = \frac{56.67 - 52.25}{5.80}\sqrt{0.6 \times 0.4} = 0.37$$

答：测验得分和性别的相关系数是 0.37。

（二）双列相关

双列相关系数适用于两列变量均为来自正态总体的等距或等比变量，而其中一列被人为地划分为两个类别的数据，其计算公式为：

$$r_{pb} = \frac{\overline{X}_p - \overline{X}_q}{S_t}\frac{pq}{Y} \qquad (3-18)$$

$\overline{X}_p$ 为等距或等比变量中与 p 对应的那部分数据的平均值

$\overline{X}_q$ 为等距或等比变量中与 q 对应的那部分数据的平均值

S_t 为全部等距或等比变量的标准差

p 为二分变量中取某一值的数据个数占二分变量数据总个数的比率

q 为二分变量中取另一值的数据个数占二分变量数据总个数的比率

Y 为标准正态曲线下 p 与 q 交界点的 Y 轴高度（通过查正态分布表得出）

【例 3-14】 表 3-6 是 10 名被试在一次智力测验的总分和其中某一道试题的得分，在这道题上得 6 分和 6 分以上为通过，得 6 分一下为不通过。试求这道题和智力测验总分的相关。

表 3-6　10 名被试在一次智力测验中的总分和其中某道试题的得分表

考生	A	B	C	D	E	F	G	H	I	J
智力测验得分	63	65	47	55	57	53	51	46	57	55
试题得分	8	7	7	4	6	4	3	4	8	7

解：该试题得分被人为划分为通过和不通过两类，本题应求双列相关。

$n = 10 \qquad S_X = 5.8$

$p = 0.6 \qquad q = 0.4 \qquad \bar{X}p = 56.33 \qquad \bar{X}q = 51.25$

查正态分布表：当 p＝0.6 时，Y＝0.3866

$$r_{pb} = \frac{\bar{X}_p - \bar{X}_q}{S_t} \cdot \frac{pq}{Y} = \frac{56.33 - 51.25}{5.8} \times \frac{0.6 \times 0.4}{0.3866} = 0.54$$

答：该试题与智力测验总分的相关系数为 0.54。

五、用 SPSS 计算相关系数

用 SPSS 计算相关系数，几种相关系数的过程几乎相同，在此仅以积差相关为例予以说明（见图 3-9，数据见本书所附光盘中例 03-10.sav）。等级相关仅在图中所示第③步中改为选择“Spearman”即可，而点二列相关本质上与积差相关相同，采用积差相关的方法可得到同样的结果。

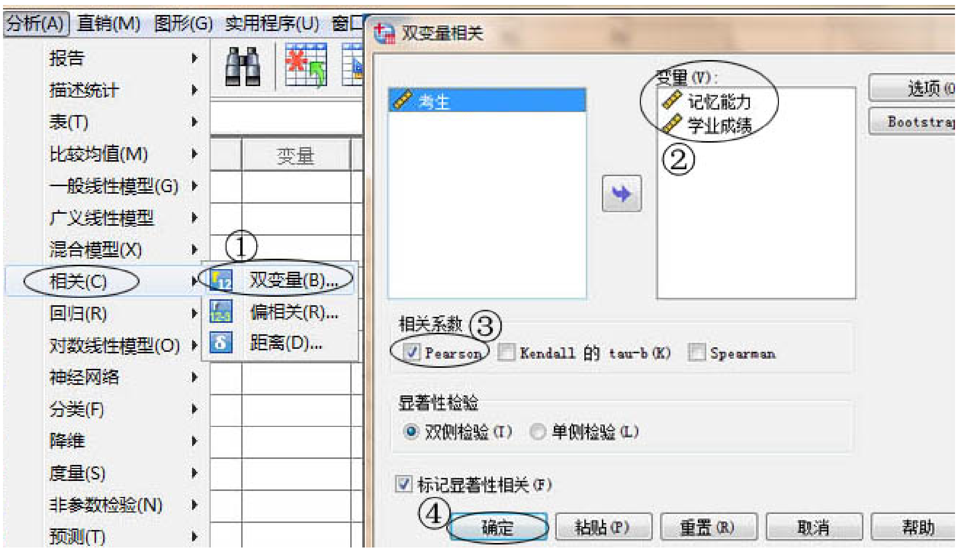


图 3-9　用 SPSS 计算积差相关系数的过程示意图

小 结

在这一章里，我们介绍了几种用以概括数据位置和变异程度的描述统计量。与表格法和图形法不同，本章介绍的用来概括数据的量度是一些数值。为了描述数据的中心位置，我们使用了平均数、中位数和众数等集中量数。为了描述数据的离散程度，我们使用了平均差、方差、标准差、全距、百分位数、百分位差、四分位数、四分位差等。结合这一部分内容，我们补充介绍了探索性数据分析技术中的箱形图的画法与使用。接下来，我们介绍了集中量数和差异量数的应用——标准分数。最后，我们介绍了分析两变量关系的统计技术——相关分析。

关键术语

集中量数：用来描述数据的集中趋势的特征量数。

差异量数：用来描述数据的离中趋势的特征量数。

算术平均数：衡量数据的总和除以数据个数的商。

中数：在按大小顺序排列的一组数据中，位置居中的那个称为这组数据的中数。中数将所有的数据分为两个相等部分，一部分的值都大于或等于它，而另一部分的值都小于或等于它。

众数：是指一组数据中出现次数最多的那个数据值。

方差：是离均差平方的算术平均数。离均差是总体中每个数据与总体平均数之差。

标准差：用方差的正的平方根。

全距：用一组数据中最大值与最小值之差。

百分位数：是指量尺上的一个点，对于一组按大小升序排列的数据，位于此点以下的数据的个数占数据总个数的一定百分比。其中第 p 百分位数（记为P_p）是这样一个值，它使得至少有 p%的数据项小于或等于这个值，且至少有（100-p）%的数据项大于或等于这个值。

四分位数：是指将一组数据从小到大排列，然后把它们划分为个数相等的四部分，与此对应出现的三个分位点。其中第一个四分位数Q_1在数据组中有 25%（1/4）的数据小于或等于它；第二个四分位数Q_2在数据组中有 50%（1/2）的数据小于或等于它；第三个四分位数Q_3在数据组中有 75%（3/4）的数据小于或等于它。

箱形图：利用数据中的五个统计量：最小值、第一四分位数、中位数、第三、第四分位数与最大值来描述数据的一种方法

标准分数（即 Z **分数**）：是原始分数与其平均数之差除以标准差所得之商。

函数关系：是指事物之间存在着严格的依存关系，其特征是现象与现象之间的数量关系是一一对应的，即对某一个变量的每一个数值，另一个变量都有唯一确定的值与之对应。它可以用数学表达式十分准确地表达出来。

相关关系：指两个事物或现象之间存在一定的数量关系，但又不能确定何者为因何者为果的数量之间的关系。

相关分析：是围绕变量之间是否存在相关、存在什么样的相关以及相关的程度如何等问题进行研究时所使用的统计分析方法，通常用相关系数来表达其分析结果。

相关系数：是指描述变量间相关的方向和相关程度的统计量。相关系数的取值范围在＋1 和－1 之间。相关系数的正负号表示相关的方向。相关系数的绝对值表示相关的程度，绝对值越大表示相关越强，绝对值越小表示相关越弱。

重要公式

算术平均数：$\overline{x} = \frac{\sum X}{N}$

方差：$S = \sqrt{\frac{\sum(X - \overline{X})^2}{N}}$

标准差：$s = \sqrt{s^2}$

标准分：$Z = \frac{X - \overline{X}}{S}$

皮尔逊积差相关：$r = \frac{\sum(X - \overline{X})(Y - \overline{Y})}{\sqrt{\sum(X - \overline{X})^2}\sqrt{\sum(Y - \overline{Y})^2}}$

等级相关：$\gamma_R = 1 - \frac{6\sum D^2}{N(N^2 - 1)}$

点二列相关：$r_{pb} = \frac{\overline{X}_p - \overline{X}_q}{S_t}\sqrt{pq}$

双列相关：$r_{pb} = \frac{\overline{X}_p - \overline{X}_q}{S_t}\frac{pq}{Y}$

思考与练习

1. 一个样本的数据如下：53，55，70，58，64，57，53，69，57，68，53。计算它们的平均数和中位数。

2．下面的样本是1 5个大学四年级学生在大四最后一学期中所获的学分数：

15 21 18 16 18 21 19 15 14 18 17 20 18 15 16

（1） 计算平均数、中位数和众数并解释它们的含义。

（2） 计算第1和第3四分位数。

（3） 计算第70百分位数并解释其含义。

3． 某生产部门利用一种抽样程序来检验新生产出来的产品的质量，该部门使用下面的法则来决定检验结果：如果一个样本中的14个数据项的方差大于0.0 05，则生产线必须关闭整修。假设搜集的数据如下：

3.43	3.45	3.43	3.48	3.52	3.50	3.39
3.48	3.41	3.38	3.49	3.45	3.51	3.50

问此时的生产线是否必须关闭？为什么？

4．高速公路损失调查机构在1998年9月的车祸和伤亡损失报告中，对于不同的车型根据事故发生后的保险金额进行评定。评定分数接近100被视为是正常情况。同时分数越低越好，它表明该车型更加安全。下面是对20辆中型轿车和20辆小型轿车的评定结果。

中型	81	91	93	127	68	81	60	51	58	75
轿车	100	103	119	82	128	76	68	81	91	82
小型	73	100	127	100	124	103	119	108	109	113
轿车	108	118	103	120	102	122	96	133	80	140

（1）画箱形图。

（2）通过你的概括，说明中型轿车和小型轿车哪种车型更为安全。

5.计算下列数据的平均差与标准差。

11.0	13.0	10.0	9.0	11.5	12.2	13.1	9.7	10.5

6.某运输部门对中型汽车的速度和里程进行了研究，其数据如下：

速度	30	50	40	55	30	25	60	25	50	55
里程	28	25	25	23	30	32	21	35	26	25

计算上述数据的样本相关系数并解释含义。

7.下列两变量为非正态，选用恰当的方法计算相关。

被试	1	2	3	4	5	6	7	8	9	10
X	13	12	10	10	8	6	6	5	5	2
Y	14	11	11	11	7	7	5	4	4	4

第四章 概率与概率分布

学习本章内容，将有助于你对以下问题的理解与思考：

1. 随机事件与概率的含义是什么？
2. 离散型概率分布和连续型概率分布主要有什么特点？
3. 利用正态分布表主要能解决哪些心理学中的问题？

一天，一对夫妇带着他们六岁的儿子到某校心理学系做了一次韦氏儿童智力测验，结果总分为 115 分。负责测验的老师告诉他们：这孩子的智商不错，同龄儿童中大约只有16%的人的智商比他好。这一结果的科学依据是什么？

第一节 随机事件与概率

一、随机试验与随机事件

日常生活中我们会遇上各种各样的现象。例如：

（1） 将一枚硬币或色子轻抛出手，通常他们不会向上飞而是会受地球引力的影响落在地面上。

（2） 让一枚硬币掉落地面，在它落地前，我们不能确定它是正面向上还是反面向上。但是,如果我们在有控制的条件下将这件事做很多次,我们会发现正面向上或反面向上的次数几乎相同。

在概率论中,通常把类似于第一种那样在一定条件下必然发生的现象称为**确定性现象**（determinant phenomenon）；把类似于第二种那样的现象，即在个别试验中其结果呈现出不确定性，但是在大量重复试验中，其结果又具有统计规律性的现象称为**随机现象**（random phenomenon）。

概率论与数理统计是研究随机现象统计规律性的一门数学学科。在概率论中，通常把“通过对随机现象进行测试后获取相应的数据结果，进而揭示该随机现象有何规律的实验”称为**随机试验**(random experiment,简称试验，通常用英文大写字母“E”表示)。随机试验具有以下一些特点：

（1）试验可以在相同条件下大量重复地进行；

（2）每次试验的可能结果不止一个，并且在试验之前能知道试验的所有可能结果；

（3）在进行一次试验之前，不能确定哪一个结果会出现，但若进行大量重复试验的话，其可能结果出现又有一定的统计规律性。

从心理研究的角度说，由于我们对随机现象的研究都属于一种人类行为，因此，我们也可以对“试验”下这样的定义：当我们在控制条件下有目的地让某种随机现象发生以便观察其统计规律性时， 一般把这样的行为过程称为“**试验**”。

例如，将一枚面值为 1 元的硬币在一米高的同一个位置上让它的圆边垂直向下落地，每落地一次，就作为一次试验。硬币落地后，可能的结果只能是以下两种情况之一（图 4-1)：

在一次试验过程中，硬币落地前并不知道会是什么结果，但是，这种试验可以重复进行，并且在大量重复试验后，试验结果会呈现出上述两种结果大约各

图 4-1 投掷硬币可能结果示意图

出现一半次数的统计规律性。因此，投掷硬币研究其落地结果的过程就是一种随机试验。

下面这些现象由于满足上述三个特点，因此都属于随机试验：

E_1：将一枚硬币连续抛掷三次，观察正面向上（在图 4-1 中指含有 1 元图案的那一面，通常用英文大写字母“H”表示）或反面向上（在图 4-1 中指含有国徽图案的那一面，通常用英文大写字母“T”表示）出现的情况；

E_2：投掷一颗均匀对称的骰子，观察出现的点数；

E_3：将一颗均匀对称的骰子投掷两次，观察出现的点数；

E_4：记录一段时间内，某城市 119 火警报警电话接到报警的次数；

E_5：从一批某同一品牌同一型号充满电的音乐播放器中，任意抽取一只看其连续播放的时间。

在随机试验中，试验可能产生的任何结果都称为“**样本点**”（sample points），如硬币落地后可能有图 4-1 所示的两种结果，这两种结果都是“让硬币落地”这一随机试验的样本点。

由某种随机现象全体样本点构成的集合称为试验的“**样本空间**”（sample space），一般用符号“Ω”表示（有的教科书用英文大写字母“S”表示），样本空间的所有样本点一般用大括号括起来。如硬币落地这一随机现象的样本空间是 $\Omega=\{H, T\}$。在上述列出的随机试验中：

E_1 的样本空间 Ω_1：{HHH，HHT，HTH，THH，HTT，THT，TTH，TTT}；

E_2 的样本空间 Ω_2：{  }，也可用数字表示为{1，2，3，4，5，6}；

E3 的样本空间Ω3 则有如图 4-2 所示：

如果我们用第一个数字表示浅色骰子的点数，用第二个数字表示深色骰子的点数，那么，投掷两个骰子的样本空间也可表示为图 4-3。

E4 的样本空间Ω4：{ 0，1，2，3……}；

E5 的样本空间Ω5：$\{\ t \mid t\geqslant 0\ \}$。

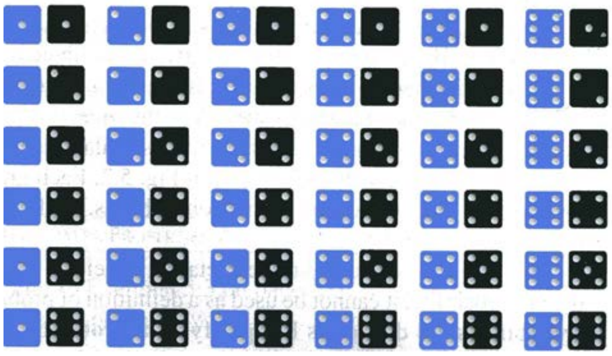


图 4-2 投掷两个骰子的样本空间示意图

$$\left\{\begin{array}{l}
1,1;\ 1,2;\ 1,3;\ 1,4;\ 1,5;\ 1,6\\
2,1;\ 2,2;\ 2,3;\ 2,4;\ 2,5;\ 2,6\\
3,1;\ 3,2;\ 3,3;\ 3,4;\ 3,5;\ 3,6\\
4,1;\ 4,2;\ 4,3;\ 4,4;\ 4,5;\ 4,6\\
5,1;\ 5,2;\ 5,3;\ 5,4;\ 5,5;\ 5,6\\
6,1;\ 6,2;\ 6,3;\ 6,4;\ 6,5;\ 6,6
\end{array}\right\}$$

图 4-3 投掷两个骰子的样本空间示意图

在一个随机试验中，具有一定特性的样本点的子集称为试验 E 的**随机事件**（random events），简称**事件**。在每次试验中，当且仅当这一子集中一个样本点出现时，称这一事件发生。样本空间 Ω 包含所有的样本点，它本身也是事件，在每次试验中它总是发生，称为**必然事件**。空集 Φ 不包含任何样本点，为了数学研究的方便，把空集 Φ 也定义为事件，它在每次试验中都不发生，称为**不可能事件**。由一个样本点组成的单点集称为**基本事件**。

二、频率与概率的含义

如前所述，概率论与数理统计是研究随机现象统计规律性的一门数学学科。随机现象的统计规律性通常要在大量的随机试验中才能表现出来。例如，“投掷一枚硬币，观察其落地后出现正反面的情况”的随机现象中，定义“事件 A”是“试验后正面向上”这一结果出现，则可以通过大量的试验来揭示这一随机现象的统计规律性。历史上曾有许多著名学者对这一随机现象的统计规律性作了大量的试验，表 4-1 列出了其中几位的试验结果。

表 4-1 不同学者投掷硬币试验记录表

试验者	投掷次数 n	出现正面朝上的次数 n_A	频率 $f_n(A)=\dfrac{n_A}{n}$
DeMorgan	2048	1061	0.5180
Buffon	4040	2048	0.5069
Feisher	10000	4979	0.4979
Pearson	12000	6019	0.5016
Pearson	24000	12012	0.5005

在表 4-1 中，第一列是进行这一试验的不同学者的名字；第二列是每位学者在一次研究中共投掷硬币的次数，在此用字母“n”表示；第三列是各学者在 n 次试验中出现正面朝上的次数 n_A，通常把 n_A 称为频数；第四列是事件 A 的频数 n_A 与试验总次数的比率，通常将这一比率值称为频率（frequency），用 $f_n(A)$ 表示。

在概率论中，**概率**（probability）是表示随机事件出现可能性大小的客观指标。通常把概率区分为“先验概率”和“后验概率”两种类型。所谓“后验概率（posterior probability）”是指：

设随机事件 A 在相同的条件下进行的 n 次试验中发生了 n_A 次，则称 n_A 是事件 A 在这 n 次试验中发生的频数，称比值 n_A/n 为事件 A 在这 n 次试验中发生的频率，记成 $f_n(A)$。

大量的试验研究表明，随机事件 A 的频率在试验次数逐渐增多时，将逐渐稳定于某个数的附近。这就是频率的稳定性，也就是我们所说的随机现象的统计规律性。例如，由表 4-1 所列出的数据可知，在投掷硬币这样的随机试验中，当投掷总次数 n 越来越大时，频率 $f_n(A)$ 呈现出越来越接近 0.5。频率的稳定性反映了事件 A 发生可能性的大小。当试验次数 n 有充分大时，可以把较为稳定的事件 A 的频率值即 $f_n(A)$ 作为事件 A 出现可能性的度量，将其视为事件 A 出现的概率（通常用“P(A)”表示）。即：

$$P(A) \approx \frac{n_A}{n} \qquad \text{（公式 4-1）}$$

公式 4-1 通常被称为“**概率的统计定义**”，显然，它是将事件 A 出现的频率视为事件 A 的概率，因为它是经过 n 次试验之后得到的，所以被称为“后验概率”。由上述定义或知，频率具有下述基本性质：

(1) $0 \le f_n(A) \le 1$

(2) $f_n(\Omega)$ = 1

(3) 若 A1，A2，…，An…是两两互不相容的事件，则

$$f_n(A_1 \cup A_2 \cup \ldots \cup A_k) = f_n(A_1) + f_n(A_2) + \ldots + f_n(A_k) \qquad \text{（公式 4-2）}$$

在实际生活中，我们不可能对每一个随机事件都做大量的试验，而后求该事件 A 的频率来表征它发生的可能性的大小。为了理论研究的需要，概率论的研

究专家通常对概率作如下定义：

在一个随机试验 E 的样本空间Ω中，对于 E 的每一事件 A，赋于一个实数 P(A)与之对应，如果集合函数 P(A)满足下列三个条件，我们就称 P(A)为事件 A 的概率。这三个条件是：

（1）非负性：对于每一个事件 A，有 P（A）≥ 0

（2）规范性：对于必然事件Ω，有 P(Ω) = 1

（3）可列可加性：设 A1，A2，…，An…是两两互不相容的事件序列，即 AiAj=Φ（i≠j；i，j=1，2，3…），则有

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) \quad （公式 4-3）$$

上述三条也是**概率的主要性质**。此外，由这此概率性质可导出的其他性质主要包括：

（1）空集Φ的概率为零，即：P(Φ)=0；

（2）随机事件 A 的概率小于或等于 1，即：P(A)≤1；

（3）随机事件 A 及其逆事件-A 的概率之和等于 1，即 P(A)+ P(-A)=1；

（4）如果事件 A 是事件 B 的子集，那么事件 A 的概率会小于或等于事件 B 的概率，即：P(A) ≤ P(B)；

当所研究的随机试验具有以下两个特点时，一般被称为“先验概率（prior probability，又称**古典概型或**等可能概型）”试验，这两个特点是：

（1）有限性：指所有可能的试验结果仅有有限种，即样本空间Ω中基本事件总数有限；

（2）等可能性：每次试验中各个基本事件出现的可能性相同。

在古典概型中，假定能够知道事件 A 中包含的基本事件数，就可以通过这个数与试验的基本事件总数之比计算出概率 P（A）。

定义：若试验结果一共由 n 个基本事件 $E_1, E_2, \cdots E_n$ 组成，并且这些事件的出现具有相同的可能性，而事件 A 由其中某 m 个基本事件 $E_{i1}, E_{i2}, \cdots E_{im}$ 组成，则事件 A 的概率可以用下式计算：

$$P(A) = \frac{A\text{中所含的基本事件数}}{\text{样本空间}\Omega\text{中的基本事件总数}} = \frac{m}{n} \quad （公式 4-4）$$

前面所提到的投掷硬币或投掷骰子的随机试验都属于古典概型试验。

在投掷硬币，观察正面或反面向上的试验中，可能结果只有两个，即其样本空间Ω={H，T}，也就是说 n=2，若事件 A 为观察其正面向上（H）出现的次数,H 是这类试验样本空间Ω的两个结果之一，于是 m=1，代入上式得：

$$P(A) = \frac{m}{n} = \frac{1}{2} = 0.5$$

三、概率的一些主要计算公式

1. 概率的加法公式

概率的加法定公式是概率的基本性质之一"概率的可列可加性"的扩展。概率的可列可加性是指：若 $A_1, A_2, \cdots A_n$ 是两两互不相容（相互独立）的事件，则有

$$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n) \qquad \text{（公式 4-5）}$$

这一性质通常亦被称为是概率的加法公式（或加法定理），其含义是，两个（或多个）互不相容的事件的和的概率等于这两个（或多个）事件的概率之和。所谓互不相容事件是指事件 A 与事件 B 不能同时发生。即 AB=Φ 。互不相容的事件 A 与事件 B 没有共同的样本点。例如，投掷骰子的随机试验中，若定义 A_1=出现 1 点；A_2=出现 2 点，由 P（A_1）和 P（A_2）的概率都是 1/6，若 A_3=出现 1 点和出现 2 点这两个事件之和，即 A_3=（A_1+A_2），根据概率的可列可加性，出现 1 点或 2 点中任何一个结果的事件的概率会等于分别出现 1 点或 2 点这两个事件的概率的和，即

$$P(A_1 + A_2) = P(A_1) + P(A_2) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}$$

2. 条件概率与乘法定理

条件概率是概率论中的一个重要而实用的概念。从广义上说，任何概率都是在一定条件下发生的，因此，要注意概率论中所讨论的**条件概率**是指两个事件相互关联时，在事件 A 已发生的条件下事件 B 出现的概率。下面我们来看看投掷一
枚骰子的随机试验。

［例 4-1］ 将一枚骰子投掷一次，设事件 A 为"掷出的点数是偶数"，事件 B 是"掷出 6 点"。求在事件 A 出现的条件下事件 B 出现的概率。

解： 根据前面的分析，可知这类试验的相关数据是：

试验的样本空间 Ω={1， 2， 3， 4， 5， 6}

事件 A={2， 4， 6}；事件 B={6}；事件 A 的概率是：P(A)=3/6；事件 B 的概率是：P(B)=1/6；

若已知事件 A 已经发生，即已知"掷出的点数是偶数"的条件下，试验所有可能的结果所构成的集合就是 A，事件 A 中有 3 个基本事件，每个基本事件出现的可能性是相等的，而其中一个基本事件就是事件 B，于是在事件 A 已发生的条件下事件 B 发生的概率（通常用 P(B/A)表示）就是 1/3。

答： 在掷出的点数是偶数的条件下，掷出的点数是 6 和概率为三分之一。

由上面的分析可知：P(B)≠P(B/A)，前者等于 1/6，后者等于 1/3，这种差异是由于我们是限制事件 A 已经发生的条件下来考虑事件 B 发生的概率的。概率发生的条件不一样，所得到的结果自然会不一样。投掷一枚骰子，"掷出 6

点”这一结果即属于事件 A 又属于事件 B，即这是一个事件 A 和事件 B 同时发生的结果，事件 B 的概率 P(B)是“掷出 6 点”这一结果与整个样本空相比得到的，因此等于 1/6；而在事件 A 已经发生的条件下，考虑事件 B 发生的概率 P(B/A)时，事件 A 就变成了新的样本空间，因此其概率就等于事件 A 和事件 B 同时发生的概率 P(AB)与事件 A 的概率 P(A)之比，即：

$$P(B/A)=\frac{P(AB)}{P(A)}=\frac{1/6}{3/6}=\frac{1}{3}$$

在概率论中，通常对条件概率（conditional probability）作如下定义：

设 A，B 是某种随机试验的两个事件，且 P(A) ＞ 0，称

$$P(B/A)=\frac{P(AB)}{P(A)}, \qquad \text{（公式 4-6）}$$

为在事件 A 发生的条件下，事件 B 发生的条件概率。

上述定义很直观，但缺点是要求 P(A) ＞ 0，为方便使用，规定：当 P(A) = 0 时，P(B/A)=0。这样，在任何情况下都有

$$P(AB)=P(A)P(B/A), \qquad \text{（公式 4-7）}$$

一般把公式 4-8 称为条件概率的乘法公式。在一些实际问题中，条件概率 P(B/A)往往根据试验背景容易得到，因此常用乘法公式计算 P(AB)。

3. 独立性与概率树

一般来说，某个事件 B 发生的“无条件概率”P(B)与已知另一事件 A 发生前提下的条件概率 P(B/A)是不同的，因为后者比前者基于更多的信息。但是，如果事件 A 发生与否不能提供关于发生与否的信息，则条件概率与无条件概率没有区别，也就是说：P(B)= P(B/A)。这时可以说：事件 B 的发生“独立”于事件 A 的发生。

【例 4-2】 设试验 E 为“将一枚硬币连续投掷两次（或一次投掷两枚硬币)，观察其正反面出现的情况”。

解：设事件 A 为第一次投掷结果为正面（H）向上，事件 B 为第二次投掷结果为正面（H）向上，那么，每次投硬币的结果有如图 4-5 所示。

答： 一枚硬币连续投掷两次包括以下四种正反面情况：正-正，正-反，反-正，反-反。

在概率论中，像上面这样来表达随机试验的图形称之为树形图。这是一种很实用的分析工具。

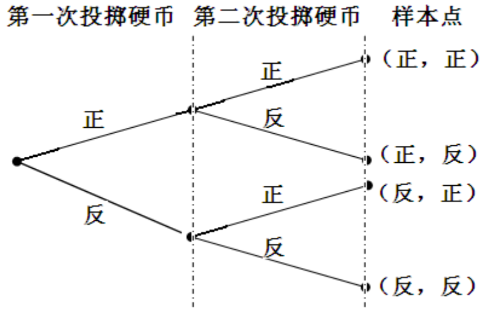


图 4-5 概率树示意图

若我们用字母“H”表示正面向上，“T”表示反面向上，先后两次投掷结果各用一个字母表示，则这种试验的样本空间为：Ω={HH，HT，TH，TT}

由公式 4-4 可得：

事件 A 在样本空间中的可能结果有两个（HH，HT），其概率为：P(A)= 2/4=1/2，事件 B 在样本空间中的可能结果也有两个（HH，TH），其概率为：P(B)=2/4=1/2，在第一次投硬后出现正面的情况下，第二次投硬币的可能结果有{HH，HT}两个，事件 B（HH）占其中一个，因此，事件 B 在事件 A 出现的情况下出现的条件概率为：P(B/A)=1/2，

事件 A 与事件 B 共同出现的结果在该试验的样本空间中只有一个，其概率为：P(AB)=1/4。

从上面的分析可知，事件 B 在事件 A 出现的情况下出现的条件概率 P(B/A)，与事件 B 本身的概率 P(B)是相等的，即都等于 1/2。也就是说，事件 B 的出现与事件 A 是否出现没有关系，事件 A 的出现与事件 B 是否出现没有影响。两个事件共同出现的概率等于两个事件各自概率的乘积，即 $P(AB) = P(A)P(B)$，它们之间是相互独立的。

在概率论中，事件的独立性被定义为：

设 A 与 B 为两个事件，若 $P(AB) = P(A)P(B)$，则称事件 A 与事件 B 相互独立，简称 A，B 独立。

根据上述定义容易得到以下结论：

（1）独立性具有对偶性，即若 A 独立于 B，则 B 独立于 A。

（2） 若 A 与 B 相互独立，且 P(A)>0，则 P(B/A)= P(B)：条件概率等于无条件概率。

（3）若 P(A)=0，则 A 与任意事件 B 相互独立。

以上讨论的是两个事件的独立性。多个事件之间也可以相互独立。例如，三个事件 A，B，C 相互独立是指下列等式同时成立：

$$P(AB) = P(A)P(B);\qquad P(AC) = P(A)P(C),$$
$$P(BC) = P(B)P(C);\qquad P(ABC) = P(A)P(B)P(C)$$

一般的讲，设 $A_1, A_2, \cdots A_n$ 是 n(n≥2)个事件，如果对于其中任意 2 个，任意 3 个，……，任意 n 个事件的积事件的概率，都等于各事件概率之积，则称事件 $A_1, A_2, \cdots A_n$ 相互独立。

由定义，可以得到以下两点推论。

（1） 若事件 $A_1, A_2, \cdots A_n$，（n≥2）相互独立，则其中任意 $k(2 \le k \le n)$ 个事件也是相互独立的。

（2） 若 n 个事件 $A_1, A_2, \cdots A_n$，（n≥2）相互独立，则将 $A_1, A_2, \cdots A_n$ 中任意

多个事件换成它们的对立事件，所得的 n 个事件仍相互独立。

须注意，三个事件之间两两相互独立，不一定意味着这三事件相互独立；反之，若三事件相互独立，则一定两两独立。从而当一组事件的个数超过了两个时，这组事件的相互独立与两两独立不是一个概念。重要的是，两事件相互独立的基本含义是指当它们中有一个事件已经发生时，对另一个事件是否发生的概率不会产生影响的。

第二节 随机变量与概率分布

一、随机变量的含义

人们在作随机试验时，所关心的常常不是试验结果本身，而是对于试验结果联系的某个数值。在有些试验中，试验的结果看起来与数值无关，但我们可以引进一个变量来表示它的结果，也就是说，把试验结果数值化。

例如，掷一枚硬币只有出现正面向上还是反面向上两种结果，对此我们可以引进一个变量 X，用它表示正面出现的次数，当出现正面时，令 X=1；当出现反面时，令 X=0。即：

$$X=\begin{bmatrix}1\ldots\text{掷出正面}\\0\ldots\text{掷出反面}\end{bmatrix}$$

则 X 是一个随机变量。试验中到底取哪个值，要等到掷硬币后才能知道。

可见，**随机变量**（random variable）是试验结果的数值化，它随试验结果的不同而取不同的值。因而在试验之前只知道它可能取值的范围，而不能肯定它将取哪个值。

在概率理论中，一般对随机变量作如下定义：

设 E 为随机试验，$\Omega=\{e\}$ 为其样本空间，若对任意 $e\in\Omega$，有唯一实数 $X(e)$ 与之对应，则称 $X(e)$ 为随机变量。这一定义如图 4-5 所示。

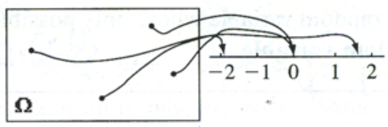


图 4-8 随机变量示意图

从某种意义上说，随机变量实际上是对于某一随机试验中，将试验结果与某一个实数联系起来的一种法则。它是试验结果的数值化，随试验结果的不同而取不同的值，因而在试验之前只知道它可能取值的范围，而不能肯定它将取哪个值。在数学上，一般以大写的字母如 X，Y，Z 等表示随机变量，而在表示随机变量所取的值时，则通常以小写字母如 x，y，z 等来表示。

例如，前面所说的投掷一枚骰子的随机试验，其样本空间是

$\Omega = \{$  $\}$；

当我们用 X 作为该变量的代表符号，把这六个基本事件用与它们的点数相对应的实数，例如，结果出现 1 点时，用“x = 1”表示，结果出现 6 点时，用“x = 6”表示，数值“1，2，3，4，5，6”都是随机变量 X 的具体的值。

引入随机变量后，随机试验中的各种事件，就可以通过随机变量表达出来。从某种意义上说，随机事件这个概念实际上是包括在随机变量这个更广的概念内。也可以说，随机事件是从静态的观点来研究随机现象，而随机变量则是一种动态的观点，就像数学分析中常量与变量的区别那样。概率论作为一门数学分支，在引入随机变量的概念后，对随机现象统计规律的研究，就由对事件及事件概率的研究扩大为对随机变量及其取值规律的研究。

二、概率分布的含义与主要类型

在进行某一种随机试验时，虽然随机变量的取值在试验之前不能预知，但是由于试验的各个结果的出现总是具有一定的概率，因而随机变量的取值也就有一定的概率。我们可以将所有随机试验的结果及与之对应的概率值列在一起，将其称之为该随机实验的**概率分布**（probability distribution）。换言之，概率分布是描述随机变量取值及其概率的变化规律的数学方法。

随机变量按其所有可能的取值的性质，通常分为两大类：

一类叫离散型随机变量。其特征是随机变量的所有取值可以逐个一一列举出来，如某地区的人口数。

另一类叫连续型随机变量，这种变量的所有可能取值不仅是无穷多的，而且不能无遗漏地列举出来，而是充满一个区间。如“电视机的寿命”，“测量误差”等。

与这两类随机变量相对应，随机试验结果的概率分布可分为离散型概率分布和连续型概率分布。

三、离散型概率分布

（一） 离散型随机变量的概率分布

如前所述，当某随机变量的所有可能取值只是有限个或可列无限多个时，我们就称这种随机变量为**离散型随机变量**（discrete random variable）。

在概率理论中，对离散型随机变量的概率分布（probability distribution of discrete random variable，又称为分布率，概率函数，分布列），通常可以定义为：

设离散型随机变量 X，其所有可能取值为 $x_1, x_2, x_3, \cdots x_k, \cdots$，且取各个可能值的概率分别为 $p_1, p_2, p_3, \cdots p_k, \cdots$，即

$$P\{X = x_k\} = p_k,\ k = 1, 2, \cdots$$ （公式 4-8）

如果每一个 p_k 都能满足以下两点，即：$p_k \ge 0,\ k = 1, 2, 3, \cdots$；$\sum p_k = 1$，则称公式 4-8 为离散型随机变量 X 的分布律。

离散型随机变量的分布律也可以象表 4-2 用表格的形式来表示：

表 4-2 离散型随机变量分布表

X	x_1	x_2	x_3	…	x_n	…
p_k	p_1	p_2	p_3	…	p_n	…

表 4-2 直观地表示了随机变量 X 取各个值的概率的规律。X 取各个值各占一些概率，这些概率合起来是 1。可以将此想象成：概率 1 以一定的规律分布在各个可能值上。这就是将上面的表格称为概率分布的缘故。

在心理研究中，常见的离散型随机变量主要包括：0 -1 分布和与二项分布。

概率理论对 0 -1 分布的界定如下：

设随机变量 X 只可能取 0，1 两个值，且取 1 的概率为 p，取 0 的概率为 1-p，（0 ＜ p ＜ 1），则称随机变量 X 服从参数为 P 的 0-1 分布，记为 X ~ B(1，p)，其概率分布为：

$$P(X = 1) = p, \qquad P(X = 0) = 1 - p$$

这种概率分布也可用表 4-3 的形式来表示。

表 4-3 0-1 型随机变量分布表

X	1	0
pk	p	1-p

这种概率分布还可以用公式表示为

$$P(X = k) = p^k (1 - p)^{1-k}$$ （公式 4-9）

若某随机试验 E 只有两个可能结果，或我们仅仅关心相互对立的两类结果，那么只要将其中的一个（或一类）结果对应于数字 1，另外的结果对应于数字 0，于是就可以用 0-1 分布的随机变量来描述有关的随机事件。

（二）贝努里试验与二项分布

设将某试验 E 重复进行 n 次，且各次试验的结果在概率上互不影响，则称这 n 次试验为 **n 重独立试验**。

若一次实验只有两个可能的结果：A 与-A，则称这种试验为贝努里试验（Bernoulli trial）。

显然一个服从 0-1 分布的随机变量，就是贝努里试验下的随机变量。将贝努里试验独立重复 n 次，称之为 n 重独立贝努里试验，简称 n 重贝努里试验。

若在 n 重贝努里试验中，用 X 表示事件 A 出现的次数，则它是一个离散型随机变量，其可能值为 0，1，2，…，n。若每次试验中事件 A 出现的概率均为 p，考虑 X 取值 k 的概率 $P(X=k), k=0,1,2,\cdots,n$，就可导出**二项分布**（binominal distribution）。

设随机变量 X 具有如下面公式所示的概率分布：

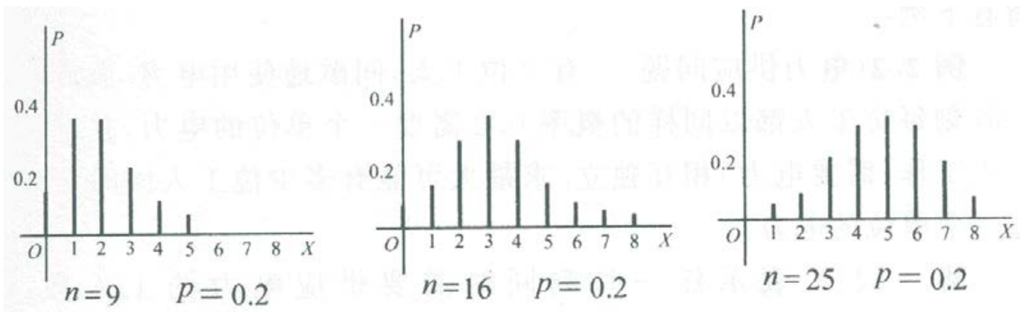
$$P(X=k)=\binom{n}{k}p^k q^{n-k}, k=0,1,2,\cdots,n \qquad \text{（公式 4-10）}$$

其中，$0<p<1, q=1-p$，则称 X 服从参数为 n，p 的**二项分布**，记为 X ~ B（n, p）。

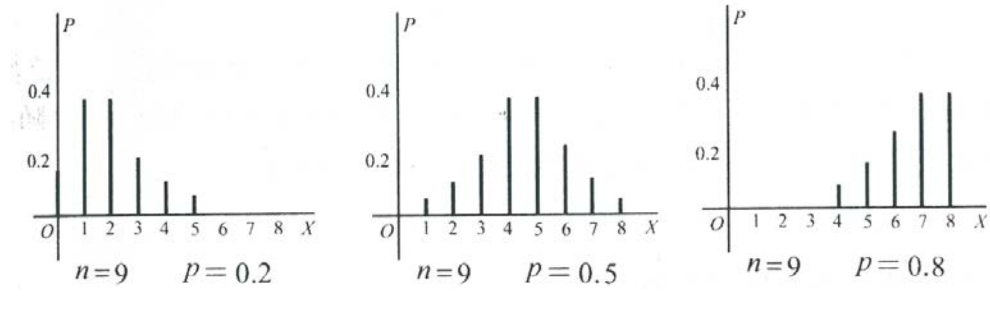
请注意，$\binom{n}{k}p^k q^{n-k}$ 正好是二项式 $(p+q)^n$ 的展开式中出现 p^k 的那一项。

显然，当 n=1 时，二项分布就化为两点分布，即 $P(X=k)=p^k q^{n-k}$，$k=$。0，1。可见，两点分布是二项分布的特例。

下面对某些 n 和 p 的取值，由公式 4-10 计算出 p^k，并作出 X 的概率分布图（图 4-6）：



（1）固定 p 变化 n 时 X 的概率分布图



（2）固定 n 变化 p 时 X 的概率分布图

图 4-6 二项分布图（转引自华中科技大）

给定 n 后，若 p=0.5 则二项分布是对称的，否则就不对称。不对称的分布

朝向它拖尾的方向偏倚：如 p=0.2 的分布朝右边偏倚，p=0.8 的分布朝左边偏倚。

给定 p 后，n 增大则二项分布的偏倚变得不太明显。

另外，诸图均显示，若将各个概率值用一条光滑曲线连接，这条曲线将出现一个单峰，由此可得到二项分布的中心项，即最可能出现次数（其整数部分一般是[（n+1）p]。

四、连续型概率分布

（一） 连续型随机变量的含义

对于离散型随机变量，可以用概率函数描述它取值的概率规律。对于连续型随机变量 *X*，由于它可以取某个区间上的所有实数，这些实数不可能逐个列举出来，而且，它取任意一个确定实数值 α 的概率都是 0，即 p{X= α =0}。由于这两点，我们需要用其他方法来描述连续型随机变量。在概率理论中，有关连续型随机变量的一般定义有如下述：

设随机变量 X 的分布函数 $F(x)$，若存在某非负可积函数 $f(x), x \in (-\infty, = \infty)$，使对一切实数 x，均有

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \text{（公式 4-15）}$$

则称 X 为**连续型随机变量**（continuous random variable），且称 $f(x)$ 为 X 的**概率密度函数**，简称概率密度或密度函数。$f(x)$ 的图形称为**概率密度曲线**。

（二） 正态分布及其特点

在心理研究领域，涉及的连续型随机变量主要是具有正态分布性质的随机变量。在概率理论中，正态分布的定义有如下述。

设连续型随机变量 X 具有概率密度函数：

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2}{2\sigma^2}} \qquad \text{（公式 4-16）}$$

则称 X 服从参数为 μ, σ 的**正态分布**（normal distribution），记为 $X \sim N(\mu, \sigma^2)$（其中 μ 是平均数，σ 是标准差，$\mu \in R$，$\sigma > 0$，这两个概念我们将在下一章参数估计部分进行介绍)。

正态分布的概率密度曲线有如图 4-7 所示：

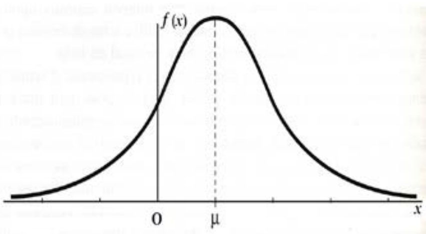


图 4-10 正态分布的概率密度曲线示意图

正态分布的密度函数 F（x）具有如下特点：

（1）正态分布的形状是以直线 $X = \mu$（即过平均数点的垂直线）为对称轴的

对称曲线。

（2）正态分布中，平均数、中数和众数三者相等，均位于对称轴所在的位置上，该点在概率密度曲线中处于最高点，其 y 值最大，然后逐渐对称性地向两侧下降，下降时先向内弯曲，然后向外弯曲，拐点位于正负 1 个标准差处。曲线两端向靠近 X 轴处无限延伸，但永不相交。由于其曲线特点是中间高两边低的对称分布，因此，通常称之为倒 U 形分布或钟形分布（即像一口大钟的外形）。

（3）正态分布是一族分布，它随随机变量的两个参数即理论平均数 μ 和理论标准差 σ 的大小与单位的不同而有不同的形态。

当 σ 固定，μ 变动时，$F(x)$ 的图形不变，但根据理论平均数 μ 的大小而向右或向左移动，这有如图 4-8 所示。

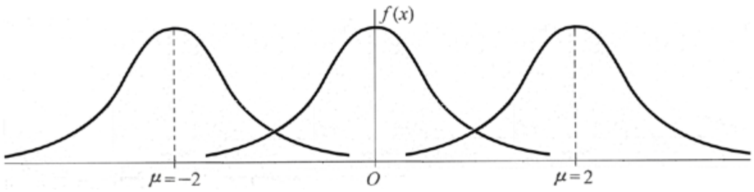


图 4-8 σ 值固定 μ 值不同的正态分布比较图

当 μ 固定，σ 变动时，$F(x)$ 的图形的位置不变，但根据理论标准差 σ 的大小而呈现出不同的形状。如图 4-9 所示。

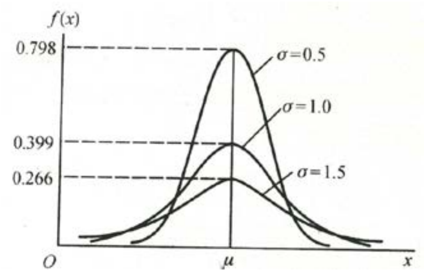


图 4-12 μ 值固定 σ 值不同的正态分布比较图

（4）所有正态分布的曲线下面积都等于 1，而且都可以根据一定的公式转换成标准正态分布。

（5） 当正态分布函数中的 $\mu=0$，$\sigma=1$ 时，称 X 服从标准正态分布。其概率密度函数就是公式 4-12，其分布函数有如下列公式所示：

$$f(x)=\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{Z^2}{2}}dz$$

（公式 4-13）

式中，$Z^2=(\frac{X-u}{\sigma})^2$，$Z$ 的大小随 X 的值而变。

（6） 在正态分布曲线下，标准差与曲线下面积（概率）有一定的数量关系，这有如图 4-10 所示。

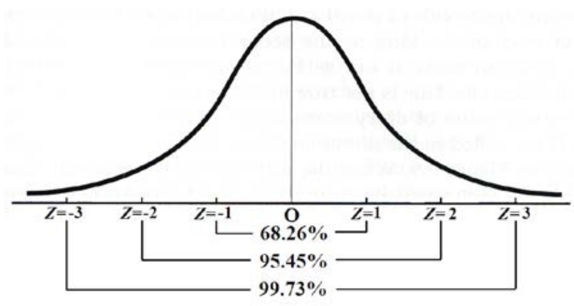


图 4-10 标准差与曲线下面积（概率）关系示意图

第三节 正态分布在心理研究中的应用

一、标准正态分布表的构成

概率论和数理统计学领域的专家已编制了标准正态分布表供使用者使用。任何一般的正态分布均可通过下列标准分数的计算公式而转换成标准正态分布，转换后其各项性质均保持不变。

$$Z = \frac{x - u}{\sigma} \qquad \text{（公式 4-14）}$$

概率论与统计学家们根据不同的需要制定出不同的正态分布表：从所包含的曲线下面积看，有的正态分布表的 Z 值所包含范围是“-4”到“0”再到“+4”，而有的 Z 值范围只包括“0”到“+4”；从所包含的内容看，有的只包含 Z 值和该 Z 值到平均数（即“Z=0”）之间的曲线下概率面积 P 值；有的除上述两列外，还包括用包括用 0.50 减去该 Z 值到平均数之间的概率值后所剩余的概率值，有的除 Z 值和 P 值外，包括 Z 值所在位置的 Y 值。

在我国心理研究领域中，大多数采用的是从 Z=0 到 Z=3.99，内容包括 Z 值、Y 值和由平均数到 Z 值的 P 值三方面数值的标准正态分布表，本书后面所附的标准正态分布表（附表 1）除包含上述三者外，还加了一列 Z 值之外的 P 值，其结构有如表 4-4 所示。

表 4-4 标准正态分布示例表

Z	Y	P（平均数到Z）	P（超过Z）
.00	.39894	.00000	50000
.01	.39892	.00399	.49601
……	……	……	……

续表

Z	Y	P（从平均数到 Z 值）	P（超过 Z 值）
1.00	.24197	.34134	.15866
1.01	.23955	.34375	.15625
……	……	……	……
2.00	.05399	.47725	.02275
2.01	02592	.47778	.02222
……	……	……	……
3.99	.00014	.49997	.00003

在心理研究中，假设很多现象是服从正态分布的，因此，通过标准正态分布表，我们就可以从所获得的研究数据中得到更多有用的结果。

二、依据 Z 分数值求概率 P 值

在心理研究中，主要有以下两类问题可以通过 Z 分数值来求概率 P 值，一类是求某一个 Z 值在它所在总体中的相对位置，另一类问题是求两个 Z 值之间的概率 P 值。

【例 4-3】 一位心理学研究者用韦氏儿童智力量表对两位同龄小学生进行测试后，确定这两位学生总的智商分数分别为 100 分和 115 分，试对这两个分数作出解释。

解： 就这两个分数而言，除了能确定分数为 115 分的学生比分数为 100 分的学生智商更高外，似乎没有其他更多的信息。假如人们的智力分数确实服从正态分布，则通过标准正态分布表，我们可以得到如下所示更多的信息。

我们知道韦氏儿童智力量表中的标准智商分数是根据各年龄段儿童抽样样本的数据经标准化后得出的，其平均数定为 100 分，一个单位的标准差定为 15 分。将上述两个学生的智商分数代入求 Z 值的公式后，就可以得到各自的 Z 分数值：

$$学生_A : Z = \frac{(100-100)}{15} = 0 \qquad 学生_B : Z = \frac{(115-100)}{15} = 1$$

与这两个 Z 值对应的正态分布图有如图 4-11 所示。

由图 4-11 可知，由于这两位学生所在群体的智商的平均数是 100 分，在正态分布曲线中，平均数处于曲线下面积中的最中间位置，因此，学生 A 得到 100 分的智商分数就意味着他处于同龄人智商的中间位置，查正态分布表，对应于 Z=0 的 P 值也是“0”（这是因为本书中所附的正态分布表是从曲线的中点起算），这意味着智商分数比他更高或比他更低的人在总人数中所占的比例都是 50%；

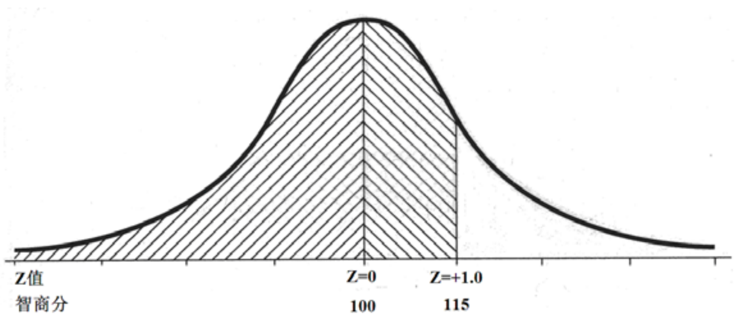


图 4-11 两位儿童智商分数与正态分布曲线下面积关系示意图

学生 B 得到 115 分的智商分数，由于一个单位的标准差是 15 分，因此，通过求 Z 值的公式可得到他的 Z 值是“1”，查正态分布表第三列，对应于 Z=1 的 P 值是“0.34134”，在正态分布曲线下面积中，这是从平均数 100 分所在点到 115 分所在点之间的面积，加上平均数所在点左边 50%的正态分布曲线下面积，则意味着智商比他低的同龄群体的人在总人数中所占的比例大约是 84%；从正态分布表第四列可知，比他智商高的同龄人在总人数中所占的比例约为 16%。显然，这样获知的信息量与没有经过查正态分布表之前的信息量相比要更大，能更清楚地表明这两位学生的智商在同龄群体中的地位。

答：学生 A 的总智商处于中等水平；学生 B 的总智商处于较好的水平，同龄学生中智商比他更高的人数不超过 16%。

在心理研究中，可以利用上述原理来确定特定分数界限内的考生人数

【例 4-4】 假定 500 个学生某科成绩分布近似正态分布，其 $\mu=70$，$\sigma=10$，问：①75 分以下有多少人？②85 分以上有多少人？

解：

（1）分别求出这两个分数的 Z 值：

$$Z_1=\frac{75-70}{10}=0.5，\qquad Z_2=\frac{85-70}{10}=1.5$$

（2）根据正态分布表求对应的曲线下面积

在 $Z_1=0.5$ 时，正态分布表第三列中对应的值是 0.19146，这是正态曲线的中线到右边 0.5 个标准差的面积，这一点左边所有的分数都低于 75 分，因此，还要加上曲线下面积左边的 50%，即：

$$P_1 = 0.50 + 0.19146 = 0.69146$$

本例第二问是希望知道高于 85 分的人数比例，其含义是 Z 值所在点右边曲线下所含的面积，因此，在 $Z_2=1.5$ 时，应该查正态分布表第四列中的数值，对应的值是 0.06681，这也是高于 85 分的人数比例。

（3）根据相应的曲线下面积求低于 75 分和高于 85 分的人数：

低于 75 分的实际人数为500 × 0.69146 = 346人。

高于 85 分的实际人数为500 × 0.06681 = 33人。

答：这 500 个学生某科成绩在 75 分以下的有 346 人，85 分以上的有 33 人。

【例 4-5】 两位大学二年级的学生在参加大学英语四级统考后，分别获得 72 分和 75 分。已知该校（或该市、该省…）全体考生的平均成绩是 80 分，标准差为 4 分，试求介于 72-75 分之间的考生占全体考生的比例。

解：

（1）分别求出与这两个学生成绩对应的 Z 值：

$$学生_A : Z = \frac{(72-80)}{4} = -2; \qquad 学生_B : Z = \frac{(75-80)}{4} = -1.25$$

（2）根据正态分布表求对应的曲线下面积，

学生 A 的分数的 Z 值是“-2”，在附表 1 的标准正态分布表中，并没有列出负 Z 值及相应的 P 值等内容。我们知道，正态分布曲线下面积的特点之一，是以中点为界两边对应相等，因此，要查负的 Z 值可以通过查相对应的正的 Z 值来求其面积。在正态分布表第三列中，对应于 Z=2 的 P 值“0.47725”，这意味着从正态分布曲线的中点到 Z=2 或 Z=-2 的面积面都是 0.47725；学生 B 的 Z 值是“-1.25”，查正态分布表第三列中，对应于 Z=1.25 的 P 值是“0.34935”，在正态分布曲线下面积中，这也是从平均数所在的中点位置到 Z=1.25 或 Z=-1.25 所在点的面积。

（3）根据相应的曲线下面积求这两个学生成绩之间的人数比例，由此可得到这两个 Z 值所对应的曲线下面积之差为：

$$P = 0.47725-0.34935 = 0.1279$$

这有如图 4-12 所示。

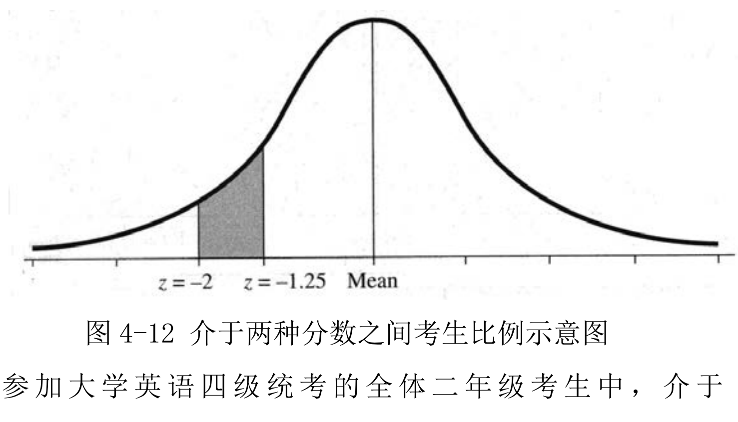


图 4-12 介于两种分数之间考生比例示意图

答：在该校参加大学英语四级统考的全体二年级考生中，介于 72-75 分之间的考生约占全体考生的 12.79%。

知道 Z 值，通过查正态分布表就可以查到相应的 P 值。它们两者的关系实际上是通过公式 4-12 计算后得来的。由于现在计算机的使用非常普遍，考虑到

用公式 4-12 的积分表达式来计算不很方便，因此，有统计学家给出了下面两个近似计算公式来计算 P 值（参见盛骤等，2001）。

（1） 当 Z ＜ 0 时，

$$P=\frac{1}{\sqrt{2\pi}}\times\frac{1}{e^{\frac{Z^2}{2}}}\times(a_1t+a_2t^2+a_3t^3) \quad \text{（公式 4-15）}$$

（2） 当 Z ＞ 0 时，

$$P=1-\frac{1}{\sqrt{2\pi}}\times\frac{1}{e^{\frac{Z^2}{2}}}\times(a_1t+a_2t^2+a_3t^3) \quad \text{（公式 4-16）}$$

在上面两个公式中，$\frac{1}{\sqrt{2\pi}}=0.3989422804$，常量 e=2.718281828，a1，a2 和 a3 也都是常量，其中：$a_1=0.4361836$，$a_2=-0.1201676$，$a_3=0.9372986$，Z 是唯一的变量，t 可以通过下面的公式求出：

$$t=\frac{1}{1+0.33267\times|Z|} \quad \text{（公式 4-17）}$$

经过公式 4-15 或公式 4-16 计算后所得到的结果，与通过正态分布表所查到的 P 值非常接近（注意，正态分布表中，P=0 意味着两边面积各占 50%。）例如：

（1） 当 Z=0 时，代入公式 4-17 可得 t=1，再将这些值代入公式 4-15，可求得相应的 P 值：

$$P=0.3989422804\times1\times(0.436183-0.1201676+0.9372986)$$
$$=0.5000001846$$

（2）当 Z=-1 时，代入公式 4-17 可得 $\frac{1}{e^{\frac{(-1)^2}{2}}}=0.6065306597$；

$$t=\frac{1}{1+0.33267\times|-1|}=\frac{1}{1.33267}=0.7503733107$$

a_1t = 0.4361836 x (0.7503733107) = 0.327300532

a_2t^2 = 0.1201676 x $(0.7503733107)^2$ = 0.06766158152

a_3t^3 = 0.9372986 x $(0.7503733107)^3$ = 0.3960136031

再将这些值代入公式 4-15，可求得相应的 P 值：

$$P=0.39894\times0.60653\times(0.3273-0.06766+0.396)=0.158645$$

（3） 当 Z=1 时，将 Z=-1 的计算结果代入公式 4-16，可得 P=1-0.158645=0.841355，与查表所求的 P=0.34134 再加上 50 等于 0.84134 的 P 值非常接近。

三、从概率 P 值求 Z 分数

有时候,我们需要从确定的人数比例中求取相应的 Z 分数,并通过 Z 分数来确定相应的分数值。

【例 4-6】 假定我们在使用韦氏儿童智力测验时，将高智商的范围定在高分端 2%之内，那么，在韦氏儿童智力量表测验后，要达到多少分才能算是高智商呢？

解：显然，这是一个与上述第一种从 Z 分数求概率问题相应的逆运算问题。在解这一问题时，我们可以从标准正态分布表第四列中找到与 2%相对应的 P 值，该列没有正好等于 0.02 的 P 值，与此接近的有：0.02018（对应的 Z 值是 2.05）和 0.197（对应的 Z 值是 2.06），显然，Z=2.05 的 P 值更为接近，这时，我们可以用计算标准分 Z 值的公式来计算区分属于 IQ 高分区 2%的 IQ 分 X，即：

由 $Z=\frac{(X-u)}{\sigma}$ ，IQ 的平均分 u=100，标准差 σ=15，得：

$$X = Z\sigma + u = 2.05 \times 15 + 100 = 130.75$$

如果需要更精确地得到与 P 值等于平均数右边 48%相对应的 Z 值，可以用以下的插值法来进行计算：

$$Z = 2.05 + \frac{0.48 - 0.47982}{0.48030 - 0.47982} \times (2.06 - 2.05) = 2.054$$

$X = Z\sigma + u = 2.05375 \times 15 + 100 = 130.806$（见图 4-13）

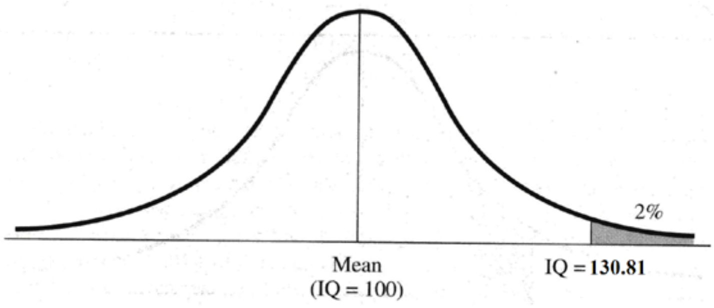


图 4-13 由概率求 Z 值示意图

答：在韦氏儿童智力量表测验中要达到 130.81 分才能算是高分端 2%之内的高智商。

在心理研究中可以利用上述原理来确定录取分数线和分析试题的难度

【例 4-10】 某县对初一年级 1000 名学生进行能力测验，结果 $\mu=75$，$\sigma=10$，现拟根据此结果选取 25 名学生作为“尖子班”重点培养，假定测验成绩近似正态分布，问测验分数在多少分以上才能被选到“尖子班”学习？

解：

（1） 求尖子班拟录取人数在总人数中的比例：

$$P=\frac{25}{1000}=0.025,$$

（2） 再用 $P=0.025$ 查正态分布表第四列求对应的 Z 值为 1.96（见图 4-14）；

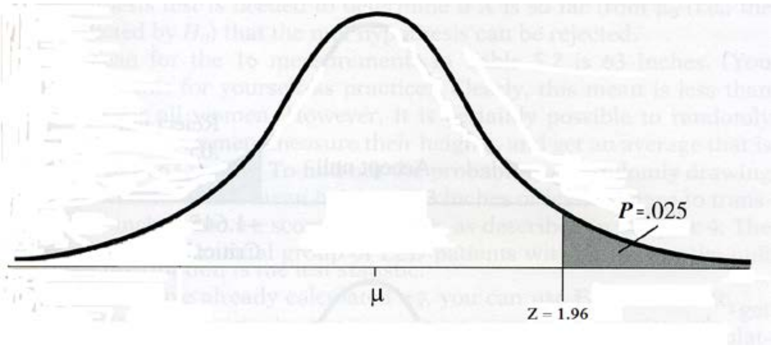


图 4-14 由概率求 Z 值示意图

（3）根据公式 4-8 可知，$X=1.96\times10+75=94.6$。

答：在本例条件下，测验分数要在 94.6 分以上的学生才可以录取到“尖子班”学习。

有时候，心理研究会碰到类似例 4-4 那种问题的逆运算问题，即需要从正态分布曲线下某部分面积的 P 值确定该 P 值上下端的两个 Z 值及相应的分数。

【例 4-8】 已知某校大学二年级的学生在参加大学英语四级统考后，全校考生的平均成绩是 80 分，标准差为 4 分，若以 10%作为高分，10%作为低分，试求中间 80%考生的成绩应该介于多少分至多少分之间。

解：假设考生的成绩服从正态分布，则这一问题是由正态分布曲线下面积求解相对应的 Z 值的问题。由于附表 1 列出的是从曲线下面积中间点向右的 50%的面积值，并且曲线下面积值两边对应相等，因此，要找中间 80%的面积，只要找到 Z 值外侧的面积是 10%（20%/2=10%）即 0.10 相对应的 Z 值就行。查正态分布表第四列的 P 值可知，与 P = 0.10 最接近的 Z 值是 1.28，因此，相应的分数是：

$$X = Z\sigma + u = 1.28 \times 4 + 80 = 85.12$$

因为正态分布曲线下面积值两边对应相等，因此，低于平均数 40%的 Z 值就是-1.28，相应的分数是：

$$X = Z\sigma + u = -1.28 \times 4 + 80 = 74.88$$

由此可知，介于平均数两边 80%的考生的分数应该在 74.88 – 85.12 分之间，低于 74.88 分的考生和高于 85.12 分的考生各占全体考生的 10%（见图 4-15）。

答：某校大学二年级学生在参加大学英语四级统考后，中间 80%考生的成

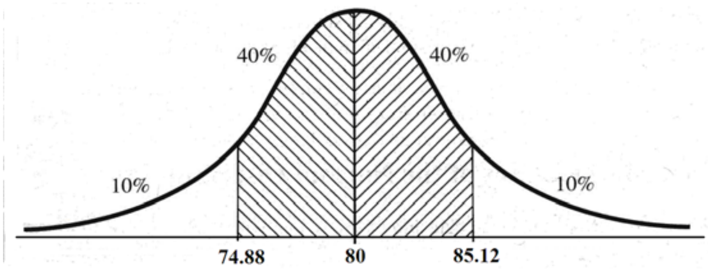


图 4-18 由预设考生比例求两端分值示意图

绩介于 74.88 分至 85.12 分之间。

四、求概率密度 Y，即正态曲线的高

在心理研究中，有时需要根据 Z 值或概率 P 值来求相应的 y 值。如果已知 Z 值，可以从正态分布表查第一列找到与该 Z 值相对应的第二列的 Y 值；如果已知概率 P 值，则可以从正态分布表查第三列找到与该概率 P 值对应的第二列 Y 值。例如：

（1）已知 Z=0，在正态分布表中根据第一列的值查到 0 后，相应的第二列的值是：Y = 0.3989；

（2）已知 Z=2，在正态分布表中根据第一列的值查到 2.00 后，相应的第二列的值是：Y = 0.05399；

（3）已知概率值 P 是 0.35314，在正态分布表中根据第三列的值查到 0.35314 后，相应的第二列的值是：Y = 0.22988。

在心理与教育研究中，概率密度 Y 的作用是帮助我们通过求两变量二列相关来进行考试试题区分度的分析。

小　结

本章第一节在讨论随机现象、随机实验及各事件相互关系的有关规律基础上，引出了概率的定义。而后对诸如概率的加法公式、条件概率与乘法公式等作了介绍。

本章第二节重点阐述有关随机变量与概率分布的基本原理，其中，离散型概率分布是进行实计数推断统计的基础，而正态分布在心理统计中的应用几乎涵盖推断统计的所有领域。因此，本章第三节对标准正态分布表的构成及正态分布在心理与教育研究中的主要应用作了较为详尽的介绍。

概率与概率分布是心理学研究中进行推断统计的基础。以后各章所阐述的所有各种参数估计或假设检验，都是在不同的概率分布基础上进行的。因此，虽然初学者对本章的内容也许会觉得太难，但还是需要通过反复的学习去理解其内容，同时，只要用心学习，在现有数学知识基础上还是可以掌握其基本原理的。

关键术语

随机现象：在个别试验中其结果呈现出不确定性，但是在大量重复试验中，其结果又具有统计规律性的现象。

随机试验：通过对随机现象进行测试后获取相应的数据结果，进而揭示该随机现象有何规律的实验。它具有以下一些特点：(1) 试验可以在相同条件
下大量重复地进行；（2）每次试验的可能结果不止一个，并且在试验之前能知道试验的所有可能结果；（3）在进行一次试验之前不能确定哪一个结果会出现，但若进行大量重复试验的话，其可能结果出现又有一定的统计规律性。

概率：是表示随机事件出现可能性大小的客观指标。通常把概率区分为“先验概率”和“后验概率”两种类型。

后验概率：设随机事件 A 在相同的条件下进行的 n 次试验中发生了 n_A 次，则称 n_A
是事件 A 在这 n 次试验中发生的频数，称比值 n_A/n 为事件 A 在这 n 次试验中发生的频率，记成 $f_n(A)$。当试验次数 n 有充分大时，可以把较为稳定的事件 A 的频率值即 $f_n(A)$ 作为事件 A 出现可能性的概率。

先验概率：若试验结果一共由 n 个基本事件 $E_1, E_2, \cdots E_n$ 组成，并且这些事件的出现具有相同的可能性，而事件 A 由其中某 m 个基本事件 $E_{i1}, E_{i2}, \cdots E_{im}$ 组成，经公式 $P(A) = \dfrac{A\text{中所含的基本事件数}}{\text{样本空间}\Omega\text{中的基本事件总数}} = \dfrac{m}{n}$ 计算后所得到的事件 A 的概率。

条件概率：在考虑两个事件相互关系时，事件 A 已发生的条件下事件 B 出现的概率。

随机变量：指试验结果的数值化，它随试验结果的不同而取不同的值。

概率分布：将所有随机试验的结果及与之对应的概率值列在一起，称之为该随机实验的概率分布。

离散型随机变量：当某随机变量的所有可能取值只是有限个或可列无限多个时，我们就称这种随机变量为离散型随机变量。

二项分布：设随机变量 X 具有如下概率分布 $P(X = k) = \dbinom{n}{k} p^k q^{n-k}$， k =0, 1, 2, …, n，其中，$0 < p < 1, q = 1 - p$，则称 X 服从参数为 n，p 的二项分布，记为 X ~ B（n, p）。

连续型随机变量：设随机变量 X 的分布函数 $F(x)$，若存在某非负可积函数 $f(x)$，$x \in (-\infty, = \infty)$，使对一切实数 x，均有 $F(x) = \int_{-\infty}^{x} f(t)dt$，则称 X 为**连续型随机变量**。

正态分布：设连续型随机变量 X 具有概率密度函数 $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2}{2\sigma^2}}$ ，则称 X 服从参数为μ,σ的正态分布，记为$X \sim N(\mu,\sigma^2)$（其中μ，σ是参数，$\mu \in R$，$\sigma > 0$）。

重要公式：

概率：$P(A) \approx \frac{n_A}{n}$

古典概率：$P(A) = \frac{A\text{中所含的基本事件数}}{\text{样本空间}\Omega\text{中的基本事件总数}} = \frac{m}{n}$

概率的加法：$P(A_1 \cup A_2 \cup \ldots \cup A_n) = P(A_1) + P(A_2) + \ldots + P(A_n)$

条件概率：$P(B / A) = \frac{P(AB)}{P(A)}$

概率的乘法：$P(AB) = P(A)P(B / A)$， $P(A) > 0$

离散型随机变量的概率分布：$P\{X = x_k\} = p_k,\ k = 1, 2, \cdots$

二项分布：$P(X = k) = \binom{n}{k} p^k q^{n-k}$， k =0, 1, 2,...,n,

离散型随机变量的概率密度函数：$F(x) = \int_{-\infty}^{x} f(t)dt$

正态分布的概率密度函数： $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-u)^2}{2\sigma^2}}$

标准分数：$Z = \frac{x-u}{\sigma}$

标准正态分布的概率密度函数：$f(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{Z^2}{2}} dz$

思考与练习：

1. 在一次智力测验中，某小组 10 位同学中有两位同学被测定为高智商。一位研究者设计的实验要求每次有两位同学参加，试求下列事件的概率：

（1）抽取的两个同学都是高智商者；

（2）抽取的两个同学都不是高智商者；

（3）抽取的两个同学中一位是高智商，另外一位不是高智商；

（4）第二次抽取的那个同学是高智商

2.对于一份具有 20 个是非选择题目的试卷，请计算下述事件的概率：

（1）答对 15 道题的概率

（2）答对 12-15 道题的概率

（3）答对 15 道题以上的概率

3.已知某年级全体考生在一次英语考试中的平均分是 86 分，标准差是 4 分。其中两位考生的成绩分别是 82 分和 90 分，请问

（1）这两位考生在该次英语考试成绩中各处于什么位置？

（2）高于 92 分的考生大概占总考生的百分比是多少？

（3）这次英语考试中，成绩位于这两位考生的成绩之间的考生占总考生人数的比例约为多少？

第五章 参数估计与假设检验的基本原理

学习本章内容，将有助于你对以下问题的理解与思考：

1. 抽样分布的含义是什么？
2. 从正态总体中抽样，其样本平均数的抽样分布有何特点？
3. 如何根据样本平均数的抽样分布构建总体平均数的置信区间？
4. 假设检验的基本原理是什么？主要包含哪几个步骤？
5. 假设检验中有哪两类错误？它们有何关系？

1. 某中学从本年度初一新生中随机抽取28名学生，测得他们的平均身高为155cm。若已知该校初一新生历年来的身高标准差为5.24cm，能否用区间估计的方法推论该校本年度初一新生的平均身高。做推论时犯错误的可能性有多大？

2. 某校历年招收新生时都要测其智商，历年新生的智商分数服从正态分布，平均分数为 110 分，标准差为 15 分。从今年新生的总体中随机抽取一个 25 人的样本进行测验，测得其智商平均数为115分，能否据此推论今年新生的总体智商与往年不同？

第一节 抽样分布的基本原理

一、抽样分布的含义与类别

本书第一章曾指出，根据样本的数字特征来推断总体的数字特征的过程，称为**统计推断**(statistical inference)。利用对样本调查的结果去推论总体是统计学的核心内容，这方面的各种统计分析方法通常被统称为**推断统计** (inferential statistics)。推断统计主要包括参数估计和假设检验两大部分内容，其基础都是抽样分布理论。

统计学中所说的**总体** (population)是指所要研究的具有某种特征的一类事物的全体。总体中的每个基本单元叫做**个体** (individual)。总体中所有个体数据的分布称为**总体分布**（population’s distribution)。从总体中抽取一定数量的个体并对其某项特征进行研究的过程叫做抽样过程，简称**抽样** (sampling)。通过抽样过程被抽取的一组个体构成该次研究的一个**样本**（sample)。该样本中所有个体在某项特征上的数据分布称为**样本分布**（sample’s distribution)。

本书第二章介绍了集中量数、差异量数、相关系数的计算方法。如果这些数据量度是根据总体数据计算得到的，则称它们为**总体参数**(population parameter，简称为参数)；如果这些数据量度是根据样本数据计算得到的，则分别被称为样本平均数、样本方差、样本比率或样本相关系数等，统称为**样本统计量**（sample’s statistic)。

在抽样过程中，无论对于有限总体还是无限总体，只要所抽取样本中包含的个体数小于总体所包含的个体数，那么可以抽取的样本就不止一个。根据每次抽样得到的不同样本也都可以计算出各种不同的样本统计量。特别是对于无限总体和包含很多个体的总体，其可能的样本和根据样本计算的统计量都近乎无穷。多次抽样后，每一种样本统计量都可以构成该种样本统计量的分布，如样本平均数的分布，样本方差的分布，样本比率的分布等，心理统计学把这些全部可能的样本统计量的分布统称为“**抽样分布**（sampling distribution)”。

在心理统计学中，通常有两种对抽样分布进行分类的方法：

一种分类的方法是根据样本统计量的内容来分类。前面所述的有关随机变量分布的平均数、中位数、众数、比率以及方差和标准差等数字特征当用做样本统计量时，因为都有相应的抽样分布并分别根据其名称来称呼，因此可以把抽样分布分为平均数的抽样分布、方差的抽样分布或其他样本统计量的抽样分布等；

另外一种分类的方法是根据不同的检验统计量的分布规律而把抽样分布分为Z分布、t分布、F分布和X^2分布等，由于它们是进行推断统计时，各种假设检验方法的数学基础，因此假设检验也就相应地区分为Z检验、t检验、F检验和X^2检验等。

当我们利用样本统计量来对总体参数进行估计或进行假设检验时，就有必要对第三章所述的那些常见的公式，在计算总体参数和样本统计量时所使用的符号做出界定。

总体参数是描述总体情况的数据量度，通常用希腊字母表示，而样本统计量是描述样本情况的数据量度，一般用大写的英文字母来表示。总体中个体的个数通常用英文大写字母“N”来表示，而抽样样本中个体的个数通常用小写字母“n”表示。下面是几种主要的总体参数与样本统计量在计算公式表达上的区别：

（1）总体平均数用 μ（读作 mu）表示，样本平均数则常用 $\overline{X}$ 表示，因此，计算算数平均数的公式在用于计算总体平均数时就成为：

$$\mu = \frac{(X_1 + X_2 + \cdots + X_N)}{N}$$ （公式 5-1）

用于计算样本平均数时就改为：

$$\overline{X} = \frac{(X_1 + X_2 + \cdots + X_n)}{n}$$ （公式 5-2）

（2）总体标准差和总体方差分别用 σ（读作 sigma）和 σ^2 表示，样本标准差和样本方差则分别用 S 和 S^2 表示，因此，计算总体方差和总体标准差的公式为

$$\sigma^2 = \frac{\sum (X - \mu)^2}{N}$$ （公式 5-3）

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$$ （公式 5-4）

而用于计算样本方差和样本标准差的公式则相应地改为：

$$S^2 = \frac{\sum (X - \overline{X})^2}{n}$$ （公式 5-5）

$$S = \sqrt{\frac{\sum (X - \overline{X})^2}{n}}$$ （公式 5-6）

在大多数的统计应用中，待分析的数据集是一个样本，由于总体均数未知，只能用样本中每个个体值与样本平均数之差代替与总体平均数之差。可以证明，这种相对于样本平均数的离均差平方之和必定小于相对于总体平均数的离均差平方之和。为了弥补这一缺点，求离均差平方和的平均数时，不除以 n，而是除以（n-1），称为样本的无偏方差，有时记为 S_{n-1}^2，其公式如下：

$$S_{n-1}^2=\frac{\sum(X-\overline{X})^2}{n-1}$$
（公式 5-7）

$$S_{n-1}=\sqrt{\frac{\sum(X-\overline{X})^2}{n-1}}$$
（公式 5-8）

其中，n-1 称为自由度（degree of freedom）。所谓自由度，就是允许自由取值的项数。公式 5-7 的分子中共有 n 项，因为 $\sum(X-\overline{X})=0$，这 n 项 $(X-\overline{X})$ 中只有 n-1 项可以自由变化，剩余的一项不能自由取值，所以，我们说自由度只有 n-1。

（3）总体相关系数通常用 ρ 表示，样本相关系数则用 r 表示，这时，计算样本相关系数的公式为：

$$r=\frac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{\sum(Y-\overline{Y})^2}}$$
（公式 5-9）

而计算作为总体参数的相关系数的公式则是：

$$\rho=\frac{\sum(X-\mu_X)(Y-\mu_Y)}{\sqrt{\sum(X-\mu_X)^2}\sqrt{\sum(Y-\mu_Y)^2}}$$
（公式 5-10）

对于总体来说，其总体参数是常量。也就是说，如果某个总体中的所有个体都参与某种心理现象的测量，例如，我们对某市所有小学一年级新生都进行智力测验，那么，测验后只能得到唯一的平均数或其他参数，因此说某个总体的参数是常量。

如前所述，由于在实际研究中对总体进行研究往往是不可行的，因此总体参数也通常是未知的，这就需要通过样本统计量所提供的有关信息对总体参数进行估计和推断。样本统计量一方面描述样本本身的分布状况和特征，另一方面也是总体参数的估计量。样本统计量随着抽取样本的不同而变化，是随机变量。例如，我们从某市所有小学一年级新生中随机抽取 100 名新生进行智力测验，并利用这 100 名新生的智力测验成绩来推断该市所有小学一年级新生的智力水平。但是，我们也可以在该市小学一年级新生的总体中，另外再抽取 150 名学生进行智力测验，同样可以根据他们的成绩来推断总体的智力水平。理论上我们还可以抽取更多的样本，而根据每个样本算出的平均智力水平通常又是不一样的。可见，样本统计量随着抽取样本的不同而变化，是随机变量。

下面，我们以正态总体样本平均数的抽样分布为例来解释样本统计量的概率分布。

二、正态总体样本平均数的抽样分布

从正态总体中，采取有放回的抽样，理论上可以抽取无数个容量为 n 的样

本。例如，如果我们想通过新生入学考试时的成绩来了解某市所有高中一年级学生的英语学习水平，我们可以预先确定抽取多少学生来反映全市高中一年级新生的英语学习水平，如果我们预先确定$n=50$，那么随机抽取的 50 名学生并计算他们入学英语考试成绩的平均分数；计算后把这些学生放回到总体中，我们再一次随机抽取 50 名学生，用同样的方法得到这 50 名学生入学英语考试的平均分数。由于第二次抽样所抽取的 50 人与第一次抽样所抽取的 50 人一般很少重叠，因此，这两个反映该市新生入学英语学习水平的平均分数也会有所不一样。从理论上说，我们可以抽取无数个容量为 50 的新生样本，那么，也就有无数多个入学英语考试的平均分数。也就是说，从同一总体中抽样，根据每个样本计算出的各种统计量通常并不完全相等，由不同抽样得到的样本平均数之间是有差异的。与此类似，不同抽样的样本标准差之间、样本比率之间以及其他样本统计量之间也总是存在一定的差异

下面以样本平均数为例,来了解一下当在某个总体中进行多次随机抽样后每次抽样计算出的样本平均数的分布情况。

从理论上可以证明，若$X_1,X_2,\cdots,X_n$是来自总体$N(\mu,\sigma^2)$的一个样本，则样本平均数$\overline{X}$的分布仍然是正态的，其均值仍为μ，其方差为$\dfrac{\sigma^2}{n}$，标准差为$\dfrac{\sigma}{\sqrt{n}}$，即：$\overline{X}\sim N(\mu,\dfrac{\sigma^2}{n})$。

要特别注意的是，根据一个抽样样本计算的统计量不仅与根据其他抽样样本计算的同一统计量有差异，而且根据每个样本计算的统计量与总体参数之间通常也会存在一定的差异，这种由抽样本身所造成的差异，叫做**抽样误差**（sampling error)。在统计学中，单个样本统计量的误差大小是很难估计的，但是可以度量全部可能的样本统计量的平均误差。例如，在“想通过新生入学考试成绩来了解某市所有高中一年级学生的英语学习水平”的例子中，假如我们抽取容量为 50 的样本，那么我们可以通过公式 5-2 即$\overline{X}=\dfrac{(X_1+X_2+\cdots+X_n)}{n}$（注意这里的$n=50$）得到一个反映这 50 人入学英语平均成绩的平均数，通常称为样本平均数；重复 30 次同样的抽样工作（理论上可以重复无数次)，就可以得到 30 个相互之间互有差别的样本平均数，我们可以把这 30 个平均数分别计为：$\overline{X}_1,\overline{X}_2,\cdots\overline{X}_{30}$。如果我们再次用公式 5-2 即$\overline{X}=\dfrac{(X_1+X_2+\cdots+X_n)}{n}$来计算$\overline{X}_1,\overline{X}_2,\cdots\overline{X}_{30}$这 30 个通过 30 次而得到的样本平均数的平均数（注意这时的样本容量是$n=30$），我们又可以得到一个新的平均数，其含义显然与这 30 个样本平均数的含义不同，这 30 个样本平均数，每一个反映的都是某一次抽样中所抽取的 50 人的英语入学考试成

绩的平均数，而这个新的平均数反映的是这 30 个抽样平均数的平均数，在此我们用符号 $\overline{X}_{总平均}$ 来表示，那么，本例中的 $\overline{X}_{总平均}$ 的计算过程就是：$\overline{X}_{总平均}=\frac{(\overline{X}_1+\overline{X}_2+\cdots+\overline{X}_{30})}{30}$。数理统计学家告诉我们，从理论上说这样的抽样可以进行无数次，这无数多个样本平均数 $\overline{X}$ 的分布仍然是正态的，其抽样平均数的平均数即 $\overline{X}_{总平均}$ 就会等于 μ。

综上所述，在某个总体中每次随机抽取样本容量为“n”的个体，计算每一次抽样的平均数，从该总体中多次随机抽样后各次抽样得出的平均数的分布就构成“**样本平均数分布**（distribution of sample’s means）”

另一方面，每次抽取容量为 $n=50$ 的样本，样本中的每个数据都与该次抽样的平均数有一定的差距，通常用公式 5-6 即 $S=\sqrt{\frac{\sum(X-\overline{X})^2}{n}}$ 计算的标准差来代表该组数据的平均差距。例如，第一次抽样样本的标准差是：$S_1=\sqrt{\frac{\sum(X-\overline{X}_1)^2}{50}}$（注意这里的 $n=50$）；同时，在 $\overline{X}_1,\overline{X}_2,\cdots\overline{X}_{30}$ 等 30 个样本平均数中，每一个都与抽样平均数的平均数 $\overline{X}_{总平均}$ 会有一定的差距，我们可以再通过公式 5-4 即 $S=\sqrt{\frac{\sum(X-\overline{X})^2}{n}}$ 的原理来计算这 30 个抽样平均数分布的标准差（注意这里的样本容量是 $n=30$）。显然，通过这种方式计算得到的样本平均数的标准差与每一次抽取 50 人的数据所计算的样本标准差即 $S_1,S_2\cdots S_{30}$ 是不一样的，与总体的标准差 σ 也是不一样的，它反映的是多个样本平均数之间的差异。为了将样本平均数分布的标准差与总体标准差做出区分，通常把样本平均数分布的标准差称为标准误差（standard error，简称标准误），一般用符号 $SE_{\overline{X}}$ 来表示，有时也记作 $\sigma_{\overline{X}}$。在本例中，$SE_{\overline{X}}$ 的计算方法为：

$$SE_{\overline{X}}=\sqrt{\frac{\sum(\overline{X}_i-\overline{X}_{总平均})^2}{n}}=\sqrt{\frac{\left(\overline{X}_1-\overline{X}_{总平均}\right)^2+\left(\overline{X}_2-\overline{X}_{总平均}\right)^2+\cdots+\left(\overline{X}_{30}-\overline{X}_{总平均}\right)^2}{30}}$$

虽然我们不可能像上述那样通过抽取无数次样本的方式来计算样本平均数 $\overline{X}$ 的标准误差，但数理统计学家告诉我们，它与总体标准差之间具有如下关系：

$$SE_{\overline{X}}=\frac{\sigma}{\sqrt{n}} \qquad （公式 5-11）$$

由公式 5-11 可知，一方面，样本平均数的标准误 $\sigma_{\overline{X}}$ 与总体标准差成正比，即总体标准差越大，样本平均数的标准误也越大；反之，总体标准差越小，样本平均数的标准误也越小。另一方面，样本平均数的标准误 $SE_{\overline{X}}$ 与样本容量 n 的算术平方根成反比，样本容量越大，样本平均数的标准误反而越小；反之，样本容量越小，样本平均数的标准误则越大。由于样本容量 n 通常总大于

1，所以样本平均数的标准误总是小于总体标准差，从图 5-1 中可以看出，随着样本容量的增大，样本平均数的标准误差逐渐缩小。

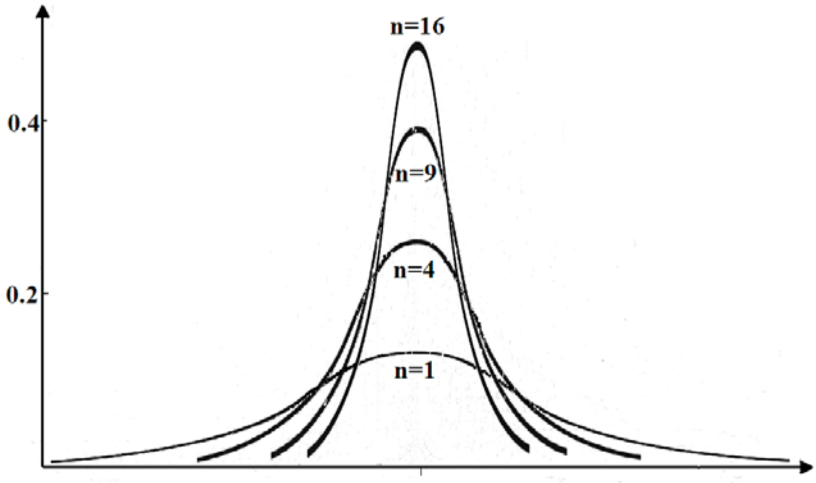


图 5-1　不同样本容量时 $\overline{X}$ 的抽样分布曲线

样本统计量的标准误差的概念，是进行推断统计最重要的数学基础，后面的内容中，针对总体平均数的各种不同的推断统计方法之所以有不同的计算公式，就在于样本统计量的抽样分布不一样。换言之，样本统计量的抽样分布不一样，就有不同的计算标准误差的方法。只有正确理解样本统计量在不同抽样分布情况下计算标准误差的方法，才能更好地达到对某种心理现象得出推断统计结论的目的。

三、Z统计量的分布

由第四章的内容可知，对于任意一个服从正态分布的随机变量，我们都可以通过线性变换把它变换为标准正态分布。例如，设随机变量 $X \sim N(\mu,\sigma^2)$，只要利用公式 4-14 令 $Z=(x-\mu)/\sigma$，所得的随机变量 Z 就服从标准正态分布，即 $Z \sim N$（0,1),转换后其各项性质均保持不变。因此,标准正态分布也称为 Z 分布(Z distribution)。

同样地，如果样本平均数 $\overline{X}$ 服从正态分布，即 $\overline{X} \sim N(\mu,\frac{\sigma^2}{n})$，则样本平均数的 Z 分数也服从标准正态分布,并且构成我们进行 Z 检验的基本指标。所谓 Z 统计量(statistic of Z test）是指**在 Z 检验中用以判断接受还是拒绝虚无假设**的检验统计量，其计算公式为：

$$Z_{\overline{X}}=\frac{\overline{X}-\mu}{SE_{\overline{X}}} \qquad （其中 SE_{\overline{X}}=\frac{\sigma}{\sqrt{n}}） \qquad （公式 5-12）$$

所谓 Z 检验（Z test），则是指根据检验统计量是否服从或渐近服从正态分布规律的假设检验方法。换言之，利用正态分布规律来进行的假设检验过程就是 Z 检验。

由于 Z 统计量服从标准正态分布，因此，第四章所述有关 Z 分数的应用原理同样可以应用到样本统计量的相应计算上。

【例 5-1】某地区成年男性的体重服从正态分布，平均体重为 75kg，标准差为 10kg。若从该地区成年男性中随机抽取 n=30 的样本，则样本平均数应该服从怎样的抽样分布？样本平均数超过 78kg 的概率有多大？

解：因为体重一般服从正态分布，且总体方差已知，样本平均数服从

$$N(\mu,\frac{\sigma^2}{n})$$

样本平均数（抽样分布）的标准误 $SE_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{10}{\sqrt{30}}=1.83$

所以样本平均数服从均值为 75，标准误差为 1.83 的正态分布，即 N（75，1.83^2）。

由 于 $Z_{\bar{x}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，所以可以通过公式 5-12 来计算平均体重在 78kg 以上的概率所对应的 Z 分数值：

$$Z_{\bar{x}}=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{78-75}{1.83}=1.64$$

查 Z 值表，当 Z=1.64 时，大于该 Z 值的概率为 0.0505。

答：样本平均数服从 N（75， 1.82^2），30 名成年男性平均体重超过 78kg 的概率为 0.0505。

【例 5-2】已知某能力测验分数总体服从正态分布（μ=100，σ=15），随机抽取 25 名适龄青少年进行该能力测验，问该样本组能力测验平均分数超过 105 分的概率是多少？

解：能力测验分数总体服从正态分布，总体方差已知，样本平均数服从

$$N(\mu,\frac{\sigma^2}{n})$$

样本平均数的标准误 $SE_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{15}{\sqrt{25}}=3$

由于 $Z_{\bar{x}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$，所以可以通过公式 5-12 来计算该样本组能力测验平均分数超过 105 分概率所对应的 Z 分数值：

$$Z_{\bar{x}}=\frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}}=\frac{105-100}{3}=1.67$$

查表可知，当 Z=1.67 时，大于该值的概率约为 0.0475。

答：该样本组能力测验的平均分数超过 105 分的概率约是 0.0475。

总之，在心理学研究中，当需要通过样本对总体进行分析时，就必须知道该样本统计量的抽样分布规律。上面所说的是样本平均数的抽样分布规律，其他样本统计量也都有各自的分布规律，只有知道样本统计量的分布规律，才可以用它们来对总体参数进行推论，即进行对总体参数的区间估计和假设检验。在心理学研究中，常用的抽样分布主要有 Z 分布、t 分布、χ^2 分布和 F 分布等，对应的假设检验方法也分别称为 Z 检验、t 检验、χ^2 检验和 F 检验。

由样本推断总体，是统计学的核心内容，其两个基本问题是参数估计和假设检验。下一节，我们将阐述有关参数估计的基本原理。

第二节　参数估计的基本原理

一、参数估计的意义与类别

利用样本统计量的信息估计总体参数的过程称为参数估计（parameter estimation)。通常把参数估计分为点估计和区间估计。下面我们分述之。

（一）点估计

点估计（point estimation）就是用样本统计量的单一数值估计未知的总体参数。例如，我们用样本的平均数 $\overline{X}$ 估计总体平均数 μ；用样本方差 S^2 估计总体方差 σ^2；用样本相关系数 r 估计总体相关系数 ρ；用样本比率 p 估计总体比率 P，都属于点估计。由于存在抽样误差，估计量（用于估计总体参数的样本统计量）是随机变量，它正好和被估计的总体参数相等的机会微乎其微。因此，我们问某一个具体的估计值与总体参数是否接近没有太大意义，而应该考虑所构造的估计量与被估计的参数是否接近，是否是一个好的估计量。对于一个未知参数，人们可以构造多个估计量去估计它，那么如何评价一个估计量的好坏呢，或者说在众多估计量中进行比较和取舍的标准应该是什么呢？统计学中经常考虑的主要有以下三条标准。

首先，良好估计量应该是一个无偏估计量（unbiased estimate)，即样本统计量抽样分布的平均数应该和总体参数相等。也就是说，估计量应该以被估计的总体参数为中心上下波动，如果用无穷多个样本统计量作为总体参数的估计值，其偏差的平均数为 0。否则，就称这种估计量是有偏的。例如 $S_n^2=\dfrac{\sum(X-\overline{X})^2}{n}$ 和 $S_{n-1}^2=\dfrac{\sum(X-\overline{X})^2}{n-1}$ 都可以用于估计总体方差 σ^2，但 S_{n-1}^2 是无偏的， S_n^2 是有偏的，

所以在实际中通常用S_{n-1}^2而不是S_n^2去估计总体方差σ^2。

其次，良好估计量的分布应该相对集中，也就是方差越小越好（minimum variance）。如果两个估计量都是无偏的，但其中一个估计量抽样分布的方差小于另一个估计量抽样分布的方差，即前一个估计量的抽样误差更小，我们称前一个估计量更为有效。例如，$\overline{X}$、M_{dn}、是M_O都是总体平均数μ的无偏估计量，但$\overline{X}$抽样分布的方差$\sigma_{\overline{X}}^2$最小，故用$\overline{X}$估计总体平均数μ最为有效。

最后，我们期望随着样本容量的增加，良好的估计量应该与被估计的总体参数越来越接近。如果样本容量趋于无穷的时候，一个估计量能够等于被估计的总体参数，我们称这个估计量与总体参数是一致的。例如，当本容量$n\to\infty$时，$\bar{X}\to\mu,\quad S_{n-1}^2\to\sigma^2$，所以$\overline{X}$和$S_{n-1}^2$分别是总体平均数$\mu$和总体方差$\sigma^2$的一致估计量。

需要指出的是，一致性和无偏性是两个不同的概念。无偏性是指样本容量固定，而抽取的样本个数趋于无穷时，估计量抽样分布的平均数和被估计的总体参数相等。一致性是样本个数不变，而样本容量无限增大的时候，估计量趋于被估计的总体参数。这两个概念没有必然联系：一个无偏的估计量可能不是一致的，一个一致的估计量可能是有偏的。

点估计的结果是一个具体的数值，看起来比较确定，但由于有抽样误差，所以估计量和总体参数相等的可能性极低，而且点估计也不能告诉我们估计量和总体参数之间的误差大小。为弥补这个缺点，可以考虑采用区间估计。

（二）区间估计

区间估计（interval estimation）是依据样本统计量，根据一定精确度的要求，推断总体参数所在的区间和范围。区间估计以点估计为基础，它用数轴上的一段距离表示总体参数可能落入的范围，这一段距离或范围称为置信区间（confidence interval），也称置信间距。置信区间的上下两端点的值叫做置信界限。总体参数被置信区间包含的概率或把握，统计上称为置信水平（confidence level）或置信度。

同点估计一样，在进行区间估计时也可能犯错误。在区间估计（或假设检验）中，犯错误（即所求的置信区间未能包括总体参数）的概率叫做显著性水平（lever of significant），用符号α表示，而$1-\alpha$就是置信度或置信水平。例如，区间估计中常说的 0.95（或 0.99）的置信区间，就是指有 95%（或 99%）的把握包含总体参数的区间；换言之，估计总体参数落在该区间之内的概率为 0.95（或 95%），犯错误的概率为 5%（或 1%）。

可见，区间估计需要同时把握总体参数所在范围（置信区间）和正确估计的概率（置信水平）两方面的问题。通常，人们都希望估计的区间尽可能小，而估计的把握或准确性尽可能大。实际上这两种要求是矛盾的。在其他条件不变的情

况下，若要提高正确估计的概率，势必要放宽估计的范围，增大置信区间；若要使估计的范围缩短，则必然会降低正确估计的概率，增大犯错误的概率。统计分析中置信水平$1-\alpha$一般取 0.95 或 0.99，相应地犯错误的概率α分别为 0.05 和 0.01，这是因为统计上认为发生概率在 0.05 或 0.01 以下的事件为小概率事件。可见，区间估计中要求估计错误应是小概率事件。

二、区间估计的原理和一般步骤

区间估计的理论基础是上一节讲过的抽样分布理论。具体来说，如何在一定置信水平的前提下估计总体参数所在的区间呢？下面我们以已知方差的正态分布$\mathrm{N}\left(\mu,\sigma_0^2\right)$总体平均数$\mu$的区间估计为例，说明如何根据样本平均数的抽样分布理论，计算置信区间并解释估计的准确性：

当总体呈正态分布时，样本平均数也呈正态分布。根据抽样分布的原理，从正态总体$N\left(\mu,\sigma^2\right)$中抽样，样本平均数$\overline{X}$的分布为$N\left(\mu,\sigma^2/n\right)$，即$\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$；样本平均数$\overline{X}$的$Z$分数的分布为$N\left(0,1\right)$，即$Z=\frac{\overline{\mathrm{X}}-\mu}{\sigma/\sqrt{n}}\sim N\left(0,1\right)$。因此可以根据 Z 统计量的分布规律，对总体平均数μ进行区间估计。具体估计步骤如下：

（1）首先，要找到被估计总体参数的一个好的点估计量。例如，对于总体平均数μ，样本平均数$\overline{X}$是一个无偏、一致的估计量，且相对于其他估计量，其抽样分布的方差最小，即更具有效性。

（2）研究点估计量的抽样分布，并在此基础上，构造一个新的统计量。这个新统计量要同时和点估计量以及待估计的总体参数都有关系，而与其他未知参数无关。由于这个新统计量和点估计量以及待估计的总体参数都有关，所以能通过它将点估计量和待估总体参数联系起来，故而又称作枢轴量，意即起到关键性连接作用的统计量。根据抽样分布的原理，正态总体平均数μ的点估计量$\overline{X}\sim N(\mu,\frac{\sigma_0^2}{n})$；通过对点估计量进行标准化，得到一个新统计量$Z_{\bar{x}}=\frac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}$。由于$Z$统计量与点估计量$\overline{X}$和待估参数$\mu$都有关系，且其计算公式中除$\mu$外没有其他未知参数（$\sigma_0^2$是已知的)，故$Z$统计量可以作为枢轴量。

（3）通过研究所构造的枢轴统计量的分布规律，确定其理论上“应该的”取值范围。按照小概率事件的原理，在一次抽样试验中，小概率事件不应该发生。按照这个逻辑，枢轴量作为样本统计量，其具体数值取在其抽样分布的众数周围的概率较高，而取在距离其抽样分布众数很远位置的概率很低，或者说不大可能出现在抽样分布的尾端位置（因其概率很小）。所以，我们应该把枢轴量限定在一个其取

值概率比较高的特定区间内，枢轴量在该区间取值的概率对应着置信水平。

根据第 1 节所述的公式 5-12，我们知道 Z 统计量服从标准正态分布 N（0，1），即 $Z_{\bar{x}}=\dfrac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}\sim N(0,1)$，从而我们可以容易地确定它“应该的”取值范围。对于给定的置信水平 $1-\alpha$（或显著性水平 α），Z 统计量的取值范围应该在 $[Z_{1-\alpha/2},Z_{\alpha/2}]$ 内。由于 Z 统计量的分布是对称的，所以这个区间也可以写成 $[-Z_{\alpha/2},Z_{\alpha/2}]$。在参数估计中，置信水平一般取 0.95 或 0.99。由 Z 分布表可知，若置信水平为 0.95，Z 统计量的取值范围应该在[-1.96，1.96] 内；若置信水平为0.99，则 Z 统计量的取值范围应该在[-2.58，2.58] 内。

（4）令枢轴量落在根据置信水平 $1-\alpha$ 确定的其应在的区间之内，建立相应的不等式。当置信水平为 $1-\alpha$ 时，可建立不等式 $-Z_{\alpha/2}\le Z\le Z_{\alpha/2}$。特别是，当置信度为 0.95 时，$-1.96\le Z\le 1.96$；当置信度为 0.99 时，$-2.58\le Z\le 2.58$。在枢轴量 $Z_{\bar{x}}=\dfrac{\overline{X}-\mu}{\sigma_0/\sqrt{n}}$ 中，只有未知参数 μ 是未知的，所以反解不等式，能容易地得到总体平均数 μ 应在的区间 $\left[\overline{\mathrm{X}}-Z_{\alpha/2}\dfrac{\sigma_0}{\sqrt{n}},\overline{\mathrm{X}}+Z_{\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}\right]$，即：

$$\overline{\mathrm{X}}-Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}\le\mu_0\le\overline{\mathrm{X}}+Z_{\alpha/2}\frac{\sigma_0}{\sqrt{n}}$$ （总体平均数估计值的表达式）

由于枢轴量应该落入的区间是根据给定的置信水平计算得出的，所以总体参数 μ 落入通过反解不等式得出的取值区间的概率，就等于置信水平 $1-\alpha$，从而区间 $\left[\overline{\mathrm{X}}-Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}},\overline{\mathrm{X}}+Z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}\right]$ 就是总体参数 μ 的置信水平为 $1-\alpha$ 的置信区间。当置信度为 0.95 时，该区间为 $\left[\overline{\mathrm{X}}-\dfrac{1.96\sigma}{\sqrt{n}},\overline{\mathrm{X}}+\dfrac{1.96\sigma}{\sqrt{n}}\right]$，通常也可表示为 $\overline{\mathrm{X}}-1.96\dfrac{\sigma}{\sqrt{n}}\le\mu\le\overline{\mathrm{X}}+1.96\dfrac{\sigma}{\sqrt{n}}$；当置信度为 0.99 时，该区间为 $\left[\overline{\mathrm{X}}-\dfrac{2.58\sigma}{\sqrt{n}},\overline{\mathrm{X}}+\dfrac{2.58\sigma}{\sqrt{n}}\right]$，通常也可表示为 $\overline{\mathrm{X}}-2.58\dfrac{\sigma}{\sqrt{n}}\le\mu\le\overline{\mathrm{X}}+2.58\dfrac{\sigma}{\sqrt{n}}$，为简便起见，上述公式中的 $\dfrac{\sigma}{\sqrt{n}}$ 有时也可以用 $SE_{\overline{X}}$ 代替。

可见，在总体参数的区间估计中，确定枢轴统计量是最为关键的一环。参数估计中对枢轴统计量有两方面的要求：第一是它要和点估计量以及待估计的总体参数都有关系，而与其他未知参数无关，从而通过该枢轴量在点估计量和总体参数之间建立起联系；第二是其分布易于确定，这样才能根据其分布规律和给定的置信水平，确定其“应该的”取值范围。有了这两条，通过反解不等式的方法，就能解

出待估总体参数的置信区间。在应用统计中，大多数情况下不需要我们去构造枢轴统计量，进而确定其分布，只需要能应用已知其分布的枢轴统计量即可。

上面阐述的是进行参数估计的理论基础，下面我们举一个实例来说明进行总体平均数的参数估计的实际操作过程。

【例 5-3】 某中学从本年度初一新生中随机抽取 28 名学生，测得他们的平均身高为 155cm。若已知该校初一新生历年来的身高标准差为 5.24cm，试在 0.95 的置信水平下，估计该校本年度初一新生的平均身高。做推论时犯错误的可能性有多大？

分析：身高、体重等体征变量一般服从正态分布或近似正态分布，根据题意可知 $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$，可以根据 Z 统计量的分布规律对总体平均数 μ 进行估计。

解：

（1）根据显著性水平 α（或置信区间 1－α）查 Z 分布表求相应的 $\dfrac{\alpha}{2}$ 的 Z 值。

之所以是根据 $\dfrac{\alpha}{2}$ 查 Z 值是因为我们并不知道总体平均数 μ 是高于样本平均数还是低于样本平均数，因此，当我们想把估计后犯错误的概率控制在 5%范围内时，将它平均分布在曲线下面积的两侧，因此每一侧在显著性水平 α 临界点之外的曲线下面积都是 $\dfrac{\alpha}{2}$。在本例中，当 $\alpha=0.05$（即置信区间是 0.95）时，$\dfrac{\alpha}{2}=0.025$。查 Z 分布表，对应于单侧面积 P 为 0.025 的 Z 值是 1.96。

（2） 根据公式 5-12 即 $Z=\dfrac{\overline{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$，建立总体平均数 μ 的显著性水平在 α = =0.05（或置信区间在 0.95）时的估计区间不等式为：

$$\overline{X}-1.96\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+1.96\frac{\sigma}{\sqrt{n}}$$

（3） 将有关数据代入上述不等式求具体的解：

$$155-1.96\frac{5.24}{\sqrt{28}}<\mu<155+1.96\frac{5.24}{\sqrt{28}}$$

$$153.06<\mu<156.94$$

答：根据这 28 学生组成的样本，可以推论该校本年度初一新生的平均身高在 153.06 厘米至 156.94 厘米之间，估计正确的概率为 0.95，错误的概率为 0.05。

如果我们把显著性水平定为 $\alpha=0.01$（相应的置信区间就是 0.99），则在上

述求法的第一步时查 Z 分布表求对应于单侧面积 P 为 $\frac{\alpha}{2}=0.005$ 的 Z 值是 2.58，第二步建立以下不等式：$\overline{X}-2.58\frac{\sigma}{\sqrt{n}}<\mu<\overline{X}+2.58\frac{\sigma}{\sqrt{n}}$；第三步代入相应数据后可得总体平均数 μ 的 0.99 的置信区间为：152.45 ≤ μ ≤157.56。据此可以推论该校本年度初一新生的平均身高在 152.45 厘米至 157.56 厘米之间，估计正确的概率为 0.99，错误的概率为 0.01。

本节以对总体平均数的估计为例，介绍了区间估计的原理和一般过程。如前文所述，不仅样本平均数存在抽样分布，各种其他的样本统计量（如样本方差、样本比率和样本相关系数等）也都有相应的抽样分布规律。因此，我们不仅可以根据样本平均数的具体值和其抽样分布规律对总体平均数进行区间估计，也同样可以根据样本方差及其抽样分布规律对总体方差进行区间估计和其他统计推断。只是在对总体方差进行区间估计时，必须借助于和样本方差以及待估计的总体方差都有关的枢轴统计量才能进行。此外，我们还可以根据样本比率及其抽样分布规律对总体比率进行区间估计。这些方面的内容，将在今后的章节中进行具体讨论。

第三节　假设检验的原理

除参数估计之外，推断统计的另外一项重要内容是假设检验。所谓假设检验（hypothesis testing），通常是指根据抽样结果，在一定可靠程度上，对一个或多个有关总体分布的原假设做出拒绝或接受该原假设的推论的统计分析过程。在假设检验过程中，用于判断接受还是拒绝原假设的统计量被称为“检验统计量（statistic of test）”。本节将从统计假设检验的两种误差、两种假设、检验的逻辑、检验的两种方法和检验的步骤等方面来介绍假设检验的基本原理。

一、统计假设检验的基本思想

（一）两种误差

在心理研究中，由观察或测量获得的样本统计量与客观存在的总体参数之间总会存在着一定的差异，这种差异可以称为误差。根据误差产生的原因及性质可以将其分为系统误差和偶然误差（或随机误差）两类。

系统误差又叫规律误差。它是在一定的测量条件下，对同一个对象进行多次重复测量时，误差值的大小和符号（正值或负值）保持不变，或者在条件变化时，按一定规律变化的误差。

在相同条件下，对同一对象进行多次测量，由于各种偶然因素（比如随机抽样）的影响，会出现测量值时而偏大，时而偏小的误差现象，这种类型的误差叫做偶然误差（随机误差），其中由随机抽样方法本身引起的差异称为抽样误差，它代表的是样本统计量与被推断的总体参数之差。

假设检验的目的就是要判断测得的样本统计量与总体参数之间的差异，是由偶然误差（随机抽样）引起的，还是由系统误差（实验条件变化）引起的。

（二）两种假设

在进行任何一项研究时，

我们都需要根据已有的理论和经验，事先对研究结果做出一种预想的希望证实的假设，这种假设叫科学假设，用统计学术语表示时叫做研究假设，记作 H_1。如在本章开头的问题 2 中，若以 μ 表示今年新生总体的平均智商分数，以 μ_0 表示历年新生智商的平均分数，根据今年测验的平均分数，我们假设今年新生的智

商分数与往年不同，则检验的目的是要证实 $\mu \neq \mu_0$，即要证实研究假设 H_1：$\mu \neq \mu_0$。

但在实际研究中，我们常常不能对 H_1 的真实性进行直接检验（因为我们一般通过抽样得到的是样本统计量 $\overline{X}$，很难对全体研究对象进行测量，无法知道确实存在的 μ），而是需要检验它的对立形式，即检验虚无假设（null hypothesis）。虚无假设也叫无差假设、零假设、原假设，记作 H_0。例如，对于前述问题 2，虚无假设 H_0：$\mu = \mu_0$，即假设今年新生总体的平均智商分数与历年新生的没有差异。请注意，这里的“没有差异”是指 μ 和 μ_0 没有实质上的差别。即虽然 $\overline{X}$ 与 μ_0 在绝对数值上有一定差别，但这种差别没有达到统计上的显著性要求，不具有统计意义；也就是说 $\overline{X}$ 与 μ_0 在数值上的差异可能是由于抽样误差引起的，所以没有足够的把握推论 $\overline{X}$ 所来自总体的平均数 μ 与 μ_0 不同。

在假设检验中，H_0 总是作为直接被检验的假设，而 H_1 与 H_0 对立，二者择一，因而 H_1 又叫做对立假设或备择假设（alternative hypothesis）。H_1 与 H_0 两者的并集为全域，交集为空集。在建立假设时，希望得到的结果或题目的问法常可作为研究假设，它的否定形式即为虚无假设。

假设检验的问题就是要通过统计方法判断拒绝还是接受 H_0。如前述的问题 2，若经过计算推论 H_0：$\mu = \mu_0$ 为真，即接受 H_0，拒绝 H_1，说明 $\overline{X}$ =115 与 μ_0 = 110 相差的 5 分，很可能是由于随机抽样（偶然误差）引起的，不是实质性的差异。若经过计算推论 H_0：$\mu = \mu_0$ 为假，即拒绝 H_0，接受 H_1，说明 $\overline{X}$ =115 与 μ_0 =110 相差的 3 分，很可能是由于今年新生总体的平均智商分数确实比历年新生的高，即是由于条件不同（系统误差）引起的实质性差异。

（三）统计假设检验的逻辑

通过总结（一）（二）两点我们可以看出，统计假设检验的基本思想是带有概率值保证的反证法。也就是说，我们想要证实研究假设，但并不是从研究假设出发进行验证，而是建立与它对立的虚无假设，并假定虚无假设为真。在虚无假设为真的前提下，通过收集实际信息并依据正确逻辑推理和数学上的计算与分析，看实际获得的资料所导致的结果是否与虚无假设成立时应出现的结果发生矛盾。如果出现了矛盾，则表明虚无假设 H_0 是错误的，应该给予否定，从而接受研究假设。如果没有出现矛盾，则表明没有充分理论否定虚无假设。

假设检验中的“反证法”思想与纯数学中的反证法是不一样的，数学中的反证法是在假设某一条件的情况下导致逻辑上的矛盾，从而否定原来的假设条件。而假设检验中的“反证法”是带有概率值保证的。也就是说，判断由虚无假设推论的结果与实际出现的结果是否矛盾即 H_0 是否成立的依据是“小概率事

件原理”，该原理认为“小概率事件在一次试验中几乎是不可能发生的”。小概率事件是否出现，是统计决策的依据。小概率事件发生，说明实际获得信息反映出的结果与 H_0 成立条件下的推论结果之间矛盾很明显，可以拒绝虚无假设，从而接受研究假设。

二、差异显著性检验的原理

心理学研究中使用假设检验的目的在于基于抽样分布的原理，在特定显著性水平下考察样本数据的相关特点是否可以推论到总体。所谓差异显著性检验（significance test of difference）是指检验某个样本统计量的数值与总体参数之间的差异是否显著，或者检验两个样本统计量之间的差异是否显著的假设检验过程，其目的在于判断样本数据中表现出来的差异是否具有统计学意义，即是否可以推论到总体。

从总体中随机抽取一个样本，其统计量与总体参数间总不会完全相等，假设两者间的差异是由偶然的抽样误差造成的，按照抽样分布理论，其差值通常不应太大，也就是说，样本统计量在总体参数附近波动的可能性（概率）大，而远离总体参数的可能性（概率）小。但是如果在一次随机抽样的样本中，样本统计量和假设的总体参数间的差值较大，且达到一定显著性水平所对应的误差限度值（临界值）时，我们就说小概率事件发生，或者说这种差异已经不太可能是由抽样误差造成的，很可能是样本所来自总体的真实参数与假设的总体参数确实有差异，这种情况叫做差异显著或有显著性差异（sig-
nificant difference），应拒绝虚无假设，从而接受研究假设。反之，若所得到差异未达到规定限度，则不能排除差异是由抽样误差造成的可能性，这时称差异不显著或无显著性差异，应接受虚无假设（图 5-2）。需要指出的是，假设检验中的差异是否显著是一种统计推断，其含义是能否根据样本资料的差异去推论总体上存在差异，如果有充分的把握作出推论，则称差异显著或差异具有统计学意义，否则称差异不显著或差异不具有统计学意义。所以，统计上所说的差异显著与不显著，与实践中常说的“效果显著”中的“显著”二字含义不同，请读者们注意。

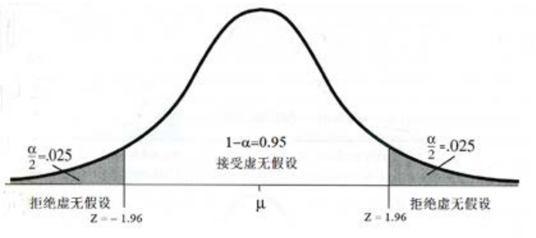


图 5-2 假设检验原理示意图

三、假设检验的两种方法

在假设检验中，如果我们关心的是 μ 与 μ_0 的差异，并不关心 μ 比 μ_0 大还是小，这时在 μ_0 两侧都需要一个临界点，临界点以外的区域为 H_0 的拒绝区域。如

果显著性水平 α 定为 0.05，则两端拒绝区域的面积比率为 $\dfrac{\alpha}{2}$ 即 0.025，这种强调差异而不强调方向性的检验方法叫做双侧检验（two tailed test）方法，双侧检验的小概率事件如图 5-2 阴影部分所示。反之，如果检验的目的在于检验 μ 是否显著低于 μ_0 或 μ 是否显著高于 μ_0，则 α 区域集中于 μ_0 的一端。这种强调某一方向的检验方法叫做单侧检验（single tailed test）方法，通常适用于检验某一参数是否“大于（或小于）”、“优于（或劣于）”、“快于（或慢于）”另一参数这一类问题。图 5-3 中，（1）所示的称为右侧单侧检验方式，（2）所示的称为左侧单侧检验方式，其小概率事件有如图中阴影部分所示。

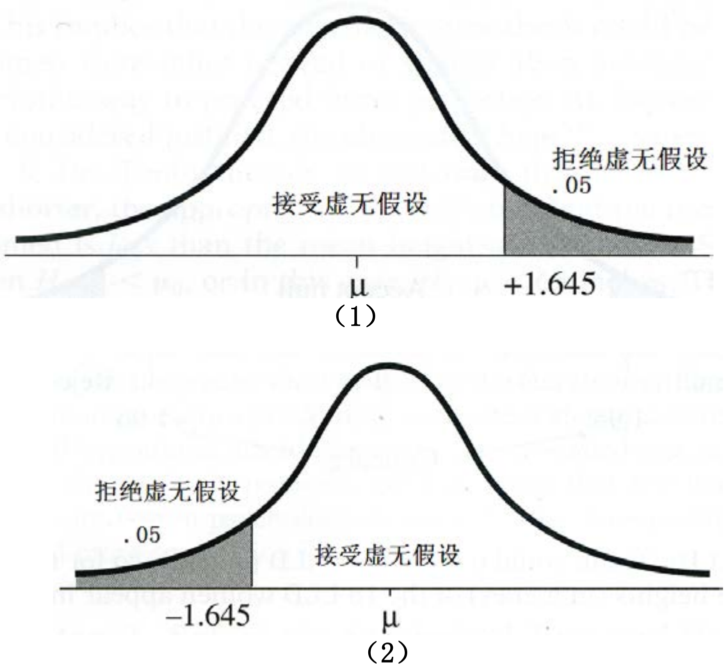


图 5-3 单侧检验示意图

从以上两种检验的定义中，可以看出它们是有一定的区别的，这可以从以下几个方面来进行比较，具体见表 5-1。

表 5-1 两种检验的区别

	目的	假设	危机域	临界值	灵敏度
双侧	有无差异	$H_0:\mu=\mu_0$ $H_1:\mu\neq\mu_0$	两块	两个	低
单侧	大于或小于	$H_0:\mu\leq\mu_0$ 或 $H_0:\mu\geq\mu_0$ $H_1:\mu>\mu_0$ 　 $H_1:\mu<\mu_0$	一块	一个	高

灵敏度高指的是更容易拒绝虚无假设。例如，当检验统计量 Z=1.87 时，在 α =0.05 的显著性水平下，如果采用单侧检验，其临界值为 1.64，则此时可拒绝虚无假设；而在双侧检验中，其临界值为 1.96，无法拒绝虚无假设。因而，在实际研究中何时用单侧检验何时用双侧检验，一定要根据研究目的所规定的问题方向性来确定，而不能按照自己所希望出现的结果随心所欲的选用。

四、假设检验的步骤

根据上述假设检验的原理和方法，可以总结出假设检验的一般步骤：

（一）建立假设

双侧检验为：$\begin{matrix} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{matrix}$；单侧检验为 $\begin{matrix} H_0: \mu \le \mu_0 \\ H_1: \mu > \mu_0 \end{matrix}$ 或 $\begin{matrix} H_0: \mu \ge \mu_0 \\ H_1: \mu < \mu_0 \end{matrix}$

（二）选择和计算检验统计量

在 H_0 成立的条件下，寻找和决定合适的统计量及其抽样分布，并计算出统计量的值。本章第一节曾指出，常用的抽样分布主要有 Z 分布、t 分布、χ^2 分布和 F 分布等，因此，对应的假设检验方法也分别有 Z 检验、t 检验、χ^2 检验和 F 检验。

（三）确定显著性水平并做出统计决策

首先规定显著性水平 α，然后根据 α 查相应的分布表来确定临界值，从而确定出 H_0 的拒绝区间或接受区间。比较临界值和统计量值，若统计量值落在拒绝 H_0 区间中，则拒绝 H_0，即推论差异达到显著性水平或差异有统计意义；若统计量值落在接受 H_0 区间中，则接受 H_0，即推论差异不显著或差异没有统计意义。

下一节，我们将以总体分布形态已知时，单总体样本平均数与总体平均数之间差异的显著性检验为例，来进一步说明如何进行假设检验的实际操作。有关与平均数差异的假设检验有关的其他问题，以及方差、比率及相关系数等样本统计量的假设检验问题，将在随后的各章中加以论述。

第四节　单总体平均数差异显著性的 Z 检验

一 、单总体平均数差异显著性检验的含义

单总体平均数的显著性检验是指对样本平均数 $\overline{X}$ 与某一已知总体平均数 μ_0

的差异进行的显著性检验。这里，样本是从某未知总体水平 μ 的总体中随机抽取的，$\overline{X}$ 是来自该总体的样本的平均数。检验的目的是确定样本平均数与已知总体平均数 μ_0 间的差异是由随机抽样误差造成的，还是由于样本并非来自已知总体造成的。例如，研究人员可能想知道一组学生（样本）的智商平均分数与常模分数之间的差异，也可能在已知历年学生的智商分数的情况下，想比较该组学生（样本）的智商分数与历年学生的智商分数（总体平均数）的差异。

二、单总体平均数差异显著性的 Z 检验

当总体的分布形态是正态分布时，在总体方差已知的条件下，其抽样分布服从 Z 分布，因此可以用公式 5-12 所表达的 Z 检验方法来解决这个问题。

【例 5-4】 某校历年招收新生时都要测其智商，历年新生的智商分数服从正态分布，平均分数为 110 分，标准差为 15 分。从今年新生的总体中随机抽取一个 25 人的样本进行测验，测得其智商平均数为 115 分，能否据此推论今年新生的总体智商与往年不同？

今年新生中抽取的样本的平均智商分数为 115 分，比历年新生的平均智商分数 110 分高出 5 分。从统计的原理来看，这增加的 5 分，有可能是由于随机抽样（偶然误差）引起的，也有可能是由于今年新生总体的平均智商分数确实比历年新生的高，即由于条件不同（系统误差）引起的。假设检验的目的就是要判断 $\overline{X}=115$ 代表的今年新生总体智商分数（假如用 μ 表示）和 $\mu_0=110$ 的历年新生总体智商分数之间的差异，是由于随机误差引起的还是由系统误差引起的。

要检验今年新生的总体智商分数是否与往年新生一样，就要看这 25 名新生的智商分数的出现是否属于小概率事件。具体的做法就是，首先要建立研究假设和虚无假设，以虚无假设为真作为出发点，进行检验，如果拒绝虚无假设，则接受研究假设。检验的目的是看从样本平均数与总体平均数之间的差异，是否能推论出这个样本所来自的总体（今年新生）与往年新生在智商分数上有差异。既然虚无假设是无差异，则在虚无假设为真的前提下，可以把今年新生总体与往年新生总体看做同一总体，即可以把这个样本看做从 $\mu_0 = 110, \sigma_0 = 15$ 的总体中随机抽取的一个样本，根据第一节中介绍的抽样分布理论，此时样本平均数的抽样分布服从正态分布，即 $\bar{X} \sim N(\mu_0, \frac{{\sigma_0}^2}{n})$，因为 $Z=\dfrac{\bar{X}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}} \sim N(0,1)$，因此可以使用 Z 检验方法来对假设进行检验。

具体而言，对例 5-4 的假设检验过程有如下述：

（1）建立统计假设：$\begin{matrix} H_0: \mu = 110 \\ H_1: \mu \neq 110 \end{matrix}$ 或 $\begin{matrix} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{matrix}$

（2）利用公式 5-12 即 $Z=\dfrac{\bar{X}-\mu_0}{SE_{\overline{X}}}$ （ $SE_{\overline{X}}=\dfrac{\sigma_0}{\sqrt{n}}$ ）计算检验统计量：

$$Z=\frac{\bar{X}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}}=\frac{115-110}{\frac{15}{\sqrt{25}}}=1.67$$

（3）确定显著性水平并做出统计决策

当 $\alpha=0.05$ 时，查正态分布表可知 $Z_{\frac{\alpha}{2}}=1.96$，也就是说，$\left|Z_{\frac{\alpha}{2}}\right|=1.96$ 是是否发生小概率事件的误差限度值（临界值）。

由于 $Z=1.67<Z_{\frac{\alpha}{2}}=1.96$，

没有超出误差限度，落在 $Z=1.96$ 和 $Z=-1.96$ 的中间（如图 5-2 中间白色区域所示），表明小概率事件没有发生，因此没有理由拒绝虚无假设，即接受两者无判别的虚无假设。

答：今年新生样本的智商分数与往年相比差异不显著，即这种差异可能是由抽样误差造成的，故（在 0.05 的显著性水平下）不能推论今年新生总体的智商分数与往年新生总体的不同。

【例 5-5】某市举行了一次标准化的语文成就测验，已知全市初三考生的成绩服从 N（$83,8^2$）。从市内某校随机抽取了 36 名初三考生，算出其 $\overline{X}=86.5$ 分，能否据此推论该校初三考生的语文成绩高于全市？（$\alpha=0.05$）

从表面看，某校初三考生样本的平均成绩为 86.5 分，比全市考生的平均分高出 3.5 分，但是并不能就此说明该校初三考生在语文成就测验中的真实水平比全市要高。我们要弄明白这种差异究竟是来自偶然误差还是来自系统误差，或者说是否真的存在显著性差异，就需要按照假设检验的步骤来进行显著性检验，从而做出决策。

解：因为学生成绩服从正态分布，而且总体标准差已知，所以可用 Z 检验。

（1）建立假设：$\begin{array}{l}H_0:\mu\le 83\\H_1:\mu>83\end{array}$ 或 $\begin{array}{l}H_0:\mu\le \mu_0\\H_1:\mu>\mu_0\end{array}$

（2）计算检验统计量：由于总体方差已知，样本来自于正态总体，故 $\bar{X}\sim N(\mu_0,\dfrac{\sigma^2}{n})$，则由抽样分布可知，检验统计量为：

$$Z=\frac{\bar{X}-\mu_0}{\frac{\sigma_0}{\sqrt{n}}}=\frac{86.5-83}{\frac{8}{\sqrt{36}}}=2.63$$

（3）确定显著性水平并做出统计决策

$\alpha=0.05$，从题意可知应该采用单侧检验，查表得 $Z_{0.05}\approx 1.64$

$Z=2.63>Z_{0.05}\approx 1.64$，即拒绝虚无假设。

答：该校初三学生语文成绩与全市的平均成绩之间有显著性差异，即可以推论

差异不是由于偶然因素造成。

三、总体平均数区间估计与单总体平均数 Z 检验的异同

如前所述，区间估计是依据样本统计量，根据一定精确度的要求，推断总体参数所在的区间和范围，而假设检验的目的则是要判断测得的样本统计量与总体参数之间的差异是由偶然误差（随机抽样）引起的，还是由系统误差（实验条件变化）引起的。下面我们以已知总体方差的平均数的区间估计和单总体平均数的显著性检验为例，来了解这两者之间的异同。

它们之间的共同点是服从同样的抽样分布，因此可以用同样的 Z 统计量进行计算。也就是说，当总体呈正态分布，样本平均数也呈正态分布。根据公式 5-11 和公式 5-12 可知，从正态总体 $N\left(\mu,\sigma^2\right)$ 中抽样，样本平均数 $\overline{X}$ 的分布为 $N\left(\mu,\sigma^2/n\right)$，样本平均数 $\overline{X}$ 的 Z 分数的分布为 $N\left(0,1\right)$，即 $\overline{X}\sim N(\mu,\frac{\sigma^2}{n})$，$Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}\sim N$（0，1）。由于它们两者都服从这一抽样分布，所以，都可以根据同一公式 $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ 来计算 Z 统计量，并根据 Z 统计量的分布规律，在“小概率事件在一次抽样中一般不会发生”的概率统计思想指导下，或者对总体平均数 μ 进行区间估计，或者对单总体平均数显著性情况进行 Z 检验。

两者之间的不同点在于：已知条件不一样，所要解决的任务也不同。区间估计的已知条件是：总体标准差 σ，样本平均数 $\overline{X}$ 和样本容量 n，在不知道总体平均数 μ 的情况下，希望通过公式 5-12 即 $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ 对总体参数 μ 可能会在什么范围之间进行推测。由于总体参数 μ 既有可能比抽样平均数更大，也有可能比抽样平均数更小，因此，将公式 5-12 进行转换后就得到两个计算总体参数 μ 的公式：当总体参数 μ 可能比抽样平均数大时，用公式 $Z=\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}}$ 的变式 $\mu=\overline{\mathrm{X}}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ 来求它在预设置信区间的高端值；当总体参数 μ 可能比抽样平均数更小时，用公式 $\mu=\overline{\mathrm{X}}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$ 来求它在预设置信区间的低端值，由此得到总体参数 μ 的置信区间为：$\overline{\mathrm{X}}-Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}<\mu<\overline{\mathrm{X}}+Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$。

但是，对单总体平均数进行 Z 检验时，已知条件除了区间估计中已知的条件即总体标准差σ、样本平均数$\overline{X}$和样本容量n之外，还知道总体参数μ，检验的目的是希望通过公式 5-12 即$Z=\dfrac{\bar{X}-\mu}{\dfrac{\sigma}{\sqrt{n}}}$所计算得到的 Z 值与预先设定的显著性水平即$\alpha$的临界值相比较后，对样本平均数$\overline{X}$与总体平均数$\mu$之间的差异是否存在统计学意义上的显著性做出判断。

还可以从另外一个角度来看两者关系：如果总体平均数落在区间估计的范围之内，则接受虚无假设，否则就拒绝虚无假设。

理解上述总体平均数区间估计与单总体平均数 Z 检验的异同，是理解其他统计量的区间估计与假设检验之间的异同的基础，因此，希望读者能对上述原理详加领会。

第五节 假设检验中的两类错误

一、两类错误

假设检验的原理和方法是用样本去推断总体，由于抽样误差的存在，因而不可能有百分之百的把握使推断无误（只是大概率保证推断正确）。在对假设进行检验时，只是以样本统计量出现的概率大小来决定假设的取舍。如在例 5-4 中，当给定的显著性水平为 0.05 时，我们算出的结果$|Z|>Z_{\frac{\alpha}{2}}$，从而拒绝了虚无假设H_0，但这个结论有可能犯错误。若实际情况是不应当拒绝虚无假设H_0，而我们在统计决策过程中拒绝了虚无假设H_0，称这种错误为第一类错误（type I error）或“弃真”错误。由于虚无假设本来正确，而统计量却落在了拒绝区域，我们依次拒绝了虚无假设，得出了错误的结论，而拒绝区域的面积（概率）为α，所以当虚无假设正确时而拒绝虚无假设所犯的第一类错误的概率正是显著性水平α。第一类错误又叫α型错误。

另一种情形是，当计算得到$|Z|\le Z_{\frac{\alpha}{2}}$时，我们接受了虚无假设，这时也有可能犯错误。若实际情况不应当接受虚无假设，而此时却接受了，把这种错误称为第二类错误（type II error）或“纳伪”错误，这类错误的概率以β表示，因而又叫β型错误。

不管H_0是真还是假，总体的状态实际上都是未知的。如果面对这种未知的不确定性做出决策的话，有四种可能的结果，其中的两种情形是我们不得不冒着犯第一类错误和第二类错误的风险。这四种结果概括如表 5-2 所示。

表 5-2 假设检验的各种可能结果

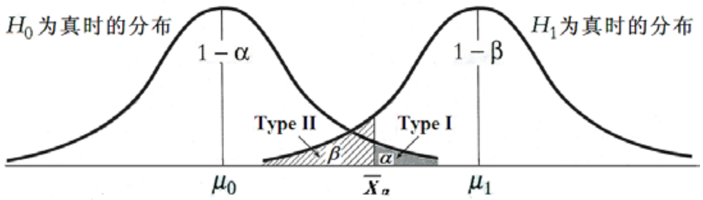
	接受 H_0	拒绝 H_0
H_0 为真	正确决策，概率=1-α =置信度	第一类错误，概率=α =检验水平
H_0 为假	第二类错误，概率=β	正确决策，概率=1-β =统计检验力

在实践中，通常主要考虑控制第一类错误的概率 α ，给 α 设定一个较小的值，即强调必须有足够的把握（正确概率为 1-α ）时才能拒绝 H_0 。在这一原则下进行的假设检验称为显著性检验，相应地 α 就称为显著性水平。考虑到不能拒绝 H_0 时，接受 H_0 也可能犯第二类错误，其概率为 β ，因此在假设检验中，在控制 α 的前提下，应尽量使用犯第二类错误概率 β 较小的方法。β 越小，则其反面 $1-\beta$ 就越大，1-β 在统计学中称之为统计检验力（power of test)，是个非常重要的概念，我们在第十一章再详述。

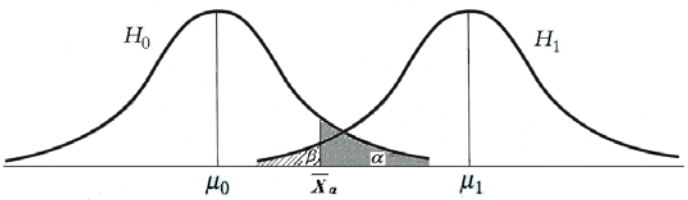
二、两类错误的关系

1. α 和 β 是在两个前提下的概率

α 型错误是指在虚无假设 H_0 为真时，拒绝 H_0 所犯错误的概率；β 型错误是指在虚无假设 H_0 为假时，接受 H_0 所犯错误的概率。两者关系可以通过图 5-4 中（1）表示出来。由于两类错误的前提不一样，所以 $\alpha+\beta$ 不一定等于 1。



（1）



（2）

图 5-4 α 和 β 关系示意图

2. 在其他条件不变的情况下，α 和不可能同时减小或增大

我们在进行假设检验时，通常是通过预设显著性水平来控制可能犯 α 型错误的概率，但在其他条件不变的情况下，α 和 β 不可能同时减小或增大，这可以通过比较图 5-4（1）和（2）中 $\bar{X}_\alpha$ 的位置来直观地理解。图 5-4（1）中的 $\bar{X}_\alpha$ 离虚无假设 H_0 为真时 μ_0 的距离更远，而图 5-4（2）中的 $\bar{X}_\alpha$ 则离虚无假设 H_0 为真时

的 μ_0 的距离更近，由两个图中以 $\overline{X}_\alpha$ 为分界线所示右边显示 α 的阴影部分的面积可知，$\overline{X}_\alpha$ 离 μ_0 的距离更远时可能犯 α 型错误的概率，比 $\overline{X}_\alpha$ 离 μ_0 的距离更近时小，但同时犯 β 型错误的概率则大。

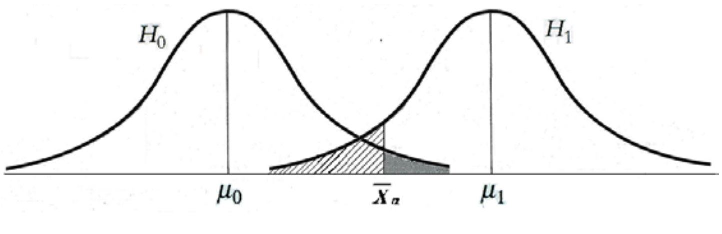
在一般情况下，β 错误的概率是无法准确推断和预先控制的，只能通过其他条件来减小犯 β 型错误的可能性，这些条件是：

（1）对于固定的 n，α 越小，则 $Z_{\frac{\alpha}{2}}$ 越大，从而接受虚无假设的区间越大，H_0 就越容易被接受，这样“纳伪”的概率 β 就越大。反之，α 越大，则 $Z_{\frac{\alpha}{2}}$ 越小，从而接受虚无假设的区间越小，这时“纳伪”的概率 β 就越小，但是增大了“弃真”的概率 α。也就是说，对于一定的样本容量 n 来说，α 和 β 不可能同时增大或减小。减小了 β 势必增大 α，反之亦然。

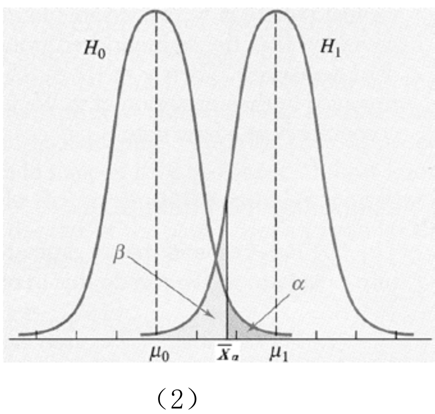
（2）β 的大小与真假值之间的距离（即 μ_0 与 μ 的距离）成反比。距离越远越容易拒绝虚无假设，这时是犯第一类错误，从而减小了第二类错误的概率。

通过（1）（2）可以看出这两点都不是一个很好的减小犯错误概率的方法，α 和 β 不可能同时增大或减小，减小了 β 势必增大 α，反之亦然。所以要减小犯错误的概率就只有通过第（3）点增大样本容量来实现。

（3）要想同时减小 α 和 β，一个方法就是增大样本容量 n。n 越大，样本越接近总体，抽样标准误（$\sigma/\sqrt{n}$）越小，则分布就越瘦长，从而减小了两种错误的概率 α 和 β。这可以通过图 5-5 很好地表示出来，图 5-5（2）所示的是样



（1）



（2）

图 5-5 不同标准误影响 β 大小示意图

本量增大后的图形。

三、统计决策中“接受 H_0” 的含义

统计学中的假设检验，通常关注的是如何运用小概率事件的原理去拒绝或证伪 H_0，因而为拒绝 H_0 设立了较严格的标准。拒绝 H_0 时，犯错误的概率要求不能大于给定的显著性水平 α，正确的概率为 1－α，而在显著性检验中 α 通常设定为一个较小的值，例如 0.05 或 0.01。所以，在显著性检验中如果拒绝 H_0，是有较大把握的（推论正确的概率超过 0.95 或 0.99）。当虚无假设 H_0 不能被拒绝时，进行统计推论时则要谨慎，有时我们会说接受 H_0。但需要指出的是，接受 H_0 并不等于 H_0 被证实了，只是说根据现有的资料，尚无足够的把握推论 H_0 不成立，只能暂时承认差异不显著的事实。统计假设检验理论的创始人——著名的统计学家费希尔本人也强调：虚无假设（H_0）不能被证明，他说“每一个实验的存在，仅仅是为了给事实一个反驳虚无假设的机会”。

在统计教材中，H_0 多以“$\mu = \mu_0$”或“$\mu_1 = \mu_2$”等形式存在，因此“接受 H_0”的说法常常会给初学者带来误解，以为“接受 H_0”就意味着 H_0 被证实，进而推论两总体的平均数是相等的或者是无差异的。事实上，差异不显著并不等同于无差异。接受 H_0，只是说样本平均数与总体平均数（或两样本平均数）的差异没有达到统计上的显著性水平，即尚无足够的把握推论“样本平均数与总体平均数（或两样本平均数）的差异不是由抽样误差照成的”或者“样本平均数与总体平均数（或两样本平均数）的差异是由实验处理造成的系统性差异”。

另外需指出的是，接受 H_0，也可能犯错误，而犯错误的概率 β 我们通常是不知道的，如果把“接受 H_0”当成是“H_0 被证实了”，推论两总体的平均数是相等的或者是无差异的，则是在缺乏概率证据的前提下进行的推论，显然是不正确的。在本书的第十一章，我们会谈到对统计检验力和效果大小的分析，其中会涉及到对 β 的估计，届时可以看到，很多时候我们接受 H_0 时，犯二类错误（其概率为 β）并不是一个小概率事件。

因为以上的原因，一些统计学者提倡在不能拒绝 H_0 时，采用“不拒绝 H_0”的说法，而不用“接受 H_0”的说法，以免引起初学者的误解。考虑到统计决策中经常要用到“拒绝域”和“接受域”的说法，为保持一致，本书中沿用了“拒绝 H_0”和“接受 H_0”的习惯说法，但需要指出的是，“接受 H_0”不等于“证实 H_0”，希望广大读者尤其是初学者不要误解。当然，如果某种假设检验不是一般意义上的显著性检验，其检验理论和方法可能和本书中讨论的有所不同，则超出了本书所讨论的范围，读者可以参阅其他统计学著作。

小 结

总体参数在很多时候是未知的，需要对其进行估计。利用样本统计量估计总体参数的过程叫做参数估计，参数估计是推断统计的核心内容之一。本章主要阐述了参数估计与假设检验的基本原理。

估计总体参数的方法有两种：点估计和区间估计。点估计就是用样本统计量的单一数值估计未知的总体参数。区间估计是根据点估计量及其抽样分布的规律，根据一定精确度的要求，推断总体参数所在的区间和范围。区间估计以点估计为基础，它用数轴上的一段距离表示总体参数可能落入的范围。

利用样本统计量对总体进行估计，找到一个好的估计量十分重要。一个好的估计量应该满足无偏性、有效性、一致性和充分性。但无论如何，由于抽样误差的存在，点估计总是以误差的存在为前提的。

总体参数可能落入的范围统计上称为置信区间或置信间距，其上下两端点值叫做置信界限。总体参数被置信区间包含的概率或把握，统计上称为置信水平或置信度。在区间估计中，犯错误（即所求的置信区间未能包括总体参数）的概率叫做显著性水平，用符号 α 表示，而 $1-\alpha$ 就是置信度或置信水平。通常置信水平的要求为 0.95 或 0.99，相应地显著性水平取 0.05 和 0.01。

大多数总体参数的区间估计步骤包括：确定点估计量；构造出包含估计量和总体参数的枢轴统计量，并明确其分布；根据枢轴统计量的分布规律和给定的估计精度（置信水平），通过查统计表确定枢轴统计量“应该的”取值范围；通过反解不等式，得出特定置信水平下总体参数取值应在的区间。

根据样本统计量对总体参数进行区间估计，要根据待估计参数及其抽样分布的不同而采取不同的计算方法。经常需要估计的总体参数有总体平均数 μ 、总体方差 σ^2、总体相关系数 ρ 和总体比率 p 等。

假设检验是抽样推断中的一项重要内容，本章主要介绍了统计假设检验的概念和原理，假设检验的方法和步骤。

假设检验的目的是判断测得的样本统计量与总体参数之间的差异是由偶然误差（随机抽样）引起的，还是由系统误差（实验条件变化）引起的。它是根据原资料做出一个总体参数是否等于某一数值，某一随机变量是否服从某种概率分布的假设，然后利用样本资料，采用一定的统计方法计算出有关检验的统计量，依据一定的概率原则，以较小的风险来判断估计数值与总体数值(或者估计分布与实际分布)是否存在显著差异，是否应当接受原假设选择的一种检验方法。统计假设检验的思想是“带有概率值保证的反证法”。

统计假设检验有两种方法：双侧检验和单侧检验，在实践中采用何类检验是由实际问题的性质来决定的。

统计决策的原理是“小概率事件原理”。在统计决策时，无论做出何种决策，都有可能犯错误的，这两类错误分别叫做第一类错误（α 型错误）和第二类错误（β 型错误）。

在各类参数的假设检验中，要根据各种不同的条件，使用相应的计算标准误差的公式，不可错用。常用的假设检验方法有 Z 检验、t 检验、F 检验和 χ^2 检验。

$1-\beta$ 反映着正确辨认真实差异的能力，统计学中称之为统计检验力。利用样本的平均数、标准差、比率和相关系数等统计量进行某种类型的假设检验后，都需要用不同方法对相应的统计检验力进行估计。

关键术语：

统计推断：是通过从某个总体中抽取一定数量的个体组成样本，而后再通过计算样本的数字特征来推断总体的数字特征的过程。

推断统计：是指利用对样本调查的结果去推论总体数字特征的各种统计分析方法的统称，主要包括参数估计和假设检验两大部分内容。

总体：是指所要研究的具有某种特征的一类事物的全体。

个体：总体中的每个基本单元。

总体分布：总体中所有个体数据的分布。

抽样：从总体中抽取一定数量的个体并对其某项特征进行研究的过程。

样本：通过抽样过程被抽取的一组个体构成该次研究的一个样本。

样本分布：样本中所有个体在某项特征上的数据分布。

样本统计量：如果集中量数、差异量数、比率或相关系数等数据量度是根据样本数据计算得到的，则称这些数据量度为样本统计量。

抽样分布：是指样本统计量的分布。例如从总体中做很多次抽样，每次抽样得到的样本数据各自都有其样本平均数、样本方差、样本比率等统计量，它们的分布就属于抽样分布。

抽样平均数分布：从某个总体中多次随机抽样后各次抽样的平均数的分布。

标准误差：样本平均数分布的标准差与总体标准差之差

Z 分布：即标准正态分布。

Z 统计量：是指在 Z 检验中用以判断接受还是拒绝虚无假设的检验统计量。

Z 检验：是指根据检验统计量是否服从或渐近服从正态分布规律的假设检验方法。

总体参数：如果集中量数、差异量数、比率或相关系数等数据量度是由来自总体中的数据计算得到的，则称这些数据量度为总体参数。

自由度：就是允许自由取值的项数。

抽样误差：在同一总体中进行多次抽取同样容量的抽样过程中，由抽样本身所造成某类样本统计量之间的差异。

参数估计：利用样本统计量的信息估计总体参数的过程称为参数估计。

点估计：用样本统计量的单一数值估计未知的总体参数

区间估计：区间估计是依据样本统计量，根据一定精确度的要求，推断总体参数所在的区间和范围。

置信区间：区间估计以点估计为基础，它用数轴上的一段距离表示总体参数可能落入的范围，这一段距离或范围，统计上称为置信区间，也称置信间距。

置信界限：置信区间的上下两端点值叫做置信界限。

置信水平：总体参数被置信区间包含的概率或把握，统计上称为置信水平或置信度。

假设检验：根据抽样结果，在一定可靠程度上，对一个或多个有关总体分布的原假设做出拒绝或接受该原假设的推论的统计分析过程。

检验统计量：在假设检验过程中，用于判断接受还是拒绝原假设的统计量

研究假设：根据已有的理论和经验事先对研究结果做出一种预想的希望证实的假设。

虚无假设：研究假设的对立形式。

小概率事件原理：认为“小概率事件在一次试验中几乎是不可能发生的”。小概率事件的小概率记为α，称为显著性水平，是虚无假设为真的情况下，研究者错误地拒绝了虚无假设的最大概率值。

差异显著性检验：是指检验某个样本统计量的数值与总体参数之间的差异是否显著，或者检验两个样本统计量之间的差异是否显著的假设检验过程，其目的在于判断样本数据中表现出来的差异是否具有统计学意义，即是否可以推论到总体。

单侧检验：要检验的假设只是μ与μ_0的差异，而不是μ 比μ_0更大还是更
小的假设检验过程。

双侧检验：要检验的假设是μ比μ_0更大还是更小的假设检验过程。

第一类错误（α 型错误）：当虚无假设为真时拒绝接受它所犯的错误。

第二类错误（β 型错误）：当虚无假设为假时接受它所犯的错误。

重要公式：

总体平均数：
$$\mu = \frac{(X_1 + X_2 + \cdots + X_N)}{N}$$

样本平均数：
$$\overline{X} = \frac{(X_1 + X_2 + \cdots + X_n)}{n}$$

总体方差：
$$\sigma^2 = \frac{\sum (X - \mu)^2}{n}$$

$$\sigma = \sqrt{\frac{\sum (X - \mu)^2}{n}}$$

总体标准差：

样本方差：
$$S^2 = \frac{\sum (X - \overline{X})^2}{n}$$

样本无偏方差： $S^2=\dfrac{\sum(X-\overline{X})^2}{n-1}$

样本标准差： $S=\sqrt{\dfrac{\sum(X-\overline{X})^2}{n}}$

样本无偏标准差： $S=\sqrt{\dfrac{\sum(X-\overline{X})^2}{n-1}}$

总体相关系数： $\rho=\dfrac{\sum(X-\mu_X)(Y-\mu_Y)}{\sqrt{\sum(X-\mu_X)^2}\sqrt{\sum(Y-\mu_Y)^2}}$

样本相关系数： $r=\dfrac{\sum(X-\overline{X})(Y-\overline{Y})}{\sqrt{\sum(X-\overline{X})^2}\sqrt{\sum(Y-\overline{Y})^2}}$

样本平均数的 Z 分数： $Z_{\overline{X}}=\dfrac{\overline{X}-\mu}{SE_{\overline{X}}}=\dfrac{\overline{X}-\mu}{\sigma/n}$

总体平均数估计的表达式： $\overline{\mathrm{X}}-Z_{\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}<\mu_0<\overline{\mathrm{X}}+Z_{\alpha/2}\dfrac{\sigma_0}{\sqrt{n}}$

单总体平均数差异显著性的 Z 检验公式： $Z=\dfrac{\bar{X}-\mu_0}{\dfrac{\sigma_0}{\sqrt{n}}}$

思考与练习

1. 某中学从重点班随机抽取 45 名学生进行比奈智力测验，结果 $\overline{X}=108$。已知比奈测验的常模 $\mu_0=100$，$\sigma_0=16$。能否认为该中学重点班学生的智力水平高于常模水平？

2. 已知某情绪智力测验的成绩服从正态分布，且其标准差为 9。现从参加该情绪智力测验的某校学生中随机抽取 16 名学生，算得他们的平均测验分数为 76 分，试估计该校学生情绪智力测验的平均成绩。

3. A 市的心理协会对该市心理咨询收费情况进行调查，随机抽取了 18 名心理咨询师，了解到他们的平均收费标准为 300 元/小时。已知全国心理咨询的平均收费标准为 360 元/小时，标准差为 120 元/小时。若该市心理咨询收费的标准差与全国相同，请估计 A 市心理咨询平均收费 0.95 的置信区间。能否推论该市心理咨询师的收费与全国水平不同？

4. 为了研究社交焦虑者是否比一般人更内向，研究者随机抽取 40 名社交焦虑者进行内外向人格测试，测得他们的平均内倾性分数为 38 分，已知普通人群的平均内倾性分数为 32 分，标准差为 7.6 分，能否推论社交焦虑者比普通人更为内向？

第六章 平均数的参数估计与差异显著性检验

学习本章内容，将有助于你对以下问题的理解与思考：

1. 什么是 t 分布与 t 检验？
2. 如何进行总体方差未知时的单总体平均数显著性检验？
3. 如何进行两总体平均数差异显著性检验？

1. 某心理学研究者试图对某校新生在智商方面的性别差异进行比较，在男女新生中各随机抽取 100 名学生进行韦氏智力量表测试后，测得女生平均智商为 115 分，男生平均智商为 111 分。根据常模，男女生总体智商的标准差都是 15 分。请检验该校男女新生在智商方面是否有显著差异。

2. 某职业院校安排学生进行技能实训，对照组（20 人）采用传统的现场实训法，而实验组（20 人）采用计算机模拟和现场实训相结合的新方法，已知两组被试是随机分配的且两组其他条件保持一致。实训一个月后对两组进行技能测试，结果对照组成绩的平均数为 83.3，标准差为 3.50；实验组成绩的平均数为 86.4，标准差为 3.66。能否推论计算机模拟和现场实训相结合的新方法效果更好？

第一节　总体平均数的估计

一、 t 统计量与 t 分布的概述

（一）t 统计量与 t 分布的含义

在第 5 章介绍参数估计的基本原理时，所介绍的是总体呈正态分布且总体标准差已知时，估计总体平均数的方法。但是，我们并非总是在知道总体标准差的情况下，来对总体平均数进行估计。另一方面，与抽样平均数相应的总体分布也并非都是正态分布，因此，我们还需要在不知道总体标准差以及总体为非正态分布的情况下，根据样本平均数的有关数据来对总体平均数进行区间估计。

当总体是正态分布且总体标准差 σ 未知时，虽然样本平均数 $\overline{X}$ 的抽样分布仍然是正态分布 $N\left(\mu,\sigma^2/n\right)$，但由于 σ 未知，无法计算 $\overline{X}$ 分布的标准误 $SE_{\overline{X}}=\dfrac{\sigma}{\sqrt{n}}$，统计量 $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 中同时含有 μ 和 σ 两个未知参数，因此也无法根据公式 5-12 计算 Z 统计量，也就不能用 Z 统计量作为枢轴量对 μ 进行区间估计。根据数理统计学的原理，这时可以考虑用样本的无偏标准差 S_{n-1} 来近似代替总体标准差 σ。 但替代之后，统计量 $\dfrac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 服从的分布不再是标准正态分布，这是它和统计量 $Z=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 的重要区别。数理统计学中规定，若（$X_1,X_2,\cdots,X_n$）是来自正态总体的一个样本, 统计量 $\dfrac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 服从的分布是自由度为 n-1 的 t 分布，记作

$$t=\frac{(\overline{X}-\mu)}{\dfrac{S_{n-1}}{\sqrt{n}}}\sim t_{n-1}\qquad\text{（公式 6-1）}$$

服从 t 分布的统计量称为 t 统计量（t score，又称 t 分数），因而统计量 $\dfrac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 是 t 统计量。

公式 6-1 中，S_{n-1} 是样本的无偏标准差，是由样本的无偏方差 $S_{n-1}^2=\dfrac{\sum(X-\overline{X})^2}{n-1}$ 开方求得。如果样本方差是由公式 $S_n{}^2=\dfrac{\sum(X-\overline{X})^2}{n}$ 求得（本书中未加声明的 S^2 均指 $S_n{}^2$，即样本的有偏方差）,则公式 6-1 也可以被表示成

$$t = \frac{(\overline{X} - \mu)}{\frac{S_n}{\sqrt{n-1}}} \sim t_{n-1}$$ （公式 6-2）

如前所述，从总体方差σ未知的正态总体中抽样，尽管样本平均数$\overline{X}$的抽样分布仍然是正态分布$N\left(\mu,\sigma^2/n\right)$。但由于$\sigma$未知，故无法计算$\overline{X}$分布的标准误$SE_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$，因此实际工作中有时也用$\frac{S_{n-1}}{\sqrt{n}}$作为$\frac{\sigma}{\sqrt{n}}$的估计值，也称为$\overline{X}$分布的（近似）标准误，记作$SE_{\overline{X}} = \frac{S_{n-1}}{\sqrt{n}}$或$SE_{\overline{X}} = \frac{S_n}{\sqrt{n-1}}$。此时，公式 6-1 或公式 6-2 可以简单地表示为

$$t = \frac{\overline{X} - \mu}{SE_{\overline{X}}}$$ （公式 6-3）

（二）t 分布的性质

t 分布（t distribution）是一族分布，其分布曲线与标准正态分布相似，都是单峰对称的分布。随自由度的不同，t 分布的曲线形状呈现出差异。不同自由度情况下，曲线下大于或小于某一特定 t 值的面积占总面积的比率不同。具体而言，t 分布具有如下性质：

（1）t 分布是单峰对称的分布，其平均值为 0。

（2）t 统计量的取值在-∞ ～+∞之间，左侧 t 为负值，右侧 t 为正值，

（3）t 分布是一族分布。当样本容量 n 趋于无穷大时，t 分布趋近于正态分布。一般当自由度 n-1 在 30 以上时，t 分布就接近正态分布，方差接近 1，如图 5-3 所示。

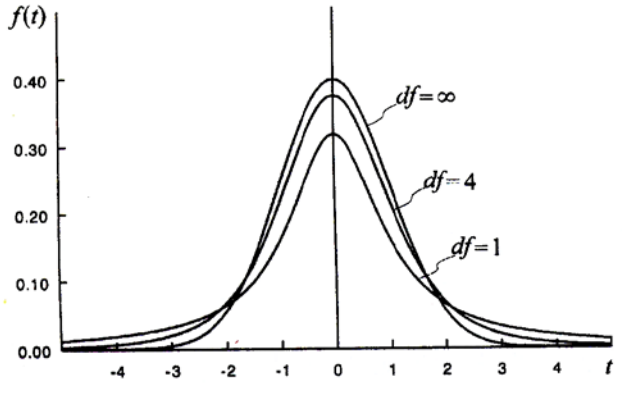


图 6-1　t 分布的概率密度曲线

（三）t 分布的分位数和分布表

在一个特定的 t 分布中，若大于某 t 值的概率为α，则称该 t 值为 t 分布的上α分位数；类似地，若小于某 t 值的概率为α，则称该 t 值为 t 分布的下α分位数。一般情况下，用t_α表示 t 分布的上α分位数。由于 t 分布是连续分布，所以 t 分布的下α分位数和上(1-α)分位数是同一个数值，故 t 分布的下α分位数一般记作$t_{1-\alpha}$，如图 5-4。由于 t 分布是对称的，所以$t_{1-\alpha} = -t_\alpha$。

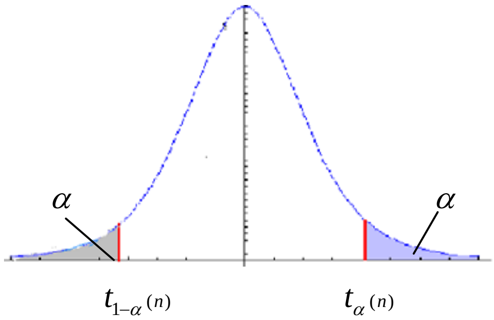


图 6-2 t 分布的分位数

例如，若某个随机变量服从自由度为 df =10 的 t 分布，其取值大于 1.812 的概率为 0.05，即 $P(t \geq 1.812)$=0.05，则称 1.812 是自由度为 10 的 t 分布的上 0.05 分位数，记作 $t_{0.05}$，有时为说明自由度，也可以记作 $t_{0.05}(10)$；由于 $-t_{\alpha}(n) = t_{1-\alpha}(n)$，所以 $P(t \leq -1.812)$= 0.05，即 $t=-1.812$ 是 t 分布的下 0.05 分位数。

附表 2 是常用的 t 分布表，表的左列为自由度，最上一行是不同自由度下大于上分位数和小于下分位数的概率和，即 P 值。表中是相应自由度和双侧概率下的 t 值，称为双侧界限，用 $t_{\alpha/2}$ 表示；表的最下一行是不同自由度下，上分位数以上的概率，称为单侧概率，相应的 t 值即单侧界限，记作 t_{α}。例如，df =16，双侧概率为 0.05 时，t 值为 2.12，表示大于 t=2.12 和小于 t=-2.12 的概率和为 0.05；而在同样自由度下，单侧概率为 0.05 时，t=1.746。

二、 总体正态分布但总体方差 σ^2 未知时，利用 t 分布估计总体平均数

用样本标准差 S 取代总体标准差 σ 之后，我们就得到与统计量 $Z_{\bar{x}} = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}}$ 相似的另一个统计量 $t = \frac{(\bar{X}-\mu)}{S_{n-1}/\sqrt{n}}$。$t$ 统计量中只有 μ 是未知的总体参数，所以可以根据 t 统计量的分布规律对 μ 进行区间估计。不同的是，总体标准差 σ 是总体参数，它是一个常量，而 S_{n-1} 是样本统计量，这是一个变量。t 统计量和 Z 统计量一样，在统计推断中占有重要的地位。

t 分布和标准正态分布 $N(0,1)$ 比较相似，都是平均数为 0、单峰对称的分布，而且当 $n \geq 30$ 时，t 分布逐渐接近标准正态分布。所以，我们有时也把 $\bar{X}$ 分布的近似标准误 $SE_{\bar{X}} = \frac{S_{n-1}}{\sqrt{n}} = \frac{S_n}{\sqrt{n-1}}$ 简称为 $\bar{X}$ 分布的标准误。比较公式 5-12 和公式 6-1 可以知道，计算 Z 统计量和计算 t 统计量的两个公式之间，只在于标准误差的计算方法不同。实际上，本章后面各种不同条件下的计算公式，其差别都在于因为抽样分布不一样，而使得计算相应的标准误差的方法不同。

当我们知道用样本标准差 S_{n-1} 替代总体标准差 σ 之后，其抽样分布服从 t 分布，那么我们就可以利用公式 6-1 即 $t = \frac{(\bar{X}-\mu)}{SE_{\bar{X}}}$ 计算 t 统计量，而后查 t 分布表（附表 2）来进行总体参数 μ 的区间估计。下面我们举例加以说明。

【例 6-1】已知某校学生每天的锻炼时间服从正态分布，从全体学生中随机抽取容量n=61的样本，调查得知样本中学生每天锻炼的平均时间为 26 分钟，方差 $S_{n-1}^2=25$（即标准差为 5 分钟），试估计全校学生平均每天锻炼时间。

分析：该题中总体方差σ^2未知，而样本方差S_{n-1}^2已知且是无偏样本量，加之样本容量 n ＞ 30，无论学生每天的锻炼时间是否为正态分布，应该考虑用 t 统计量来对全校学生平均每天锻炼的时间进行估计。

解：

（1）根据公式 6-3 即 $t=\dfrac{\overline{X}-\mu}{\frac{S_{n-1}}{\sqrt{n}}}$，建立总体平均数$\mu$在 0.05 显著性水平上（$\alpha=0.05$）的估计区间不等式：$\overline{X}-t_{\frac{\alpha}{2}}\dfrac{S_{n-1}}{\sqrt{n}}<\mu<\overline{X}+t_{\frac{\alpha}{2}}\dfrac{S_{n-1}}{\sqrt{n}}$

（2）根据显著性水平α查 t 分布表求相应的$\dfrac{\alpha}{2}$的 t 值。

当$\alpha=0.05$时，$\dfrac{\alpha}{2}=0.025$。查 t 分布表，对应于$\alpha=0.05$的双侧检验，自由度为 $n-1=61-1=60$的 t 值是 2.00。

（3）将有关数据代入上述不等式求具体的解：

$$26-2.00\frac{5}{\sqrt{61}}<\mu<26+2.00\frac{5}{\sqrt{61}}$$

$$24.72<\mu<27.28$$

如果把总体平均数的置信水平$1-\alpha$设定为 0.99，则有$\dfrac{\alpha}{2}=0.005$。查 t 分布表，对应于$\alpha=0.01$的双侧检验，自由度为 $n-1=61-1=60$的 t 值是 2.66。将有关数据代入上述不等式后可得：

$$26-2.66\frac{5}{\sqrt{61}}<\mu<26+2.66\frac{5}{\sqrt{61}}$$

$$24.30<\mu<27.70$$

答：全校学生平均每天锻炼的时间有 95%的可能性落在$24.72\sim27.28$分钟之间；有 99%的可能性落在$24.30<\mu<27.70$分钟之间。

在上例中，由于样本容量n=61，因此 t 分布的自由度为 60（远大于 30），此时 t 分布近似正态分布，因此也可以采用 Z 统计量为枢轴统计量构建能够包含总体平均数μ的置信区间$\overline{X}-Z_{\frac{\alpha}{2}}\dfrac{S_{n-1}}{\sqrt{n}}\le\mu\le\overline{X}+Z_{\frac{\alpha}{2}}\dfrac{S_{n-1}}{\sqrt{n}}$。以$\alpha=0.05$为例，Z 值为 1.96，而自由度为 60 时 t 值是 2.00，二者相差很小，因而所得置信区间也非常接近。

【例 6-2】某校对高中二年级学生进行英语水平测试，测试后从中随机抽取 12 个考生，考生的成绩分别为 83，91，62，50，74，68，70，65，85，71，58，63。试对该年级考生的该次测试成绩的平均数进行区间估计（取α=0.05）。

解：

（1）根据已知数据求样本平均数和样本标准差得：$\overline{X}=70$，$S=11.43$（这是有偏标准差）；

（2）根据公式 6-2 即 $t=\dfrac{\overline{X}-\mu}{\dfrac{S}{\sqrt{n-1}}}$，建立总体平均数 μ 在 0.05 显著性水平（或置信区间在 0.95）上的估计区间不等式：$\overline{X}-t_{\frac{\alpha}{2}}\dfrac{S}{\sqrt{n-1}}<\mu<\overline{X}+t_{\frac{\alpha}{2}}\dfrac{S}{\sqrt{n-1}}$。

（3）根据显著性水平 $\alpha=0.05$ 查 t 分布表，对应于 $\alpha=0.05$ 的双侧检验，自由度为 $n-1=12-1=11$ 的 t 值是 2.201；

（4）根据题目给定条件求得 $\overline{X}$ =70， S =11.43，将有关数据代入上述不等式求具体的解：

$$70-2.201\frac{11.43}{\sqrt{12-1}}<\mu<70+2.201\frac{11.43}{\sqrt{12-1}}$$

$$62.416<\mu<77.584$$

答：该校高二年级学生该次英语测试的成绩在 62.416～77.584 分之间，估计正确的概率为 0.95，错误的概率为 0.05。

请读者注意：在使用 Z 统计量进行 Z 检验时，若要使推断误差控制在显著性水平 α =0.05，只需要将判定接受或拒绝虚无假设的 Z 统计量的临界点值确定为 Z=1.96 就能达目的；而在使用 t 统计量进行 t 检验时，需要根据自由度来确定判定接受或拒绝虚无假设的 t 统计量的临界点值：在例 6-1 的条件下， t 统计量的临界点值要达到 t=2.00，在例 6-2 的条件下，t 统计量的临界点值要达到 t=2.201 才能将显著性水平控制在 $\alpha=0.05$。因此，学习心理统计学时，一定要牢记不同条件下的参数估计或假设检验问题一定要选用合适的统计量计算方法，否则就会出错。

三、非正态总体样本平均数的抽样分布及总体平均数的估计

现实当中也有很多随机变量的总体分布不明确，或者是非正态分布，那么它们的样本均值的分布情况如何确定呢？

理论上可以证明，当总体分布非正态时，样本平均数的分布与样本容量有关。当样本容量足够大（一般要求 $n\geq30$）时，样本平均数的分布近似正态分布。这就是中心极限定理的内容。中心极限定理是指：从平均数为 μ，方差为 σ^2 的任意总体中抽取容量为 n 的样本，当 n 充分大时（一般要求 $n\geq30$），样本平均数 $\overline{X}$ 的抽样分布近似服从平均数为 μ，标准差为 $\sigma/\sqrt{n}$ 的正态分布，即 $\overline{X}\sim N(\mu,\dfrac{\sigma^2}{n})$。其中，要求总体方差 σ^2 的值是有限的。

通过研究各种不同样本容量下，样本平均数 $\overline{X}$ 的抽样分布发现，如果总体分布偏离正态越远，要使样本平均数 $\overline{X}$ 的分布接近正态所需的样本容量就越大。在样本容量大于 30 时，样本平均数 $\overline{X}$ 的抽样分布都近似符合正态分布。如图 6-1 所

示，图中第一行的三个图形分别是三种总体分布形态，很明显它们都不是正态的。从这三个非正态总体中，随机抽取样本，随着样本容量的不同，样本平均数的分布曲线发生了明显的变化。当 $n=2$ 时，样本平均数 $\overline{X}$ 的分布与总体分布曲线颇不一样；当 n=5 时，样本平均数 $\overline{X}$ 的分布已经变成单峰分布了；当样本容量 n=30 时，

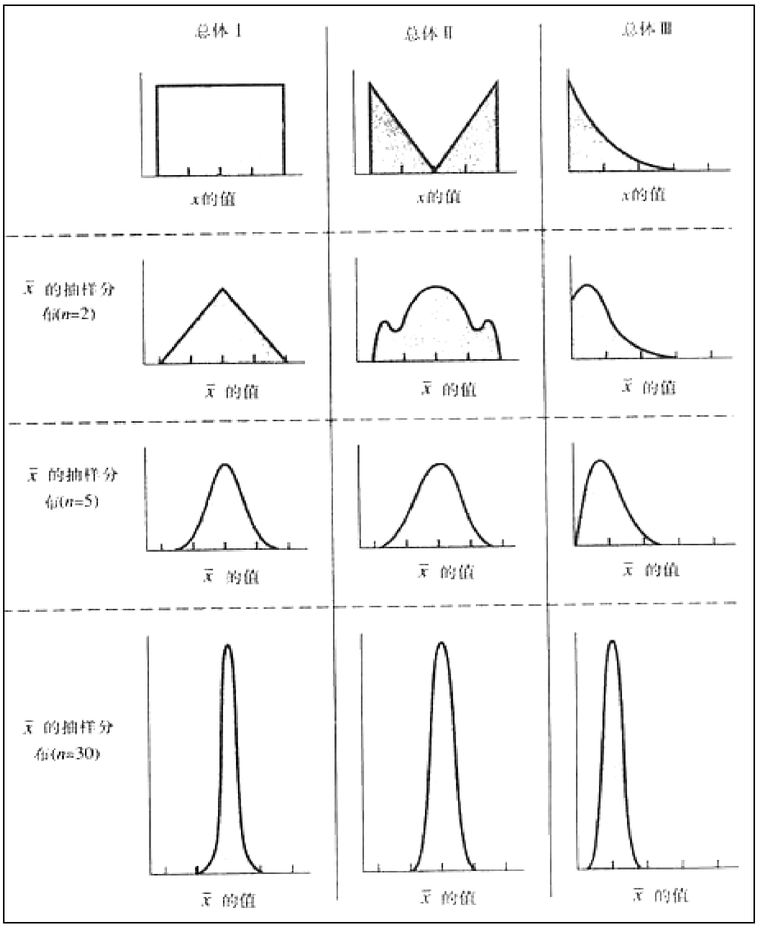


图 6-3 样本容量对非正态总体 $\overline{X}$ 抽样分布的影响

来自三个不同总体的样本平均数 $\overline{X}$ 的分布都近似正态分布。

总之，若总体为非正态分布，样本容量较小（ $n<30$ ）的情况下，一般不能对总体平均数进行估计。但是，若总体为非正态分布但样本容量有足够大（$n\geq 30$）时，根据中心极限定理可知，样本平均数 $\overline{X}$ 近似服从正态分布，其 Z 分数近似服从标准正态分布，即 $\overline{X} \sim N(\mu,\frac{\sigma^2}{n})$； $Z=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$。此时亦可以根据 Z 统计量的分布规律，对总体平均数 μ 进行区间估计。因为是近似估计，我们用符号 Z' 来代替原来的 Z 统计量。在实际中，由于总体方差 σ^2 也常常是未知的，因此可以样本的无偏标准差 S_{n-1} 代替 σ 计算 $\overline{X}$ 抽样分布的标准误，由于样本容量较大，此时统计量 $\frac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 也近似服从 $N(0,1)$。与这种计算相应的总体平均数的置信区间为：

$$\overline{X} - Z'_{\alpha/2}\frac{S_{n-1}}{\sqrt{n}} \leq \mu \leq \overline{X} + Z'_{\alpha/2}\frac{S_{n-1}}{\sqrt{n}}$$

【例 6-3】 从一个 μ=30，σ=5 的偏态分布中抽取一个容量为 35 的样本，则该样本平均数大于 32 的概率是多少？

解：总体分布非正态，样本容量为 n=35 > 30，样本平均数的分布近似正态分布
样本标准误 $SE_{\bar{x}}=\sigma/\sqrt{n}=5/\sqrt{35}$=0.84

由 $Z_{\bar{x}}=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}\sim N(0,1)$ 得出相应的 Z 统计量为：$Z'_{\overline{X}}=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}=\dfrac{32-30}{0.84}=2.38$

查附表 1 可知当 Z=2.38 时，大于该值以上的概率为 0.0087。

答：样本平均数高于 32 的概率是 0.0087。

【例 6-4】已知某学校初三年级语文会考成绩的分布是负偏态的。从该校初三学生总体中抽取 n=40 的样本，得到样本的平均成绩 $\overline{X}$ 为 79 分，标准差 S_{n-1} 为 12 分。试估计该学校初三语文会考成绩的总体平均数。

解：此题总体分布不是正态，但样本容量 $n\geq 30$，故此时统计量 $\dfrac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 也近似服从 $N(0,1)$，故可以根据 Z 统计量的分布规律进行推论。设置信水平为 0.95（或显著性水平为 0.05），则总体平均数 μ 的 0.95 的置信区间为：

$$\overline{X}-1.96\frac{S_{n-1}}{\sqrt{n}}\leq\mu\leq\overline{X}+1.96\frac{S_{n-1}}{\sqrt{n}}，即 \overline{X}-1.96\times\frac{12}{\sqrt{40}}\leq\mu\leq\overline{X}+1.96\times\frac{12}{\sqrt{40}}$$

$$75.281\leq\mu\leq 82.719$$

答：估计该校高三语文成绩的真实分数在 75.281~82.719 之间，估计正确的概率为 0.95，犯错误的概率为 0.05。

第二节　总体方差未知时的单总体平均数差异显著性检验

在第五章，我们用总体为正态分布、总体方差已知情况下的单总体平均数显著性为例，阐述假设检验的具体操作过程。但是，当总体分布形态是否正态、总体方差是否已知，抽取样本容量是大还是小等条件不一样时，相应的平均数的抽样分布原理也不同，因此对单总体平均数进行显著性检验的方法也不一样。下面我们分述之。

一、总体服从正态分布

当总体服从正态分布，总体方差 σ^2 未知时，一般使用 t 分布对抽样平均数与总体平均数之间，或者两个抽样平均数之间进行差异显著性检验，这种检验方法称为 **t 检验**（t test）。在实际操作时又分为两种情况：一种是抽取的样本是大样本即当 n≥30 时，由于 t 分布会近似正态分布。因而，当总体服从正态分布，总体方差未知时，只要是大样本，检验样本平均数和总体平均数间的差异可近似的用 Z 检验。由于总体方差未知，可用样本方差代替总体方差，则此时样本平均数的标准误为 $SE_{\overline{X}}=\dfrac{S_{n-1}}{\sqrt{n}}$，为区别起见，检验统计量可写作

$$Z'=\frac{\overline{X}-\mu}{\frac{S_{n-1}}{\sqrt{n}}} \qquad （公式 6-4）$$

当所抽取的样本又是小样本时，根据平均数的抽样分布理论可知，当总体服从正态分布，σ^2 未知时，方差 σ^2 可用其无偏估计 $S^2 = \frac{\sum(X-\overline{X})^2}{n-1}$ 来代替，这时 $\frac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 服从自由度 $df = n-1$ 的 t 分布，则此时的假设检验可由 t 检验来进行。

从理论上说，服从 Z 分布的统计量都可以叫做 Z 统计量，服从 t 分布的统计量也都可以叫做 t 统计量，但在心理统计中应用最多的 Z 统计量 $\left(Z = \frac{(\overline{X}-\mu)}{\frac{\sigma}{\sqrt{n}}}\right)$ 和 t 统计量 $\left(t = \frac{(\overline{X}-\mu)}{\frac{S}{\sqrt{n}}}\right)$，它们都与样本平均数的分布有关。利用 t 分布的规律来进行的假设检验过程称为 t 检验。下面我们用一个实际例子来阐述 t 检验的实际操作过程。

【例 6-5】 某教师试验一种新的英语教学法，经过一学期教学后，其所教班级 26 名学生的平均成绩为 74 分，标准差 S_{n-1} 为 10 分，而该学期全年级的平均分为 70 分。假设这 26 名学生在其他方面和全年级学生同质，试分析采用这种新教学方法是否提高了教学效果？（该年级学生考试成绩服从正态分布，$\alpha = 0.05$）

上述数据在总体上服从正态分布，由于不知道总体方差 σ^2 是多少，且所抽取的样本小于 $n = 30$，因此我们可以用 t 分布来对其进行单总体平均数差异显著性检验。具体操作步骤如下：

解：根据问题要求，采用单侧检验。

（1）建立假设：$H_0: \mu \le 70$
$H_1: \mu > 70$

（2）利用公式 6-1 计算统计量：$t = \frac{\overline{X}-\mu}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{74-70}{\frac{10}{\sqrt{26}}} = 2.04$

（3）确定显著性水平的临界值并做出决策：根据显著性水平 $\alpha = 0.05$、自由度 $df = n-1$，查 t 分布表得 $t_{\alpha(25)} = 1.708$。根据 t 检验的要求，由于实得的 $t = 2.04 > t_{\alpha(25)} = 1.708$，因此拒绝虚无假设 H_0，即认为新教学方法提高了教学效果。

答：新教学方法提高了教学效果。

二、总体非正态分布

上述所讨论的 Z 检验和 t 检验都是以总体服从正态分布为前提条件的。在心理学研究中，大部分连续变量在总体上都可看作是正态分布的，但如果在某些研究中，事先有理由认为某一变量其总体不是服从正态分布，就不能进行 Z 检验和 t 检验，而应该采用非参数检验。

但实际上，当样本容量较大的情况下（n≥30），也可以近似地应用 Z 检验。因为数学上已证明：从平均数为 μ，标准差为 σ 的总体（无论正态与否）中随机

抽样，则样本平均数 $\overline{X}$ 的分布将随着样本容量的增大而趋近于正态分布，即 $\overline{X}$ ~ $N(\mu, \frac{\sigma^2}{n})$。所以当 n≥30 时，尽管总体分布未知或非正态，对于平均数的显著性检验仍可以近似用 Z 检验，即 $Z' = \frac{(\overline{X}-\mu)}{\frac{\sigma}{\sqrt{n}}}$，当总体方差 σ 未知时，可用 S 代替 σ，即：

$$Z' = \frac{(\overline{X}-\mu)}{\frac{S_{n-1}}{\sqrt{n}}} \qquad \text{（公式 6-5）}$$

【例 6-6】 某省进行物理竞赛，结果分数的分布不是正态分布。竞赛的总平均分是 61 分。随机抽取某市参加竞赛的学生 168 人，算得他们竞赛成绩的平均分 $\bar{X}$=59.4，S_{n-1}=18.7，问该市物理竞赛的成绩与全省是否有显著差异？

解：总体不是正态分布，但 n=168，是大样本，则可用近似 Z 检验。

（1）建立假设：$H_0: \mu = 61$
$H_1: \mu \neq 61$

（2）计算统计量：$Z' = \frac{(\overline{X}-\mu)}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{59.4-61}{\frac{18.7}{\sqrt{168}}} = -1.11$

（3）确定显著性水平并做出决策：由于 $|Z'|$=1.11＜1.96，故接受虚无假设 H_0，即认为该市平均分与省平均分没有显著性差异。

答：该市平均分与省平均分没有显著差异。

在单总体均值的显著性检验中，需要强调的是：当总体呈正态分布，总体方差已知时，无论是大样本还是小样本，都可以用 Z 检验；当总体正态分布，总体方差未知时，小样本时，只能用 t 检验，而如果是大样本，用 t 检验和近似 Z 检验均可；当总体非正态，小样本时，不能用 Z 检验和 t 检验，而只能用非参数检验，只有在大样本情况下，才可以近似地采用 Z 检验。

第三节 两总体平均数差异显著性检验

一、两总体平均数差异显著性检验的含义及其抽样分布原理

第五章及本章前面所讨论的假设检验仅是对一个样本平均数与总体平均数之间差异的检验。其中涉及的差异显著性检验的基本逻辑、检验目的、原理和过程等，即可以用于单总体，也可以用于对两个总体平均数之间是否具有显著差异进行检验。

（一）“两总体平均数差异显著性检验”的含义

所谓“两总体平均数差异显著性检验”是指“从分别命名为‘X’和‘Y’的两个总体中各抽取一个样本，对这两个样本平均数之间的差异（$\bar{X}-\bar{Y}$）是否具有统计学意义所进行的假设检验过程”。这种检验的目的在于通过对两个样本平均数之间的差异（$\bar{X}-\bar{Y}$）进行考察，看看能否在特定的显著性水平下推论两个总体平均数之间存在差异。

与单总体平均数差异显著性检验过程一样，两总体平均数差异显著性检验过程也是根据其抽样分布原理，通过以下三个步骤来进行统计推断：（1）建立假设（2）求取检验统计量；（3）确定显著性水平并做出统计推断的决策。

（二）“两变量之差的方差”的含义

由于在对“两总体平均数差异显著性检验”的求解计算过程中将用到两变量之差的方差。因此我们首先要了解两变量之差的方差（σ_{X-Y}^2）的计算过程。

假设有两个总体分别为 X 和 Y，它们的平均数分别为 μ_X 和 μ_Y，方差分别为 σ_X^2 和 σ_Y^2，$X_1, X_2, \cdots, X_n$ 是来自 X 的样本容量为 n 的随机样本，$Y_1, Y_2, \cdots, Y_n$ 是来自 Y 的样本容量为 n 的随机样本。若两样本之差记作 D，则 $D_i = X_i - Y_i$，由多个 D 构成的总体的平均数 $\mu_D = \mu_X - \mu_Y$。根据公式 3-6 即方差计算公式 $\sigma^2 = \frac{\sum (X-\mu)^2}{n}$，可得两样本之差 D 的方差 σ_{X-Y}^2：

$$
\begin{aligned}
\sigma_{X-Y}^2 &= \frac{\sum (D_i - \mu_D)^2}{n} = \frac{\sum [(X_i - Y_i) - (\mu_X - \mu_Y)]^2}{n} \\
&= \frac{\sum (X_i - Y_i - \mu_X + \mu_Y)^2}{n} = \frac{\sum [(X_i - \mu_X) - (Y_i - \mu_Y)]^2}{n} \\
&= \frac{\sum [(X_i - \mu_X)^2 - 2(X_i - \mu_X)(Y_i - \mu_Y) + (Y_i - \mu_Y)^2]}{n} \\
&= \frac{\sum (X_i - \mu_X)^2}{n} - \frac{2\sum (X_i - \mu_X)(Y_i - \mu_Y)}{n} + \frac{\sum (Y_i - \mu_Y)^2}{n} \\
&= \sigma_X^2 - 2r\sigma_X\sigma_Y + \sigma_Y^2
\end{aligned}
$$

（r 为两样本的相关系数）（公式 6-6）

同理，可得两变量和（σ_{X+Y}^2）的方差的计算公式：

$$\sigma_{X+Y}^2 = \sigma_X^2 + 2r\sigma_X\sigma_Y + \sigma_Y^2 \quad \text{（公式 6-7）}$$

当两个样本独立时，即 $r=0$，$\sigma_{X\pm Y}^2 = \sigma_X^2 + \sigma_Y^2$ （公式 6-8）

（三）两样本平均数之差的抽样分布

对平均数差异的显著性进行检验需要了解其抽样原理，在具体的求解过程中还需要注意对这两个样本所属的总体分布情况、两样本是独立样本还是相关样本、总体方差是否已知、总体方差是否齐性以及两个样本容量大小是否相同等不同条件加以区别，并根据不同条件采用不同的假设检验方法和公式。

下面在“两总体正态分布、两总体方差均已知、并且两个样本相互独立”的条件下对两样本平均数之差的抽样分布进行讨论。

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $X \sim N(\mu_X, \sigma_Y^2)$ 的样本容量为 n_1 的随机样本，该样本的平均数为 $\overline{X}$；$Y_1, Y_2, \cdots, Y_n$ 是来自正态总体 $Y \sim N(\mu_Y, \sigma_Y^2)$ 的样本容量为 n_2 的随机样本，该样本的平均数为 $\overline{Y}$，且两组样本相互独立。由于两个样本数据的取样过程是独立进行的，即两样本中的数据不存在对应性，这样抽取的两个样本称为独立样本（independent sample）。我们要通过这两个样本带来的信息，推断两总体平均数 μ_X 和 μ_Y 是否有差异。

根据平均数的抽样分布理论可知，当总体正态，σ_X^2 已知时，$\overline{X} \sim N(\mu_X, \frac{\sigma_X^2}{n_X})$，

总体正态，σ_Y^2 已知时，$\overline{Y} \sim N(\mu_Y, \frac{\sigma_Y^2}{n_Y})$，则 $\overline{X}-\overline{Y}$ 也服从正态分布，即 $(\overline{X}-\overline{Y}) \sim N(\mu_X-\mu_Y, \frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y})$（根据上述两变量之差的方差推导原理，很容易推算出 $\sigma^2_{\overline{X}-\overline{Y}} = \frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}$），其抽样分布的平均数是两总体平均数之差 $\mu_X-\mu_Y$，其标准误是 $SE_{\overline{X}-\overline{Y}} = \sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}$。任何普通的正态分布都可以转换成标准正态分布，在这里，$Z = \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}} \sim N(0,1)$，此时对两总体均值的检验就可以用 Z 检验，其统计量的计算公式为：

$$Z = \frac{(\overline{X}-\overline{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}} \qquad \text{（公式 6-9）}$$

上述是对两个不同平均数的抽样分布的理论阐述。后面举例说明时，为了方便，我们用 $\overline{X}_1$ 表示前面所述的 $\overline{X}$，用 $\overline{X}_2$ 表示前面所述的 $\overline{Y}$，相应的两个总体参数也分别用 μ_1 和 μ_2 表示，两个抽样的样本容量分别用 n_1 和 n_2 表示，这样，公式 6-9 就可以相应地表示为：

$$Z = \frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \qquad \text{（公式 6-10）}$$

（四）“虚无假设分布”的含义 （P241）：

第五章曾指出：“统计假设检验的基本思想是带有概率值保证的反证法。也就是说，我们想要证实研究假设，但并不是从研究假设出发进行验证，而是建立与它对立的虚无假设，并假定虚无假设为真。在虚无假设为真的前提下，通过收集实际信息并依据正确逻辑推理和数学上的计算与分析，看实际获得的资料所导致的结果是否与虚无假设成立时应出现的结果发生矛盾。如果出现了矛盾则表明原先的假设 H_0 是错误的，应该给予否定，从而接受研究假设。如果没有出现矛盾，则表明没有充分理论否定虚无假设”。

下面我们要结合前面所述的“两样本平均数之差的抽样分布”来理解“虚无假设分布”这一概念（关于“备择假设分布”的概念将在本书第十一章论述）。

在对两总体**平均数差异显著性进行**假设检验过程中，通常将检验的虚无假设设为 $H_0: \mu_1 = \mu_2$，这时，所谓虚无假设分布（null hypothesis dis- tribution,NHD）就是指当虚无假设 H_0 为真时，$\overline{X}_1 - \overline{X}_2$ 的分布。

当把检验的虚无假设设为 $H_0: \mu_1 = \mu_2$ 时，“虚无假设 H_0 为真”意味着 $\mu_1 - \mu_2$=0，两个总体的正态分布图几乎完全重叠在一起（如图 6-4 所示）。

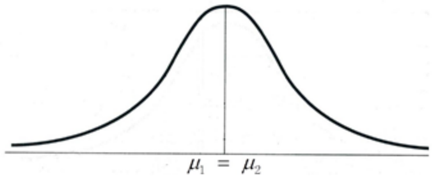


图 6-4 虚无假设 $H_0: \mu_1 = \mu_2$ 的正态分布图

如果我们在上述“μ_1” 和“μ_2”两个正态总体中随机各抽取同样数量的样本，比如说50 人（即$n_1 = n_2 = 50$），计算其某项心理特征的平均指标，则可以分别得到“$\overline{X}_1$”和“$\overline{X}_2$”两个样本统计量，由于存在随机性抽样误差（记为“ε_1”），因此，虽然这两个样本相应的总体之差即“$\mu_1 - \mu_2$=0”，但这次抽取的两个样本统计量之差（记为“D_1”）不一定会正好等于零，也许比零大一点，也许比零小一点，换言之：“$D_1 = \overline{X}_1 - \overline{X}_2 = \varepsilon_1$”；按同样方法我们可以在这两个总体中第二次随机各抽取 50 人计算同一项心理特征的平均指标，又可以再次分别得到“$\overline{X}_1$”和“$\overline{X}_2$”两个样本统计量，以及这两个样本统计量之差“D_2”（$D_2 = \overline{X}_1 - \overline{X}_2 = \varepsilon_2$）……我们可以用同样的抽样方法重复无数多次，比如重复 K 次，则可以得到 K 次“$\overline{X}_1$”和“$\overline{X}_2$”两个样本统计量，以及这两个样本统计量之差“D_k”（$D_k = \overline{X}_1 - \overline{X}_2 = \varepsilon_k$）。

由于每个“D值”都围绕“0”左右摇摆，故这“k”个 D 值也有自己的分布，所有 D 值的分布就构成“虚无假设分布（即‘$\overline{X}_1 - \overline{X}_2$的分布’”）；又因为“$\varepsilon_1 + \varepsilon_2 + \cdots + \varepsilon_k = 0$”，因此，“虚无假设分布”也是以零为中心的正态分布，同时，在虚无假设为真的前提下计算得到的虚无假设差异显著性检验的 Z 统计量分布（或 t 统计量分布）也是以零为中心的分布（a central Z or t distribution）。

二、两总体正态分布，并且相互独立的平均数差异显著性检验的具体求解过程

（一）两总体方差已知

【例 6-7】某心理学研究者试图对某校新生在智商方面的性别差异进行比较，在男女新生中各随机抽取 100 名学生进行韦氏智力量表测试后，测得女生平均智商为 115 分，男生平均智商为 111 分。根据常模，男女生总体智商的标准差都是 15 分。请检验该校男女新生在智商方面是否有显著差异。

解：已知
$$n_1 = 100, \overline{X}_1 = 115, \sigma_1 = 15$$
$$n_2 = 100, \overline{X}_2 = 111, \sigma_2 = 15$$

（1）建立假设：
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

（2）用公式 6-10 计算检验统计量：

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{115 - 111}{\sqrt{\frac{15^2}{100} + \frac{15^2}{100}}} = \frac{4}{2.12} = 1.89$$

（3）确定显著性水平并做出决策：在$\alpha = 0.05$的水平下，$Z = 1.89 < 1.96$，故接受虚无假设，即认为**该**校男女生在智商方面没有显著性差异。

答：该校男女生在智商方面没有显著性差异。

（二）两总体方差未知

与单总体平均数假设检验一样，当总体正态分布，总体方差σ^2未知时，可用样本方差S^2来代替，且需要用到 t 检验。由于这里是两个总体方差σ_1^2和σ_2^2都未知，因此需要考虑的条件更多，除了要考虑两个总体方差是否相等外，还要考虑两个样本的大小等条件。

1. **两总体方差一致或相等**

首先要说明的是，这里的“一致（或相等）”，不一定是两总体方差在数值上完全相同，而是进行方差齐性检验时（方差齐性检验方法参见第七章），两总体的方差差异不显著。

根据平均数的抽样分布理论可知，当总体正态，σ^2 未知时，方差 σ^2 可用其无偏估计 $S_{n-1}^2=\frac{\sum(X-\overline{X})^2}{n-1}$ 来代替。$S_{n_1-1}^2=\frac{\sum(X-\overline{X}_1)^2}{n-1}$ 和 $S_{n_2-1}^2=\frac{\sum(X-\overline{X}_2)^2}{n-1}$ 分别为各自总体方差的无偏估计，在这一条件下，$\overline{X}_1-\overline{X}_2$ 的抽样分布服从 $df=n_1+n_2-2$ 的 t 分布，其平均数为 $\mu_1-\mu_2$， 其标准差 $SE_{\overline{X}_1-\overline{X}_2}=\sqrt{\frac{(n_1-1)S_{n_1-1}^2+(n_2-1)S_{n_2-1}^2}{n_1+n_2-2}\cdot(\frac{1}{n_1}+\frac{1}{n_2})}$ 或

$SE_{\overline{X}_1-\overline{X}_2}=\sqrt{\frac{n_1S_{n_1}^2+n_2S_{n_2}^2}{n_1+n_2-2}\cdot(\frac{1}{n_1}+\frac{1}{n_2})}$（两样本平均数之差的标准误差），当 $n_1=n_2$ 时，

$SE_{\overline{X}_1-\overline{X}_2}=\sqrt{\frac{S_{n_1-1}^2+S_{n_2-1}^2}{n}}$ 或 $SE_{\overline{X}_1-\overline{X}_2}=\sqrt{\frac{S_{n_1}^2+S_{n_2}^2}{n-1}}$。此时检验统计量为：

$$t=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{SE_{\overline{X}_1-\overline{X}_2}} \qquad \text{（公式 6-11）}$$

【例 6-8】 两区进行小学六年级作文竞赛。为比较结果，从甲区随机抽取了 80 名学生，其竞赛成绩的平均分为 82 分，标准差为 11.5 分；从乙区随机抽取了 60 名学生，其竞赛成绩的平均分为 78 分，标准差为 10.5 分。问两区作文竞赛成绩有无显著差异？（ α＝.01）

解：已知 $n_1=80,\overline{X}_1=82,S_1^2=11.5^2$
$n_2=60,\overline{X}_2=78.42,S_2^2=10.5^2$

（1）建立假设：$H_0:\mu_1=\mu_2$
$H_1:\mu_1\neq\mu_2$

（2）用公式 6-10 计算检验统计量：

$$SE_{\overline{X}_1-\overline{X}_2}=\sqrt{\frac{(n_1-1)S_{n_1-1}^2+(n_2-1)S_{n_2-1}^2}{n_1+n_2-2}\cdot(\frac{1}{n_1}+\frac{1}{n_2})}$$

$$=\sqrt{\frac{80\times11.5^2+60\times10.5^2}{80+60-2}\cdot(\frac{1}{80}+\frac{1}{60})}=1.906$$

$$t=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{SE_{\overline{X}_1-\overline{X}_2}}=\frac{(82-78)-0}{1.906}=2.10$$

（3）确定显著性水平并做出决策：因为 $t = 2.10 < t_{\frac{0.01}{2}(138)} \approx 2.62$，不能拒绝虚无假设 H_0。在 0.01 的显著性水平下无法推论两区作文竞赛成绩不同。

答：在 0.01 的显著性水平下无法推论两区作文竞赛成绩不同。

本题解题过程中有几个方面值得注意。首先，采用本题中公式计算 $SE_{\overline{X}_1-\overline{X}_2}$ 的前提是两总体方差齐性，理论上要先进行两总体方差的齐性检验（本题中 S_1^2 和 S_2^2 相差不大，经检验两总体方差为齐性）。其次，本题中对显著性水平的规定较为严格，即要求拒绝 H_0 犯错误的概率不能超过 0.01。如果将 α 设定为 0.05，由于 $t = 2.10 > t_{\frac{0.05}{2}} = 1.98$，则可以在 0.05 的显著性水平下拒绝 H_0，接受 H_1。可见假设检验是基于特定样本信息和概率标准的统计决策和推论，当对犯错误的概率要求不同时，可能得出不同的决策结果。此外，本题由于在两区中抽样的样本容量均较大，$\overline{X}_1 - \overline{X}_2$ 的抽样分布也近似 Z 分布，因而也可以参考 Z 分布的临界值做统计决策。

【例 6-9】 某职业院校安排学生进行技能实训，对照组（20 人）采用传统的现场实训法，而实验组（20 人）采用计算机模拟和现场实训相结合的新方法，已知两组被试是随机分配的且两组其他条件保持一致。实训一个月后对两组进行技能测试，成绩如表 6-1 所示。能否推论计算机模拟和现场实训相结合的新方法效果更好？（本题原始数据见本书所附光盘中例 6-9.sav 。）

表 6-1 两组被试测试成绩表

序　号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
对照组	80	87	85	78	88	80	86	82	83	79	81	88	86	79	89	81	87	83	84	80
实验组	77	88	90	85	89	83	87	86	88	86	78	89	91	86	90	84	88	87	89	87

解：
$$n_1 = 20, \overline{X}_1 = 83.3, S_{n_1-1} = 3.50$$
$$n_2 = 20, \overline{X}_2 = 86.4, S_{n_2-1} = 3.66$$

（1）建立假设：
$$H_0: \mu_1 \geq \mu_2$$
$$H_1: \mu_1 < \mu_2$$

（2）计算检验统计量：由于 S_1^2 与 S_2^2 相差不多，故可以认为两总体方差相等（符合方差齐性检验结果），根据公式 6-11 得

$$t = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{SE_{\overline{X}_1-\overline{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2)}{\sqrt{\dfrac{S_{n_1-1}^2 + S_{n_2-1}^2}{n}}} = \frac{83.3 - 86.4}{\sqrt{\dfrac{3.50^2 + 3.66^2}{20}}} = -2.738$$

（3）确定显著性水平并作出决策：对给定的 $\alpha = 0.05$，且 $df = 20 + 20 - 2 = 38$，查 t 分布表得 $t_{0.05(38)} \approx 1.684$，$|t| = 2.738 > 1.684$，应拒绝虚无假设 H_0，接受 H_1

答：可以推论计算机模拟和现场实训相结合的新方法效果更好，作此推论犯错误的概率不超过 0.05。

2. 两总体方差不齐性

若从经验上或经过方差齐性检验发现，两总体方差差异显著，即$\sigma_1^2 \neq \sigma_2^2$，这时统计量$t$不再服从自由度为$df = n_1 + n_2 - 2$的$t$分布，用公式 6-11 进行$t$检验是不合适的。对于这个问题不少人提出过各种检验方法，其中阿斯平-威尔士（Aspin-Welch）检验和柯克兰-柯克斯（Cochran-Cox）检验最常用。

（1）阿斯平-威尔士（Aspin-Welch）检验

设

$$k = \frac{\frac{S_1^2}{n_1}}{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} \qquad df = \frac{1}{\frac{K^2}{n_1} + \frac{(1-K)^2}{n_2}} \quad \text{（df 取整数）} \qquad \text{（公式 6-12）}$$

此时，

$$t' = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad \text{（公式 6-13）}$$

将近似地服从自由度为 df 的 t 分布，从而可以根据给定的显著性水平及自由度查 t 分布表，得到临界值，然而通过比较做出结论。

【例 6－10】 仍以例 6－9 的技能实训方法问题为例说明。其它条件不变，只是实验组和对照组的标准差分别为 5 和 3，检验新实训方法的效果是否更好。

解：

$$n_1 = 20, \overline{X}_1 = 83.3, S_1 = 5$$
$$n_2 = 20, \overline{X}_2 = 86.4, S_2 = 3$$

因为$S_{n_1-1}^2$和$S_{n_2-1}^2$相差较大，方差齐性检验结果显示两总体方差不等，则采用阿斯平-威尔士检验。由

$$k = \frac{\frac{S_{n_1-1}^2}{n_1}}{\frac{S_{n_1-1}^2}{n_1} + \frac{S_{n_2-1}^2}{n_1}} = \frac{\frac{5^2}{20}}{\frac{5^2}{20} + \frac{3^2}{20}} = 0.74$$

$$df = \frac{1}{\frac{k^2}{n_1} + \frac{(1-k)^2}{n_1}} = \frac{1}{\frac{0.74^2}{20} + \frac{(1-0.74)^2}{20}} = 32.51$$

（1）建立假设：

$$H_0 : \mu_1 \geq \mu_2$$
$$H_1 : \mu_1 < \mu_2$$

（2）计算检验统计量，根据公式 6-12 得

$$t' = \frac{(\overline{X_1} - \overline{X_2})}{\sqrt{\frac{S_{n_1-1}^2}{n_1} + \frac{S_{n2\_1}^2}{n_2}}} = \frac{83.3 - 86.4}{\sqrt{\frac{5^2}{20} + \frac{3^2}{20}}} = \frac{-3.1}{1.30} = -2.38$$

（3）确定显著性水平并做出决策：对给定的 $\alpha = 0.05$，且 df = 32.51，查 t 分布表得 $t_{0.05(30)} \approx 1.697$，$|t| = 2.38 > 1.697$，应拒绝虚无假设。

答：计算机模拟和现场实训相结合的新方法效果更好，作此推论犯错误的概率不超过 0.05。

（2）柯克兰-柯克斯（Cochran-Cox）检验

【例 6-11】也可以用柯克兰-柯克斯检验，其检验公式还是和公式 6-13 一样，只是对于给定的显著性水平，t' 的临界值可以近似地由两个 $t_{\frac{\alpha}{2}(n_1-1)}$、$t_{\frac{\alpha}{2}(n_2-1)}$ 的加权平均数计算出来。

$$t'_{\frac{\alpha}{2}} = \frac{\frac{S_{n_1-1}^2}{n_1} t_{\frac{\alpha}{2}(n_1-1)} + \frac{S_{n_1-1}^2}{n_1} t_{\frac{\alpha}{2}(n_1-1)}}{\frac{S_{n_1-1}^2}{n_1} + \frac{S_{n_2-1}^2}{n_2}} \quad \text{（公式 6-14）}$$

在总体方差不齐性的情况下，除上述两种方法外还可以采用非参数检验的方法，非参数检验对总体分布没有要求，而且简便易行。具体内容可参见第十二章“非参数统计”。

3.两总体方差未知，当两个样本都是大容量样本

由抽样分布的理论可知，当 n_1 和 n_2 都是大样本容量时，此时 $\bar{X} - \bar{Y}$ 的抽样分布近似正态分布，因而可以不管两总体方差是否相等，仍然可以用 Z 检验，其检验统计量为

$$Z = \frac{(\overline{X_1} - \overline{X_2})}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \quad \text{（公式 6-15）}$$

由于其检验过程与前述相同，这里不再赘述。

三、两总体正态分布，并且相关的平均数差异显著性检验

此时需要用两相关样本的平均数之间的差异，来推断两总体平均数之间是否存在差异。两相关样本（correlated sample）是指两个样本的数据之间存在一一对应的关系。主要有两种情况：一是同一批被试在不同条件下形成的两组样本数据间存在相关，二是一一严格配对的两组被试其测量值是相关的。相关样本的差异显著性检验不需要事先进行方差齐性检验，因为相关样本两组数据存在对应关系，这样可以求出对应数据的差（d_i），把对（$\bar{X} - \bar{Y}$）的显著性检验转化为对 $\overline{d}$ 的显著性检验。下面我们仍然用 $\overline{X}_1$ 表示上述 $\overline{X}$，用 $\overline{X}_2$ 表示上述 $\overline{Y}$ 的方式来介绍具体的计算方法。

（一）给出原始配对数据的两总体平均数差异显著性检验

在心理研究中，若按照条件相当的原则将被试对象一一配对后分成两组，然后对两组被试施以不同的处理。比较不同处理的效果差异是否显著，就构成了对配对样本数据均值的比较问题。

设两个配对样本的观测值为 (X_{1i}, X_{2i})，i =1，2，…，n，则其差数 $d_i = X_{1i} - X_{2i}$

也是正态总体的一个随机样本，该总体均值为 $\mu_d=\mu_1-\mu_2$。则 $\bar{d}_i=\frac{\sum d_i}{n}=\frac{\sum(X_{1i}-X_{2i})}{n}=\frac{\sum(X_{1i})}{n}-\frac{\sum(X_{2i})}{n}=\bar{X}_1-\bar{X}_2$；$S_d^2=\frac{\sum(d_i-\bar{d})^2}{n}$。由平均数的抽样分布理论可知，这时抽样误差的计算公式为：$SE_{\bar{X}}=\frac{S_d}{\sqrt{n}}$，t 统计量服从 $df=n-1$ 的 t 分布，计算公式为：

$$t=\frac{(\bar{X}_1-\bar{X}_2)-(\mu_1-\mu_2)}{SE_{\bar{X}}}=\frac{\bar{d}_i-\mu_d}{\frac{S_d}{\sqrt{n}}} \quad \text{（公式 6-16）}$$

应用 t 检验的步骤和方法，就可以得到假设检验的结果。

【例 6-12】 某校开展一项教育实验，比较甲乙两种教学方法的效果。根据条件，将 16 名学生一一配对，随机分成两组，随机接受其中一种教学方法。一学期后，进行一次测验，成绩如表 6-2 所示。问甲教学法是否显著优于乙教学法。（α =0.05）

表 6-2 甲乙两种教学方法的效果表

甲教学法	82	58	88	83	79	94	74	77
乙教学法	72	61	69	83	85	87	57	49
d	10	-3	19	0	-6	7	17	28

解：由原始数据计算得 $\bar{d}_i$ 和 S_d 分别为 9 和 11.8322

(1)建立假设：$H_0:\mu_{甲}\le\mu_{乙}$

$H_1:\mu_{甲}>\mu_{乙}$

(2)计算检验统计量，根据公式 6-16 得

$$t=\frac{\bar{d}_i}{\frac{S_d}{\sqrt{n}}}=\frac{9}{\frac{11.8322}{\sqrt{8}}}=2.151$$

（3） 确定显著性水平并做出决策：对给定的 $\alpha=0.05$，且 $df=7$，查 t 分布表得 $t_{\alpha(7)}$=1.896，t=2.151＞$t_{\alpha(7)}$=1.896，故应拒绝虚无假设，即认为甲教学法要明显优于乙教学法。

答：甲教学法要明显好于乙教学法。

（二）给出相关系数的两总体平均数差异显著性检验

当已知两个样本相关系数 r 时，根据公式 6-6 可知：$S_d^2=S_{n_1-1}^2-2rS_{n_1-1}^2\cdot S_{n_2-1}^2+S_{n_2-1}^2$，其中 $S_{n_1-1}^2=\frac{\sum(X_{1i}-\bar{X}_1)^2}{n-1}$ 是 μ_1 的无偏估计，$S_{n_2-1}^2=\frac{\sum(X_{2i}-\bar{X}_2)^2}{n-1}$ 是 μ_2 的无偏估计。根据抽样分布理论，这时抽样误差的计算公式为：$SE_{\bar{X}}=\sqrt{\frac{S_{n_1-1}^2+S_{n_2-1}^2-2rS_{n_1-1}S_{n_2-1}}{n}}$，t 统计量服从 $df=n-1$ 的 t 分布，计算公式为：

$$t=\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{S_{n_1-1}^2+S_{n_2-1}^2-2rS_{n_1-1}\cdot S_{n_2-1}}{n}}}$$ （公式 6-17）

应用 t 检验的步骤和方法，就可以得到假设检验的结果。

【例 6-13】 有 10 名小学生，他们在冬训前的一分钟跑的平均成绩为 300 米，标准差为 16 米，冬训后他们的一分钟跑的测验成绩平均数为 330 米，标准差为 18 米。已知两次测验成绩的相关系数为 0.60，请在 α=0.05 的水平上检验一下冬训后成绩是否有了明显的进步？

解：因为相关系数已知，故可用公式 6-17 进行解答。

（1）建立假设： $H_0: \mu_{后} \le \mu_{前}$

$H_1: \mu_{后} > \mu_{前}$

（2）计算检验统计量，根据公式 6-17 得

$$t=\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{S_{n_1-1}^2+S_{n_2-1}^2-2rS_{n_1-1}\cdot S_{n_2-1}}{n}}}=\frac{300-330}{\sqrt{\frac{16^2+18^2-2\times 0.60\times 16\times 18}{10}}}=-6.20$$

（3）确定显著性水平并做出决策：根据显著性水平 α 及自由度 df=n-1=9，查 t 分布表得 $t_{\alpha(n-1)}=1.833$ ，由于 $|t|=6.20>t_{\alpha(9)}=1.833$ ，所以拒绝虚无假设，从而认为冬训后成绩有了明显进步。

答：冬训后成绩有了明显进步。

四、两总体非正态分布的平均数差异显著性检验

在第二节中曾指出，当总体分布非正态时，可取大样本（n≥30 或 n≥50）进行 Z' 检验。这种方法同样适用于两个总体非正态分布的平均数差异检验。就是说，当两个样本容量都大于 30（或都大于 50 时）也可以用 Z' 检验。

（一）独立样本的平均数差异检验

$$Z'=\frac{(\overline{X_1}-\overline{X_2})}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$ （总体方差已知时） （公式 6-18）

或者 $$Z'=\frac{(\overline{X_1}-\overline{X_2})}{\sqrt{\frac{S_{n_1-1}^2}{n_1}+\frac{S_{n_2-1}^2}{n_2}}}$$ （总体方差未知时） （公式 6-19）

（二）相关样本的平均数差异检验

$$Z'=\frac{(\overline{X_1}-\overline{X_2})}{\sqrt{\frac{\sigma_1^2+\sigma_2^2-2r\sigma_1\sigma_2}{n}}}$$ （总体方差已知时） （公式 6-19）

或者 $$Z'=\frac{(\overline{X_1}-\overline{X_2})}{\sqrt{\frac{S_{n_1-1}^2+S_{n_2-1}^2-2rS_{n_1-1}^2\cdot S_{n_2-1}^2}{n}}}$$ （总体方差未知时） （公式 6-20）

五 使用 SPSS 计算平均数差异显著性假设检验的结果

【例 6-14】下面我们以例 6-9 中的数据来说明如何利用 SPSS 软件包求平均数差异显著性假设检验的结果（原始数据见本书所附光盘中例 6-9.sav）。

运用 SPSS18.0 软件包，在已打开的数据文件窗口中，将例 6-9 中的数据正确输入到 SPSS 数据框，其中在“分组”变量中分别用“1”代表对照组，“2”代表实验组；在“测试成绩”变量中按序录入每组 20 名被试的测试成绩。然后按序点击图 6-5 所示的操作步骤可以求解本例的因素分析结果（见表 6-3。注意：在把“分组”放入“分组变量框”时，计算机会在“分组”一词后面自动显示“（？？）”，继续点击该栏下面的“定义组”后会出现一个新的菜单）。

1. 分析→比较均值→独立样本 T 检验
2. 检验变量框：测试成绩
3. 分组变量框：（1）选入“分组”
   （2）点击“定义组”按组
   （3）在“使用指定值”框中的“组 1”框中输入 1
   “组 2”框中输入 2
   （4）$\boxed{\text{继续}}$
4. $\boxed{\text{确定}}$

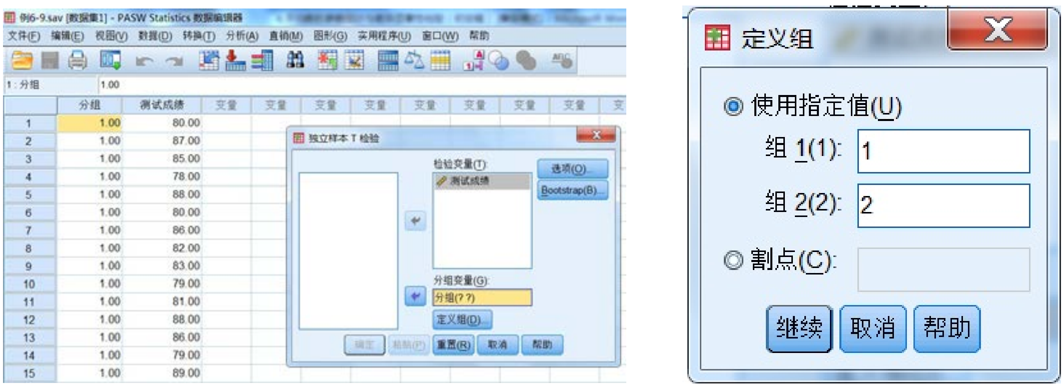


图 6-5 在 SPSS 中设置独立样本 T 检验分析程序各步骤示意图

表 6-3 T 检验主要结果表

		方差方程的 Levene 检验		均值方程的 t 检验						
									差分的 95% 置信区间	
		F	Sig.	t	df	Sig.（双侧）	均值差值	标准误差值	下限	上限
测试成绩	假设方差相等	.385	.539	-2.738	38	.009	-3.10000	1.13207	-5.39175	-.80825
	假设方差不相等			-2.738	37.918	.009	-3.10000	1.13207	-5.39191	-.80809

由表中结果可知，方差齐性检验结果为：$F=0.385$，$P=0.539>0.05$，说明两组方差之间差异未达显著水平；$t=-2.738$，$P=0.009<0.01$，计算结果与正文中例6-9的结果完全一样。

小　结

上一章，我们介绍了总体参数已知条件下，平均数的参数估计和假设检验的基本原理，在此基础上，本章对其他抽样分布条件下的平均数参数估计和假设检验方法作了介绍。

用样本标准差 S 取代总体标准差 σ 之后就可以得到与统计量 $Z_{\bar{x}}=\dfrac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ 相似的另一个统计量 $t=\dfrac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$。本章首先对 t 统计量、t 分布和 t 检验的含义、特点、t 分布表作了介绍，然后对如何用 t 分布来进行平均数的区间估计，对不同总体分布情况、两样本是独立样本还是相关样本、总体方差是否已知、总体方差是否齐性以及不同样本容量等不同条件下如何进行平均数差异检验进行了讨论，给出了不同条件下的各种抽样平均数分布的标准误差，强调在不同条件下要采用不同的假设检验方法和计算公式。以便更为准确地贯彻“小概率事件在一次抽样中不太可能出现”的统计推断思想。

关键术语

t 分布：是均值为 0、单峰对称的分布，当自由度 $n-1$ 在 30 以上时，t 分布就接近正态分布，方差接近 1；当样本容量 n 趋于无穷大时，t 分布趋近于标准正态分布。

t 统计量：若（$X_1, X_2, \cdots, X_n$）是来自正态总体的一个样本,统计量 $t=\dfrac{(\overline{X}-\mu)}{S_{n-1}/\sqrt{n}}$ 服从的分布是自由度为 $n-1$ 的 t 分布。服从 t 分布的统计量称为 t 统计量。

t 检验：当总体服从正态分布，总体方差 σ^2 未知时，一般使用 t 分布对抽样平均数与总体平均数之间，或者两个抽样平均数之间进行差异显著性检验，这种检验方法称之为 t 检验。

两总体平均数差异显著性检验：指从两个总体中各自抽取一个样本，对这两个样本平均数之间的差异是否具有统计学意义所进行的假设检验过程。其目的在于通过对两个样本平均数之间的差异进行考察，看看能否在特定的显著性水平下推论两个总体平均数之间存在差异。

两独立样本：两个样本数据的取样过程是独立进行的，即两样本中的数据不存在对应性。

两相关样本：两相关样本是指两样本数据间存在一一对应的关系。主要有两种情况：一是同一批被试在不同条件下形成的两组样本数据间存在相关；二是一一严格配对的两组被试其测量值是相关的。

虚无假设分布：在对两总体**平均数差异显著性进行**假设检验过程中，通常将检验的虚无假设设为 H_0 ：$\mu_1=\mu_2$，这时，所谓虚无假设分布就是指当虚无假设 H_0 为真时，我们在这两个总体中每次各抽取一定数量的样本进行差异显著性检验，在进行无数多次抽样后，每次抽取的两个样本平均数之差（即“$\overline{X}_1-\overline{X}_2$”之差）的分布。它是一种以零为中心的正态分布。

重要公式

总体平均数的区间估计和单总体平均数的显著性检验

$$t = \frac{\overline{X} - \mu}{SE_{\overline{X}}} = \frac{(\overline{X} - \mu)}{\frac{S_{n-1}}{\sqrt{n}}} = \frac{(\overline{X} - \mu)}{\frac{S_n}{\sqrt{n-1}}}$$

$$Z' = \frac{\overline{X} - \mu}{S_{n-1}\big/\sqrt{n}} = \frac{(\overline{X} - \mu)}{S_n\big/\sqrt{n-1}}$$

两总体正态分布，两独立样本的平均数差异检验

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ （两总体正态分布，且方差均已知）

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{SE_{\bar{X} - \bar{Y}}}$$ （$SE_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{(n_1 - 1)S_{n_1-1}^2 + (n_2 - 1)S_{n_2-1}^2}{n_1 + n_2 - 2} \cdot (\frac{1}{n_1} + \frac{1}{n_2})}$ ）

两总体正态分布，两相关样本的平均数差异检验

$$t = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{S_d\big/\sqrt{n}} = \frac{\bar{d}_i - \mu_d}{S_d\big/\sqrt{n}}$$ （给出原始配对数据）

$$t = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{S_{n_1-1}^2 + S_{n_2-1}^2 - 2rS_{n_1-1} \cdot S_{n_2-1}}{n}}}$$

思考与练习

1. 随机抽取 $n = 75$ 的样本，算得 $\overline{X} = 65$ ，$S_{n-1} = 18$，试求总体参数 μ 的 0.99 的置信区间。

2. 某地区对两年内犯罪受害者进行应激性心理障碍调查，测试后从中随机抽取 12 名受害者，了解到他们应激水平的分值分别为 83，91，62，50，74，68，70，65，85，71，58，63。试对该地区犯罪受害者应激水平的均值作区间估计（取 α =0.05）。

3. 某心理学家认为一般人的听觉反应时平均为 170 毫秒。有人随机抽取 37 名短跑运动员进行测试，结果其平均听觉反应时为 160 毫秒，标准差为 20 毫秒。能否推论短跑运动员的听觉反应时与一般人不同？

4. 某省进行化学竞赛，结果分数的分布不是正态，总平均分为 67。从某市参加竞赛的学生中随机抽取 69 人，$\overline{X} = 75, S = 12$，能够推论该市平均分与全省平均分不同？

5. 为调查某地区不同工种的工资水平状况，随机抽取 A 工种 45 人，计算得到平均工资为 5680 元，抽取 B 工种 36 人，计算得到平均工资为 6595 元。根据历年调查资料，该地区 A 工种的工资标准差 $\sigma_1 = 800$元，B 工种的工资标准差 $\sigma_2 = 950$元。能否推论 B 工种工人的工资高于 A 工种工人？

6. 在一项记忆实验研究中，70 名被试随机分成两组。其中精加工策略组记忆任务的成绩为: $\overline{X}_1 = 140$，$S_1 = 13.5$；复述策略组记忆任务的成绩为: $\overline{X}_1 = 120$，$S_2 = 15$。能否推论两组的记忆成绩有显著差异？

7. 为了比较不同认知方式（场独立和场依存）的被试在心理旋转操作方面的差异，共抽取 80 名被试，先通过《镶嵌图形测验》确定被试的认知方式，测验结果表明，其中 43 人为场独立型，37 人为场依存型。所有被试都进行心理旋转实验，结果表明，场独立型被试的平均旋转效率（综合了反应时和正确率）$\overline{X}_1$ = 764.7，$S_1 = 171.5$；场依存型被试的平均旋转效率 $\overline{X}_2 = 985.3$，$S_1 = 168.9$。试问场独立型被试和场依存型被试在心理旋转效率上是否存在显著差异？

8. 随机抽取 49 名青少年进行了认知能力测试（$\sigma = 16$），平均分数为 $\overline{X}_1 = 106$。一年后再对同组被试施测该测验，评分分数为 $\overline{X}_2 = 110$。已知两次测验结果的相关系数 $r = 0.74$。能否据此推论青少年的认知能力有了提高？

9. 某研究拟考察汽车刹车灯颜色与后车驾驶员反应时间的关系。随机选取了 9 名被试，在红光和黄光两种情况下测试他们的视觉反应时（单位：毫秒），所测数据如下。问红光和黄光情况下被试的视觉反应时是否有差异？

被试	1	2	3	4	5	6	7	8	9
红光	175	170	165	174	167	170	175	168	172
黄光	170	172	169	175	167	171	179	170	172

第七章 方差的参数估计与假设检验

学习本章内容，将有助于你对以下问题的理解与思考：

1. 正态总体样本方差和两样本方差之比的抽样分布有何特点？

2. 怎样对总体方差和标准差进行估计？

3. 如何对关于总体方差的假设进行检验？

1. 已知某地区小学生的某项生理指标服从正态分布，今从该地区小学生中随机取出一个 n=16 的样本，算得他们该项生理指标的方差 S_{n-1}^2=4，求该地区小学生该项生理指标标准差的 0.95 的置信区间。

2. 某项研究中，40 名被试完全随机的被分为实验组和对照组，对照组数据的标准差 S_{n_1-1} 为 5，实验组数据的标准差 S_{n_1-1} 为 3，试对实验组和对照组进行方差齐性检验。

第一节　方差与标准差的区间估计

一、χ^2 分布概述

χ^2 分布是心理统计中非常重要的抽样分布，下面我们对 χ^2 分布的有关内容作简要介绍。

（一）χ^2 分布的定义

设 $X_1, X_2, \cdots, X_n$ 相互独立，且都服从标准正态分布 $N(0,1)$，则称统计量 $\chi^2 = \sum_{i=1}^{n} X_i^2$ 所服从的分布是自由度为 n 的 χ^2 分布（chi-square distribution），记作 $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi_{(n)}^2$ 。

显然，如果 $X_1, X_2, \cdots, X_n$ 是来自标准正态分布 $N(0,1)$ 的一个样本，则 $\sum_{i=1}^{n} X_i^2 \sim \chi_{(n)}^2$ 。

（二）χ^2 分布的性质

（1）χ^2 是正偏态分布，随样本容量 n 的不同，其分布曲线的形状也不同，n 越小，分布越偏斜。当 n 趋向于无穷大时，χ^2 分布趋近于正态分布，如图

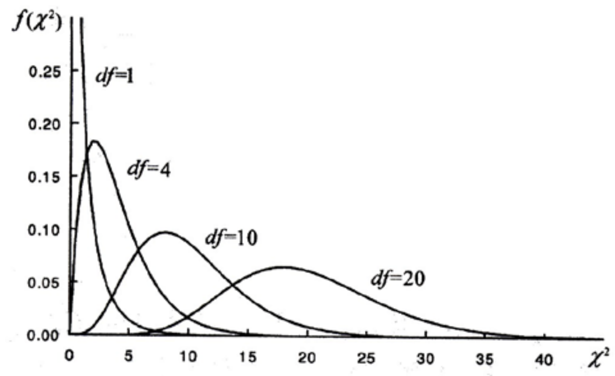


图 7-1　χ^2 分布的概率密度曲线

7-1 所示。

（2） χ^2 值都是正值。

（3） χ^2 分布具有可加性。设 $X \sim \chi^2_{(n_1)}$， $Y \sim \chi^2_{(n_2)}$，且 X 与 Y 相互独立，则：$X + Y \sim \chi^2_{(n_1+n_1)}$。

（4）若 $X \sim \chi^2_{(n)}$，则 $\mu_X = n$， $\sigma^2_X = 2n$。

（三） χ^2 分布的分位数和分布表

在一个特定的 χ^2 分布中，若大于某 χ^2 值以上的概率为α，则称该 χ^2 值为 χ^2 分布的上α分位数；类似地，若小于某 χ^2 值以下的概率为α，则称该 χ^2 值为 χ^2 分布的下α分位数。如果不做特殊声明，一般用 χ^2_α 表示 χ^2 分布的上α分位数。由 χ^2 分布的定义可知， χ^2 分布的下α分位数和上(1-α)分位数是同一个数值，故 χ^2 分布的下α分位数一般记作 $\chi^2_{1-\alpha}$。

如图 7-2， χ^2_α 是 χ^2 分布的上α分位数，而 $\chi^2_{1-\alpha}$ 是其下α分位数。

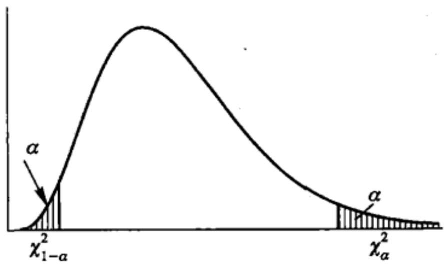


图 7-2 χ^2 分布的分位数

例如，若某一个随机变量服从自由度 df =40 的 χ^2 分布，其取值大于 55.8 以上的概率为 0.05，即 $P(\chi^2 \geq 55.8)$=0.05，则称 55.8 是自由度为 40 的 χ^2 分布的上 0.05 分位数，记作 $\chi^2_{0.05}$，有时为了说明自由度，也可以记作 $\chi^2_{0.05}(40)$。同时 55.8 也是该 χ^2 分布的下 0.95 分位数。

χ^2 分布是一族分布，随着自由度的不同其分布形状也不同。但是 χ^2 分布曲线下的面积都为 1，在不同的自由度下，同一 χ^2 值以上或以下所含的面积与总面积的比值不同。附表 3 是 χ^2 的分布数值表，其结构有如表 7-1 所示。

表 7-1 χ^2 分布表的结构示意表

df	0.10	0.05	0.25	0.01	0.005
1	2.71	3.84	5.02	6.63	7.88
2	4.61	5.99	7.38	9.21	10.6
3	6.25	7.81	9.35	11.3	12.8

续表

df	0.10	0.05	0.25	0.01	0.005
……					
40	51.8	55.8	59.3	63.7	66.8
50	63.2	67.5	71.4	76.2	79.5
60	74.4	79.1	83.3	88.4	92.0

由表 7-1 可知，χ^2 分布表的左列是自由度，最上一行是不同自由度下大于某 χ^2 值以上的概率。例如当 df =15，χ^2 =7.261 时，大于该值的概率为 P=0.95，则小于该值的概率为 1-0.95=0.05。而当 df =30 时，只有 χ^2 值大于 18.493，大于该值的概率才是 0.95。可见 χ^2 分布是随自由度不同而变化的。如果要查的 χ^2 值在表中恰好没有，要确定其概率，可以用临近值代替或用内插法近似计算。

二、样本方差的抽样分布

在第六章中，我们讨论了样本平均数的抽样分布。除了样本均值外，样本方差也是重要的样本统计量，它的分布情况如何呢？

设 $X_1, X_2, \cdots, X_n$ 是来自正态总体 $N(\mu, \sigma^2)$ 的一个样本，$\overline{X}$ 为样本平均数，$S_{n-1}^2 = \dfrac{\sum(X-\overline{X})^2}{n-1}$ 为样本无偏方差，则 $\overline{X}$ 与 S^2 相互独立且 $\dfrac{(n-1)S_{n-1}^2}{\sigma^2}$ 服从 χ^2 分布，即 $\dfrac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$。

证：因为 $S_{n-1}^2 = \dfrac{\sum(X-\overline{X})^2}{n-1}$，移项后有 $(n-1)S_{n-1}^2 = \sum(X-\overline{X})^2$，方程两边同除以总体方差得：$\dfrac{(n-1)S_{n-1}^2}{\sigma^2} = \dfrac{\sum(X-\overline{X})^2}{\sigma^2}$，将方程右边整理后得：$\dfrac{\sum(X-\overline{X})^2}{\sigma^2} = \sum\left(\dfrac{X-\overline{X}}{\sigma}\right)^2 = \sum Z^2$，由于 $\sum Z^2$ 服从 χ^2 分布，所以 $\dfrac{(n-1)S_{n-1}^2}{\sigma^2}$ 服从 χ^2 分布，即 $\dfrac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$。

三、方差的区间估计

根据前面所说的方差抽样分布理论可知，从正态分布 $N(\mu, \sigma^2)$ 中抽取样本容量为 n 的随机样本，样本平均数 $\overline{X}$ 与样本方差 S_{n-1}^2 相互独立，且统计量 $\dfrac{(n-1)S_{n-1}^2}{\sigma^2} \sim \chi_{n-1}^2$。由于统计量 $\dfrac{(n-1)S_{n-1}^2}{\sigma^2}$ 中包含了 S_{n-1}^2 和 σ^2，且不含其它任何

未知总体参数，所以可以利用 X^2 分布规律对 σ^2 进行区间估计。这时，X^2 统计量的计算公式为：

$$X^2 = \frac{\sum (X-\overline{X})^2}{\sigma^2} = \frac{(n-1)S_{n-1}^2}{\sigma^2} \qquad \text{（公式 7-1）}$$

当总体正态，置信水平为 $1-\alpha$ 时，由于 $P\left\{\chi^2_{1-\frac{\alpha}{2}} \le \frac{(n-1)S_{n-1}^2}{\sigma^2} \le \chi^2_{\frac{\alpha}{2}}\right\} = 1-\alpha$，

所以 $P\left\{\frac{(n-1)S_{n-1}^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)S_{n-1}^2}{\chi^2_{1-\alpha/2}}\right\} = 1-\alpha$。这时，利用抽样方差估计总体方差的区间不等式的表达形式为：

$$\frac{(n-1)\times S_{n-1}^2}{X^2_{\frac{\alpha}{2}}} \le \sigma^2 \le \frac{(n-1)\times S_{n-1}^2}{X^2_{1-\frac{\alpha}{2}}} \qquad \text{（公式 7-2）}$$

【例 7-1】 某学校学生身高服从正态分布，从中取样本 n=35，样本标准差 S_{n-1}= 6cm，问该学校全体学生身高方差 σ^2 在 0.95 置信水平上的置信区间。

解：

根据题意可知，$n=35$，显著性水平 $\alpha=1-$置信区间$=1-0.95=0.05$，查 χ^2 分布表附表 3 中的 $df=n-1=34$，确定 α=0.05，df=30 时 $\chi^2_{\alpha/2}$ 与 $\chi^2_{1-\alpha/2}$ 的值分别为：$\chi^2_{0.025}$=46.98，$\chi^2_{0.975}$=16.79。将相关数据填入到公式 7-2 所示的不等式后可得总体方差 σ^2 的 0.95 置信区间为：

$$\frac{34\times 6^2}{46.98} \le \sigma^2 \le \frac{34\times 6^2}{16.79}，即 \qquad 26.05 \le \sigma^2 \le 72.90$$

答：总体方差在 0.95 的置信水平上的置信区间为[26.05，72.90]。

如果要求对总体标准差进行区间估计时，可先对其方差进行估计，得到方差的置信区间后对其开方，取正平方根，即为标准差的置信区间。

【例 7-2】 已知某地区小学生的某项生理指标服从正态分布，今从该地区小学生中随机取出一个 n=16 的样本，算得他们该项生理指标的方差 S_{n-1}^2=4，问该地区小学生该项生理指标标准差的 0.95 的置信区间。

解：根据题意可知，$n=16$，显著性水平 $\alpha=1-$置信区间$=1-0.95=0.05$，查附表 3 χ^2 分布表中的 $df=15$，可得 $\chi^2_{0.025}$=27.49，$\chi^2_{0.975}$=6.26，所以，将相关数据填入到公式 7-2 所示的不等式后可得总体方差 σ^2 的在 0.95 置信水平上的置信区间为：

$$\frac{15\times 4}{27.49} \le \sigma^2 \le \frac{15\times 4}{6.26}；\quad 2.182 \le \sigma^2 \le 9.585$$

$$1.477 \le \sigma^2 \le 3.096$$

答：总体的离散程度（用标准差表示）在 1.477～3.096 之间，此推论正确的概率为 0.95，错误概率为 0.05。

第二节　总体方差的假设检验

总体方差的假设检验包括单总体方差的显著性检验和两总体方差的齐性检验。在上一章两总体平均数差异的假设检验中，当总体方差未知时，需要利用样本方差来推断总体方差是否相等（齐性），这个问题将在本节中可以得到解决。

一、单总体方差的显著性检验

单总体方差的显著性检验是指对从某一总体中抽取的样本计算得到的 S^2 与该总体 σ^2 进行的比较。根据抽样分布理论，由公式 7-1 可知，从总体中抽取样本容量为 n 的随机样本，其样本方差 S_{n-1}^2 与总体方差 σ^2 的比值服从 $df=n-1$ 的 χ^2 分布。

根据公式 7-1 计算得到的 χ^2 值若落在图 7-2 中阴影区，即 $\chi^2 < \chi^2_{(1-\frac{\alpha}{2})(n-1)}$ 或 $\chi^2 > \chi^2_{\frac{\alpha}{2}(n-1)}$，则表明小概率事件发生，$S_{n-1}^2$ 与 σ^2 有显著差异。若 $\chi^2_{(1-\frac{\alpha}{2})(n-1)} \le \chi^2 \le$，$\chi^2_{\frac{\alpha}{2}(n-1)}$ 则认为 S_{n-1}^2 与 σ^2 差异不显著。

【例 7-3】 大学新生入学要参加心理健康测试，全校新生测验你分数的方差为 9^2，从某学院的新生中随机抽取 25 人，算得他们该测验得分的方差 S_{n-1}^2 为 4^2，能否推论该学院新生的心理健康测验得分的方差与全校存在显著差异（α=0.05）？

解：（1）建立假设：$\begin{matrix} H_0:{\sigma_0}^2=\sigma^2 \\ H_1:{\sigma_0}^2\ne\sigma^2 \end{matrix}$ 或 $\begin{matrix} H_0:{\sigma_0}^2=9^2 \\ H_1:{\sigma_0}^2\ne 9^2 \end{matrix}$

（2）计算检验统计量，根据公式 7-1 可得

$$X^2=\frac{(n-1)S_{n-1}^2}{\sigma^2}=\frac{(25-1)4^2}{9^2}=4.74$$

（3）与临界值比较并做出决策：查 χ^2 分布表，当 df=25-1=24 时，$\chi^2_{(1-\frac{\alpha}{2})(n-1)}$=12.40，$\chi^2_{\frac{\alpha}{2}(n-1)}$=39.37，而 $\chi^2=4.74<\chi^2_{(1-\frac{\alpha}{2})(n-1)}$=12.4，故拒绝 H_0，接受 H_1。

答：该学院新生的心理健康测验得分的方差与全校新生存在显著性差异。

二、　F 分布概述

（一）*F* 分布的定义

设 $\overline{X}_1$ 和 $\overline{X}_2$ 分别服从于自由度为 n_1-1 和 n_2-1 的 χ^2 分布，即 $X_1 \sim X^2(n_1-1)$，$X_2\sim X^2(n_2-1)$，且 $\overline{X}_1$ 和 $\overline{X}_2$ 相互独立，则称统计量 $F=\dfrac{\frac{X_1^2}{n_1-1}}{\frac{X_2^2}{n_2-1}}$ 服从自由

度为（n_1-1）和（n_2-1）的 F 分布（F distribution），即 $F=\dfrac{\dfrac{X_1^2}{n_1-1}}{\dfrac{X_2^2}{n_2-1}}\sim F(n_1-1,n_2-1)$。

（二）F 分布的性质

（1） F 分布是一族正偏态分布，随分子、分母的自由度不同分布曲线不同，分子、分母趋近于无穷大时，F 分布趋近正态分布。图 7-3 是自由度分别为（2，6），（6，10）和（10，20）的 F 分布曲线，从图 7-3 可知，分子分母自由度越大，越趋近于正态分布。

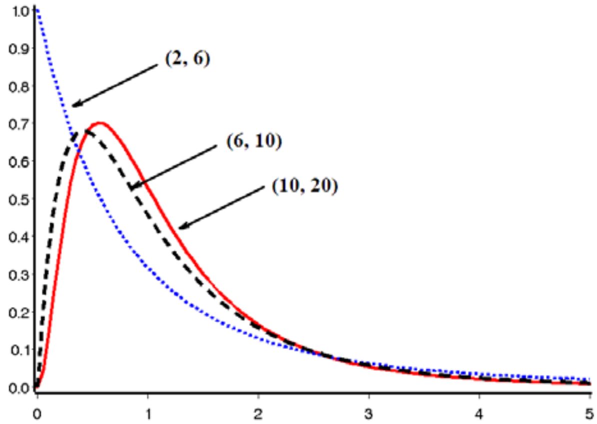


图 7-3 F 分布的密度曲线

（2） F 值均为正值。根据 F 分布的定义可知，如果 F 服从 F 分布，则 $\dfrac{1}{F}$ 也服从 F 分布，而且 $F_{\alpha(n,m)}=\dfrac{1}{F_{1-\alpha(m,n)}}$。

（3）当分子自由度为 1，分母自由度为任意值时，F 值与相同自由度的 t 值（双侧概率）的平方相等。例如，当分子自由度为 1，分母自由度为 15 时，单侧概率为 0.05 时的 F 值为 4.54；而 df =15 时，双侧概率为 0.05 的 t =2.131，$(t_{0.05})^2$=4.54，与 F 值相同。

（三）F 分布的分位数和分布表

1. F 分布的分位数

在一个特定的 F 分布中，若大于某 F 值的概率为α，则称该 F 值为 F 分布的上α分位数；与此类似，若小于某 F 值的概率为α，则称该 F 值为 F 分布的下α分位数。如果没有特殊声明，一般用 F_α 表示 F 分布的上α分位数（也称为右侧α分位数），如图 7-4。由于 F 分布是连续分布，所以 F 分布的下α分位数

和上 1-α分位数是同一个数值，故 F 分布的下α分位数一般记作 $F_{1-\alpha}$。

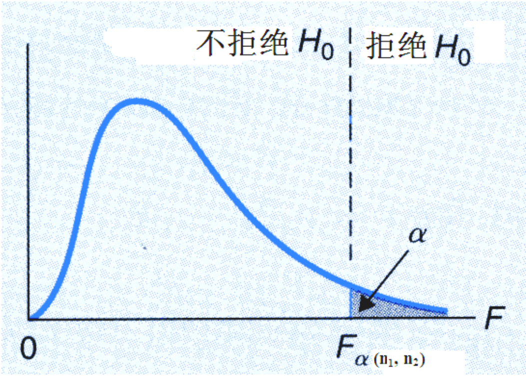


图 7-4 F 分布的分位数

例如，若某随机变量服从 F 分布，当分子自由度为 9，分母自由度为 12 时，其取值大于 F = 2.8 的概率为 0.05，所以 2.8 即为 F 分布的上 0.05 分位数，记作 $F_{0.05}$ 或 $F_{(0.05)(9,12)}$；同时，2.8 也是该 F 分布的下 0.95 分位数。

2.F 分布的分布表

F 分布是一种小样本分布，它是一族随着分子和分母的自由度变化而变化的偏态分布曲线，曲线下的面积为 1。在 F 分布表中（参见附表 4 和附表 5），列出了不同的分子和分母自由度，及相应的概率为 0.05 和 0.01 的上下分位 F 值。

附表 4-A 和附表 4-B 分别是 $\alpha=0.05$ 和 $\alpha=0.01$ 双侧检验的 F 值表，附表 5-A 和附表 5-B 分别是 $\alpha=0.05$ 和 $\alpha=0.01$ 单侧检验的 F 值表。四个表的结构大致相同，但用于单侧检验的附表较为常用，其结构有如表 7-2 所示。四个表的左边一列均为分母的自由度，表的最上一行是分子自由度，表中数据是不同自由度下的 F 值。

表 7-2 单侧检验 F 值表的结构示意表

分母		分子 df									
df	α	1	2	…	8	…	10	…	20	…	∞
1	.05	161	200		239		242		248		254
…											
8	.05	5.32	4.46		3.44		3.34		3.15		2.93
…											
10	.05	4.96	4.10		3.07		2.97		2.77		2.54
…											
20	.05	4.35	3.49		2.45		2.35		2.12		1.84
…											
∞	.05	3.84	2.99		1.94		1.83		1.57		1.00

例如，当 $\alpha=0.05$ 时，分子自由度为 8，分母自由度为 10，查附表 4-A，$F_{\alpha(n,m)}=F_{0.05(8,10)}=3.85$。由于 $F_{\alpha(n,m)}=\frac{1}{F_{1-\alpha(m,n)}}$（注意括号中的自由度应该先换位置），所以在分子自由度为 10，分母自由度为 8 时，$F_{1-\alpha(m,n)}=F_{1-0.05(10,8)}=\frac{1}{3.85}$。这两个结果表示在上述条件下的双侧检验中，F 值大于 3.85 和小于 $\frac{1}{3.85}$ 的概率之和为 $p=0.05$。附表 5-A 和附表 5-B 是单侧概率，当 $\alpha=0.05$ 时，与上例相同的自由度（即分子自由度为 8，分母自由度为 10）的上分位数是 $F=3.07$。

二、两样本方差之比的抽样分布

对来自两个不同抽样样本的方差进行差异显著性检验，需要先了解它们之间差值的抽样分布情况。理论上可以证明，若 $X_1,X_2,\cdots,X_{n1}$ 是来自正态总体 $N(\mu_1,\sigma_1^2)$ 的一个样本；$Y_1,Y_2,\cdots,Y_{n2}$ 是来自正态总体 $N(\mu_2,\sigma_2{}^2)$ 的一个样本，且 X 和 Y 相互独立，$S_{n_1-1}^2$ 为第一个抽样样本的无偏方差，$S_{n_2-1}^2$ 为第二个抽样样本的无偏方差。若 $\sigma_1^2=\sigma_2^2$，则 $\frac{S_{n_1-1}^2}{S_{n_2-1}^2}$ 服从第一自由度为 $df_1=n_1-1$ 和第二自由度为 $df_2=n_2-1$ 的 F 分布，即 $\frac{S_{n_1-1}^2}{S_{n_2-1}^2}\sim F_{(n_1-1,n_2-1)}$。因此，对两个不同抽样样本的方差，可以用 F 统计量对其进行差异显著性检验，所谓 **F 统计量（statistic of F test）是指在 F 检验中用以判断接受还是拒绝虚无假设**的检验统计量，其计算公式为：

$$F=\frac{S_{n_1-1}^2}{S_{n_2-1}^2} \quad \text{（公式 7-3）}$$

根据检验统计量是否服从或渐近服从 F 统计量的参数检验方法被称为 F 检验（F test）。

当 $\sigma_1^2=\sigma_2^2$ 时，样本方差差异检验结果 $F=\frac{S_{n_1-1}^2}{S_{n_2-1}^2}$ 应该围绕着 1 上下波动，而且 $F=\frac{S_{n_1-1}^2}{S_{n_2-1}^2}\sim F_{(n_1-1,n_2-1)}$，$\frac{1}{F}=\frac{S_{n_2-1}^2}{S_{n_1-1}^2}\sim F_{(n_2-1,n_1-1)}$。

四、两总体方差差异的显著性检验

（一）两个独立总体的方差齐性检验

当两个独立总体的方差未知时，需要利用两样本方差来检验它们所来自的两总体方差是否相等，这种检验称为方差齐性检验（the test for homogeneity of variance）。

前面已经指出，根据两个样本方差的抽样分布理论，我们可以用公式 $F=\frac{S_{n_1-1}^2}{S_{n_2-1}^2}$ 所得到的 F 统计量对两个不同抽样样本的方差进行差异显著性检验。对于

给定的显著性水平 α，$df_1 = n_1 - 1$，$df_2 = n_2 - 1$，当 $F_{(1-\frac{\alpha}{2})} \le F \le F_{\frac{\alpha}{2}}$ 时，说明两个方差之间差异不显著，因此接受 H_0，可以认为 $\sigma_1{}^2 = \sigma_2{}^2$；当 $F < F_{(1-\frac{\alpha}{2})}$ 或 $F >$ $F_{\frac{\alpha}{2}}$ 时，说明两个方差之间的差异显著，因此拒绝 H_0，可以认为 $\sigma_1{}^2 \ne \sigma_2{}^2$。

由于 F 分布是左右不对称的，因此 $F_{\frac{\alpha}{2}}$ 与 $F_{1-\frac{\alpha}{2}}$ 两个值之间的绝对值并不相等，但是，前面提到过，这两者之间具有互为倒数的关系，即 $F_{\alpha(n,m)} = \frac{1}{F_{1-\alpha(m,n)}}$（单侧检验）或者 $F_{\frac{\alpha}{2}(n,m)} = \frac{1}{F_{1-\frac{\alpha}{2}(m,n)}}$（双侧检验），因此，只要知道了其中的一个值，就可以通过其倒数求另外一个值。因而一般书中都只给出各种不同自由度下的 $F_{\frac{\alpha}{2}}$ 值，在做双侧检验时，$F_{(1-\frac{\alpha}{2})}$ 的值可由 $F_{\frac{\alpha}{2}}$ 的倒数求得。我们在实际应用中需要解决不同总体方差的差异显著性检验问题时，为了避开计算倒数的麻烦，通常总是将更大值的方差作为分子，让更小值的方差作为分母，即用 $F = \frac{S_{大}^2}{S_{小}^2}$ 的公式来计算 F 值（公式中的 S^2 理论上还是 $S_{n_2-1}^2$，但实际上当 n_1 与 n_2 相差不大时，可以用 S^2 代替 $S_{n_2-1}^2$）。其计算结果也总是会有 $F \ge 1$，因而其否定域也总会在右侧，由此从 F 分布表中就可以较为容易地找到临界值 $F_\alpha(df_1, df_2)$。在此，df_1 是分子的自由度，df_2 是分母的自由度。需要注意的是，这时的检验方式是单侧检验。

【例 7-4】 对第六章例 6-10 中的方差进行齐性检验。（α=0.05）

解：已知 $\begin{array}{l} n_1 = 20, S_{n_1-1} = 5 \\ n_2 = 20, S_{n_2-1} = 3 \end{array}$，根据公式 7-21，用单侧检验。

（1）建立假设：$\begin{array}{l} H_0 : \sigma_1^2 = \sigma_2^2 \\ H_1 : \sigma_1^2 \ne \sigma_2^2 \end{array}$

（2）计算检验统计量：$F = \frac{S_{大}^2}{S_{小}^2} = \frac{5^2}{3^2} = 2.78$

（3）确定显著性水平并做出决策：根据显著性水平 $\alpha = 0.01$，查 F 单侧分布表（附表 5-A）得 $F_{\alpha(19,19)}$ 约为 2.41，F=2.78＞$F_{\alpha(29,29)}$=2.41，所以拒绝原假设，从而认为实验组与对照组成绩的方差不齐性。

答：实验组与对照组方差不齐性。

（二）两相关总体的方差齐性检验

两个样本相关时，对其方差的差异检验需要按下面的公式进行 t 检验

$$t = \frac{s_1^2 - s_2^2}{\sqrt{\frac{4s_1^2 s_2^2 (1-r^2)}{n-2}}} \qquad (df = n-2) \qquad \text{（公式 7-4）}$$

式中，s_1^2 与 s_2^2 分别为两个样本方差；r 为两个样本之间的相关系数；n 为样本容量。用上式算出 t 值，并查 $df=n-2$ 时 t 值表中的临界值，若 t 值大于临界值，则两个方差之间有显著差异，反之，则没有显著性差异。

五、多个总体方差差异的显著性检验

前面所说的是如何用 F 统计量对两个独立样本的方差进行差异显著性检验，或者如何用 t 统计量对两个相关样本的方差进行差异显著性检验。如果需要检验的抽样方差多于两个时，则上述两种方法均不适用，需要用其他方法。所谓多于两个需要检验的抽样方差是指如下情况：

假定有 K 个正态总体 $N(\mu_i,\sigma_i^2)$，$i=1,2,\cdots,K$，从第 i 个总体中得到样本容量为 n_i 的样本，样本均值为 $\overline{Y}_i$，样本方差为 S_i^2，现要在显著性水平 α 上检验假设：

$$\begin{cases} H_0:\sigma_1^2=\sigma_2^2=\cdots=\sigma_K^2 \\ H_a:\sigma_1^2,\sigma_2^2,\cdots,\sigma_K^2\text{不全相等} \end{cases}$$

在上述情况下，当 $n_1=n_2=\cdots=n_K$ 即各组样本容量相等时，常用哈特莱（Hartley）最大 F 比率（maximum F-ratio）检验或柯赫伦（Cochran）最大方差检验。这里仅介绍哈特莱检验。

哈特莱检验用的统计量为：

$$F_{\max} = \frac{\max\{S_1^2,\ S_2^2,\ \cdots\ S_K^2\}}{\min\{S_1^2,\ S_2^2,\ \cdots\ S_K^2\}} \qquad \text{（公式 7-5）}$$

式中：“max”意为“最大”，“min”意为最小，该公式的实际含义就是 $F=\dfrac{S_{\text{大}}^2}{S_{\text{小}}^2}$，表示要做方差齐性检验的一组 K 个方差中，把最大值的方差作为分子，把最小值的方差作为分母，由此计算出的商称为“$F_{\max}$”。

在 α 显著性水平上的拒绝域为

$$F_{\max} \geq F_{\max,\alpha(K,df)} \qquad \text{（公式 7-6）}$$

式中：K 是组数；$df=n-1$（n 为样本容量）；$F_{\max,\alpha(K,df)}$ 是依据 α、K 和 df 查附表 6“$F_{\max}$的临界值（哈特莱方差齐性检验）”而得到的临界值。

需要检验的抽样方差多于两个，但 n_1，$n_2,\cdots$，n_K 不全相等时，可采用巴特莱特（Bartlett）检验，其统计量为

$$\chi^2=\frac{1}{c}[df_E\ln MS_E-\sum_{i=1}^{K}(n_i-1)\ln S_i^2] \qquad \text{（公式 7-7）}$$

式中：

$c=\dfrac{1}{3(K-1)}(\sum_{i=1}^{K}\dfrac{1}{n_i-1}-\dfrac{1}{df_E})+1$，其中 K 为自变量的水平数，n_i 为各组的样本容量，df_E 是误差自由度（也就是组内自由度）；MS_E 为误差均方（即组内均方 MS_W）；S_i^2 为各组的样本方差。

在 α 显著性水平上的拒绝域为

$$\chi^2 \geq \chi^2_{\alpha,(df=K-1)} \qquad \text{（公式 7-8）}$$

式中，$\chi^2_{\alpha,(df=K-1)}$ 是显著性水平为 α，自由度为 $K-1$，查附表 3“χ^2 临界值表”可以得到相应的临界值。

六、用 SPSS 软件包进行方差齐性检验

下面将如何通过 SPSS 软件包对本书第十章中例 10-1 相应的方差齐性检验过程介绍如下。

运用 SPSS18.0 软件包，在已打开的数据文件窗口中，将例 10-1 中的数据正确输入到 SPSS 数据框（原始数据文件见本书所附光盘中例 10-1.sav），其中一列变量为“反馈类型”，另一列变量为“自尊水平得分”。然后按序点击图 7-5 所示的操作步骤可以求解本例的方差齐性检验结果（见表 7-3）。

1. 分析→比较均值→单因素 ANOVA
2. 因变量列表框：自尊水平得分
3. 因子框：反馈类型
4. 选项：（1）统计量框中选择“方差同质性检验”
   （2）继续
5. 确定

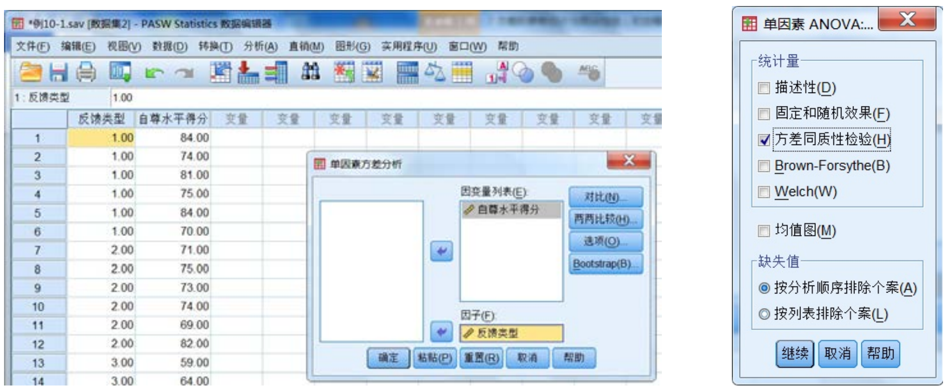


图 7-5 在 SPSS 中设置方差齐性检验分析程序各步骤示意图

表 7-3 例 10-1 的方差齐性检验结果

Levene 统计量	*df1*	*df2*	*显著性*
.824	2	15	.458

表 7-3 给出的是方差齐性检验结果，Levene 方法检验统计量为.824，在当前自由度下对应的 P 值为 0.458（大于 0.05），可认为样本所来自的总体满足方差齐性的要求。

小 结

本章介绍了对总体方差和标准差进行区间估计的计算方法，以及对“样本方差与总体方差”和“两样本方差”的差异显著性进行假设检验的方法。

从正态总体中进行随机抽样，样本方差的相关统计量服从 χ^2 分布，因此我们可以利用 χ^2 分布来对总体的方差和标准差进行区间估，还可以对样本方差与总体方差差异的显著性进行检验。

从两个正态总体中分别抽样，如果两总体方差相等，则两样本方差之比服从 F 分布，因此我们可以利用 F 分布来对两个样本方差之间的差异进行显著性检验，即对两总体的方差是否齐性进行检验。

前两章介绍的，Z 分布、t 分布，和本章介绍的 χ^2 分布、F 分布，是心理统计学中进行统计推断所依赖的四大抽样分布，是对样本平均数、方差、相关系数、比率与实计数等样本统计量进行区间估计与假设检验的数学基础，我们需要很好地加以掌握。

关键术语

χ^2 分布：若 $X_1, X_2, \cdots, X_n$ 相互独立，且都服从标准正态分布 $N(0,1)$，则称统计量 $\chi^2 = \sum_{i=1}^{n} X_i^2$ 所服从的分布是自由度为 n 的 χ^2 分布，记作 $\chi^2 = \sum_{i=1}^{n} X_i^2 \sim \chi^2_{(n)}$ 。

F 分布：若 $X_1, X_2, \cdots, X_{n1}$ 是来自正态总体 $N(\mu_1, \sigma_1{}^2)$ 的一个随机样本，$Y_1, Y_2, \cdots$，Y_{n2} 是来自正态总体 $N(\mu_2, \sigma_2{}^2)$ 的一个随机样本，且 X 和 Y 相互独立，$S^2_{n_1-1}$ 为第一个样本的无偏方差，$S^2_{n_2-1}$ 为第二个样本的无偏方差。则 $\frac{S^2_{n_1-1} / \sigma_1{}^2}{S^2_{n_2-1} / \sigma_2{}^2}$ 服从第一自由度为 n_1-1 和 n_2-1 的 F 分布。若 $\sigma_1^2 = \sigma_2^2$，则 $\frac{S^2_{n_1-1}}{S^2_{n_2-1}}$ 服从第一自由度为 $df_1 = n_1 - 1$ 和第二自由度为 $df_2 = n_2 - 1$ 的 F 分布，即 $F = \frac{S^2_{n_1-1}}{S^2_{n_2-1}} \sim F_{(n_1-1, n_2-1)}$。

F 统计量：在 F 检验中利用公式 $\frac{S^2_{n_1-1}}{S^2_{n_2-1}}$ 求取的用以判断接受还是拒绝虚无假设的检验统计量被称为 F 统计量。

***F*检验**：根据检验统计量是否服从或渐近服从 F 分布的参数检验方法。

方差齐性检验：当两个独立总体的方差未知时，需要利用两样本方差来检验它们所来自的两总体方差是否相等，这种检验称为方差齐性检验。

重要公式

χ^2 统计量的计算公式：

$$\chi^2 = \frac{\sum (X-\overline{X})^2}{\sigma^2} = \frac{(n-1)S^2}{\sigma^2}$$

利用样本方差估计总体方差的区间不等式：

$$\frac{(n-1\times S_{n-1}^2}{X_{\frac{\alpha}{2}}^2} < \sigma^2 < \frac{(n-1\times S_{n-1}^2}{X_{1-\frac{\alpha}{2}}^2}$$

两独立总体的方差齐性检验公式

$$F = \frac{S_{n_1-1}^2}{S_{n_2-1}^2}$$

$$F = \frac{S_{大}^2}{S_{小}^2}$$

两相关总体的方差齐性检验公式：

$$t = \frac{s_1^2 - s_2^2}{\sqrt{\frac{4s_1^2 s_2^2(1-r^2)}{n-2}}}$$

思考与练习：

1. 从某高校中随机抽取一个 n=160 的大学生样本，算得他们每天上网时间的标准差为 0.6 小时，试估计该校学生总体上网时间标准差的 0.95 置信区间。

2. 在习题 1 中，若样本中 75 名男生每天上网时间的标准差为 0.7 小时，85 名女生每天上网时间的标准差为 0.5 小时，能否推论该校女生上网时间的离散程度更小？

3. 试对例 6-9 中的总体方差进行齐性检验。

4. 有研究者认为小学生的数学成绩的离散程度随年龄增加而增加，为此在某市小学六年级学生中随机抽取 360 人，算得他们在标准化数学测试中成绩的方差为 9^2，而他们在小学三年级的标准化数学测试中成绩的方差为 6^2。已知两次测验成绩的相关系数为 0.65，调查结果是否能够支持该研究者的观点？

第八章　比率的参数估计及比率与分布的假设检验

学习本章内容，将有助于你对以下问题的理解与思考：

1. 如何进行总体比率的区间估计和假设检验？
2. 应如何使用 χ^2 检验方法来对计数数据进行假设检验？
3. 如何进行拟合良度和同质性的 χ^2 检验？

1. 为调查学生对某种语文阅读教学方法的评价，随机抽取了 85 名学生，其中男生 40 名，女生 45 名。男生中有 19 人对该教学方法持支持态度，女生中有 25 人对该教学方法持支持态度。问这些调查结果能否说明，男女学生对该教学方法所持的态度的比率有显著差异？

2. 从某大学随机抽取 200 名学生参加心理健康测试，实施测验后，有 26 人的成绩为优，94 人为良，75 人为中等，5 人成绩为差。问：学生在心理健康测验上的表现是否符合正态分布？

第一节　比率与计数数据的含义

在研究工作中，面临的数据是多种多样的。有些特征是用等距量表或比率量表测量，如学生成绩、身高等；有些特征则是用等级量表或称名量表测量，如按优、良、中、差评定各被试等级，调查每个被试对某事件持赞成或反对态度。当用等级量表或称名量表来测量时，可以将原始数据资料整理成在每个等级的人数或在每种属性分类特征上的人数，这就形成了计数资料（有时也称为实计数）。所谓计数数据 (count data) 是指在心理研究中，对研究对象通过计算个体数目而获取的数据。当用各种实际得到的计数数据除以它们的总次数时，所获得的商称为比率（ratio）。各种结果的比率之和等于 1。

【例 8-1】 以本章开头第二个问题为例，“有 26 人的成绩为优，94 人为良，75 人为中等，5 人为差”，这样表达调查结果的方式属于使用计数资料的表达方式；如果将结果表达为：“在对 200 名学生进行的心理测验后，成绩为优的占总人数的 13%（即 $\frac{26}{200}$），成绩为良的占 47%（$\frac{94}{200}$），成绩中等的占 37.5%（$\frac{75}{200}$），成绩为差的占 2.5%（$\frac{5}{200}$）”，则称其为使用比率的表达方式。上述使用实计次数或比率的表达方式也可以相应的用表 8-1 来表示。

表 8-1 某年级 200 次名学生在某心理测验上的评定结果统计表

测验成绩	优	良	中	差	总和
人数	26	94	75	5	200
比率	0.13	0.47	0.375	0.025	1

与例 8-1 相似的可以用两种表达方式表达结果的研究还有很多，下面是其中一例：

【例 8-2】 有人想知道一枚骰子是否六面都均匀，将骰子投掷了 120 次，出现 1、2、3、4、5、6 点的次数分别是 10、15、20、25、30、20。这些结果也

可以相应的用表 8-2 表示为实计次数或比率。

表 8-2 投掷 120 次骰子的结果统计表

投掷结果	1 点	2 点	3 点	4 点	5 点	6 点
实计次数	16	25	20	24	12	23
比　　率	0.133	0.208	0.167	0.20	0.10	0.192

本章开头第一个研究问题是个带有两个变量的更为复杂点的例子，我们可对其分析如下。

【例 8-3】 有关 85 名男女生对某种语文阅读教学方法的态度调查结果，如果表达为“接受调查的 85 名学生中，赞成的有 44 人，不赞成的有 41 人。其中，40 位男生中赞成人数为 19 人；45 位女生中赞成人数为 25 人”，则这样表达调查结果的方式属于使用计数资料的表达方式；如果将结果表达为“接受调查的 85 名学生中，赞成者占总人数的 51.76%（$\frac{44}{85}$，不赞成者则占总人数的 48.24%，$\frac{41}{85}$）；其中，男生赞成人数占男生总数的 47.5%（$\frac{19}{40}$），女生赞成人数占女生总数的 55.6%（$\frac{25}{45}$）”，则称其为使用比率的表达方式。上述结果也可以用表 8-3 的形式来表达。

表 8-3　85 名男女学生对某种教学方法的态度调查结果

性　别	男　生（n=40）		女　生（n=45）		合　并	
态　度	赞成	不赞成	赞成	不赞成	赞成	不赞成
人　数	19	21	25	20	44	41
比　率	0.475	0.525	0.556	0.444	.5176	.4824

在心理研究中，有时需要对某种比率数据进行区间估计或假设检验，有时则需要对实际获得的计数资料进行直接的检验或分析。本章第二和第三节分别阐述有关比率的区间估计和假设检验方法；第四至第七节则阐述如何用 χ^2 统计量来对实计数据进行检验或分析。

第二节　总体比率的区间估计

在本书中，我们约定用大写字母 P_0 表示某种事件的总体比率，用大写字母 Q_0 表示该事件之外其他事件出现的总体比率；用 x 表示该种事件在抽样中出现的

次数，用 n 表示抽样的总次数，用小写字母 p_1 表示样本比率， $p_1=\frac{x}{n}$ ，抽样中该事件之外其他事件出现的比率用小写字母 q_1 表示，$q_1=1-p_1$。

比率的分布为二项分布。若在总体中，某事件出现的比率（也即概率）为 P_0，则这一事件之外其他事件 Q_0 出现的概率为 $Q_0=1-P$。从这样的总体中抽取容量为 n 的样本，利用实际获得的该现象出现的次数 X 与样本容量 n 所求得的样本比率 p_1 来估计总体比率 P_0 的所在范围，称为总体比率的区间估计（interval estimation of population ratio)。

在大样本即 $np_1\geq 5$ 且 $P_0<Q_0$（或者 $nq_1\geq 5$ 且 $Q_0<P_0$）的条件下，二项分布近似为正态分布，这时，n 次独立试验中某事件发生的次数 $X\sim N(nP_0,nP_0Q_0)$，发生的比率 p_1 $\sim N(P_0,\frac{1}{n}P_0(1-P_0))$。有关比率的抽样标准差 SE_{p_1} 的计算公式有如公式 8-1 所示。

$$SE_{p_1}=\sqrt{\frac{P_0(1-P_0)}{n}}\approx\sqrt{\frac{p_1(1-p_1)}{n}}=\sqrt{\frac{p_1q_1}{n}} \qquad \text{（公式 8-1）}$$

公式 8-1 中，样本比率 p_1 的抽样标准差 SE_{p_1} 理论上等于 $\sqrt{\frac{P_0(1-P_0)}{n}}$，若总体比率 P_0 未知，则可以用 $\sqrt{\frac{p_1(1-p_1)}{n}}$ 作为对抽样标准差 SE_{p_1} 的近似估计。根据第五章中参数估计的基本原理，易知总体比率 P_0 的置信水平为 1-α 的置信区间为：

$$p_1-Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1(1-p_1)}{n}}\leq P_0\leq p_1+Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1(1-p_1)}{n}} \qquad \text{（公式 8-2）}$$

在实际进行总体比率的区间估计时，还要注意在不同的条件下要选用适当的
方法来进行估计。

一、 当 $np_1\geq 5$ 或 $nq_1\geq 5$ 时总体比率的区间估计

【例 8-4】 对某年级学生进行某项技能训练后，从全年级学生的总体中选出 80 人样本进行测验，结果有 65 人通过，15 人未通过。求整个总体中通过测验的人数比率在 95%置信水平上的置信区间。

解：因为 $p_1=\frac{x}{n}=0.8125$， $q_1=1-p_1=0.1875$， $np_1>5$，样本比率近似正态分布，因此可以用 Z 统计量来进行参数估计。具体步骤为：

（1） 列出总体比率 P_0 在置信水平为 1-α 时的估计区间不等式：

$$p_1-Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1(1-p_1)}{n}}\leq P_0\leq p_1+Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1(1-p_1)}{n}}$$

（2）计算比率的抽样误差 SE_{p_1} 和给定置信水平 0.95 时的 $Z_{\frac{\alpha}{2}}$ 值。

$$SE_{p_1}=\sqrt{\frac{p_1(1-p_1)}{n}}=\sqrt{\frac{0.8125\times0.1875}{80}}=0.0436；\ Z_{\frac{0.05}{2}}=1.96$$

（3）将有关数据代入上述不等式求具体的解：

$$0.8125-1.96\times0.0436\le P_0\le0.8125+1.96\times0.0436$$

$$0.727\le P_0\le0.898$$

答：根据这 80 位学生组成的样本，可以推论该年级全体学生通过该项技能测验的比率在 0.727 ~ 0.898 之间，或者说，通过测验人数的比率介于 72.7%～89.8%之间，估计正确的概率为 0.95，错误的概率为 0.05。

【例 8-5】 随机抽查某高校 500 名大学生，发现有 342 名大学生来自城市，问该学校城市大学生比率 **0.99 的置信区间**？

解：$p_1=\frac{x}{n}=\frac{342}{500}=0.684$，$q_1=1-p_1=0.316$，$np_1>5$，样本比率近似正态分布，因此可以用 Z 统计量来进行参数估计。具体步骤为：

（1）根据置信水平 1-$\alpha=0.99$ 查 Z 分布表求得 $Z_{\frac{0.05}{2}}=2.58$。

（2）计算比率的抽样误差为$SE_{p_1}=\sqrt{\frac{p_1(1-p_1)}{n}}=\sqrt{\frac{0.684\times0.316}{500}}=0.021$；

（3）将有关数据代入上述不等式（公式 8-2）求具体的解：

$$0.684-2.58\times0.021\le P_0\le0.684+2.58\times0.021$$

$$0.63\le p\le0.738$$

答：该学校**来自城市的大学生**的比率在 0.630～0.738 之间，作此推论正确的概率为 0.99，错误的概率为 0.01。

二、当 $np_1\le5$ 或 p_1 很小时总体比率的区间估计

当 $np_1\le5$ 或 p_1 很小时，其二项分布不是近似正态分布，这时的区间估计要直接查附表 21 即“二项分布上下置信界限表”来进行计算。

【例 8-6】从某中学初一年级随机抽取 30 名学生，测得英语成绩优秀者为 4 人，试估计该校初一学生此次测验英语成绩优秀的比率是多少？

解：$p_1=0.13$，$np_1<5$，不能用正态分布法求置信区间。应查附表 21，样本容量 n=30 这一列与实计数 $X=4$ 这一行的相交处，查得两个数字 4、31，即初一学生英语成绩优秀者的比率在 95%置信水平上的置信区间为[4%，31%]，作此推论错误的可能性为 5%。

答：该校初一学生此次测验英语成绩优秀的比率约在 4%至 31%之间。

第三节　总体比率的假设检验

总体比率的假设检验（hypothesis testing of population ratio）是指是指通过样本比率推断总体比率的统计推断过程。包含以下两种情况：一种情况是指已知某个总体比率 P_0，要检验样本比率 p_1 与总体比率 P_0 之间的差异是否显著所进行的差异显著性检验过程，这被称为**单总体比率差异显著性检验**；另外一种情况是对两个比率之间的差异是否显著所进行的差异显著性检验过程，这被称为**两总体比率差异显著性检验**。下面我们分述之。

一、单总体比率差异显著性检验

单总体比率显著性检验是已知总体比率 P_0，要检验样本比率 p_1 与总体比率 P_0 的差异。也就是说，如果在 n 次独立的观察或试验中，事件 A 发生了 K 次，那么在这 n 次观察或试验中，事件 A 的发生频率为 $p_1=\frac{k}{n}$，其分布属于二项分布。当 $p_1<q_1,np_1\geq 5$（或 $p_1>q_1,nq_1\geq 5$）时，$\frac{k}{n}$ 接近正态分布。则此时要检验 p_1 与 P_0 的差异，其检验统计量为：

$$Z=\frac{p_1-P_0}{\sqrt{\frac{p_1(1-p_1)}{n}}} \qquad \text{（公式 8-3）}$$

【例 8-7】 已知某市高一学生会考优秀率为 45%，从该市某区随机抽取了 500 名高一学生中，会考成绩优秀的有 268 人。问该区高一学生会考优秀率与全市的优秀率是否有显著差异？（$\alpha=0.01$）

解：根据题意计算 p_1 值：$p_1=\frac{268}{500}=0.536$

（1）建立假设：$H_0{:}P_0=0.45$

$H_1{:}P_0\neq 0.45$

（2）根据公式 8-3 计算检验 Z 统计量得：

$$Z=\frac{0.536-0.45}{\sqrt{\frac{0.45(1-0.45)}{300}}}=\frac{0.086}{0.0287}=2.997$$

（3）确定显著性水平并做出决策：

根据显著性水平 $\alpha=0.01$，查正态分布表得 $Z_{0.01/2}=2.58$，$Z=2.997>Z_{0.01/2}=2.58$，所以拒绝虚无假设 H_0。

答：该区高一学生会考优秀率与全市的优秀率有显著差异。

这里应当注意的是，只有当 p_1 不接近 0 或 1，np_1 和 nq_1 中数值小者在 5 以上时，才可用公式 8-2，因为此时的二项分布接近正态分布。

二、两总体比率差异显著性检验

如果两个比率 p_1、p_2 分别表示两个独立样本中同一事件发生的比率，且 $np_1 \geq 5$，$np_2 \geq 5$，此时 p_1 和 p_2 都近似服从正态分布，即 p_1 来自近似正态总体 $N(P_{0-1},\frac{P_{0-1}Q_{0-1}}{n_1})$，$p_2$ 来自近似正态总体 $N(P_{0-2},\frac{P_{0-2}Q_{0-2}}{n_2})$，则 p_1-p_2 的分布也服从正态分布，即 p_1-p_2 应来自近似正态总体 $N(P_{0-1}-P_{0-2},\frac{P_{0-1}Q_{0-1}}{n_1}+\frac{P_{0-2}Q_{0-2}}{n_2})$。在上述表述中，符号 P_{0-1} 代表抽样比率 p_1 的总体比率，符号 P_{0-2} 代表抽样比率 p_2 的总体比率，n_1，n_2 是两个样本的容量。如果我们假设这两个总体比率之间的差异为零，即 $H_0\text{:}P_{0-1}-P_{0-2}=0$，则两个比率间差异抽样标准误差为：$SE_{p_1-p_2}=\sqrt{P_0Q_0(\frac{1}{n_1}+\frac{1}{n_2})}$。那么我们可以利用公式 8-3 来计算比较两个比率间差异是否显著的 Z 统计量。

$$Z=\frac{p_1-p_2}{SE_{p_1-p_2}}=\frac{p_1-p_2}{\sqrt{P_0Q_0(\frac{1}{n_1}+\frac{1}{n_2})}} \qquad \text{（公式 8-4）}$$

由于 P_0 和 Q_0 都是未知数，因而我们无法通过计算得到 Z 值。当样本容量 n_1 和 n_2 都比较大时，则可以采用以样本的概率估计值 $\hat{p}$ 代替 P_0 的方法来计算，样本的概率估计值 $\hat{p}$ 可以通过计算这两个抽样概率值的加权平均数来求得，其计算方法有如公式 8-4 所示：

$$\hat{p}=\frac{n_1}{n_1+n_2}p_1+\frac{n_2}{n_1+n_2}p_2=\frac{p_1n_1+p_2n_2}{n_1+n_2} \qquad \text{（公式 8-5）}$$

而后，可以用公式 8-5 来计算比较这样两个比率间差异是否显著的 Z 统计量。

$$Z=\frac{p_1-p_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1}+\frac{1}{n_2})}} \qquad (\hat{q}=1-\hat{p}) \qquad \text{（公式 8-6）}$$

【例 8-8】 为调查学生对某种语文阅读教学方法的评价，随机抽取了 85 名学生，其中男生 40 名，女生 45 名。男生对该教学方法持支持态度的占 47.5%，女生对该教学方法持支持态度的占 55.6%。能否说明男女学生对该

教学方法所持的态度的比率有显著差异？（$\alpha=0.05$）

解：根据题意得

（1）建立假设：$\begin{matrix} H_0: P_{0-1}=P_{0-2} \\ H_1: P_{0-1}\neq P_{0-2} \end{matrix}$

（2）计算检验统计量：

先根据公式 8-4 计算 p 值：

$$\hat{p}=\frac{40\times 0.475+45\times 0.556}{40+50}=\frac{19+25}{85}=0.518$$

$$\hat{q}=1-0.518=0.482$$

再根据公式 8-5 计算检验统计量，

$$Z=\frac{0.475-0.556}{\sqrt{(0.518\times 0.482)(\frac{1}{40}+\frac{1}{45})}}=\frac{-0.081}{0.1086}=-0.746$$

（3）确定显著性水平并做出决策：根据显著性水平 $\alpha=0.05$，查正态分布表得 $Z_{0.05/2}=1.96$，实际得到的检验统计量 $|Z|=0.746<Z_{0.05/2}=1.96$，所以接受虚无假设 H_0，即认为男女学生对该教学方法所持的态度的比率没有显著差异。

答：男女学生对该教学方法所持的态度的比率没有显著差异。

第四节 拟合良度的 χ^2 检验

一、χ^2 检验的含义与计算方法

（一） χ^2 统计量、χ^2 检验的含义与适用资料

如前所述，当实得的数据不去除以相应的总体数据以求它们的比率时，我们得到的就是计数数据。前面各章节学习的统计方法，大多用于计量数据的分析，这样的数据能计算平均数和方差，例如用 t 检验或方差分析来进行均值差异显著性检验。当研究面临的数据不是计量数据，而是不同分类特征中的计数数据时，一般用 χ^2 检验的方法来达到推断统计的目的。所谓 χ^2 检验（chi-square test）是指检验统计量是否服从或渐近服从 χ^2 分布的非参数检验方法。从某种意义上说，所谓非参数检验也就是对总体分布不做严格假定的检验（详见 11 章）。应用 χ^2 检验分析计数数据，对计数数据总体的分布形态不作任何假设，因此，χ^2 检验被视为是一种非参数检验。

另外，比率的差异检验一般只能对两个比率之间的差异显著性进行检验，如果要检验的比率数超过 2 个时，则可以通过用 χ^2 统计量对拟合良度的检验方法来达到推断统计的目的。

第七章我们曾对 χ^2 分布的定义、性质、χ^2 分布的分位数和 χ^2 分布表等作了介绍，现在我们对有如公式 8-6 所示的 χ^2 统计量的一般表达式作进一步的解释。

所谓χ^2**统计量**（statistic of χ^2 test）**是指在χ^2检验中用以判断接受还是拒绝虚无假设**的检验统计量，其计算公式为：

$$X^2=\sum\frac{(f_0-f_e)^2}{f_e} \qquad \text{（公式 8-7）}$$

在上式中，f_o表示实际次数（actual frequencies），是指在实验或调查中得到的计数资料，又称为观察次数（observed frequencies）。f_e表示理论次数(theoretical frequencies)，是指根据概率原理、某种理论、某种理论次数分布或经验次数分布计算出来的次数，又称为期望次数（expected frequencies）。χ^2是实际次数与理论次数之间差异程度的指标。以本章第一节例 8-2 为例，可以说明χ^2为什么是这样的一个指标。

根据概率论，当骰子各面均匀时，掷骰子每个点数出现的概率都是相同的。掷骰子 120 次，出现每个点数的次数理论上都是 20。可将这个例子的情况列成表 8-4。

表 8-4　掷骰子 120 次各种点数的 f_o 和 f_e

点数	1	2	3	4	5	6	Σ
f_o	16	25	20	24	12	23	120
f_e	20	20	20	20	20	20	120
f_o-f_e	-4	5	0	4	-8	3	0
$\frac{(f_o-f_e)^2}{f_e}$	0.8	1.25	0	0.8	3.2	0.45	6.5

从表 8-4 看到，实际次数与理论次数存在一定的差异。 如果把掷骰子 120 次当作一次实验，这个差异是属于抽样误差，还是有实质性的差异？要回答这个问题， 首先需要确定一个统计量用以表示实际次数与理论次数偏离的程度；然后判断这一偏离程度是否属于抽样误差，即进行显著性检验。为了度量实际次数与理论次数偏离的程度，最简单的办法是求出实际次数与理论次数的差数之和。但从表 8-4 中，可以看到$\sum(f_o-f_e)$=0，在这个例子中，掷得每个点数的理论次数 20 其实就是掷得各种点数的实际次数的平均数，离均差之和总是为 0。为了避免实际次数与理论次数差数之和为 0，可将差数平方后再相加。$\sum(f_o-f_e)^2$的值越大，表示实际次数与理论次数的偏离程度越大；$\sum(f_o-f_e)^2$的值越小，表示实际次数与理论次数的偏离程度越小。但此时还有一个问题，当不同组的理论次数有差异时，即使f_o-f_e的值相同，所代表的偏离程度也不同。如两个组的实际次数分别为 500 和 50，这两个组的理论次数分别为 495 和 45，两个组中实际次数和理论次数的差值都是 5，但两组实际次数和理论次数的偏离程度是不同的。为了解决这个问题，可先将各差数平方除以相应的理论次数后再相加，并记之为χ^2。χ^2越小，表明实际次数与理论次数越接近；χ^2=0，表示两者完全吻合；

χ^2越大，表示实际次数与理论次数两者相差越大。

（二） χ^2检验的基本逻辑

χ^2检验的目的是检验实际次数和理论次数的偏离是抽样误差还是存在实质性的差异。为了完成这种检验，首先要找到一个统计量，这个统计量要能描述样本中实际次数和理论次数的偏离程度。χ^2统计量就属于这样的一个统计量。χ^2统计量要能够作为假设检验中待检验的统计量，还要求χ^2统计量的概率分布是已知的。当我们从一个总体中多次抽取样本时，根据每个样本的实际次数与理论次数，都可以计算出一个χ^2统计量值，由于每个样本中的个体不同，每个样本的χ^2统计量值往往是不同的，多次抽样研究得到的多个χ^2统计量近似地服从统计学中一种连续型随机变量的概率分布——χ^2分布，实验次数或样本容量越大，这种近似程度越好。根据χ^2统计量近似服从的χ^2分布，就可以知道取得某χ^2统计量的概率。如果是小概率，则可以根据小概率事件原理推翻虚无假设；如果不是小概率，则无充分理由拒绝虚无假设。此处的虚无假设是实际次数分布符合计算期望次数时所依据的某种理论概率原理、某种理论次数分布或经验次数分布。

以上述掷骰子的资料为例，如果要知道掷骰子出现各种点数的次数是否符合出现每个点数的概率相等的理论，根据假设检验的方法，首先要建立假设。虚无假设是掷骰子出现各种点数的次数符合出现每个点数的概率相等的理论。之所以要将这种说法作为虚无假设，是因为可以基于这个虚无假设计算出现每个点数的期望次数。如果虚无假设是掷骰子出现各种点数的次数不符合出现每个点数的概率相等的理论 ，那么就没有办法计算出现每个点数的期望次数。接下来计算χ^2统计量，用来描述掷骰子出现各种点数的实际次数和根据每个点数出现的概率相等的理论计算得的理论次数间的偏离程度。如果χ^2统计量在它近似服从的χ^2分布中是小概率事件，就说明，χ^2统计量表现出的偏离程度是很少见的，此时，作出掷骰子出现各种点数的次数不符合出现每个点数的概率相等的理论是符合逻辑的。

使用χ^2检验的方法能解决诸如拟合良度检验、独立性检验、同质性检验等问题。在χ^2检验方法解决这些问题时，都是先建立假设。在虚无假设成立的条件下，计算各类别或各单元格的理论次数，然后计算出表示实际次数与理论次数的偏离程度的χ^2统计量，最后在χ^2分布中判断计算得的χ^2统计量是否是小概率事件，进而判断有无充分理由拒绝虚无假设。

（三） χ^2检验的假设

1. 分类相互排斥，互不包容

χ^2检验中的分类要做到相互排斥，互不包容，这样每一个观测值就会被划分到一个类别或另一个类别中，不会出现某一个观测值同时划分到更多的类别中

去的情况。

2. 观测值相互独立

各个被试的观测值之间要彼此独立，观测值可能来自于不同被试和同一被试。来自不同被试的不同观测值彼此独立相对容易满足，来自同一被试的观测值要保持独立较难。如5位男生和5位女生对5位教师进行评价，可能有某位学生对所有教师的评价都很低。因此在实际研究中，要求每个被试只有一个观测值是确保观测值相互独立的最安全的做法。

3. 期望次数的大小

由公式8-6式计算的χ^2统计量只是近似地服从连续型随机变量χ^2分布，为了使χ^2统计量更接近连续型随机变量χ^2分布，每一个分类下的理论次数应该至少在5个以上。当分类下的人数过少时，可以用下述几种方法加以处理，以提高每一分类下的次数。

方法一，类别并法。即适当调整变量的分类方式，将部分类别进行合并。例如在学历层次中，如果博士生过少，可以将博士生与硕士生合并成为研究生计算。

方法二，增加样本数。样本数增加了，则每一分类下的人数都可能增加，这样可提高每一分类下的理论次数。

方法三，去除样本法。即去除单元格人数过少的类别。但研究的结论不能推论到这些被去除的类别的母总体中。

二、拟合良度检验

拟合良度检验（goodness of fit test）是指实际观察的属性类别分配是否符合已知属性类别分配理论或学说所进行的统计检验过程。

（一） 拟合良度检验的一般步骤

1. 建立假设

H_0：实际观察的属性类别分配符合已知属性类别分配的理论或学说。

H_A：实际观察的属性类别分配不符合已知属性类别分配的理论或学说。

2. 计算理论次数和χ^2统计量

在虚无假设成立的条件下，按已知属性类别分配的理论或学说计算各属性类别的理论次数。所计算得的各个属性类别理论次数的总和应等于各个属性类别实际观察次数的总和。将计算得的理论次数代入公式(8-6)计算出χ^2统计量。

3. 查表及做决断

查χ^2值表得到临界值$\chi^2_{0.05}$或$\chi^2_{0.01}$，若$\chi^2<\chi^2_{0.05}$或$\chi^2_{0.01}$，则 $P \leq 0.05$ 或 0.01，表明实际次数与理论次数差异不显著，可以认为实际观察的属性类别分配符合已

知属性类别分配的理论或学说；若$\chi^2 > \chi^2_{0.05}$ 或$\chi^2_{0.01}$，则 $P < 0.05$ 或 0.01，表明实际观察次数与理论次数差异显著，实际观察的属性类别分配显著不符合已知属性类别分配的理论或学说。

（二）已知理论概率的拟合良度检验

在拟合良度检验中，有时各分类变量中的概率相等，有时不同分类变量有特定的概率，有时各分类变量中的个数服从特定的分布。下面的三个例子就分别说明了上述情况。

因为计算得的各个属性类别的理论次数的总和应等于各个属性类别实际观察次数的总和，因此拟合良度检验的自由度等于属性类别分类数减 1。若属性类别分类数为 k，则拟合良度检验的自由度为 $k-1$。

【例 8-9】 以例 8-2 中的问题和表 8-4 中所列数据，检验实际掷骰子的结果是否符合每个点数出现的概率相同的理论?

解：

（1）建立假设

H_0：掷骰子的结果符合每个点数出现的概率相同的理论。

H_A：掷骰子的结果不符合每个点数出现的概率相同的理论。

（2）计算理论次数与χ^2统计量

按照每个点数出现的概率相同的理论，掷骰子 120 次，出现每个点数的次数理论上都是 20，将实际次数和理论次数代入公式 8-6 得：

$$\chi^2 = \sum\frac{(f_o - f_e)^2}{f_e} = \frac{(16-20)^2}{20} + \frac{(25-20)^2}{20} + \frac{(20-20)^2}{20} + \frac{(24-20)^2}{20} + \frac{(12-20)^2}{20} + \frac{(23-20)^2}{20} = 6.5$$

（3）确定显著性水平并做出决策

当 df= k-1=5-1=4 时，$\chi^2_{0.05}$=9.49， $X^2 < X^2_{0.05}$，因此接受虚无假设，即掷骰子的结果符合每个点数出现的概率相同的理论。

答：掷骰子的结果符合每个点数出现的概率相同的理论。

【例 8-10】 一位研究者从某大学随机抽取 200 名学生参加心理健康测试。实施测验后，有 26 人的成绩为优，94 人为良，75 人为中等，5 人为差。问：学生在该心理测验上的表现是否符合正态分布？

解：

（1）建立假设

H_0：学生在该心理测验上的表现是否符合正态分布。

H_A：学生在该心理测验上的表现是否不符合正态分布。

（2）计算理论次数与 χ^2 统计量

计算学生在该心理测验上的表现是否符合正态分布时，200 名学生在每个等级上人数的理论次数。

首先将正态曲线下横轴上的 $[-3\sigma, 3\sigma]$ 分为四等分，每等分 1.5σ。查正态分布表，求出正态曲线下每等分的面积比率。将总人数（200 人）乘以每等分的面积比率得到每等分的理论次数。

表 8-5 学生人数（N=200 人）在四种等级中的理论次数表

等级	正态曲线下横轴上的区间	面积比率	理论次数
优	$+1.5\sigma$ 至 $+3\sigma$	0.07	14
良	0σ 至 $+1.5\sigma$	0.43	86
中	-1.5σ 至 0σ	0.43	86
差	-3σ 至 -1.5σ	0.07	14

$$\chi^2 = \frac{(26-14)^2}{14} + \frac{(94-86)^2}{14} + \frac{(75-86)^2}{14} + \frac{(5-14)^2}{14}$$

$$= 10.286 + 0.744 + 1.407 + 5.786 = 18.223$$

（3）确定显著性水平并做出决策：当 df= k-1=4-1=3 时，$\chi^2_{0.05}$=7.82，$\chi^2 > \chi^2_{0.05}$，因此拒绝虚无假设，即学生在该心理测验上的表现不符合正态分布。

答：学生在该心理测验上的表现不符合正态分布。

【例 8-11】 国际色觉障碍讨论会宣布，平均每 12 个男子中，有一个是先天性色盲。从某校抽取的 132 名男生中有 4 人是色盲，问该校男子色盲比率与上述比率是否有显著差异？

解：

（1）建立假设

H_0：该校男子色盲比率与国际色觉障碍讨论会的统计结果无显著差异。

H_A：该校男子色盲比率与国际色觉障碍讨论会的统计结果有显著差异。

（2）计算理论次数与 χ^2 统计量

按国际色觉障碍讨论会的统计结果，132 人应该有 $11\left(\frac{132}{12}\right)$ 人是色盲，剩下的 121 人非色盲，代入公式有：

$$\chi^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(4-11)^2}{11} + \frac{(128-121)^2}{121} = 4.86$$

（3）查表及做决断

当 df= k-1=2-1=1 时，$\chi^2_{0.05}$=3.84，$\chi^2 > \chi^2_{0.05}$，因此拒绝虚无假设，该校

男子色盲比率与国际色觉障碍讨论会的统计结果有显著差异。

答：该校男子色盲比率与国际色觉障碍讨论会所宣布的色盲比率有显著差异。

（三）正态分布的拟合良度检验

连续随机变量的计量数据有多种理论分布，如正态分布、二项分布、泊松分布等。在实际研究中，有时并不知道所获得的连续随机变量的计量数据总体分布状况，而进一步的研究工作需要将总体分布确定下来，此时就可以根据样本的次数分布进行假设检验，来判断总体是否服从某种指定的具有明确表达式的理论次数分布。这种假设检验被称为分布的拟合良度检验。关于分布的假设检验方法有很多，最常用的是运用χ^2值进行的拟合良度检验。

χ^2的拟合良度检验中，理论分布的确定有两种途径。一种是将测量数据整理成次数分布表，画出次数分布曲线图，根据次数分布曲线选择恰当的理论分布；一种是直接选择某一直线或曲线为理论分布。选择好理论分布之后，在实际分布符合理论分布的假设条件下计算每组的理论次数，然后将实际分组次数和理论次数代入计算χ^2统计量的基本公式，将计算得的χ^2统计量和查表得的χ^2值比较，如果χ^2统计量小于查表得的χ^2值，说明实际分布符合选择的理论分布，如果χ^2统计量大于查表得的χ^2值，说明实际分布不符合选择的理论分布，可再选择其它的理论分布进行假设检验，直到找到合适的理论分布。

心理研究中常用到的分布是正态分布，下面通过一个例子来说明正态分布拟合良度检验如何进行。

【例 8-12】 有 240 名学生在某心理测验上的得分如表 8-6 所示，问这些学生在该心理测验上的得分是否服从正态分布。

解：

（1）建立假设

H_0：学生在该心理测验上的得分服从正态分布。

H_A：学生在该心理测验上的得分不服从正态分布。

（2）计算理论次数与χ^2统计量

每组的理论次数在学生在该心理测验上的得分服从正态分布的假设条件下计算，有两种方式。

方法一：

①将各组组中值转化为 Z 分数

②根据 Z 分数查正态分布表求 y 值；

③将 y 值乘以$\frac{i}{s}$（以 Z 分数为单位的组间距），得到按正态分布每个分组区间的概率 P；

表 8－6　240 名学生在某心理测验上得分的次数分布表及正态分布 χ^2 检验

组别	组限	次数 f_o	精确下限 Z 值	精确上限 Z 值	精确下限 P 值	精确上限 P 值	各组概率	各组理论次数 f_e	$\frac{(f_o-f_e)2}{f_e}$
90－95	[89.5,94.5）	8	1.37	1.89	0.41466	0.47062	.05596	13.4304	2.1957
85—90	[84.5,89.5）	52	0.86	1.37	0.30511	0.41466	.10955	26.2920	25.1370
80—85	[79.5,84.5）	32	0.34	0.86	0.13307	0.30511	.17204	41.2896	2.0900
75—80	[74.5,79.5）	64	-0.18	0.34	0.07142	0.13307	.20449	49.0776	4.5373
70—75	[69.5,74.5）	24	-0.70	-0.18	0.25804	0.07142	.18662	44.7888	9.6492
65—70	[64.5,69.5）	32	-1.22	-0.70	0.38877	0.25804	.13073	31.3752	.0124
60—65	[59.5,64.5）	8	-1.73	-1.22	0.45818	0.38877	.06941	16.6584	4.5003
55—60	[54.5,59.5）	16	-2.25	-1.73	0.48778	0.45818	.02960	7.1040	12.1012
50—55	[49.5,54.5）	4	-2.77	-2.25	0.49720	0.48778	.00942	2.2608	
	N=240 $\overline{X}=76.2333$ $S=9.6552$							$\chi^2=60.22$	

④求各组的理论次数 $f_e = p \times N$ 。

方法二：

①将各组精确上、下限转化为 Z 分数；

②查正态分布表求各 Z 分数的概率；

③求各分组区间的概率。此时需利用上精确上限的概率值和精确下限的概率值；

④求各组的理论次数 $f_e = p \times N$ 。

在表 8-6 中应用方法二来计算各组的理论次数。

在表 8-6 中计算各组概率时，如果某组的精确上限和精确下限都在平均数以上，则该组的概率等于精确上限 P 值减精确下限 P 值，如上表中 80-85 及以上各组均这样计算各组概率；如果精确下限和精确上限一个在平均数以下，一个在平均数以上，则该组的概率等于精确上限 P 值加上精确下限 P 值，如上表中 75-80 组这样计算概率；如果某组的精确上限和精确下限都在平均数以下，则该组的概率等于精确下限 P 值减去精确上限 P 值，如上表中 70-75 及以下各组均这样计算各组概率。

在计算 χ^2 统计量时，如果有小于 5 的理论次数组会影响到使用 χ^2 分布进行检验，因此要将该组和附近组合并，如上表中 50-55 组和 55-60 组合并为一组计算实际次数和理论次数。

（3）查表及做决断

χ^2 的拟合良度检验的自由度等于组数减去用到的统计量的数目。表 8-6 经合并后分 8 组，在计算理论次数的过程中用到了平均数、标准差、总数三个统计量，因此自由度 df =8-3=5。查 χ^2 分布表，当自由度等于 5 时， $\chi^2_{0.05}$=11.1，$\chi^2 > \chi^2_{0.05}$，因此拒绝虚无假设，即学生在该心理测验上的得分不服从正态分布。

答：学生在该心理测验上的得分不服从正态分布。

其它分布的拟合良度检验的方法和正态分布拟合良度检验方法相同。

第五节 同质性 χ^2 检验

一 同质性检验的含义与一般步骤

同质性检验（test for homogeneity）实际上是对我们所关心的几个研究对象之间是否同质进行的检验。其目的在于检验不同人群母总体在某一个变量的反应是否具有显著差异。它是对几个样本同一变量的分布状况的检验，是对几个样本数据是否同质作出统计决断。进行同质性检验的一般步骤有如下述。

（1）建立假设。

H_0：来自不同样本的资料同质。

H_A：来自不同样本的资料不同质。

（2）计算三种 χ^2 值。

累计 χ^2 值：指各个样本组的 χ^2 值的总和，其自由度为各个样本组自由度的总和；

合并 χ^2 值：将各样本组原始数据按相应类别合并，产生一个总的数据表，合并 χ^2 值即这个总数据表的 χ^2 值，其自由度为合并后数据表的自由度；

异质 χ^2 值：累计 χ^2 值与合并 χ^2 值之差，其自由度为各样本组累计自由度与总数据表的自由度之差。

（3）查表与判断异质 χ^2 值的显著性。

如果显著，表明几个样本组之间异质；如果不显著，表明同质。从实验设计角度讲，这几个样本组数据可以合并到一起。

二、单因素分类数据的同质性检验

【例 8-13】 在一项消费心理学的研究中，调查大学生对某品牌的态度，态度分喜欢、一般、不喜欢三种，大学生来自两个不同类型的大学，调查结果如下（表 8-7）。问来自这两所大学的调查结果是否同质？

表 8-7 两所大学的学生对某品牌的态度调查结果表

学校	喜欢	一般	不喜欢	总和
1	48	32	10	90
2	38	35	17	90
总和	86	67	27	180

例 8-13 中的单因素分类数据的同质性检验其实包含两个问题。问题一是大学生在对某品牌的三种态度上有无显著差异。问题二是检验来自两所大学的学生的态度是否同质。

解：

（1）建立假设

H_0：来自两所大学的调查结果同质。

H_A：来自两所大学的调查结果不同质。

（2） 计算三种 χ^2 值。

① 计算累计 χ^2 值

先计算各个样本组的 χ^2 值和自由度

如果大学生在对某品牌的三种态度上无显著差异，则持每种态度的人占总人数的$1/3$，总人数为 90，则持每种态度的人数的理论次数是 30 人（表 8-8）。

表 8-8 大学生对某品牌的态度调查结果的理论次数表

学校	喜欢	一般	不喜欢	总和
1	48（30）	32（30）	10（30）	90
2	38（30）	35（30）	17（30）	90
总和	86（60）	67（60）	27（60）	180

计算学校 1 在三种态度上实际次数和理论次数的χ^2值：

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(48-30)^2}{30} + \frac{(32-30)^2}{30} + \frac{(10-30)^2}{30}$$

$$= 24.267$$

计算学校 2 在三种态度上实际次数和理论次数的χ^2值：

$$X^2 = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(38-30)^2}{30} + \frac{(35-30)^2}{30} + \frac{(17-30)^2}{30}$$

$$= 8.6$$

因为态度分为三类，所以两所学校各自的自由度均为$df = k - 1 = 3 - 1 = 2$。

计算累计χ^2值：$\chi_{累计}{}^2 = \chi_1{}^2 + \chi_2{}^2 = 24.2667 + 8.6 = 32.8667$

累计χ^2值的自由度是：$df = 2 + 2 = 4$

② 计算合并χ^2值

将两学校的人数累加总人数为 180，则持每种态度的人数的理论次数是 60 人，计算合并χ^2值：

$$X^2_{合并} = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(86-60)^2}{60} + \frac{(67-60)^2}{60} + \frac{(27-60)^2}{60}$$

$$= 30.233$$

合并χ^2值后的自由度是：$df = 3 - 1 = 2$

③ 计算异质性χ^2值

$$X^2_{异质} = X^2_{累计} - X^2_{合并} = 32.8667 - 30.2333 = 2.633$$

异质性χ^2值的自由度是：$df = 4 - 2 = 2$

（3）查表与判断异质χ^2值的显著性

当 df=2 时，$\chi^2{}_{0.05}$=5.99，$\chi^2 < \chi^2{}_{0.05}$，因此无充分理由拒绝虚无假设，即来自两所大学的调查结果同质。

答：两所大学的学生态度调查结果是同质的。

三、列联表形式的同质性检验

【例 8-14】 在一项消费心理的研究中，在两所不同类型大学中调查不同性别的学生对某品牌的态度，调查结果如表 8-9 所示，问来自这两所大学的调查结果是否同质？

表 8-9 两所大学不同性别的学生对某品牌的态度调查结果表

性别	学校 1			学校 2		
	喜欢	一般	不喜欢	喜欢	一般	不喜欢
男	32	18	6	33	17	6
女	26	14	4	15	18	11

解：

（1）建立假设

H_0：来自两所大学的调查结果同质。

H_A：来自两所大学的调查结果不同质。

（2）计算三种 χ^2 值。

① 计算累计 χ^2 值

使用 $r\times c$ 列联表计算 χ^2 统计量的简捷公式先计算各个样本组的 χ^2 值，这需要先计算学校 1 和学校 2 各自数据的边缘小计和总人数（表 8-10 和表 8-11）。

表 8-10 学校 1 不同性别的学生对某品牌的态度调查结果表

性别	学校 1			边缘小计
	喜欢	一般	不喜欢	
男	32	18	6	56
女	26	14	4	44
边缘小计	58	32	10	100

表 8-11 学校 2 不同性别的学生对某品牌的态度调查结果表

性别	学校 2			边缘小计
	喜欢	一般	不喜欢	
男	33	17	6	56
女	15	18	11	44
边缘小计	48	35	17	100

$$X^2_{学校1} = N(\sum_{(i,j)} \frac{n_{ij}^{\ 2}}{a_i b_j} - 1)$$

$$= 100(\frac{32^2}{56 \times 58} + \frac{18^2}{56 \times 32} + \frac{6^2}{56 \times 10} + \frac{26^2}{44 \times 58} + \frac{14^2}{44 \times 32} + \frac{4^2}{44 \times 10} - 1)$$

$= 0.1$

$$X^2_{学校2} = N(\sum_{(i,j)} \frac{n_{ij}^{\ 2}}{a_i b_j} - 1)$$

$$= 100(\frac{32^2}{56 \times 48} + \frac{17^2}{56 \times 35} + \frac{6^2}{56 \times 17} + \frac{15^2}{44 \times 48} + \frac{18^2}{44 \times 35} + \frac{11^2}{44 \times 17} - 1)$$

$= 4.49$

两所学校的自由度均为$df = (r-1)(c-1) = (2-1)(3-1) = 2$ 。

计算累计χ^2值：$\chi_{累计}^{\ 2} = \chi_1^{\ 2} + \chi_2^{\ 2} = 0.1 + 4.49 = 4.59$

累计χ^2值的自由度是：$df = 2 + 2 = 4$

② 计算合并χ^2值：合并两个学校的数据，得到如下的数据表：

表 8-12　两所学校不同性别的学生对某品牌的态度调查结果合并表

性别	学校 1 和学校 2			边缘小计
	喜欢	一般	不喜欢	
男	65	35	12	112
女	41	32	15	88
边缘小计	106	67	27	200

计算合并χ^2值：

$$X^2_{合并} = N(\sum_{(i,j)} \frac{n_{ij}^{\ 2}}{a_i b_j} - 1)$$

$$= 100(\frac{65^2}{112 \times 106} + \frac{35^2}{112 \times 67} + \frac{12^2}{112 \times 27} + \frac{41^2}{88 \times 106} + \frac{32^2}{88 \times 67} + \frac{15^2}{88 \times 27} - 1)$$

$= 1.53$

合并χ^2值的自由度是：$df = (r-1)(c-1) = (2-1)(3-1) = 2$。

③ 计算异质性χ^2值

$$\chi_{异质}^{\ 2} = \chi_{累计}^{\ 2} - \chi_{合并}^{\ 2} = 4.59 - 1.53 = 3.06$$

异质性χ^2值的自由度是：$df = 4 - 2 = 2$。

（3） 查表与判断异质χ^2值的显著性

当 df=2 时，$\chi^2_{0.05}$=5.99，$\chi^2 < \chi^2_{0.05}$，因此无充分理由拒绝虚无假设，即来自两所大学的调查结果同质。

答：两所大学的调查结果是同质的。

四、SPSS 统计软件包在 χ^2 检验中的应用

【例 8-15】利用 SPSS 软件包求取例 8-9 数据的 χ^2 拟合良度检验结果。

在已打开的数据文件窗口中，将例 8-9 中的数据正确输入到 SPSS 数据框，其中一列变量为“类别”，另一列变量为“次数”。然后按序点击图 8-1 所示的操作步骤可以求解本例的 χ^2 拟合良度检验结果（见表 8-19 和表 8-20）（注意：在“期望值（Expected Values）”单选框中默认的是“所有类别相等”。如果各类别的期望次数不同，则要选“值”并输入各类别的理论次数）。

1. 数据→加权个案
2. 选中“加权个案”
3. 将“次数”变量选入“频率变量框”
4. 确定
5. 分析→非参数检验→卡方
6. 将“次数”变量选入“检验变量列表框”
7. 确定

图 8-1 在 SPSS 中设置方差齐性检验分析程序各步骤示意图

表8-19 不同投掷结果的观察次数与理论次数间差异表

	观察数	期望数	残差
1.00	16	20.0	-4.0
2.00	25	20.0	5.0
3.00	20	20.0	.0
4.00	24	20.0	4.0
5.00	12	20.0	-8.0
6.00	23	20.0	3.0
总数	120		

表8-20 χ^2检验表

	类别
卡方	6.500[a]
df	5
渐近显著性	.261

运算结果显示各类别的实际次数（Observed）、理论次数（Expected）、各类别的实际次数与理论次数的差值（Residual）； χ^2 （Chi-Square）= 6.5，自由度（df）= 5 ，P = 0.261 ，结果表明没有出现小概率事件。

小　结

比率的分布为二项分布。若在总体中，某事件出现的概率为P_0，则这一事件之外的其他事件出现的概率为Q_0为 $1-P_0$。从这样的总体中抽取容量为 n 的样本，利用实际获得的该事件出现的次数×与样本容量所估计的样本比率 p_1来估计总体比率P_0所在的范围，称为总体比率的区间估计。

总体比率的假设检验包括两方面内容，一是从某总体中抽取的样本的比率与该总体比率的差异是否显著的问题，即已知总体比率 P_0，要检验样本比率 p_1与总体比率 P_0 的差异；二是通过两个抽样样本的比率来比较它们各自的总体比率之间差异是否显著的问题。

χ^2检验主要用于计数资料的检验。χ^2检验的基本思路是：在假设条件成立的前提下计算不同类别的理论次数，用χ^2统计量来描述理论次数和实际次数的偏离程度。因为χ^2统计量近似服从χ^2分布，因此在χ^2分布中判断χ^2统计量是否是小概率事件，若是，则拒绝虚无假设。

χ^2检验方法在实计数数据的处理中主要用于对拟合良度进行检验和对两类变量是否同质性进行检验。

拟合良度检验主要用来判断实际观察的属性类别分配是否符合已知属性类别分配理论或学说。已知属性类别分配理论可以是无差理论、某种特定概率、正态分布等等。在检验中要根据已知属性类别分配理论计算各类别理论次数。

同质性检验在于检验不同人群母总体在某一个变量的反应是否具有显著差异。

关键术语

计数数据：是指在心理研究中，对研究对象通过计算个体数目而获取的数据。

比率：在试验中某事件出现的理论次数除以试验总次数所获得的商。

总体比率的区间估计：若在总体中，某事件出现的比率（也即概率）为P_0，则这一事件之外其他事件出现的概率Q_0为 $1-P_0$。从这样的总体中抽取容量为n的样本，利用实际获得的该现象出现的次数 X 与样本容量n所求得的样本比率

p_1来估计总体比率P_0的所在范围，称为总体比率的区间估计。

总体比率的假设检验：是指通过样本比率推断总体比率的统计推断过程。包含两种情况：一种是已知某个总体比率P_0，要检验样本比率p_1与总体比率P_0之间的差异是否显著；第二种情况是对两个比率之间的差异是否显著所进行的差异显著性检验。

比率的差异显著性检验：是指已知某个总体比率P_0，要检验样本比率p_1与总体比率P_0之间的差异是否显著；或者是对两个比率之间的差异是否显著所进行的差异显著性检验。

拟合良度检验：是指实际观察的属性类别分配是否符合已知属性类别分配理论或学说所进行的统计检验过程。

同质性检验：是对我们所关心的几个研究对象之间是否同质进行的检验。其目的在于检验不同人群母总体在某一个变量的反应是否具有显著差异。

重要公式

比率的抽样标准差：$SE_{p_1}=\sqrt{\frac{p_1(1-p_1)}{n}}=\sqrt{\frac{p_1q_1}{n}}$

总体比率的区间估计公式：

$$p_1-Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1(1-p_1)}{n}}\le P_0\le p_1+Z_{\frac{\alpha}{2}}\sqrt{\frac{p_1(1-p_1)}{n}}$$

单总体比率差异显著性检验统计量：

$$?\ Z=\frac{p_1-P_0}{\sqrt{\frac{p_1(1-p_1)}{n}}}$$

两总体比率差异显著性检验统计量：

$$Z=\frac{p_1-p_2}{\sqrt{\hat{p}\hat{q}(\frac{1}{n_1}+\frac{1}{n_2})}}$$

χ^2**统计量的一般表达式**：

$$\chi^2=\sum\frac{(f_0-f_e)^2}{f_e}$$

列联表同质性检验的统计量：

$$\chi^2=n(\sum_{(i,j)}\frac{n_{ij}{}^2}{a_ib_j}-1)$$

思考与练习

1．某省调查了应届大学毕业生的求职意向，随机抽取 1200 名大四学生，发现其中 280 人首选的求职意向为公务员，试求该省大四学生中首选求职意向为公务员的比率的 0.95 的置信区间？

2．根据以往经验，某高校的学生中参加志愿者活动的比率为 30%。现从该校随机抽查了 200 名大学生，发现其中参加志愿者活动的有 85 名。能否推论该校大学生中参加志愿者活动的比率有所上升？

3．某中学进行了心理健康测评。从参加测评的男生中随机抽取 90 名，发现有抑郁情绪的有 9 人；从女生中随机抽取 80 名，发现有抑郁情绪的有 16 人。能否推论该校男生和女生中有抑郁情绪的学生比率不同？

4．某企业对员工的心理资本进行测评，测评结果分为 A、B、C、D、E 五个等级。从该企业员工中随机抽取 100 人，其心理资本对应于上述五个等级的人数分别为 6 人、27 人、44 人、21 人、2 人。能否推论该企业全体员工的心理资本服从正态分布？

5. 某研究者通过随机抽样，调查了 A、B 两所学校教师的工作满意度，调查结果如下表所示。请分析来自这两所大学的调查结果是否同质？

	A 大学			B 大学		
	满意	一般	不满意	满意	一般	不满意
老教师	12	11	19	24	12	10
青年教师	10	21	26	23	20	11

第九章 相关系数的参数估计与假设检验

学习本章内容，将有助于你对以下问题的理解与思考：

1. 应该如何对总体积差相关系数进行区间估计？
2. 应该如何对积差相关系数的差异显著性进行假设检验？
3. 应该如何对列联表进行独立性检验或计算其品质相关？

1． 在一项双生子研究中，17 对同卵双生子比耐智力测验得分的相关系数为 0.85，24 对异卵双生子比耐智力测验得分的相关系数为 0.76。假设抽样都是随机的，问这两个相关系数是否存在显著差异？

2． 在一个关于幼儿从众行为的研究中，4、5、6 岁儿童从众回答人数与不从众回答人数见表 9-1，问年龄和从众行为有关吗？

表 9-1 儿童的年龄和从众行为调查结果表

回答	年龄			边缘小计
	4 岁	5 岁	6 岁	
从众回答	249	231	145	625
不从众回答	151	169	255	575
边缘小计	400	400	400	1200

第一节　总体相关系数的区间估计

一、积差相关系数的抽样分布

总体的相关系数 ρ 为-1.00～1.00 之间的任意值。从总体中抽取 n 对数据，其相关系数 r 的样本分布随两总体间的相关程度而异。

当总体相关系数 $\rho<0$ 且 n 一定时，样本相关系数 r 的分布呈不同程度的正偏态；$\rho>0$ 时；样本相关系数 r 的分布呈不同程度的负偏态；当 $\rho\neq0$ 时，只有 $n\geq500$ 的情况下才逐渐接近正态分布，而且趋于正态很慢。此时样本相关系数分布的标准误 σ_r 为

$$\sigma_r=\frac{1-r^2}{\sqrt{n-1}} \quad \text{（公式 9-1）}$$

当总体相关系数 $\rho=0$ 时，样本相关系数 r 的分布服从自由度为 $n-2$ 的 t 分布，标准误为：

$$\sigma_r=\frac{\sqrt{1-r^2}}{\sqrt{n-2}} \quad \text{（公式 9-2）}$$

当总体相关系数 $\rho\neq0$ 时，只有 $n\geq500$ 时逐渐接近正态分布，条件非常严格，应用受到很大限制。在一般情况下，利用费希尔(Fisher)的 Z 分布：

$$Z=\frac{1}{2}\ln\frac{1+r}{1-r} \quad \text{或} \quad Z=1.1513\lg\frac{1+r}{1-r} \quad \text{（公式 9-3）}$$

将 r 值转换为 Z 值（n 不受条件限制），这些 Z 值逐渐接近服从正态分布，即费希

尔 Z 分布，其标准误为：$SE_Z=\frac{1}{\sqrt{n-3}}$ 。其中，n 为成对数据的数目，即使 n 比较小也可视为近似正态分布。

二、积差相关系数的区间估计

无论样本容量大小，不论总体相关是否为 0，Z_r 函数的分布都近似正态分布。因此可用 Z_r 的置信区间来估计相关系数 ρ 的置信区间。具体步骤如下：

（1） 根据公式 $Z=\frac{1}{2}\ln\frac{1+r}{1-r}$ 或 $Z=1.1513\lg\frac{1+r}{1-r}$ 将样本相关系数 r 转换成 Z_r，或查 r-Zr 转换表（附表 8）进行转换。

（2）根据给定的显著性水平 α，计算 Z_r 的置信区间 $Z_r \pm Z_{\alpha/2}SE_Z$，其中 $SE_Z=\frac{1}{\sqrt{n-3}}$。

（3）利用公式或查表，将 Z_r 的置信区间转换成总体相关系数 ρ 的置信区间。

【例 9-1】 某位研究者随机从目标群体中选择 200 名被试，进行两种问卷的测验，两问卷测验结果的相关系数为 r =0.36，问这两种问卷在目标群体总体中的相关系数 ρ 的 0.99 置信区间是多少?

解：查 r-Z_r 转换表，r =0.36 时，Z_r =0.377

$$SE_Z=\frac{1}{\sqrt{n-3}}=\frac{1}{\sqrt{200-3}}=0.071$$

$Z_{0.01/2}=2.58$，因此相关系数 ρ 在置信水平 0.99 置信区间为：0.377-2.58×0.071 ≤ $Z_{r\le 0.377+2.58\times 0.071}$，即 0.194≤ Z_r ≤0.560

再查 r-Z_r 转换表，当 Z_r =0.194，ρ =0.190；当 Z_r =0.560，ρ =0.510。

答：总体相关系数 ρ 的 0.99 置信水平上的置信区间为[0.190，0.510]，作此结论犯错误的概率为 0.01。

三、等级相关系数的区间估计

斯皮尔曼等级相关系数,在 $9\le n\le 20$ 时，r_R 的分布近似为 $df=n-2$，$SE_r=\frac{\sqrt{1-r_R^2}}{\sqrt{n-2}}$ 的 t 分布。若符合这个条件，则所求的置信区间为：$r_R \pm t_{\frac{\alpha}{2}}\cdot\frac{\sqrt{1-r_R^2}}{\sqrt{n-2}}$（$df=n-2$）。若 n >20，r_R 的分布近似正态分布，标准误仍为：$SE_r=\frac{\sqrt{1-r_R^2}}{\sqrt{n-2}}$，

则所求的置信区间为:。$r_R \pm Z_{\frac{\alpha}{2}}\cdot\frac{\sqrt{1-r_R^2}}{\sqrt{n-2}}$

【例 9-2】 对同一组被试进行两种实验，已知样本 n =37，两种实验结果的相关系数 r =0.40，问被试所代表的总体的相关系数是多少（0.95 置信水平上的置信区间）？

解：$SE_r = \dfrac{\sqrt{1-r_R^2}}{\sqrt{n-2}} = \dfrac{\sqrt{1-0.4^2}}{\sqrt{18-2}} = 0.229$，查 t 表，$t_{\frac{0.05}{2}(16)} = 2.120$，

故 ρ 在 0.99 置信区间为：$0.4 - 2.120 \times 0.229 \le \rho \le 0.4 + 2.120 \times 0.229$，即

$$-0.085 \le \rho \le 0.085$$

答：总体相关系数 ρ 在 0.99 置信水平上的置信区间为[-0.085，0.885]。

第二节 积差相关系数的假设检验

总体相关系数的假设检验（hypothesis testing of population ratio）是指通过样本相关系数推断总体相关系数的统计推断过程。包含以下两种情况：一种情况是指已知某个样本相关系数 r 和总体相关系数 ρ，要检验 r 和 ρ 之间的差异是否显著所进行的假设检验过程，这被称为**单总体相关系数差异显著性检验**；另外一种情况是通过对两个样本相关系数的差异（$r_1 - r_2$）是否显著进行差异显著性检验后，对这两个样本相关系数各自的总体 ρ_1 和 ρ_2 是否有差异进行推论的假设检验过程，这被称为**两总体相关系数差异显著性检验**。下面我们分述之。

一、单总体积差相关系数差异显著性检验

单总体相关系数的显著性检验即检验样本相关系数 r 和总体相关系数 ρ 之间是否有显著差异。样本相关系数 r 的分布随着总体相关系数 ρ 和样本容量 n 的大小而变化，当 ρ =0 时，样本相关系数 r 的分布是对称的，大样本时，r 服从正态分布；小样本时，r 服从 t 分布。而当 $\rho \neq 0$ 时，样本相关系数 r 的分布一般是偏态分布。因此在实际检验时，需要分 ρ =0 和 $\rho \neq 0$ 两种情况说明。

1. 虚无假设为 ρ =0 时

此时对样本相关系数 r 与 ρ 之间差异的检验可用 t 检验，即

$$t = \frac{r-\rho}{\sqrt{\dfrac{1-r^2}{n-2}}} \qquad (df = n-2) \qquad \text{（公式 9-4）}$$

若 $t > t_{\frac{\alpha}{2}(n-2)}$，则拒绝 H_0，说明所得到的 r 不是来自 ρ =0 的总体，即 r 是显著的；若 $t \le t_{\frac{\alpha}{2}(n-2)}$，则说明所得的 r 具有偶然性，即 r 是不显著的。

【例 9-3】 对 25 名儿童进行韦氏智力测验，得到其言语智力和操作智力得分，发现二者的相关系数 $r = 0.35$。试问这两种智力的相关是否显著？（α =0.05）

解：

（1）建立假设：

$$H_0: \rho = 0$$
$$H_1: \rho \neq 0$$

（2）　计算检验统计量，根据公式 9-4 得

$$t = \frac{r-\rho}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0.35-0}{\sqrt{\frac{1-0.35^2}{25-2}}} = 1.792$$

（3）确定显著性水平并做出决策：根据显著性水平$\alpha = 0.05$，自由度$df = 25-2 = 23$，查t分布表得，$t_{0.05/2(23)} = 2.069$。由于$t = 1.792 < t_{0.05/2(23)} = 2.069$，所以接受虚无假设$H_0$，说明这两种智力得分的相关不显著。

答：这两种智力得分之间的相关不显著。

值得注意的是，上例中的言语智力和操作智力得分相关不显著的结论只是一种统计决策，其含义是基于$n = 25$和$r = 0.35$的样本信息，尚无充分把握拒绝$\rho = 0$的虚无假设，即没有充分把握推论$\rho \neq 0$。这与理论上言语智力和操作智力的实际相关性没有必然关系。如果有充分证据支持言语智力和操作智力实际上是存在相关的，则表明本次抽样研究没能支持上述结论，其原因可能是样本容量小、样本代表性不强以及抽样误差等等。之所以在相关系数的计算中要求样本容量n不能太小（一般要求n不少于 30)，正是为了降低抽样误差的影响。

在实际应用中，通常是直接查积差相关系数（r）显著性临界值表（附表 7）来进行判断。如例 9-3 中，$r = 0.35$，$n = 25$，则可从附表 7 中查得当df =25-2=23 时，$\alpha = 0.05$所对应的临界值为 0.396。这表示当自由度为 23 时，r只有达到 0.396 时才能认为在$\alpha = 0.05$的显著性水平上总体相关$\rho \neq 0$。而例 9-3 中，r=0.35 < 0.396，因而可判断相关不显著。

2. 虚无假设为$\rho \neq 0$时

ρ=0 的这种情况在实际中用得较多，但它只是解释了r是否来自ρ=0 的总体。如果已知相关系数不为 0，而是某一数值，这时需要了解r是否来自ρ为某一特定值的总体，即当$\rho \neq 0$的显著性检验。

当$\rho \neq 0$时，r的样本分布是偏态的，但是经过一定转换后可得到费希尔Z_r值，Z_r值是服从正态分布的。因此如果将r与ρ都转换成费希尔Z_r（可查附表 8 获得)，则可用 Z 检验，即：

$$Z = \frac{Z_r - Z_\rho}{SE_{Z_r}} = \frac{Z_r - Z_\rho}{\sqrt{\frac{1}{n-3}}} \quad \text{（公式 9-5）}$$

【例 9-4】 根据相关研究，某学者估计双生子的智力相关系数为 0.70。为验证这一估计是否合理，今随机抽取双生子 50 对进行了韦氏智力测验，结果相关系数$r = 0.54$。试问实测结果与该学者的估计是否一致?。（$\alpha = 0.05$）

解：

(1)建立假设：

$$H_0: \rho = 0.70$$
$$H_1: \rho \neq 0.70$$

（2）计算检验统计量：查附表 8 得，$r = 0.54$ 时 $Z_r = 0.604$，$\rho = 0.70$ 时 $Z_\rho = 0.867$

根据公式 9-5 得

$$Z = \frac{Z_r - Z_\rho}{\sqrt{\frac{1}{n-3}}} = \frac{0.604 - 0.867}{\sqrt{\frac{1}{50-3}}} = \frac{-0.263}{0.146} = -1.80$$

（3）确定显著性水平并做出决策：根据显著性水平 $\alpha = 0.05$，查正态分布表得 $Z_{0.05/2} = 1.96$； 由于 $|Z| = 1.80 < Z_{0.05/2} = 1.96$，所以接受虚无假设 H_0。说明实测结果支持该学者的估计。

答：实测结果与该学者的估计一致。

二、两总体积差相关系数差异显著性检验

两总体相关系数差异的显著性检验指的是，如果两个样本相关系数来自两对性质相同的测量，要通过这两个样本相关系数的差异（$r_1 - r_2$）来推论其各自代表的总体相关系数 ρ_1 和 ρ_2 是否有差异。

假设 r_1 和 r_2 是分别由两组彼此独立的样本得到的积差相关系数，这时，必须先将 r_1 和 r_2 分别进行费希尔 Z_r 转化。由于 Z_r 的分布近似正态分布，则同样 $Z_{r1} - Z_{r2}$ 的分布仍为正态分布，此时假设检验的公式为

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}}$$

（公式 9-6）

【例 9-5】 在一项双生子研究中，17 对同卵双生子比耐智力测验得分的相关系数为 0.85，24 对异卵双生子比耐智力测验得分的相关系数为 0.76。假设抽样都是随机的，问这两个相关系数是否存在显著差异？（$\alpha = 0.05$）

解：根据题意得

（1）建立假设：

$$H_0: \rho_1 = \rho_2$$
$$H_1: \rho_1 \neq \rho_2$$

（2）计算检验统计量：查附表 8 得，$r_1 = 0.85$ 时 $Z_{r1} = 1.256$，$r_2 = 0.76$ 时 $Z_{r2} = 0.996$

根据公式 9-6 得

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{n_1 - 3} + \frac{1}{n_2 - 3}}} = \frac{1.256 - 0.996}{\sqrt{\frac{1}{17-3} + \frac{1}{24-3}}} = \frac{0.26}{0.345} = 0.754$$

（3）确定显著性水平并做出决策：根据显著性水平 $\alpha = 0.05$，查正态分布表得 $Z_{0.05/2} = 1.96$；因为 $Z = 0.754 < Z_{0.05/2} = 1.96$，所以接受虚无假设 H_0，说明两个相关系数不存在显著差异。

答：这两个相关系数之间不存在显著差异。

第三节 其它相关系数的假设检验

由于$\rho=0$条件下的相关系数显著性检验应用最多，因而对其它类型相关系数的显著性检验只介绍这种情况。

一、等级相关系数差异显著性检验

1. 斯皮尔曼等级相关系数的显著性检验

对斯皮尔曼等级相关系数进行显著性检验时，在利用实际观测值计算出样本相关系数后，直接用计算出的相关系数与一定α水平下的等级相关系数显著性临界值（附表9）比较，可确定该相关系数是否显著，或者说是否来自零相关总体。

2. 肯德尔W系数的显著性检验

（1）当$3\le N\le 7$时（N 表示被评定者的数目）：此时检验肯德尔 W 系数是否显著可直接查W系数临界值表（附表22）。具体做法是

首先利用公式9-7求W系数

$$W=\frac{12\sum R_i^2}{K^2N(N^2-1)}-\frac{3(N-1)}{N-1} \qquad \text{（公式 9-7）}$$

R_i表示评价对象获得的K个等级之和

N表示等级评定对象的数目

K表示等级者的数目

然后按照一定的显著性水平α及 K（K 为评定者的数目）和 N 的数目，在附表 22 中找到其对应的 W 系数临界值，用实际求得的 W 系数值与该临界值比较，然后作出决策。

（2）当$N>7$时： 此时用χ^2检验，检验的公式是

$$\chi^2=K(N-1)W \qquad df=N-1 \qquad \text{（公式 9-8）}$$

如果χ^2值达到一定显著性水平的临界值，说明W系数也达到了该显著性水平。

二、点二列相关系数的显著性检验

1. 采用两个平均数间差异的t检验法

即点二列相关系数计算公式（公式 3-22）中的$\overline{X_p}$和$\overline{X_q}$之间的差异进行 t 检验。如果差异显著，则表明点二列相关系数也显著，反之，则点二列相关系数也不显著。

2. 利用积差相关系数临界值表进行检验

在一定的 α 水平下，按 $df=n-2$ 查积差相关系数临界值表（附表 7），n 为求点二列相关时连续变量的数据个数。如果实测数据根据公式 3-22 所求得的相关系数 r_{pb} 大于积差相关系数显著性临界值表中临界值所对应的相关系数 r_{α}，则说明 r_{pb} 与零相关有显著差异，即 r_{pb} 是显著的。若 $r_{pb} \le r_{\alpha}$，则表明 r_{pb} 不显著。

3. 近似法 如果样本容量较大（n>50），也可以用下面近似的方法计算相关系数。

$\left|r_{pb}\right| > \dfrac{2}{\sqrt{n}}$ 时认为 r_{pb} 在 0.05 的水平上显著；$\left|r_{pb}\right| > \dfrac{3}{\sqrt{n}}$ 时认为 r_{pb} 在 0.01 的水平上显著。

第四节 列联表的独立性检验

一、独立性检验的含义与一般步骤

前三节所述有关相关系数的研究，其数据都属于计量数据。对于计数数据，除了在第八章所述需要进行拟合良度检验或同质性检验之外，有时也需要分析两类因素之间是相互独立还是彼此相关的。

独立性检验实际上是基于计数数据对因素间的相关关系进行研究。适用独立性检验的数据总是可以用双向分类列联表呈现。如果我们把事物同时按两个特征进行分类，就可以用双向分类列联表来整理计数数据。例如根据事物的某一特征分为两类，同时根据另一特征也分为两类，这样就可以用 2×2 列联表来整理数据。假如事物的某一特征不只包括两类，则可以用 $r\times c$ 列联表来整理数据（符号“r”表示行因素的属性类别数，符号“c”表示列因素的属性类别数）。独立性检验就是检验双向分类数据结构中两个分类特征或属性之间是相互独立还是彼此相关的检验方法。

独立性检验的一般步骤为：

（1）建立假设

H_0：两个分类变量独立无关。

H_A：两个分类变量有关。

（2）计算理论次数与 χ^2 统计量

在虚无假设成立（即两个分类变量无关）的条件下来计算每一单元格的理论次数。

（3）查表及作决断

根据自由度查 χ^2 值表得到临界 χ^2 值 $\chi^2_{0.05}$ 或 $\chi^2_{0.01}$，若 $\chi^2 \le \chi^2_{0.05}$ 或 $\chi^2_{0.01}$，表明实际次数与理论次数差异不显著，可以认为两个分类变量独立无关；若 $\chi^2 > \chi^2_{0.05}$ 或 $\chi^2_{0.01}$，

表明实际观察次数与理论次数差异显著，两个分类变量相关。

独立性检验与前一章所述的拟合良度检验和同质性检验即有联系又有区别。它们所用的检验方法基本相同，都是使用χ^2检验方法来达到推断统计的目的。但它们之间又有明显的差别：

（1）检验的目的不同。独立性检验是对同一样本的若干变量间关联情形的检验，目的在于判明数据资料是相互关联还是彼此独立。而同质性检验目的则在于检验不同人群母总体在某一个变量的反应是否具有显著差异。同质性检验是对几个样本在同一变量的分布状况的检验，是对几个样本数据是否同质作出统计决断。

（2）分组方法不同。独立性检验的计数数据是按两因素属性类别进行归组，而拟合良度检验的计数数据则是按某一因素的属性类别归组。

（3）计算理论次数的基础不同。独立性检验在计算理论次数时是在两因素相互独立的假设下进行计算。拟合良度检验则按已知的属性分类理论或学说计算理论次数。

（4） 约束条件不同。在拟合良度检验中确定自由度时，只有一个约束条件：各理论次数之和等于各实际次数之和，自由度为属性类别数减 1。而在 $r\times c$ 列联表的独立性检验中，共有 $r\times c$ 个理论次数，但受到以下条件的约束：（1）$r\times c$ 个理论次数的总和等于 $r\times c$ 个实际次数的总和；（2）r 个横行中，每一个横行理论次数总和等于该行实际次数的总和。但由于 r 个横行实际次数之和的总和应等于 $r\times c$ 个实际次数之和，因而独立的行约束条件只有 $r-1$ 个；（3）与此类似，独立的列约束条件有 $c-1$ 个。因而在进行独立性检验时，自由度为 $r\times c-1-(r-1)-(c-1)=(r-1)(c-1)$，即等于（横行属性类别数-1）×（竖列属性类别数-1）。

二、2×2 列联表的独立性检验

（一）2×2 列联表独立性检验的一般问题

1. 2×2 列联表的的含义和一般结构

2×2 列联表是指把样本数据同时按两种特征进行双向分类后，所构成的两行两列的四单元格数据列联表。其一般结构有如表 9-2 所示。

表 9-2　2×2 列联表一般结构

A	B		边缘小计
	B1	B2	
A1	a	b	$N_{A1}=a+b$
A2	c	d	$N_{A2}=c+d$
边缘小计	$N_{B1}=a+c$	$N_{B2}=b+d$	$N_T=a+b+c+d$

2. 2×2 列联表中四格理论次数的求取

在虚无假设成立（即两个分类变量无关）的条件下来计算每一单元格的理论次数。

在独立性检验中，当两个分类变量独立无关时，各单元格的理论次数为列联表中各个单元格对应的两个边缘次数的积除以总次数(表 9-3)。

表 9-3　　2×2 列联表中四格理论次数

A	B		边缘小计
	B1	B2	
A1	$N_{A1}*N_{B1}/N_T$	$N_{A1}*N_{B2}/N_T$	N_{A1}=a+b
A2	$N_{A2}*N_{B1}/N_T$	$N_{A2}*N_{B2}/N_T$	N_{A2}=c+d
边缘小计	N_{B1}=a+c	N_{B2}=b+d	N_T=a+b+c+d

（二）独立样本 2×2 列联表的独立性检验

独立样本 2×2 列联表的独立性检验中，当各单元格的理论次数 f_e ≥5 时，可以用计算 χ^2 统计量的基本公式（公式 8-7）计算 χ^2 值，此时各单元格的理论次数依照表 9-2 求取。也可以用化简基本公式而得到的简捷公式计算。简捷公式如下：

$$\chi^2=\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)} \quad \text{（公式 9-9）}$$

式中 a、b、c、d 的含义见表 9-1，N 为总人数。

2×2 列联表中，χ^2统计量的自由度 df =（2-1）×(2-1)=1。

【例 9-6】在一个关于幼儿从众行为的研究中，不同性别儿童从众回答人数与不从众回答人数见表 9-4，问性别和从众行为有关吗？

表 9-4　　儿童的性别和从众行为调查结果表

性别	回答		边缘小计
	从众回答	不从众回答	
男	203	197	400
女	173	227	400
边缘小计	376	424	800

解：

（1）建立假设

H_0：性别和从众行为无关。

H_A：性别和从众行为有关。

（2）计算理论次数与 χ^2 统计量

如果根据基本公式计算 χ^2 统计量，就要先求出各单元格的理论次数。

$$f_{ea}=\frac{400\times 376}{800}=188$$

$$f_{eb}=\frac{400\times 424}{800}=212$$

$$f_{ec}=\frac{400\times 376}{800}=188$$

$$f_{ed}=\frac{400\times 424}{800}=212$$

将计算得的各单元格的理论次数代入 χ^2 统计量的基本公式：

$$\chi^2=\sum\frac{(f_o-f_e)^2}{f_e}=\frac{(203-188)^2}{188}+\frac{(197-212)^2}{212}+\frac{(173-188)^2}{188}$$

$$+\frac{(227-212)^2}{212}=4.5162$$

如果根据简捷公式计算 χ^2 统计量，只要将各单元格数据和总人数代入公式即可：

$$\chi^2=\frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}=\frac{800(203\times 227-197\times 173)^2}{400\times 400\times 376\times 424}=4.5162$$

（3）查表及做决断

当 df=1 时，$\chi^2_{0.05}$=3.84，$\chi^2=4.52>\chi^2_{.05}=3.84$，因此拒绝虚无假设，即性别和从众行为有关。

答：性别和从众行为有关。

2×2 列联表的独立性检验也可以用比率差异显著性检验方法进行。对于例 9-6，可以先分别计算出男生、女生从众人回答的人数比率，然后根据本书介绍的两比率差异显著性检验的方法检验男女生从众回答人数比率差异是否显著，当差异显著时，推论性别和从众行为有关；当差异不显著时，推论性别和从众行为无关。因此，χ^2 独立性检验和比率差异显著性检验的目的是一样的。当碰到比率差异显著性检验时，也可以用 χ^2 独立性检验进行。

（三）相关样本 2×2 列联表的独立性检验

相关样本是同一批被试先后测试两次，或一一匹配的两批被试测试一次。相关样本 2×2 列联表的独立性检验公式为：

$$\chi^2=\frac{(A-D)^2}{A+D} \qquad \text{（公式 9-10）}$$

上式中，A、D 为 2×2 列联表中，两次实验或调查中分类项目不同的那两个单元格的实际次数，自由度为 1。

【例 9-7】 让 100 名学生先后接受两项运动技能达标测试，结果如下表 9-5 所示，问两个测验成绩之间有无关联？

表 9-5 100 名学生先后两次在同一测验上的测试结果表

测试 1	测试 2		边缘小计
	通过	未通过	
通过	55 （A）	5 （B）	60
未通过	15 （C）	25（D）	40
边缘小计	70	30	100

解：

（1）建立假设

H_0：两项运动技能达标测试无关。

H_A：两项运动技能达标测试有关。

（2）计算 χ^2 统计量

$$\chi^2=\frac{(A-D)^2}{A+D}=\frac{(55-25)^2}{55+25}=11.25$$

（3）查表及作决断

当 df=1 时，$\chi^2_{0.05}$=3.84，$\chi^2>\chi^2_{0.05}$，因此拒绝虚无假设，即两项运动技能达标测试的结果是相关的。

答：两项运动技能达标测试的结果是相关的。

本质上 χ^2 分布是个连续型随机变量的分布形式。按公式（8-6）算出的 χ^2 统计量只是近似分布，这种近似的条件是计算出的理论次数比较大。χ^2 检验中，当某分类理论次数小于 5 时，χ^2 统计量不能很好地满足 χ^2 分布，此时需要对 χ^2 统计量进行校正，统计学家亚茨（Frank Yates）提出了一个矫正公式，称为 χ^2 的连续性校正，其公式如下：

$$X^2=\sum\frac{(|f_0-f_e|-0.5)^2}{f_e} \qquad \text{（公式 9-11）}$$

尽管采用此方法校正后，$\chi 2$ 统计量能较为接近卡方分布，不过仍然建议在实际中最好增大样本的容量，尽量减少出现这种不大服从理论分布的情况。

当自由度为 1 时，即使期望次数不小于 5，按公式 8-7 计算 χ^2 统计量进行 χ^2 检验误差也比较大。因此，不少统计学家建议，当自由度为 1 时都进行连续性矫正较好。

（四）2×2 列联表的的品质相关

独立性检验能回答变量间是否有关联性的问题，进一步地，如果发现有关联的话，那么相关程度怎样呢？列联表的相关程度的度量称为品质相关。四格表的品质相关一般用 Φ 相关作指标，其计算公式为：

$$\Phi = \sqrt{\frac{X^2}{N}} = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad \text{（公式 9-12）}$$

Φ 相关系数依分子的正负号可取正负值。Φ 相关的公式实际上是根据四格表的χ^2值变换而来的，通过变换使得其取值大约在正负 1 之间，这样便于联系一般的相关系数的含义进行解释。χ^2值的大小反映了实际次数与理论次数之间差异的大小，而独立性检验中的理论次数是根据两变量独立的假设计算出来的，因此χ^2值的大小也就反映了两变量距独立有多远，离独立越远就越相关，可以说χ^2值本身就反映了两变量间相关的程度。由于所有的品质相关几乎都不是独立构造的，而都是对χ^2检验中χ^2统计量的变换。因此实际上，只要进行了χ^2独立性检验，这两步过程就一次解决了。计算品质相关系数只是为了更好地理解两变量间关系的密切程度。

三、$r \times c$ 列联表的独立性检验

（一）$r \times c$ 列联表的含义和一般结构

当样本数据同时按 A 和 B 两个特征进行双向多项分类时，我们可得到一个多行多列的表格。假定一批数据以特征 A 分类可分成 r 个小类，记为 $A_1, A_2 \cdots A_r$；以特征 B 分类又可分为 c 个小类，记为 $B_1, B_2 \cdots B_k$；则同时按 A 和 B 对数据进行分类的话，可构成 r 行 c 列的单元格数据，简称 $r \times c$ 列联表。其一般结构有如表 9-6 所示。

表 9-6　$r \times c$ 列联表的一般数据结构

特征 A	特征 B				边缘小计
	B_1	B_2	…	B_k	
A_1	n_{11}	n_{12}	…	n_{1k}	a_1
A_2	n_{21}	n_{22}	…	n_{2k}	a_2
…	…	…	…	…	…
A_r	n_{r1}	n_{r2}	…	n_{rk}	a_r
边缘小计	b_1	b_2	…	b_k	N

（二）$r \times c$ 列联表 χ^2 统计量及自由度的计算

对 $r \times c$ 列联表进行计算时，要先用公式 9-13 计算各单元格的理论次数。

$$f_{ij} = \frac{a_i b_j}{N} \tag{公式 9-13}$$

然后将计算的各单元格理论次数代入 χ^2 统计量基本计算公式来计算 $r \times c$ 列联表 χ^2 统计量：

$$\chi^2 = \sum_{(i,j)} \frac{(n_{ij} - f_{ij})^2}{f_{ij}} \tag{公式 9-14}$$

$r \times c$ 列联表 χ^2 统计量也有简捷公式：

$$\chi^2 = n\left(\sum_{(i,j)} \frac{n_{ij}^{\ 2}}{a_i b_j} - 1\right) \tag{公式 9-15}$$

$r \times c$ 列联表的自由度的计算方法见公式 9-16。

$$df = (r-1)(k-1) \tag{公式 9-16}$$

【例 9-8】 在一个关于幼儿从众行为的研究中，4、5、6 岁儿童从众回答人数与不从众回答人数见下表，问年龄和从众行为有关吗？

表 9-7　儿童的年龄和从众行为调查结果表

回答	年龄			边缘小计
	4 岁	5 岁	6 岁	
从众回答	249	231	145	625
不从众回答	151	169	255	575
边缘小计	400	400	400	1200

解：

（1）建立假设

H_0：年龄和从众行为无关。

H_A：年龄和从众行为有关。

（2）计算理论次数与 χ^2 统计量

如果根据基本公式（9-14 ）计算 χ^2 统计量，就要先求出各单元格的理论次数。

$$f_{11} = \frac{a_1 b_1}{N} = \frac{400 \times 625}{1200} = 208.3333$$

$$f_{21} = \frac{a_2 b_1}{N} = \frac{400 \times 625}{1200} = 208.3333$$

$$f_{31} = \frac{a_3 b_1}{N} = \frac{400 \times 625}{1200} = 208.3333$$

$$f_{12} = \frac{a_1 b_2}{N} = \frac{400 \times 575}{1200} = 191.6667$$

$$f_{22} = \frac{a_2 b_2}{N} = \frac{400 \times 575}{1200} = 191.6667$$

$$f_{32} = \frac{a_3 b_2}{N} = \frac{400 \times 575}{1200} = 191.6667$$

将计算得的各单元格的理论次数代入 χ^2 统计量的基本公式：

$$\chi^2 = \sum_{(i,j)} \frac{(n_{ij} - f_{ij})^2}{f_{ij}} = \frac{(145 - 208.3333)^2}{208.3333} + \frac{(231 - 208.3333)^2}{208.3333}$$

$$+ \frac{(249 - 208.3333)^2}{208.3333} + \frac{(255 - 19106667)^2}{191.6667} + \frac{(169 - 19106667)^2}{191.6667}$$

$$+ \frac{(151 - 19106667)^2}{191.6667} = 61.8941$$

如果根据简捷公式（10- 6）计算 χ^2 统计量，只要将各单元格数据和总人数代入公式即可：

$$\chi^2 = N\left(\sum_{(i,j)} \frac{n_{ij}^{\ 2}}{a_i b_j} - 1\right)$$

$$= 1200\left(\frac{145^2}{400 \times 625} + \frac{231^2}{400 \times 625} + \frac{249^2}{400 \times 625} + \frac{255^2}{400 \times 575} + \frac{169^2}{400 \times 575} + \frac{151^2}{400 \times 575} - 1\right)$$

$$= 61.8941$$

（3）查表及作决断

当 df=(3-1)(2-1)=2 时，$\chi^2_{0.05}$=5.99，$\chi^2 > \chi^2_{0.05}$，因此拒绝虚无假设，即年龄和从众行为有关。

答：年龄和从众行为之间有关联。

（三）$r \times c$ 列联表的品质相关

$r \times c$ 列联表的品质相关一般用 C 系数作指标。列联相关实际上是将 Φ 相关的适用情况从四格表扩展到一般的列联表。列联相关公式的来历也基本上与 Φ 相关相同，是根据列联表的 χ^2 值变换而来的。C 相关公式为：

$$C = \sqrt{\frac{X^2}{n + X^2}} \qquad \text{（公式 9-17）}$$

该系数的取值也在 0 和 1 之间，当双向分类划分的类型数目都较多时，C 才能接近 1。与使用 Φ 相关一样，使用列联相关之前，最好先检验两变量是否相关，只有两变量相关时，这一系数才有意义。

四、 SPSS 统计软件包在独立性检验中的应用

【例 7-5】利用 SPSS 软件包求取【例 9-6】数据的 χ^2 独立性检验结果。

解：在已打开的数据文件窗口中，将例 9-6 中的数据正确输入到 SPSS 数据框，其中第一列变量为“性别”（数字“1”代表男性，“2”代表女性），第二列变量为“回答类型”（数字“1”代表从众回答，“2”代表不从众回答），第三列变量为“人数”。然后按序点击图 9-1 所示的操作步骤可以求得与例 9-6 同样的 χ^2 独立性检验结果（即 χ^2=4.516）。

1. 数据→加权个案
2. 选中“加权个案”
3. 频率变量框：人数
4. $\boxed{\text{确定}}$
5. 分析→描述统计→交叉表
6. 行框：性别
7. 列框：回答类型
8. 统计量：（1）选择“卡方”
   （2）$\boxed{\text{继续}}$
9. $\boxed{\text{确定}}$

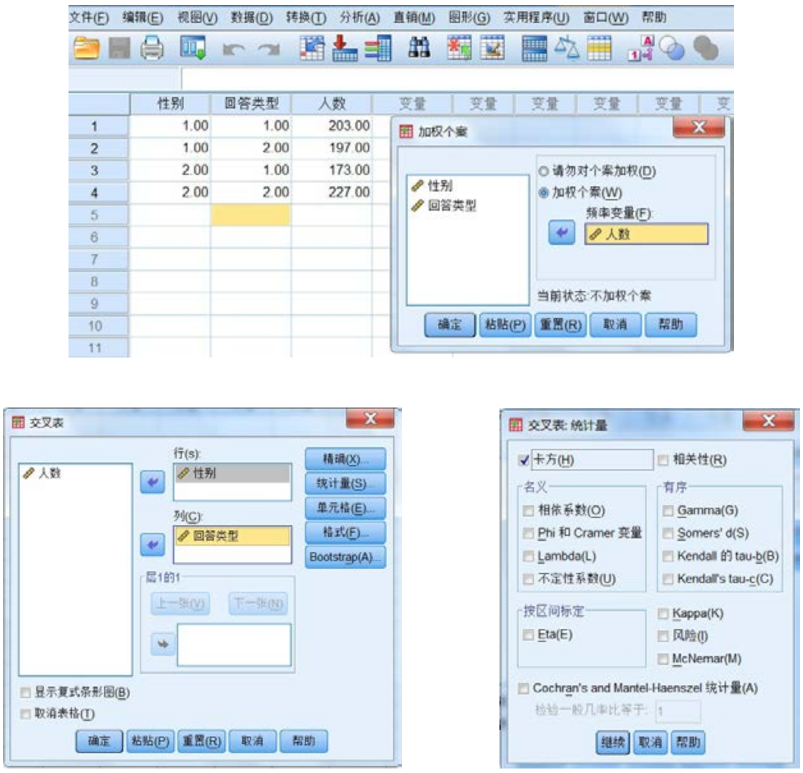


图 9-5 在 SPSS 中设置方差齐性检验分析程序各步骤示意图

小　结

本章论述了相关系数的区间估计和假设检验问题。

在各种相关系数中，积差相关是最重要和最常用的相关系数。总体的相关系数 ρ 为-1.00～1.00 之间的任意值。从总体中抽取 n 对数据，其相关系数 r 的抽样分布随两总体间的相关程度而异。

当用样本相关系数来估计总体相关系数的区间时，一般先将 r 转换成 Z_r，根据 Z_r 的抽样分布求出 Z_ρ 的置信区间，然后再通过查表或转换公式将 Z_ρ 的置信区间转换成总体相关系数 ρ 的置信区间。

单总体相关系数的显著性检验即检验样本相关系数 r 和总体相关系数 ρ 之间差异是否显著问题。两总体相关系数差异的显著性检验指的是，如果两个样本相关系数来自两对性质相同的测量，要通过这两个样本相关系数的差异（r_1-r_2）来推论其各自代表的总体 ρ_1 和 ρ_2 是否有差异。

独立性检验用来检验对事物的两个分类特征是相互独立还是彼此相关。其基本计算方法和同质性检验相同，只是检验目的不同。

当事物同时按两个分类特征来分类时，可以用双向分布列联表来整理数据。具体的数据结构可以分为 2×2 列联表和 $r\times c$ 列联表。在 2×2 列联表和 $r\times c$ 列联表中做 χ^2 检验需根据两分类特征独立无关的假设来计算每单元格的理论次数。χ^2 统计量的计算可以根据 χ^2 统计量基本公式计算，也可以根据适用于 2×2 列联表和 $r\times c$ 列联表的简捷公式计算。χ^2 检验只能说明两分类特征是否有关，如果有关，具体相关程度要用列联系数计算。

关键术语

总体相关系数的假设检验:是指通过样本相关系数推断总体相关系数的统计推断过程。

单总体相关系数的显著性检验：是指已知某个样本相关系数 r 和总体相关系数 ρ，要检验 r 和 ρ 之间的差异是否显著所进行的假设检验过程。

两总体相关系数差异的显著性检验：通过对两个样本相关系数的差异（r_1-r_2）是否显著进行差异显著性检验后，对这两个样本相关系数各自的总体 ρ_1 和 ρ_2 是否有差异进行推论的假设检验过程。

独立性检验：是用来检验分析两类因素是相互独立还是彼此相关的检验方法。

2×2 列联表：是指把样本数据同时按两种特征进行双向分类后，所构成的两行两列的四格数据列联表。

r×*c* 列联表：当样本数据同时按 A 和 B 两个特征进行双向多项分类时，我们可得到一个多行多列的表格。假定一批数据以特征 A 分类可分成 r 个小类，记为 $A_1, A_2 \cdots A_r$；以特征 B 分类又可分为 k 个小类，记为 $B_1, B_2 \cdots B_k$；则同时按 A 和 B 对数据进行分类的话，可构成 r 行 k 列的单元格数据，简称 $r \times c$ 列联表。

重要公式

$\rho \neq 0$ 时，样本相关系数分布的标准误：

$$\sigma_r = \frac{1-r^2}{\sqrt{n-1}}$$

$\rho = 0$ 时，样本相关系数分布的标准误：

$$\sigma_r = \frac{\sqrt{1-r^2}}{\sqrt{n-2}}$$

$\rho \neq 0$，n≥500 时的 Z 渐进正态分布的 Z_r 统计量：

$$Z = \frac{1}{2}\ln\frac{1+r}{1-r} \quad 或 \quad Z = 1.1513\lg\frac{1+r}{1-r}$$

$\rho = 0$ 时，单总体样本积差相关系数显著性检验的统计量：

$$t = \frac{r-\rho}{\sqrt{\frac{1-r^2}{n-2}}}$$

$\rho \neq 0$ 时，单总体样本积差相关系数显著性检验的统计量：

$$Z = \frac{Z_r - Z_\rho}{SE_{Z_r}} = \frac{Z_r - Z_\rho}{\sqrt{\frac{1}{n-3}}}$$

两总体样本积差相关系数显著性检验的统计量：

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{n_1-3} + \frac{1}{n_2-3}}}$$

独立样本 2 X 2 列联表的独立性检验统计量：

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$

相关样本 2 X 2 列联表的独立性检验统计量：

$$\chi^2=\frac{(A-D)^2}{A+D}$$

2 X 2 列联表的品质相关系数：

$$\Phi=\sqrt{\frac{X^2}{N}}=\frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}$$

r X c 列联表的 χ^2 统计量：

$$\chi^2=\sum_{(i,j)}\frac{(n_{ij}-f_{ij})^2}{f_{ij}}$$

***r X c* 列联表的品质相关系数：**

$$C=\sqrt{\frac{X^2}{n+X^2}}$$

思考与练习

1．某项目为测量自尊与主观幸福感之间的关系，选取 200 名被试接受两种问卷施测，计算两问卷相关分数 r =0.38，问此两种问卷总体相关系数 ρ 的 0.95 置信区间。

2. 某单位人力资源部门从该单位的在职员工中随机抽取 48 名员工，算出性别与薪水的相关为 0.25，问该单位员工的薪水与性别是否存在相关？

3. 在某年龄段女生中随机抽取 30 人进行 A、B 两种测验，两种测验成绩间的积差相关系数为 r=0.6；在同龄男生中随机抽取 36 人进行同样的两种测验，其成绩的相关系数为 r=0.48。能否推论男生和女生两种测验成绩间的相关程度不同？

4．有研究者想了解幼儿气质类型是否和性别有关，通过在幼儿园进行的随机抽样调查获得如下资料。研究者据此能得出什么结论？

性别	气质类型	
	容易型	困难型
男	48	22
女	36	34

5. 40 名学生进行 1000 米跑达标测试，先测一次，训练一个月后重测一次。两次测验均达标的有 14 人，初测达标而重测不达标的有 4 人；初测不达标而重测达标的有 16 人；两次均不达标的有 6 人。问训练是否有效果？

6．某项研究随机调查了男女生各 35 人，结果发现男生中喜欢体育、文娱和社团活动的人数分别为 15、8、12；而女生中喜欢这三种课外活动的人数分别为 8、13、14；试问男生和女生对课外活动的偏爱是否有显著差异？

7. 有研究者调查了某高校教师的职业倦怠情况。在调查样本中，讲师中有倦怠症状的 17 人，无倦怠症状的 15 人，副教授中有倦怠症状的 22 人，无倦怠症状的 16 人，教授中有倦怠症状的 6 人，无倦怠症状的 24 人。该校教师的职业倦怠水平是否与职称存在关联？

第十章 方差分析

学习本章内容，将有助于你对以下问题的理解与思考：

1. 方差分析的原理与步骤是什么？
2. 如何进行单因素完全随机设计的方差分析？
3. 涉及两个影响因素时该如何进行方差分析？
4. 其他常用的方差分析模型主要有哪些？

1． 有人研究了对个人表现的反馈类型对其自尊的影响。让 18 名被试参加一项知识测验，每组各 6 名被试。不管被试在测验中的实际表现如何，对积极反馈组，都告诉他们水平很高；对消极反馈组，都告诉他们表现很差；对控制组，不提供任何反馈信息。最后，让所有被试参加一个自尊测验，测验总分为 100 分，得到的分数越高，表明自尊越强。实验结果如表 10-1 所示，问不同反馈类型的各组被试的自尊水平是否存在显著差异？

表 10-1 不同反馈类型条件下被试自尊水平测试得分表

积极反馈组	控制组	消极反馈组
84.0	71.0	59.0
74.0	75.0	64.0
81.0	73.0	62.0
75.0	74.0	69.0
84.0	69.0	75.0
70.0	82.0	67.0

2． 有人想研究专业知识对记忆的影响。为此，他让被试记忆象棋棋盘上棋子的位置。被试共 20 人，一半是象棋新手，很少玩象棋，另一半被试是经验丰富的象棋大师。所有被试都会看到有 12 个棋子的棋盘。只不过，新手和专家各自有一半人看到的 12 个棋子，是专业水平比赛到中盘时的棋局；各自的另一半人看到的 12 个棋子，是人工随机摆放在棋盘上的。实验时，所有被试看棋局图片 2 分钟，然后移走图片，被试使用真实的棋盘和棋子进行复盘，成绩为被试将棋子正确放到它在棋局中的位置上的个数。实验结果如表 10-2 所示。请检验专业知识和棋局类型对记忆成绩是否有显著影响，二者是否存在交互作用。

表 10-2 专家与新手对不同棋局棋子位置的记忆成绩

棋局类型 / 人员类型	I（随机摆放）	II（比赛中盘）
I（新手）	2，4，4，3，4	8，1，2，4，4
II（专家）	1，5，6，3，6	9，9，11，11，11

第一节 方差分析的原理与步骤

一、方差分析的含义与基本条件

本书第五章曾指出，假设检验的目的在于检验不同组数据之间的差异主要是由抽样误差引起的，还是由实验所操纵的自变量引起的。当检验的对象是两组数据时，使用第五章和第六章所阐述的主要针对平均数差异进行的 Z 检验或 t 检

验方法就可以达到检验的目的。但是，心理学实验中经常有像本章开始时所列的两个问题那样，多于两个处理的实验设计，其中的问题 2 还是两个自变量的实验设计。

当检验的对象多于两个组时，需要检验的“总体性虚无假设（omnibus null hypothesis)”就成为“任何一对平均数之间是否有显著性差异”，比如，本章开始时的问题 1 中有三个实验分组，其虚无假设一般记为：$H_0: \mu_1 = \mu_2 = \mu_3$；其他两两之间的虚无假设如 $\mu_1 = \mu_2$ 或 $\mu_1 = \mu_3$ 等就称为“部分性虚无假设（parts null hypothesis)”。由于总体性虚无假设和部分性虚无假设之间差异很大，因此，第六章所论述的 Z 检验方法和 t 检验方法就难于完成这样的任务，这时，需要用方差分析的方法来进行差异显著性检验，因此，在一定意义上说，在检验目的方面方差分析是对两个平均数差异的假设检验向多个平均数差异的假设检验的延伸，从方法论上是 Z 检验方法和 t 检验方法的扩展。

方差分析（analysis of variance，ANOVA）就是使用 F 检验方法来检验若干个具有相同方差的正态总体的期望（即总体平均数）是否相等的一种假设检验方法（注意：所谓“相同方差”不是指不同组的方差完全一样，而是经方差齐性检验后没有显著差异）。方差分析的主要功能是分析实验数据中，不同来源的变异对总变异的贡献大小，从而确定实验中的自变量是否对因变量有重要影响。如果总体性虚无假设被拒绝，再通过事后检验方法来确定是哪两组之间存在显著差异。与其他统计方法一样，应用方差分析时也有一定的前提条件：

（1）正态性。即在同一条件下的试验结果是来自正态分布的一个样本。

（2）方差齐性。即指不同条件（处理）下的试验结果所来自的总体的方差相等。

（3）独立性。即观察对象是来自于各处理条件下的独立随机样本。

二、平方和的分解与方差分析的基本逻辑

（一）平方和的分解

在方差分析中，我们常以所有实验数据与总平均数的离均差平方和（简称平方和）作为总变异的统计量。平方和（sum of square)，以符号 SS 表示；用 SS_T 表示总变异，称为总平方和（the sum of squares total）。即

$$SS_T = \sum_{j=1}^{K}\sum_{i=1}^{n}(y_{ij} - \bar{Y})^2 \qquad \text{（公式 10-1）}$$

各组平均分与总平均分的加权离均差平方和就反映了不同处理带来的变异，称为组间平方和（sum of squares between groups，记为 SS_B）。即

$$SS_B = n \cdot \sum_{j=1}^{K}(\bar{Y}_j - \bar{Y})^2 \qquad \text{（公式 10-2）}$$

将各组内的实验数据与各组平均数的离均差平方和相加，称为组内平方和（sum of squares within group，记为 SS_W）。即

$$SS_w = \sum_{j=1}^{K}\sum_{i=1}^{n}(y_{ij}-\overline{Y}_j)^2 \qquad \text{（公式 10-3）}$$

由于 $y_{ij}-\overline{Y}=(y_{ij}-\overline{Y}_j)+(\overline{Y}_j-\overline{Y})$，所以

$$\sum_{j=1}^{k}\sum_{i=1}^{n}(Y_{ij}-\overline{Y})^2=\sum_{j=1}^{k}\sum_{i=1}^{n}\left[\left(Y_{ij}-\overline{Y}_j\right)+\left(\overline{Y}-\overline{Y}\right)\right]^2$$

$$=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(Y_{ij}-\overline{Y}_j\right)^2+2\cdot\sum_{j=1}^{K}\sum_{i=1}^{n}\left(Y_{ij}-\overline{Y}_j\right)\left(\overline{Y}_j-\overline{Y}\right)+\sum_{j=1}^{K}\sum_{i=1}^{n}\left(\overline{Y}_j-\overline{Y}\right)^2$$

（公式 10-4）

由于离均差的和等于零这一特性，（公式 10-4）等式右边的中间项就等于零，（公式 10-4）就可以简化为：

$$\sum_{j=1}^{k}\sum_{i=1}^{n}(Y_{ij}-\overline{Y})^2=\sum_{j=1}^{k}\sum_{i=1}^{n}\left(Y_{ij}-\overline{Y}_j\right)^2+\sum_{j=1}^{K}\sum_{i=1}^{n}\left(\overline{Y}_j-\overline{Y}\right)^2 \qquad \text{（公式 10-5）}$$

即：

$$SS_T = SS_W + SS_B \qquad \text{（公式 10-6）}$$

这样，总变异就被分解为组内变异和组间变异两个部分。

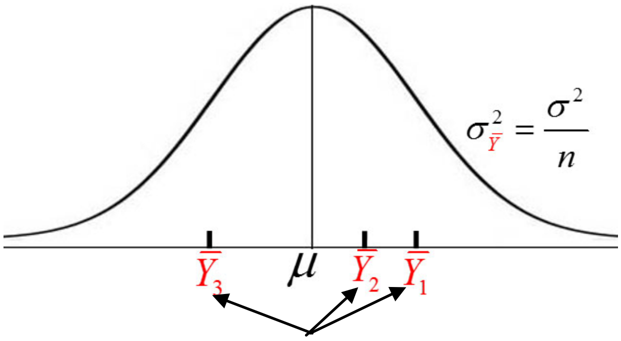
（二）方差分析的基本逻辑

通过平方和的分解，我们将总变异分解为组内变异和组间变异两个部分。组内变异则是由组内各被试得分的差异范围决定的，主要指由实验误差和组内被试间的个体差异造成的变异。组间变异反映的是由于接受不同的实验处理而造成的的差异，各组的平均数差异越大，组间变异也就越大。组间变异可以看做组间平均数差异大小的一个指标。如果三个总体平均数相等，我们可以期望三个样本平均数很接近。事实上，三个样本平均数相互越接近，我们推断总体平均数相等的证据就越充分。相对应地，样本平均数差异越大，我们推断总体平均数不等的证据就越充分。

那么，组间变异大到多少才能说明处理效应存在显著差别呢？为此，我们以本章开头的问题 1 为例。先假设各组总体平均数相等（即 H_0：$\mu_1=\mu_2=\mu_3$），那么，样本平均数的变异性“小”，则支持 H_0；如果样本平均数的变异性“大”，则不支持 H_0。在此假设下我们作如下几方面的分析

（1）如果 H_0：$\mu_1=\mu_2=\mu_3$ 为真，结合 σ^2 齐性（方差分析的前提条件），则每一个样本都是来自平均数为 μ、方差为 σ^2 的同一总体。在第五章我们曾讲过，来自正态总体的样本容量为 n 的一个简单随机样本的样本平均数 $\overline{Y}$ 的抽样分布服从

正态分布，其平均数为 μ，方差为σ^2/n。图 10-1 用图示可以说明这一分布。



当 H_0 为真时，因为只有一个抽样分布，所以几个样本平均数都“比较接近”

图 10-1　H_0 为真时 $\bar{Y}$ 的抽样分布

于是，根据本章开头的问题 1 的数据（见表 10-1）所计算的三个样本平均数 $\bar{Y}_1=78$、$\bar{Y}_2=74$ 和 $\bar{Y}_3=66$，可以被视为来自图 10-1 所示的抽样分布的值。每一个样本平均数 $\bar{X}$ 都是总体平均数 μ 的点估计量，但显然，如果用三个样本平均数 $\bar{X}$ 的平均数作为总体平均数 μ 的估计值要优于单个样本平均数。于是，$\bar{X}$ 抽样分布的平均数$\dfrac{78+74+66}{3}=72.67$ ，该统计量可称为总样本平均数。$\bar{Y}$ 抽样分布的方差$\sigma_{\bar{Y}}^2$ 可由三个样本平均数的方差给出：

$$S_{\bar{Y}}^2=\frac{(78-72.67)^2+(74-72.67)^2+(66-72.67)^2}{3-1}=37.33$$

由 $\sigma_{\bar{X}}^2=\sigma^2/n$，可得 $\sigma^2=n\sigma_{\bar{X}}^2$。由于 $\sigma_{\bar{X}}^2$ 实际上无法获得，只能用 $S_{\bar{X}}^2$ 作为其估计值：

$$\hat{\sigma}^2=nS_{\bar{X}}^2=6\times 37.33=224$$

式中，$\hat{\sigma}^2$ 是 σ^2 的估计量，所得结果 $nS_{\bar{X}}^2=258$ 称为 σ^2 **的组间估计**。

假设 H_0：$\mu_1=\mu_2=\mu_3$ 为假时，且三个总体平均数都不等时，那么三个样本平均数分别来自平均数不同的正态总体，因此必然有三个抽样分布。图 10-2 表明在这种情形下，样本平均数不再像 H_0 为真时那么接近了，所以 $S_{\bar{Y}}^2$ 将会变大，从而 σ^2 的组间估计变大。也就是说，当 H_0 不成立时，组间估计将会高估总体方差 σ^2。

（2）由于方差分析前提条件中有方差齐性的要求，因此，每个样本方差都给出了 σ^2 的一个无偏估计，于是我们可以将 σ^2 的个别估计组合或合并成一个总的估计。这种方法得到的 σ^2 的估计值称为 σ^2 的合并或**组内估计**。

因为每个样本方差给出的 σ^2 估计仅与每个样本内部变异有关，故 σ^2 的组内估计不受总体平均数是否相等的影响。若样本容量相等，σ^2 的组内估计可能通过计算各个样本方差的算术平均数得到。以本章开始时的问题 1 为例，经计算可得

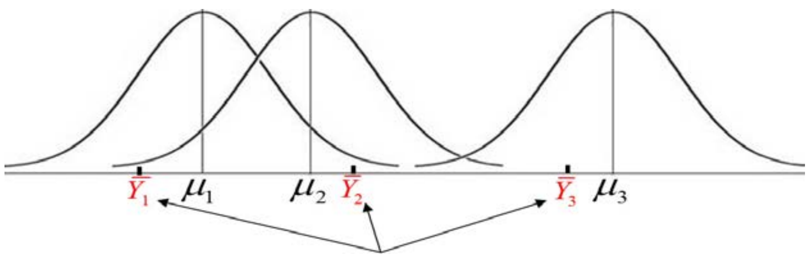


图 10-2　H_0 为假时 $\bar{Y}$ 的抽样分布

到：三个样本方差分别为：$S_1^2=34$，$S_2^2=20$，$S_3^2=32$，因此有：

$$\hat{\sigma}^2=\frac{34+20+32}{3}=28.67$$

（3）当虚无假设 H_0 为真时，总体方差 σ^2 的组内估计值和组间估计值实质上反映的是相同的误差，即都是由个体差异等随机误差带来的变异，如表 10-3 所示。在这种条件，总体方差的组间估计是 σ^2 的比较好的估计，组内估计也是 σ^2 的比较好的估计，因此，组间估计值与组内估计值之比将近似等于 1（近似，是因为二者都是估计值）。

当虚无假设 H_0 为假，而研究假设为真时，组间估计包含了两个变异来源：一是由个体差异等随机误差带来的变异，另一是由处理效应带来的变异。而组内估计仍只反映了个体差异等随机误差带来的变异。见表 10-3。在这种条件下，组间估计得到的 σ^2 估计值就会偏大，而组内估计仍是 σ^2 估计值的较好估计。因此，组间估计值与组内估计值之比就会大于 1。如果认为这个比值显著大于 1，我们就能够拒绝虚无假设。那么，如何判定这个比值是否显著大于 1 呢？

表 10-3　组内和组间方差估计的变异来源

	由个体差异等随机因素带来的变异	由处理效应带来的组间变异
虚无假设 H_0 为真		
组内估计	√	
组间估计	√	
虚无假设 H_0 为假		
组内估计	√	
组间估计	√	√

总体方差 σ^2 的组间估计值与组内估计值之比是两个方差之比，在第七章我们曾讲过，方差之比服从 F 分布。因此，我们将这个比值放到 F 分布中予以考察，就能判定这个比值是否显著大于 1，从而作出相应的统计决策。

三、方差分析的一般步骤

方差分析最重要的目的就是要得到总体方差σ^2的组间估计值与组内估计值之比的 F 值，并将其放到 F 分布中予以考察，以确定各处理对因变量的影响是否存在显著差异。一般可以把方差分析过程分为如下六步：

（一）方差分析前提条件的检验

前面所述方差分析有三个前提条件（正态性、独立性、方差齐性），这些应用条件得不到满足时，会对方差分析结果产生一定的影响，因此，方差分析之前应该检验前提条件是否满足。一般情况下，只要是严格地随机抽样，随机分派被试，独立性是能够得到保证的；正态性检验可以通过峰度、偏度检验，x^2检验，P-P 图和 Q-Q 图等统计图检验，以及有关的非参数检验来进行。正态性得不到满足时，方差分析的结论并不会受到太大的影响。在实际进行方差分析时，主要是要对方差齐性可以通过专门的方差齐性检验进行。

（二）建立假设

在方差分析过程中，对于不同的统计模型，或模型的不同表达方式，假设的形式不同。一般而言，单因素方差分析的假设形式可以是：

$$\begin{cases} H_0: \mu_1 = \mu_2 = \cdots = \mu_K \\ H_a: \mu_1, \mu_2, \cdots, \mu_K \text{不全相等} \end{cases} \qquad \text{（假设 10-1）}$$

（三）计算平方和、自由度、均方及F值

单因素方差分析中总平方和SS_T的计算公式见公式 10-1 和公式 10-6，组间平方和SS_B的计算见公式 10-2、组内平方和SS_W的计算公式见公式 10-3。

单因素方差分析的自由度包括总的自由度df_T、组间变异的自由度df_B和组内变异的自由度df_W，总自由度等于组间自由度与自由度之和。其计算公式分别为：

$$df_T = n_T - 1 \qquad \text{（公式 10-7）}$$

$$df_B = K - 1 \qquad \text{（公式 10-8）}$$

$$df_W = n_T - K \qquad \text{（公式 10-9）}$$

式中n_T表示所有被试人数，即$n_T = \sum_{i=1}^{K} n_i$；K 是自变量的水平数。

在方差分析中，常常使用均方（mean square 或 MS）来作为对总体方差的估计量，计算方法为平方和除以相应的自由度。组间均方MS_B和组内均方MS_W的计算公式分别为：

$$MS_B = \frac{SS_B}{df_B} \qquad \text{（公式 10-10）}$$

$$MS_W = \frac{SS_W}{df_W}$$ （公式 10-11）

式中， df_B、df_W 分别是组间自由度和组内自由度。

在均方的基础上，我们可以得到所需要的 F 检验统计量：

$$F = \frac{MS_B}{MS_W}$$ （公式 10-12）

（四）统计决策

方差分析的 F 检验统计量在 α 显著性水平上的拒绝域为

$$F \ge F_{\alpha,(df_1,df_2)}$$ （公式 10-13）

式中，$F_{\alpha,(df_1,df_2)}$ 是根据 α、df_1 和 df_2 查 “F 分布表”而得到的临界值；df_1 是组间自由度 df_B（分子自由度），df_2 是组内自由度 df_W（分母自由度）。

当 F 统计量大于临界值时，则拒绝 H_0，说明不同的实验处理效应存在显著差异。如果是在 α＝.05 水平上拒绝 H_0，则在 F 统计量的右上角标上一个“*”号；如果是在 α＝.01 水平上拒绝 H_0，则在 F 统计量的右上角标上两个“*”号。

当 F 统计量小于临界值时，则无充分理由拒绝 H_0，这表明，数据的总变异由分组造成的变异只占很小的部分，大部分是是由实验误差和个体差异所致，也就说明不同的实验处理效应不存在显著差异。

（五）列方差分析表

方差分析的第（3）、（4）步完成后，可以将其归纳成一个方差分析表。所谓**方差分析表（analysis of variance table）是**一种用来汇总方差分析计算和结果的表，包括显著变异来源、平方和、自由度、均方和 F 值等列。其一
般格式（以单因素方差分析为例）有如表 10-4 所示：

表 10-4 单因素完全随机设计方差分析表

变异来源	平方和	自由度	均方	F
组间	SS_B	$df_B = K - 1$	$MS_B = \frac{SS_B}{df_B}$	$F = \frac{MS_B}{MS_W}$
组内	SS_W	$df_W = n_T - K$	$MS_W = \frac{SS_W}{df_W}$	
总变异	SS_T	$df_T = n_T - 1$		

（六）方差分析 F 检验之后的步骤

在方差分析 F 检验之后，根据实际情况，还有一些后续的工作要做。主要包括：

（1）方差分析的效应大小和统计检验力。

（2）平均数的逐对比较。

（3）参数估计。也就是根据某处理的样本平均数，估计该处理来自的总体

的平均数。

这三个方面的工作相对较为复杂，依据统计模型不同，其方法也不同，因此，这些内容在后文的相应模型的方差分析中予以介绍。

第二节　单因素完全随机设计的方差分析

一、设计思想与模型

所谓完全随机设计（complete random design）是指处理被随机地指派给实验单元的一种实验设计。

单因素完全随机设计是这样一种设计：在实验或观察中只有一个因素 A，假定因素 A 有 K 个水平（K>2），在 A_i 水平下获得 n_i 个被试的反应值（因变量的值）y_{i1}，y_{i2}，…，y_{in_i}，这是来自总体 $N(\mu_i, \sigma^2)$ 的一个样本，i=1, 2, …, K。这种设计将被试随机分成 K 组，每组随机接受一种实验处理。由于各组是不同的被试，因此这种设计也称为被试间设计或组间设计。该设计的数据模型如表 10-5 所示：

表 10-5　单因素完全随机设计的数据模型

处理 1	处理 2	…	处理 K
y_{11}	y_{21}	…	y_{K1}
y_{12}	y_{22}	…	y_{K2}
…	…	…	…
y_{1n_1}	y_{1n_2}	…	y_{1n_K}

表中的 n_1，n_2，…，n_k可能都相等，也可能不全相等。

该设计模型有两种表述方式，因此假设方式也有两种，但这两种方式实质上是等价的。

1. 平均数模型

$$\begin{cases} y_{ij} = \mu_i + \varepsilon_{ij}\ ,\quad i = 1, 2, \cdots, K\ ,\quad j = 1, 2, \cdots, n_i \\ \text{各}\varepsilon_{ij}\text{相互独立，均服从}N(0,\sigma^2)\text{ 的正态分布} \end{cases} \qquad \text{（公式 10-14）}$$

要检验的假设就是假设 10-1，即

$$\begin{cases} H_0 : \mu_1 = \mu_2 = \cdots = \mu_K \\ H_a : \mu_1, \mu_2, \cdots, \mu_K\text{不全相等} \end{cases} \qquad \text{（假设 10-1）}$$

2. 主效应模型

如果记 $\mu = \frac{1}{n_T}\sum_{i=1}^{K} n_i\mu_i$ ，称为**一般平均**；又记 $a_i = \mu_i - \mu$ ，称其为因素 A 的

第 i 水平的效应，i=1, 2, …, K；并且 $\sum_{i=1}^{K} n_i a_i = 0$；其中 $n_T = \sum_{i=1}^{K} n_i$ 。此时，模型可表示为

$$\begin{cases} y_{ij} = \mu + a_i + \varepsilon_{ij}\ , \quad i = 1, 2, \cdots, K\ , \quad j = 1, 2, \cdots, n_i \\ \sum_{i=1}^{K} n_i a_i = 0 \\ \text{各}\varepsilon_{ij}\text{相互独立，均服从}N(0, \sigma^2)\text{的正态分布} \end{cases} \qquad \text{（公式 10-15）}$$

要检验的假设为

$$\begin{cases} H_0 : a_1 = a_2 = \cdots = a_K = 0 \\ H_a : a_1, a_2, \cdots, a_K \text{不全为} 0 \end{cases} \qquad \text{（假设 10-2）}$$

若拒绝 H_0 则称因素 A 显著，否则称因素 A 不显著。

二、计算实例

（一）各实验处理组样本容量相等（等重复）

【例 10-1】现在，我们可以根据方差分析的基本原理和一般步骤来对本章开头的问题 1 进行假设检验了。经计算可得：三个样本平均数分别为：$\overline{Y}_1 = 79$、$\overline{Y}_2 = 74$ 和 $\overline{Y}_3 = 66$；三个样本方差分别为：$S_1^2 = 34$，$S_2^2 = 20$，$S_3^2 = 32$。（本例题原始数据见本书所附光盘中例 10-1.sav）

解：

（1）用公式 7-6 对三组方差进行方差齐性检验（α=.05）

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2$$
$$H_a : \sigma_1^2, \sigma_2^2, \sigma_3^2 \text{不全相等}$$

$$F_{\max} = \frac{S^2_{\text{最大}}}{S^2_{\text{最小}}} = \frac{S_1^2}{S_2^2} = \frac{34}{20} = 1.7$$

查附表 6 得，$F_{\max,.05(3,5)} = 10.8$

因为 $F_{\max} = 1.7 < F_{\max(.05)(3,5)} = 10.8$，因此没有充分理由拒绝 H_0，即认为各组方差是齐性的。

用 SPSS 软件包进行方差齐性检验的方法可参见第七章，结果也是各组方差齐性。

（2）建立假设

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_a : \mu_1, \mu_2, \mu_3 \text{不全相等}$$

（3）计算平方和、自由度、均方及 F 值

利用公式 10-1、公式 10-2 和公式 10-3 计算各种平方和：

$$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij} - \overline{Y})^2 = (84-72.67)^2 + \cdots + (67-72.67)^2 = 878$$

$$SS_B = n \cdot \sum_{j=1}^{k}(\overline{Y}_j - \overline{Y})^2 = 6 \times [(78-72.67)^2 + (74-72.67)^2 + (66-72.67)^2 = 448$$

$$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{n}(y_{ij} - \overline{Y}_j)^2$$

$$= [(84-78)^2 + \cdots + (70-78)^2] + [(71-74)^2 + \cdots + (82-74)^2]$$

$$+ [(59-66)^2 + \cdots + (67-66)^2]$$

$$= 170 + 100 + 160 = 430$$

且：$SS_T = SS_W + SS_B$（878＝448+430）

利用公式 10-7、公式 10-8 和公式 10-9 计算各种自由度：

$df_T = n_T - 1 = 15 - 1 = 14$； $df_B = K - 1 = 3 - 1 = 2$； $df_W = n_T - K = 15 - 3 = 12$

利用公式 10-10 和公式 10-11 计算组间均方和组内均方：

$$MS_B = \frac{SS_B}{df_B} = \frac{448}{3-1} = 224\text{；} \quad MS_W = \frac{SS_W}{df_W} = \frac{430}{18-3} = 28.67$$

利用公式 10-12 计算我们所需要的 F 检验统计量：

$$F = \frac{MS_B}{MS_W} = \frac{224}{28.67} = 7.814$$

（4）统计决策

查附表 5-B 的 F 值表得，$F_{0.01,(2,12)} = 6.93$，因为 $F = 7.814 > F_{0.01(2,12)} =$ 6.93，所以在 $\alpha = 0.01$ 水平上拒绝 H_0，即认为不同反馈类型对被试的自尊水平存在显著影响，这样下结论犯错误的概率不超过 0.01。

（5）列方差分析表

表 10-6 不同反馈类型是否影响被试自尊水平方差分析表

变异来源	平方和	自由度	均方	F
组间	448	2	224	7.814^{**}
组内	430	15	28.67	
总变异	878	17		

（6）方差分析 F 检验之后的步骤

有关本例平均数之间的逐对比较问题我们在本章后面再讨论，统计检验力评估见下一章。

答：不同的训练方法对初中学生的阅读能力具有不同的影响作用。

（二）各实验处理组样本容量不同（不等重复）

【例 10-2】 为了比较四种不同的计算机辅助教学方案，研究人员将学生随机分为四组，每组接受一种计算机辅助教学方案，其他的教学条件均相同。经过两个月的教学，测验成绩如表 10-7 所示。问这四种不同教学方案对测验成绩是否有显著影响？（本例题原始数据见本书所附光盘中例 10-2.sav）

表 10-7　四种计算机辅助教学方案的测验成绩

	I	II	III	IV
	30	50	18	88
	74	38	56	78
	46	66	34	60
	58	62	24	76
	62	44	66	
	38	58	52	
		80		
$\overline{Y}_i$	51.33	56.86	41.67	75.5
S_i^2	266.67	202.48	367.07	134.33

解：

（1）**方差分析前提条件的检验**（α＝.05）

根据公式 10-3 可计算得组内平方和 $SS_W = 4786.57$

根据公式 10-11 可计算得 $MS_E(MS_W) = 4786.57/19 = 251.92$

利用公式 7-7 和公式 7-8 对本例中四个分组的方差齐性问题进行检验：

$$H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2$$

$$H_a : \sigma_1^2, \sigma_2^2, \sigma_3^2, \sigma_4^2 \text{不全相等}$$

$$c = \frac{1}{3(K-1)}\left(\sum_{i=1}^{K}\frac{1}{n_i-1} - \frac{1}{df_E}\right)+1$$

$$= \frac{1}{3\times(4-1)}\left(\frac{1}{6-1}+\frac{1}{7-1}+\frac{1}{6-1}+\frac{1}{4-1}-\frac{1}{19}\right)+1 = 1.09$$

$$\chi^2 = \frac{1}{c}\left[df_E \ln MS_E - \sum_{i=1}^{K}(n_i-1)\ln S_i^2\right]$$

$$= \frac{1}{1.09}[19\times\ln 251.92 - (5\times\ln 266.67 + 6\times\ln 202.48 + 5\times\ln 367.07 + 3\times\ln 134.33)]$$

$$= 0.94$$

因为 $\chi^2 = 0.94 < \chi^2_{.05,3} = 7.81$，因此没有充分理由拒绝 H_0，即认为各组方差是齐

性的。

（2）建立平均数差异显著性检验的假设

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4$$
$$H_a: \mu_1, \mu_2, \mu_3, \mu_4 \text{不全相等}$$

（3）计算平方和、自由度、均方及 F 值

$$SS_T = n_T \sigma_T^2 = 23 \times 18.2219^2 = 7636.87$$

$$SS_W = \sum_{j=1}^{k} (n_j - 1) S_j^2 = 5 \times 266.67 + 6 \times 202.48 + 5 \times 367.07 + 3 \times 134.33 = 4786.57$$

$$SS_B = SS_T - SS_W = 7636.87 - 4786.57 = 2850.3$$

$df_T = n_T - 1 = 23 - 1 = 22$ ； $df_B = K - 1 = 4 - 1 = 3$ ； $df_W = n_T - K = 23 - 4 = 19$

$$MS_B = \frac{SS_B}{df_B} = \frac{2850.3}{3} = 950.1 \quad ; \quad MS_W = \frac{SS_W}{df_W} = \frac{4786.57}{19} = 251.92$$

$$F = \frac{MS_B}{MS_W} = \frac{950.1}{251.92} = 3.77$$

（4）统计决策

查附表 5-A 的 F 值表得，$F_{0.05,(3,19)} = 3.13$，因为 $F = 3.77 > F_{0.05,(3,19)}$，所以在 $\alpha = 0.05$ 水平上拒绝 H_0，即认为不同教学方案对成绩有显著影响，这样下结论犯错误的概率不超过 0.05。

（5）列方差分析表

表 10-8 四种计算机辅助教学方案的方差分析表

变异来源	平方和	自由度	均方	F
组间	2850.30	3	950.10	3.77*
组内	4786.57	19	251.92	
总变异	7636.87	22		

答：这四种不同教学方案对测验成绩有显著影响

（三）利用样本统计量进行方差分析

【例 10-3】 某研究者研究了五种训练方案对训练成绩的影响，最后测试结果如表 10-9 所示。问五种训练方案的效果是否存在显著差异？

表 10-9　五种训练方案的训练成绩情况

训练方案	I	II	III	IV	V
样本容量	5	5	5	5	5
样本平均数	11.5	14.8	7.6	19.4	18.0
样本方差	22.2	18.6	20.8	19.4	27.5

解：

（1）方差分析前提条件的检验（α＝0.05）

$$F_{\max}=\frac{S^2_{最大}}{S^2_{最小}}=\frac{S_1^2}{S_2^2}=\frac{27.5}{18.6}=1.48$$

查附表 6 得，$F_{\max(0.05)(4,4)}=20.6$

因为 $F_{\max}=1.48<F_{\max(0.05)(4,4)}=20.6$，因此没有充分理由拒绝 H_0，即认为各组方差是齐性的。

（2）建立假设

$$H_0:\mu_1=\mu_2=\mu_3=\mu_4=\mu_5$$

$$H_a:\mu_1,\mu_2,\mu_3,\mu_4,\mu_5 不全相等$$

（3）计算平方和、自由度、均方及 F 值

$$\bar{Y}=\frac{11.5\times5+14.8\times5+7.6\times5+19.4\times5+18\times5}{5\times5}=14.26$$

$$SS_B=n\cdot\sum_{j=1}^{K}(\bar{Y}_j-\bar{Y})^2$$

$$=5\times[(11.5-14.26)^2+(14.8-14.26)^2+(7.6-14.26)^2+(19.4-14.26)^2+(18-14.26)^2]=463.36$$

$$SS_w=\sum_{j=1}^{K}(n_j-1)S_j^2=(5-1)\times(22.2+18.6+20.8+19.4+27.5)=434$$

$$SS_T=SS_B+SS_W=463.36+434=897.36$$

$df_T=n_T-1=5\times5-1=24$ ；　$df_B=K-1=5-1=4$ ；　$df_W=n_T-K=25-5=20$

$$MS_B=\frac{SS_B}{df_B}=\frac{463.36}{4}=115.84\quad；\quad MS_W=\frac{SS_W}{df_W}=\frac{434}{20}=21.7$$

$$F=\frac{MS_B}{MS_W}=\frac{115.84}{21.7}=5.34$$

（4）列方差分析表

表 10-10　五种训练方案的方差分析表

变异来源	平方和	自由度	均方	F
组间	463.36	4	115.84	5.34^{**}
组内	434	20	21.7	
总变异	897.36	24		

查附表 5-B 的 F 值表得，$F_{0.01,(4,20)}=4.43$，因为 $F=5.34>F_{0.01(4,20)}=4.43$，所以在 α＝0.01 水平上拒绝 H_0，即认为不同训练方案对训练成绩有显著影响，这样下结论犯错误的概率不超过 0.01。

答：五种训练方案的效果是否存在显著差异

三、平均数之间的多重比较

当因素 A 效应显著时，只是表明总体平均数不全相等，但并未说明究竟是哪些平均数之间不等，为此，常常需要进一步对一切 $i \neq j$ 同时检验如下假设：

$$H_0 : \mu_i = \mu_j \qquad i \neq j$$
$$H_1 : \mu_i \neq \mu_j$$

这个检验称为平均数间的**多重比较**。换言之，所谓**多重比较（multiple comparison）是**指在方差分析结果表明处理效应存在显著差异时，对成对的总体平均数进行统计检验的过程。

因为进行平均数间的多重比较时，需分别进行 C_K^2 次检验。如果对 K 个（K>2）以上的平均数反复进行 t 检验，在原定临界值 t_α 不变的情况下，差异较大的一对平均数犯 α 型错误的概率增大为

$$P_K = 1-(1-\alpha)^{C_K^2} \qquad \text{（公式 10-16）}$$

因此，进行平均数间的多重比较必须采用专用的两两比较的方法，而不能用 t 检验。

一般而言，可以把多重比较分为两种类型：计划好的和非计划的。计划好的多重比较（Planned Comparisons），是指在收集数据之前便决定了要通过多重比较来考察多个组与某个特定组之间的差异，或者是某几个特定组间彼此的差异。非计划的多重比较（Unplanned Comparisons，Post-hoc Comparisons）是在方差分析得到有显著意义的 F 值后才有必要进行的，是一种探索性的分析。

目前，关于多重比较的具体方法有多种，如：Scheffé 检验法、费希尔的最小显著差异法（Least significant difference，LSD）、Tukey 的可靠显著差异法（honest significant difference，HSD）、Newman-Keuls 检验法、Duncan 的多距检验法、Bonferroni 检验法，等等。

众多的方法并不说明统计学的繁荣，恰恰相反，这说明到目前为止还没有令人信服的方法。在实际应用中，究竟选哪个方法，统计学界存在较大争议。事实上，没有任何一种方法对所有类型的问题都是最优的。但一般而言，如果是事先计划好的比较，不论方差分析结果如何，均应进行比较，此时可采用费希尔的 LSD 检验法或 Bonferroni 检验法；如果是未计划的多重比较，一般可选择 Tukey 的 HSD 检验法，但样本容量不等时，则倾向于选择 Scheffé 检验法。在此，主要介绍 LSD 检验法和 Scheffé 检验法。

1. 费希尔的 LSD 检验法

LSD 检验法的检验统计量为

$$t = \frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{MS_W\left(\frac{1}{n_i}+\frac{1}{n_j}\right)}} \qquad \text{（公式 10-17）}$$

式中，MS_W 是方差分析的组内均方。由于组内均方反映的是个体误差和抽样误差等随机误差，因此MS_W也可记为MS_E，那么公式 10-17 也可写为

$$t=\frac{\overline{Y}_i-\overline{Y}_j}{\sqrt{MSE(\frac{1}{n_i}+\frac{1}{n_j})}}$$ （公式 10-18）

在 α 显著性水平上的拒绝域为

$$|t|\geq t_{\alpha/2,df=n_T-K}$$ （公式 10-19）

式中，$t_{\alpha/2,df=n_T-K}$ 是自由度为n_T-K时，使 t 分布上侧面积为$\alpha/2$的 t 值。

【例 10-4】 以本章开头的问题 1（即例 10-1）的资料为例，请检验三种反馈类型对补试自尊水平的影响两两之间是否存在显著差异。

解：

$H_0:\mu_i=\mu_j$ $\quad i\neq j$

$H_1:\mu_i\neq\mu_j$

积极反馈组和控制组平均数差异的检验统计量为

$$t_{12}=\frac{\overline{Y}_1-\overline{Y}_2}{\sqrt{MSE(\frac{1}{n_1}+\frac{1}{n_2})}}=\frac{78-74}{\sqrt{28.67(\frac{1}{6}+\frac{1}{6})}}=1.29$$

积极反馈组和消极反馈组平均数差异的检验统计量为

$$t_{13}=\frac{\overline{Y}_1-\overline{Y}_3}{\sqrt{MSE(\frac{1}{n_1}+\frac{1}{n_3})}}=\frac{78-66}{\sqrt{28.67(\frac{1}{6}+\frac{1}{6})}}=3.88$$

控制组和消极反馈组平均数差异的检验统计量为

$$t_{23}=\frac{\overline{Y}_2-\overline{Y}_3}{\sqrt{MSE(\frac{1}{n_2}+\frac{1}{n_3})}}=\frac{74-66}{\sqrt{28.67(\frac{1}{6}+\frac{1}{6})}}=2.59$$

查附表 2“t 值表”得，$t_{.05/2,df=18-3}=2.131$，因为t_{13}和t_{23}均大于 2.131，所以认为积极反馈组与消极反馈组、控制组与消极反馈组平均数之间存在显著差异。而t_{12}小于 2.131,所以认为积极反馈组与控制组平均数之间不存在显著差异。

答：积极反馈组与消极反馈组、控制组与消极反馈组平均数之间存在显著差异。积极反馈组与控制组平均数之间不存在显著差异。

2、Scheffé 检验法

Scheffé 检验法的检验统计量为

$$F=\frac{MS_{ij}}{MS_W},\quad i\neq j$$ （公式 10-20）

式中，MS_{ij} 是要检验的两个样本平均数的“组间均方”；MS_W 是方差分析中的组内均方。$MS_{ij}=SS_{ij}/df_B$，其中 df_B 是方差分析中的组间自由度，SS_{ij} 是要检验的两个样本平均数的“组间平方和”。SS_{ij} 的计算方法为

$$SS_{ij}=(n_i+n_j)\sigma^2_{\bar{Y}_i,\bar{Y}_j} \qquad \text{（公式 10-21）}$$

式中，n_i 和 n_j 分别是要检验的两组的样本容量；$\sigma^2_{\bar{Y}_i,\bar{Y}_j}$ 是要检验的两个样本平均数 $\bar{Y}_i$ 和 $\bar{Y}_j$ 的方差。

Scheffé 检验法的检验统计量在 α 显著性水平上的拒绝域为

$$F\geq F_{\alpha,(df_1,df_2)} \qquad \text{（公式 10-22）}$$

式中，$F_{\alpha,(df_1,df_2)}$ 是方差分析中的临界值。

【例 10-5】 以例 10-2 的资料为例（数据见表 10-7），问这四种方案的教学效果两两之间是否存在显著差异？

解：共四组，两两比较需要比较 6 次，本例题中仅比较第 I 组和第 II 组、第 II 组和第 IV 组、第 III 组和第 IV 组之间的差异。

（1）第 I 组和第 II 组平均数差异比较

$H_0:\mu_1=\mu_2$

$H_1:\mu_1\neq\mu_2$

$$SS_{ij}=(n_i+n_j)\sigma^2_{\bar{Y}_i,\bar{Y}_j}=13\times2.7568^2=98.80$$

$$MS_{ij}=SS_{ij}/df_B=98.80/3=32.93$$

$$F_{12}=MS_{ij}/MS_W=32.93/251.94=0.13$$

（2）第 II 组和第 IV 组平均数差异比较

$H_0:\mu_2=\mu_4$

$H_1:\mu_2\neq\mu_4$

$$SS_{ij}=(n_i+n_j)\sigma^2_{\bar{Y}_i,\bar{Y}_j}=11\times8.9667^2=884.42$$

$$MS_{ij}=SS_{ij}/df_B=884.42/3=294.81$$

$$F_{24}=MS_{ij}/MS_W=294.81/251.94=1.17$$

（3）第 III 组和第 IV 组平均数差异比较

$H_0:\mu_3=\mu_4$

$H_1:\mu_3\neq\mu_4$

$$SS_{ij}=(n_i+n_j)\sigma^2_{\bar{Y}_i,\bar{Y}_j}=10\times16.5732^2=2746.73$$

$$MS_{ij}=SS_{ij}/df_B=2746.73/3=915.58$$

$$F_{34}=MS_{ij}/MS_W=915.58/251.94=3.63$$

查附表 5-A 的 F 值表得，$F_{.05,(3,19)}=3.13$，因为 F_{12} 和 F_{24} 均小于 3.13，所以认为第 I 组和第 II 组、第 II 组和第 IV 组的平均数之间不存在显著差异。而 F_{34} 大

于 3.13，所以认为第 III 组和第 IV 组的平均数存在显著差异。

答：第 I 组和第 II 组、第 II 组和第 IV 组的平均数之间不存在显著差异。第 III 组和第 IV 组的平均数存在显著差异。

3. 在 SPSS 中进行多重比较

SPSS 提供的多重比较的方法有十余种。这里，以例 10-1 为例（原始数据文件见本书所附光盘例 10-1.sav）说明 LSD 方法的应用，以例 10-2 为例（原始数据文件见本书所附光盘例 10-2.sav）说明 Scheffé 检验法的应用。

（1）分析→比较均值→单因素 ANOVA，将“自尊水平”选入因变量列表，将“反馈类型”选入因子框，点击“两两比较”选中“LSD”→继续→确定（如图 10—3 所示），即可得到 LSD 检验结果（见表 10-11）。

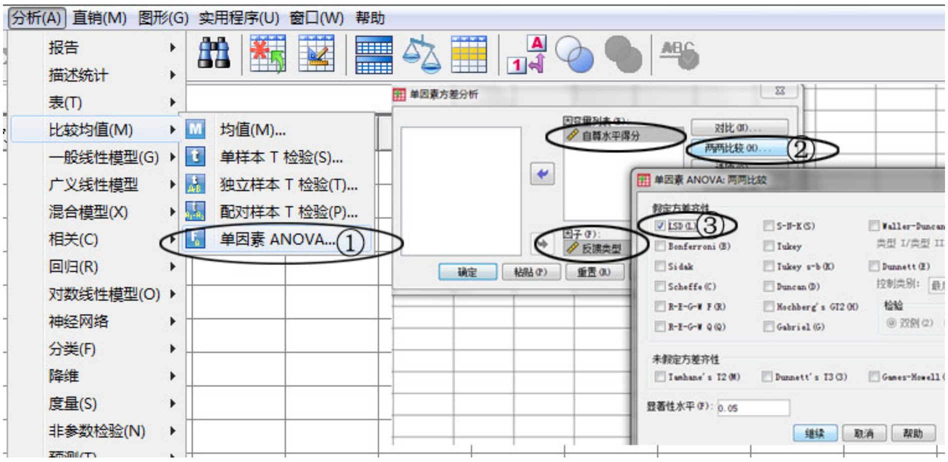


图 10-3　单因素方差分析多重比较 LSD 方法 SPSS 使用示意图

表 10-11　不同反馈类型影响被试自尊水平的多重比较结果表（LSD 法）

(I) 反馈类型	(J) 反馈类型	平均差异 (I-J)	标准误差	显著性	95% 置信区间	
					最小值	最大值
积极	控制	4.00000	3.09121	.215	-2.5887	10.5887
	消极	12.00000**	3.09121	.001	5.4113	18.5887
控制	积极	-4.00000	3.09121	.215	-10.5887	2.5887
	消极	8.00000*	3.09121	.021	1.4113	14.5887
消极	积极	-12.00000**	3.09121	.001	-18.5887	-5.4113
	控制	-8.00000*	3.09121	.021	-14.5887	-1.4113

表 10-11 第一列给出的是对照组，让其他各组（表 10-11 第二列）和它进行比较。表中第三列后依次给出的是两组间均数差值、差值的标准误、P 值以及差值的.95 置信区间。其中，如果均数差值达到显著水平，则在差值后面加上“*”作为标记。由表 10-11 可见，积极反馈组和消极反馈组之间、控制组和消极反馈组之间的均值存在显著差异，而积极反馈组和控制组之间不存在显著差异。

（2）分析→比较均值→单因素 ANOVA，将“测验成绩”选入因变量列表，

将“辅助教学方案”选入因子框，点击“两两比较”选中“Scheffé”→继续→确定，参见图 10—3，把选中“LSD”改为“Scheffé”，即可得到 Scheffé 检验结果。

Scheffé 检验会输出两张表格，其中一张结构与 LSD 方法输出表格完全一样，含义也一样，不再重复。另一张表格是寻找同质子集（Homogeneous Subsets）的结果（见表 10-12）。

表 10-12 “同类子集”检验结果

辅助教学方案	N	alpha = 0.05 的子集	
		1	2
3	6	41.67	
1	6	51.33	51.33
2	7	56.86	56.86
4	4		75.50
显著性		.488	.130

寻找同质子集的含义，简单地说，各组首先在表格的纵向上，均数按大小排列，然后根据多重比较的结果将所有组分为若干个子集，子集之间的各组平均数差异显著（P 值小于.05），子集内的各组间无差别。根据表 10-12 可见，第 III、IV 种辅助教学方案的效果存在显著差异，而其他各教学方案之间不存在显著差异。表格中的最后一行，给出的是子集内部各组比较的 P 值。

4. 多重比较出现矛盾时的解释

多重比较经常出现表 10-12 这种模糊的结论，即辅助教学方案 I、II 与方案 III 之间差异不显著，与方案 IV 之间差异也不显著，但方案 III、IV 之间却差异显著。对于这种情形，只能说两两比较还不能判明方案 I、II 的样本来自何总体。对这种情况常用的错误解释有两种：(1)“方案 I、II 样本所代表的总体水平介于总体 III 和总体 IV 之间”。这种情形实际上默认了四组分别来自三个不同的总体。(2)“方案 III 与方案 I、II 之间不存在显著差异，方案 I、II 和方案 IV 之间无显著差异，所以，方案 III 与方案 IV 之间也不存在显著差异”。这种递推是荒谬的。一个经典例子能充分说明这一点：头上有一根头发的人和一根头发都没有的秃子没有什么差别，有一根头发和有两根头发无差别，依次类推，最后会得到一个满头浓发与一个秃子没有差别的荒谬结论。

有时，方差分析时拒绝 H_0，但方差分析后的两两比较却没有任何两个样本之间存在显著差异。这时下结论应格外谨慎，这可能是统计检验力不够，最好的办法是增加样本容量重新实验。

四、参数估计

方差分析后，我们会得到各组平均水平是否存在显著差异。但有时，我们还希望得到某组样本来自的总体水平的情况，这时，就需要用到参数估计了。

置信水平为 $1-\alpha$ 的 μ_i 的置信区间为

$$[\overline{Y}_i - t_{\alpha/2,df_E}\sqrt{MSE/n_i}, \overline{Y}_i + t_{\alpha/2,df_E}\sqrt{MSE/n_i}], \quad i=1,2,\cdots,K$$ （公式 10-23）

式中

$\overline{Y}_i$ 是要进行参数估计的样本组的平均数；

$t_{\alpha/2,df_E}$ 是自由度为 df_E 时，使 t 分布上侧面积为 $\alpha/2$ 的 t 值，其中 df_E 是方差分析时的组内自由度（即误差自由度）；

MS_E 是方差分析时的组内均方（即误差均方）；

n_i 是要进行参数估计的样本组的样本容量。

【例 10-6】 求例 10-1 自尊水平测试得分最高组在 0.95 置信水平上的置信区间。

解：由表 10-1 可知，自尊水平测试得分最高组是积极反馈组，$\overline{Y}_i = 78$，$n_i = 6$。

由对例 10-1 的求解过程可知，组内均方 MS_E＝28.67，df_E ＝12，查附表 2 “t 值表”得，$t_{.05/2,12} = 2.179$

因此，置信水平为 1-α 的 μ_i 的置信区间为

$$78 - 2.179 \times \sqrt{\frac{28.67}{6}} < \mu_1 < 78 + 2.179 \times \sqrt{\frac{28.67}{6}}$$

$$73.24 \le \mu_1 \le 82.76$$

答：自尊水平测试得分最高组的 1-α＝.95 的置信区间介于 73.24 分至 82.76 分之间。

第三节 两因素方差分析

一、两因素方差分析的数学模型

设在实验或观察中有两个因素 A 和 B，假定因素 A 有 p 个水平，因素 B 有 q 个水平，在因素 A 取第 i 水平、因素 B 取第 j 水平时（以后简记为条件 A_iB_j），因变量服从 $N(\mu_{ij},\sigma^2)$ 的正态分布。记**一般平均**为

$$\mu = \frac{1}{pq}\sum_{i=1}^{p}\sum_{j=i}^{q}\mu_{ij}$$ （公式 10-24）

式中，μ_{ij} 是条件 A_iB_j 下的总体平均数。

又记 $\mu_{i.} = \frac{1}{q}\sum_{j=1}^{q}\mu_{ij}$，$\mu_{.j} = \frac{1}{p}\sum_{i=1}^{p}\mu_{ij}$，则对 $i=1,2,\cdots,p$，称

$$a_i = \mu_{i.} - \mu$$ （公式 10-25）

为因素 A 的第 i 水平的效应，且 $\sum_{i=1}^{p} a_i = 0$；对 $j = 1, 2, \cdots, q$，称

$$b_j = \mu_{\cdot j} - \mu \qquad \text{（公式 10-26）}$$

为因素 B 的第 j 水平的效应，且 $\sum_{j=1}^{q} b_j = 0$。

如果对于一切 $i = 1, 2, \cdots, p$ 和 $j = 1, 2, \cdots, q$，都有 $\mu_{ij} = \mu + a_i + b_j$，则此时的模型是**效应可加模型**。若对某些 i、j 存在 $\mu_{ij} \neq \mu + a_i + b_j$，则记

$$(ab)_{ij} = \mu_{ij} - \mu - a_i - b_j \qquad \text{（公式 10-27）}$$

称此为因素 A 的第 i 水平和因素 B 的第 j 水平的**交互效应**，$i = 1, 2, \cdots, p$，$j = 1, 2, \cdots, q$，且 $\sum_{j=1}^{q} (ab)_{ij} = 0$，$i = 1, 2, \cdots, p$；$\sum_{i=1}^{p} (ab)_{ij} = 0$，$j = 1, 2, \cdots, q$，这时的模型称为**有交互作用的模型**。

二、效应可加模型的方差分析方法

（一）模型

正如前文公式所示，所谓效应可加模型（effect additivity model）是指两因素模型中各处理均值等于一般平均数加上该处理对应的两个因素水平效应之和的一种模型。对效应可加模型而言，每种水平组下（即实验处理 A_iB_j）只要进行一次试验即可进行数据分析。若记条件 A_iB_j 的试验结果为 y_{ij}，则此时的模型可以表示为

$$\begin{cases} y_{ij} = \mu + a_i + b_j + \varepsilon_{ij} \\ \sum_{i=1}^{p} a_i = 0,\ \sum_{j=1}^{q} b_j = 0 & i = 1, 2, \cdots, p;\ j = 1, 2, \cdots, q \\ \text{各}\varepsilon_{ij}\text{相互独立，均服从}N(0, \sigma^2)\text{的正态分布} \end{cases} \qquad \text{（公式 10-28）}$$

为了检验所有 μ_{ij} 是否相等，可以转化为检验如下两个假设：

$$\begin{cases} H_{0A}: a_1 = a_2 = \cdots = a_p = 0 \\ H_{aA}: a_1, a_2, \cdots, a_p \text{不全为} 0 \end{cases} \qquad \text{（假设 10-3）}$$

$$\begin{cases} H_{0B}: b_1 = b_2 = \cdots = b_q = 0 \\ H_{aB}: b_1, b_2, \cdots, b_q \text{不全为} 0 \end{cases} \qquad \text{（假设 10-4）}$$

该模型的数据模式一般如表 10-13 所示。

表 10-13　在每种水平组合下进行一次试验的数据模式

A \ B	B_1	B_2	…	B_j	…	B_q	$\bar{Y}_{i\cdot}$
A_1	y_{11}	y_{12}	…	y_{1j}	…	y_{1q}	$\bar{Y}_{1\cdot}$
A_2	y_{21}	y_{22}	…	y_{2j}	…	y_{2q}	$\bar{Y}_{2\cdot}$
…	…	…	…	…	…	…	…

续表

A \ B	B_1	B_2	$\cdots$	B_j	$\cdots$	B_q	$\bar{Y}_{i\cdot}$
A_i	y_{i1}	y_{i2}	$\cdots$	y_{ij}	$\cdots$	y_{iq}	$\bar{Y}_{i\cdot}$
$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$	$\cdots$
A_p	y_{p1}	y_{p2}	$\cdots$	y_{pj}	$\cdots$	y_{pq}	$\bar{Y}_{p\cdot}$
$\bar{Y}_{\cdot j}$	$\bar{Y}_{\cdot 1}$	$\bar{Y}_{\cdot 2}$	$\cdots$	$\bar{Y}_{\cdot j}$	$\cdots$	$\bar{Y}_{\cdot q}$	$\bar{Y}_{\cdot\cdot}=\dfrac{\sum_{i=1}^{p}\sum_{j=1}^{q}y_{ij}}{pq}$

（二）方差分析

检验（假设 1）和（假设 2）两个假设的方差分析表见表 10-14。

表 10-14　在每种水平组合下进行一次试验的两因素方差分析表

变异来源	平方和	自由度	均方	F
因素 A	SSA	$df_A=p-1$	$MSA=\dfrac{SSA}{df_A}$	$F_A=\dfrac{MSA}{MSE}$
因素 B	SSB	$df_B=q-1$	$MSB=\dfrac{SSB}{df_B}$	$F_B=\dfrac{MSB}{MSE}$
误差 E	SSE	$df_E=df_T-df_A-df_B$	$MSE=\dfrac{SSE}{df_E}$	
总变异 T	SST	$df_T=n_T-1$		

表中，$n_T=pq$，由表 10-14 可见，进行有关的方差分析关键在于求取平方和。

1. 各平方和的定义公式

$$SST=\sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij}-\bar{Y}_{\cdot\cdot})^2 \qquad \text{（公式 10-28）}$$

$$SSA=q\cdot\sum_{i=1}^{p}(\bar{Y}_{i\cdot}-\bar{Y}_{\cdot\cdot})^2 \qquad \text{（公式 10-29）}$$

$$SSB=p\cdot\sum_{j=1}^{q}(\bar{Y}_{\cdot j}-\bar{Y}_{\cdot\cdot})^2 \qquad \text{（公式 10-30）}$$

$$SSE=SST-SSA-SSB \qquad \text{（公式 10-31）}$$

其中，$\bar{Y}_{i\cdot}$ 表示因素 A 第 i 水平下的数据的平平均数，$\bar{Y}_{\cdot j}$ 表示因素 B 的第 j 水平下的数据的平平均数，$\bar{Y}_{\cdot\cdot}$ 表示所有数据的总平平均数。

2. 拒绝域

对给定的显著性水平 α，当 $F_A \ge F_{\alpha,(df_A,df_E)}$ 时拒绝假设 10-3，当 $F_B \ge$

$F_{\alpha,(df_B,df_E)}$时拒绝假设 10-4。

3. 方差分析实例

【例 10-7】 为比较三种教材的效果，某研究者随机抽取了省重点中学、省建设重点中学、市重点中学和普通中学各一所，每所学校随机抽取三个班，每班随机接受一种教材的教学实验。假设其他影响因素得到了较好的控制，经过一段时间的教学后，通过统一考试得到每个班的平均成绩如表 10-15 所示。问四种教材的教学效果是否一致。（本例中原始数据见本书所附光盘中例 10-7.sav。）

表 10-15　四种教材实验成绩表

学校 教材	I	II	III	IV	$\bar{Y}_{i\cdot}$
I	82	62	45	47	59
II	98	83	83	55	79.75
III	66	65	55	45	57.75
$\bar{Y}_{\cdot j}$	82	70	61	49	$\bar{Y}_{\cdot\cdot}$＝65.5

解：

利用假设 10-3 和假设 10-4 建立假设组。

运用 SPSS 软件包，根据下列操作步骤（如图 10—4 所示）可以对例 10-7 进行效应可加模型的方差分析：

1.分析→一般线性模型→单变量
2.因变量框：成绩
3.固定因子框：教材、学校
4.模型：（1）选择“设定”　　（第 4 步是指定无交互效应的模型）
（2）“构建项”下拉列表：主效应
（3）“模型”框：教材、学校
（4）继续
5.两两比较：（1）“两两比较检验”框：教材、学校
（2）☑ Scheffe　　（第 5 步是为了进行均值的多重比较）
（3）继续
6.确定

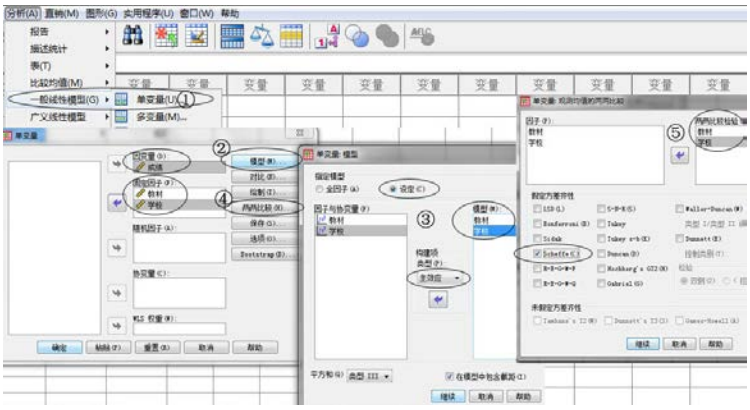


图 10-4　效应可加模型方差分析 SPSS 使用示意图

分析结果主要包括方差分析表（见表 10-16）和多重比较的结果（见表 10-18 和表 10-19）

表 10-16　组间效应检验

变异来源	平方和	自由度	均方	F	显著性
Corrected Model	2976.500[a]	5	595.300	9.387	.008
Intercept	51483.000	1	51483.000	811.821	.000
教材	1221.500	2	610.750	9.631	.013
学校	1755.000	3	585.000	9.225	.012
Error	380.500	6	63.417		
Total	54840.000	12			
Corrected Total	3357.000	11			

a. R squared=.887（Adjusted R Squared=.792）

表 10-16 的结果含义如下：

（1）第一行的“Corrected Model”是对所用方差分析模型的整体检验，其 H_0 是模型中所有影响因素均无作用，即教材、学校（有交互作用模型中，还包括因素的交互作用）对成绩均无影响（a_i、b_j 均为 0)。该检验的 P 值（表中显著性列）为.008 小于.05，说明所用模型有统计学意义，即教材、学校至少有一个因素对成绩有影响，究竟是哪一个还是两者都有影响则要看表中其他分析结果。

（2）第二行是截距，它在方差分析中没有实际意义，可以忽略。

（3）第三、四、五、七的含义同例 8-1 中的方差分析表。这里，教材、学习时长对应的 P 值均小于 0.05（显著性列），表明二者对成绩都有显著影响，即 a_i、b_j 中均至少有一个不为 0。

（4）第六行“Total”是“Intercept”与“Corrected Total”之和。在实际分析中也没有实际意义。

（5）表格最下方还给出了决定系数（R squared）和校正的决定系数（Adjusted R Squared)，其含义参照回归分析一章。

表 10-16 的结果可以简化为表 10-17 的形式，表 10-18 和表 10-19 也作了简单的处理。

表 10-17 组间效应方差分析表

变异来源	平方和	自由度	均方	F
教材（因素 A）	1221.5	2	610.75	9.63^*
学校（因素 B）	1755	3	585	9.22^*
误差 E	380.5	6	63.4167	
总变异 T	3357	11		

查附表 5－A 的 F 值表得，$F_{0.05,(2,6)}=5.14$，$F_{0.05,(3,6)}=4.76$。因为 F_A ＝9.63 大于 $F_{0.05,(2,6)}=5.14$，所以教材主效应显著；F_B ＝9.22 大于 $F_{0.05,(3,6)}=4.76$，所以学校之间也存在显著差异。

表 10-18　多重比较检验

(I) 教材	(J) 教材	均值差值(I-J)	标准误差	显著性	95%置信区间	
					下限	上限
教材 I	教材 II	-20.75	5.631	.029	-38.81	-2.69
	教材 III	1.25	5.631	.976	-16.81	19.31
教材 II	教材 I	20.75	5.631	.029	2.69	38.81
	教材 III	22.00	5.631	.022	3.94	40.06
教材 III	教材 I	-1.25	5.631	.976	-19.31	16.81
	教材 II	-22.00	5.631	.022	-40.06	-3.94

表 10-19　同质子集检验

教材	N	子集	
		1	2
教材 III	4	57.75	
教材 I	4	59.00	
教材 II	4		79.75
显著性		.976	1.000

表 10-18 和表 10-19 中内容的含义与表 10-11 和表 10-12 的含义相同，不再重复。另外，SPSS 结果给出了教材、学习时长的多重比较结果，此处仅列出教材的多重比较结果。从表 10-18 和表 10-19 来看，教材 II 的效果要好于教材 I 和 III，但教材 I、III 之间没有显著差异。

三、交互作用模型的方差分析方法

（一）模型

交互作用(interaction effect)是指，当一个因素的水平与另一个因素的水平发生作用时，对因变量产生的影响。**交互作用模型**(interaction effect model)：指每种水平组合下进行多次试验，可以分析因素之间交互作用的实验设计。对有交互作用的因素而言，每种水平组合（处理）下必须进行多次试验才能进行分析，这是实验设计中的重复原则。重复的主要作用在于，一方面这是估计实验误差的必要条件，另一方面它能让实验者获得因素效应的更精确的估计量。一般来讲，在每种处理条件下重复试验的次数应该相等，这称为**等重复**。实际工作中，也会出现各处理内重复试验次数不等的情况，这称为**不等重复**。本章第二节单因素完全随机设计方差分析中，例 7-5 就是等重复情况，例 7-6 就是不等重复情况。工作中，不等重复的情况不值得提倡，这是因为在实验总次数 n_T 一定的情况下，等重复的比不等重复的精度要高。举一个极端例子，

当 A 因素有两个水平时，$n_T=100$，取 $n_1=1$，$n_2=99$ 和 $n_1=n_2=50$ 的情况相比，后者的精度要高得多。

设在每个 A_2B_1 条件下进行 r 次试验（$r\geqslant2$）。若记 A_2B_1 条件下的试验结果为 y_{ijk}，$k=i,j,\cdots,r$，则模型可表示为

$$\begin{cases} y_{ijk}=\mu+a_i+b_j+(ab)_{ij}+\varepsilon_{ijk}\ , \\ \qquad i=1,2,\cdots,p\ ,\ j=1,2,\cdots,q\ ,\ k=1,2,\cdots,r \\ \sum\limits_{i=1}^{p}a_i=0\ ,\ \sum\limits_{j=1}^{q}b_j=0 \\ \sum\limits_{i=1}^{p}(ab)_{ij}=0\ ,\ j=1,2,\cdots,q \\ \sum\limits_{j=1}^{q}(ab)_{ij}=0\ ,\ i=1,2,\cdots,p \\ \text{各}\varepsilon_{ijk}\text{相互独立，均服从}N(0,\sigma^2)\text{ 的正态分布} \end{cases} \tag{公式 10-32}$$

为检验一切 μ_{ij} 是否相等，可以转化为检验如下三个假设，其中前两个假设的表达形式分别是假设 10-3 和假设 10-4，第三个假设的表达形式是：

$$\begin{cases} H_{0A\times B}:\text{所有}(ab)_{ij}=0 \\ H_{aA\times B}:\text{至少有一个}(ab)_{ij}\neq0 \end{cases} \tag{假设 10-5}$$

如果拒绝假设 1，则说因素 A 主效应（main effect）显著；如果拒绝假设 10-4，则说因素 B 主效应显著。如果拒绝假设 10-5，则说因素 A 与因素 B 的交互作用（interaction）显著。

该模型的数据模式一般如表 10-20 所示。

表 10-20　有交互作用的两因素方差分析的数据模式

A \ B	B_1	B_2	…	B_j	…	B_q	$\bar{Y}_{i\cdot\cdot}$
A_1	$\bar{Y}_{11\cdot}$	$\bar{Y}_{12\cdot}$	…	$\bar{Y}_{1j\cdot}$	…	$\bar{Y}_{1q\cdot}$	$\bar{Y}_{1\cdot\cdot}$
A_2	$\bar{Y}_{21\cdot}$	$\bar{Y}_{22\cdot}$	…	$\bar{Y}_{2j\cdot}$	…	$\bar{Y}_{2q\cdot}$	$\bar{Y}_{2\cdot\cdot}$
…	…	…	…	…	…	…	…
A_i	$\bar{Y}_{i1\cdot}$	$\bar{Y}_{i2\cdot}$	…	$\bar{Y}_{ij\cdot}{}^{*}$	…	$\bar{Y}_{iq\cdot}$	$\bar{Y}_{i\cdot\cdot}$
…	…	…	…	…	…	…	…
A_p	$\bar{Y}_{p1\cdot}$	$\bar{Y}_{p2\cdot}$	…	$\bar{Y}_{pj\cdot}$	…	$\bar{Y}_{pq\cdot}$	$\bar{Y}_{p\cdot\cdot}$
$\bar{Y}_{\cdot j\cdot}$	$\bar{Y}_{\cdot1\cdot}$	$\bar{Y}_{\cdot2\cdot}$	…	$\bar{Y}_{\cdot j\cdot}$	…	$\bar{Y}_{\cdot q\cdot}$	$\bar{Y}_{\cdot\cdot\cdot}=\dfrac{\sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}y_{ijk}}{pqr}$

$^{*}\bar{Y}_{ij\cdot}=\dfrac{1}{r}\sum_{k=1}^{r}y_{ijk}$

表 10-20 中，每个处理条件下（表中的每一格）的 $\bar{Y}_{ij\cdot}$ 都是本格中重复的 r 次试验数据的平平均数。例如，假设 r＝3，条件 A_1B_2 下的三次试验结果是：y_{121}＝23，y_{122}＝28，y_{123}＝27。那么，$\bar{Y}_{12\cdot}$=(23+28+27)=26。

（二）方差分析

检验这三个假设的方差分析表可参见表 10-21。

表 10-21　在每种水平组合下进行 r 次试验的两因素方差分析表

变异来源	平方和	自由度	均方	F 值
因素 A	SSA	$df_A = p-1$	$MSA = \dfrac{SSA}{df_A}$	$F_A = \dfrac{MSA}{MSE}$
因素 B	SSB	$df_B = q-1$	$MSB = \dfrac{SSB}{df_B}$	$F_B = \dfrac{MSB}{MSE}$
交互作用 $A \times B$	$SSAB$	$df_{AB} = (p-1)(q-1)$	$MSAB = \dfrac{SSAB}{df_{AB}}$	$F_{AB} = \dfrac{MSAB}{MSE}$
误差 E	SSE	$df_E = pq(r-1)$	$MSE = \dfrac{SSE}{df_E}$	
总变异 T	SST	$df_T = n_T - 1$		

表中，$n_T = pqr$，且 $df_T = df_A + df_B + df_{AB} + df_E$。由表 10-23 可见，进行有关的方差分析关键在于求取平方和。

1 各平方和的定义公式

$$SST = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2 \qquad \text{（公式 10-33）}$$

$$SSA = qr \cdot \sum_{i=1}^{p}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 \qquad \text{（公式 10-34）}$$

$$SSB = pr \cdot \sum_{j=1}^{q}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2 \qquad \text{（公式 10-35）}$$

$$SSAB = r \cdot \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2 \qquad \text{（公式 10-36）}$$

$$SSE = SST - SSA - SSB - SSAB \qquad \text{（公式 10-37）}$$

其中，$\bar{Y}_{i\cdot\cdot}$ 表示因素 A 第 i 水平下的数据的平平均数，$\bar{Y}_{\cdot j\cdot}$ 表示因素 B 的第 j 水平下的数据的平平均数，$\bar{Y}_{ij\cdot}$ 表示条件 A_iB_j 下的数据的平平均数，$\bar{Y}_{\cdot\cdot\cdot}$ 表示所有数据的总平平均数。

2 拒绝域

对给定的显著性水平 α，当 F_A ＞ $F_{\alpha,(df_A,df_E)}$ 时拒绝假设 10-3，当 F_B ＞ $F_{B,(df_B,df_E)}$ 时拒绝假设 10-4，当 F_{AB} ＞ $F_{\alpha,(df_{AB},df_E)}$ $F_{AB} \geq F_{\alpha,(df_{AB},df_E)}$ 时拒绝假设 10-5。

3. 方差分析实例

【例 10-8】下面我们以本章开头的问题 2 作为进行交互作用模型的方差分析例子来求解。与两因素效应可加模型相比，在效应可加模型中，只指定主效应，没有交互效应，而在交互作用模型中，应包括交互作用。通常我们会认为，对比赛到中盘的棋局的记忆会受专业知识的影响，专家的记忆成绩会好于新手；而对随机摆放的棋子的记忆，新手与专家的水平不存在显著差异。但这需要通过假设检验来得以证明。（本例题原始数据见本书所附光盘中例 10-8.sav）

解：利用假设 10-3、假设 10-4 和假设 10-5 建立假设组。

运用 SPSS 软件包对例 10-8（数据见例 10-8. sav）求解的操作步骤如下（参见图 10-5）：

1. 分析→一般线性模型→单变量
2. “因变量”框：记忆成绩
3. “固定因子”框：人员、棋局
4. 模型：（1）选择“全因子”　　（这是 SPSS 默认选项）
　　　　（2）继续
5. 确定

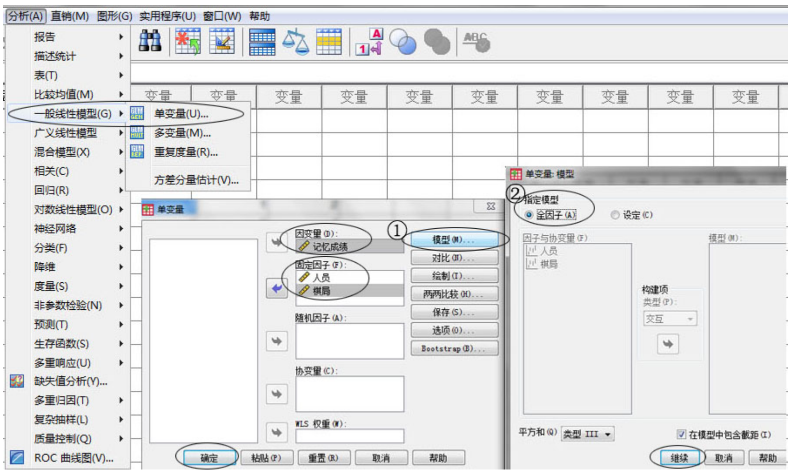


图 10-5　交互效应模型方差分析 SPSS 使用示意图

此处，没有选择其他选项（包括多重比较），所以输出结果中主要的只有方差分析结果表，其解释可以参见对表 10-16 的解释。经整理后，例 10-8 的方差分析表有如表 10-22 所示。

表 10-22　　方差分析表

变异来源	平方和	自由度	均方	F
因素 A(人员)	64.8	1	64.8	18.65^{**}
因素 B(棋局)	51.2	1	51.2	14.73^{**}
交互作用 $A\times B$	39.2	1	39.2	11.28^{**}
误差 E	55.6	16	3.475	
总变异 T	210.8	19		

查附表 5-B 的 F 值表得，$F_{0.01,(1,16)}=8.53$。由于因素 A、因素 B 和二者交互作用的自由度均为 1，因此 F 检验的临界值相同。因为 $F_A=18.65$、$F_B=14.73$ 和 $F_{AB}=11.28$，都大于 $F_{0.01,(1,16)}=8.53$，所以人员类型、棋局类型主效应显著，二者的交互作用也显著。

（三）简单效应分析

在因素实验中，简单效应（simple effects）是指其他因素水平固定时，某因素不同水平之间的变异。例如，在一个 2×2 的两因素实验中，因素 A 的两个水平在 B_1 水平上的变异，称为因素 A 在 B_1 水平的简单效应；因素 A 的两个水平在 B_2 水平的变异，称为因素 A 在 B_2 水平的简单效应。同样地，因素 B 的两个水平在 A_1 水平或 A_2 水平的变异也是简单效应。交互作用从直观上讲是指一个因素不同的水平的好坏或好坏的程度受到另一因素水平的制约。

方差分析结果表明存在交互作用时，常常需要进一步做简单效应检验，以探明两个因素之间交互作用的实质。在一个两因素的交互效应模型中，包括了 A 因素在 B 因素各水平上的简单效应，和 B 因素在 A 因素各水平上的简单效应。实际工作中，是否需要做两个方面的简单效应检验，还是只做其中一种，与研究的理论假设有对实验结果的解释有关，视具体情况而定。

进行简单效应检验，一般是固定一个因素（例如因素 B）的水平，然后检验因素 A 各水平的差异，因素 B 有多少个水平，就要进行多少次单因素（A 因素）的方差分析，以确定因素 A 各水平在因素 B 的各个特定水平上的差异是否显著。同时，还应结合简单效应分析图，才能对交互作用有较直观地了解。

【例 10-9】 对例 10-8 中的交互作用进行简单效应分析。

解：（1）绘制简单效应分析图

图 10-7 可在 SPSS 进行方差分析时，选择绘制并进行相应设置而获得，操作步骤如下（如图 10-6 所示，结果见图 10-7）：

绘制

（1）“水平轴”框：人员；“单图”框：棋局；添加

（2）“水平轴”框：棋局；“单图”框：人员；添加

继续

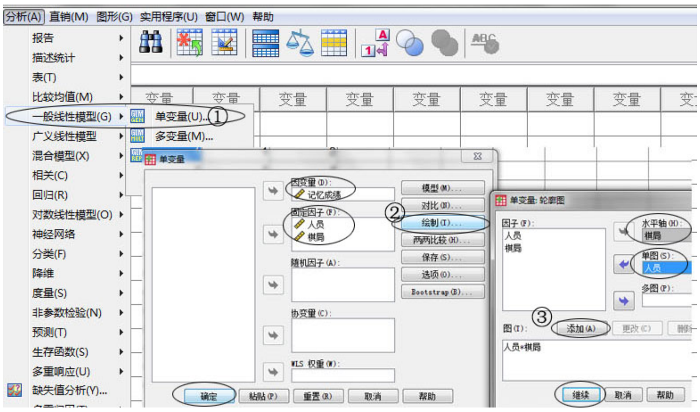


图 10-6　交互作用简单效应绘图 SPSS 使用示意图

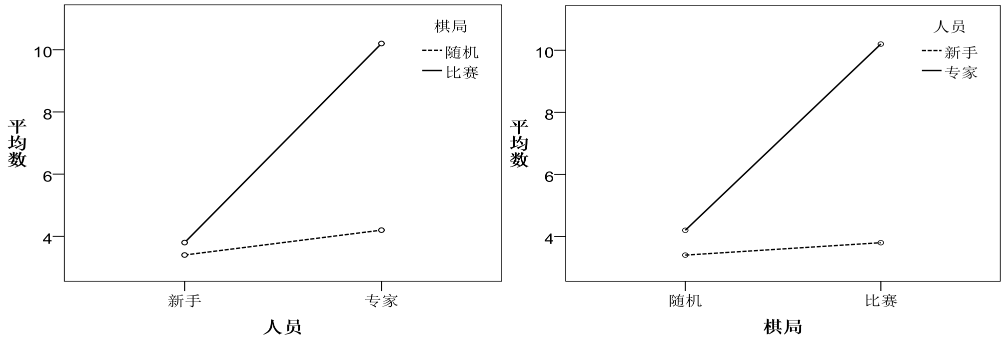


图 10-7　因素 A 与因素 B 交互作用的图解

（2）因素 A 在因素 B 各个水平上的简单效应检验

SPSS 操作步骤如图 10-8 所示：

1.数据→拆分文件

（1）“比较组”框：棋局

（2）$\boxed{\text{确定}}$

2. 分析→比较均值→单因素 ANOVA

（1）“因变量列表”框：记忆成绩

（2）“因子”框：人员

（3）$\boxed{\text{确定}}$

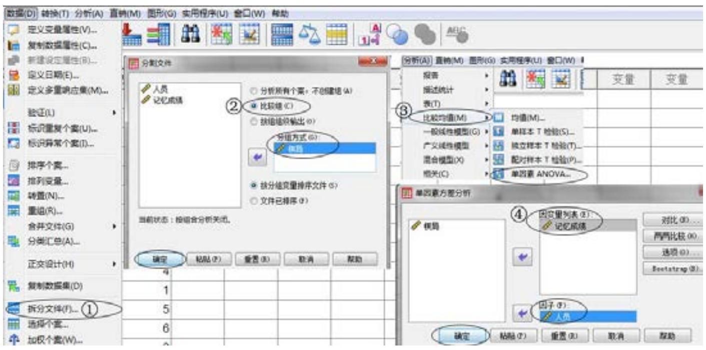


图 10-8　交互作用简单效应分析 SPSS 使用示意图

输出结果如表 10-23。

表 10-23　因素 A 在因素 B 不同水平上的简单效应检验方差分析表

棋局		平方和	自由度	均方	*F* 值	显著性
随机	组间	1.600	1	1.600	.582	.467
	组内	22.000	8	2.750		
	总体	23.600	9			
比赛	组间	102.400	1	102.400	24.381	.001
	组内	33.600	8	4.200		
	总体	136.000	9			

（3）因素 B 在因素 A 各个水平上的简单效应检验

操作步骤与第（2）步相似。可得结果表 10-24。

表 10-24　因素 B 在因素 A 不同水平上的简单效应检验方差分析表

人员		平方和	自由度	均方	*F* 值	显著性
新手	组间	.400	1	.400	.100	.760
	组内	32.000	8	4.000		
	总体	32.400	9			
专家	组间	90.000	1	90.000	30.508	.001
	组内	23.600	8	2.950		
	总体	113.600	9			

结合图 10-7，表 10-23 和表 10-24 可知，在随机棋局上，专家与新手的差异不显著；而在比赛棋局上专家得分显著高低新手得分。新手在两种棋局上的差异不显著，而专家在比赛棋局上的得分显著高于随机棋局的得分。

第四节　其他常用模型的方差分析

在本节，我们将简单介绍随机效应模型、相关样本和混合模型的方差分析，但不对其数学模型进行介绍，重点在讲清概念，并给出实例的 SPSS 应用方法。

一、随机效应模型的方差分析

（一）固定因素与随机因素

固定因素（Fixed Factor）指的是因素水平的选取是出于设计者的“偏好”，而不是随机选定的。例如，有人想研究不同的记笔记技术对学习成绩的影响，他精心设计了三种性能可能较好的记笔记技术，但不知哪个更好，于是开展有关实验研究。在他的研究中，三种记笔记的技术，纯粹属于研究者的个人“偏好”，而不是有很多现成的笔记技术，然后从中随机抽取三种技术进行对比研究。甚至可以说，在该研究中连第四种笔记技术都不存在。所以，一般来说，如果某因素是固定因素，那么该因素的所有可能的水平都会出现。如果自变量是固定因素，那么方差分析就是固定效应的方差分析（fixed factor analysis of variance）。

随机因素（Random Factor）指的是因素水平是从大量的因素处理中随机选取的。例如，一个关于青少年心理发展特点的调查研究，想比较不同城市青少年的心理发展特点是否不同，为此随机抽取了 20 个城市进行研究。显然，研究者希望研究结果能够推

及全国的所有城市，此时就涉及到将分析结果外推到未包括的城市的问题，这种情况下，城市就应当是一个随机因素了。如果自变量是随机因素，那么方差分析就是随机效应的方差分析（random factor analysis of variance）。

由于固定效应与随机效应方差分析的数学模型不同，因此在分析时应加以区分。但是，在许多时候，判断一个因素究竟是固定因素还是随机因素并不是一件容易的事情。区别一个因素是固定因素或是随机因素，并不是该因素本身特性所决定，而是取决于我们的分析目的。若设计是固定效应的，则据此得出的结论只能局限于所采用的那些特定的水平值；若设计是随机效应的，则其结论的适用范围会广泛些，它可以外推到所采用的水平的源头——因素处理上去的。同一个因素，如果研究目的不同，既可能是固定因素，也可能是随机因素。例如，前面提到的随机抽取 20 个城市的调查研究，如果希望结论能推及全国各城市，则其是随机因素；如果我们仅对这 20 个城市感兴趣，结论只推及这 20 个城市，那么这就是固定因素。

（二）随机效应模型的 SPSS 分析实例

【例 10-10】 有人想知道哪种宣传途径向大学生宣传某一流行疾病预防知识的效果更好。为此，他设计了两种宣传方式：召开班会集中讲解和发放材料在宿舍张贴。他随机抽取了四所大学，每所大学随机抽取 10 个班，接受两种宣传方式的各 5 个班。假定其他因素得到较好控制，最后考试每个班的平均成绩，如表 10-25 所示。问两种宣传方式的效果是否存在差异？四所大学的成绩是否存在差异？（本例题的原始数据见本书所附光盘中例 10-10.sav）

表 10-25　不同宣传方式下四所大学学生的成绩表

大学 / 宣传方式	大学 I	大学 II	大学 III	大学 IV
集中讲解	80，83，82，85，77	78，82，80，84，77	79，82，83，81，78	80，77，82，84，79
分发材料	74，76，78，77，75	72，76，73，77，73	72，74，76，74，71	78，76，79，81，78

解： 本例是一个两因素方差分析模型的分析问题。其中四所大学是从百余所大学中随机抽取的，并且，很显然研究者希望这一结果推及的范围不仅仅是这四所大学，而是包括其他大学。因此，此例中的大学是一个随机因素。相应 SPSS 的操作如下（见图 10-9 所示）：

1. 分析→一般线性模型→单变量
2. “因变量列表”框：成绩
3. “固定因子”框：宣传方式
4. “随机因子”框：大学
5. 确定（由于 SPSS 默认的模型是“全因子”，因此可以直接按确定键）

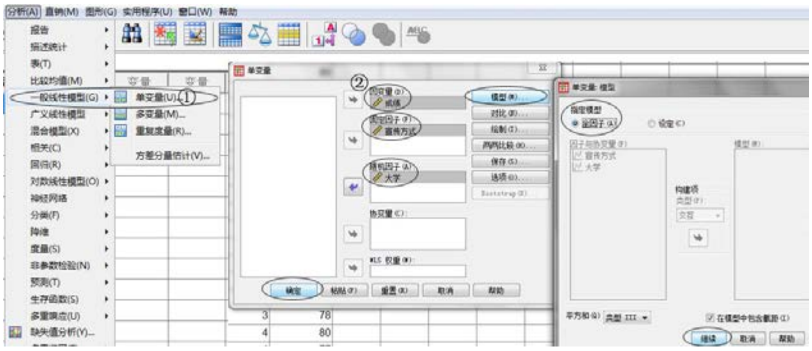


图 10-9　随机效应方差分析 SPSS 使用示意图

分析结果中的主要部分是方差分析表，见表 10-26。从表中可以看出不同宣传方式差异显著，而不同大学之间差异不显著，宣传方式与大学的交互作用也不显著。表中最下面有注释，告诉人们对不同假设进行检验时所采用的误差项。值得注意的是，交互作用检验的误差项与主效应检验的误差项不一样。如果需要进一步进行多重比较，其结果形式与只含有固定因素的模型完全相同，这里不赘述。

表 10-26　Tests of Between-Subjects Effects

变异来源		平方和	自由度	均方	*F* 值	显著性
Intercept	Hypothesis	243828.225	1	243828.225	1.798E4	.000
	Error	40.675	3	13.558[a]		
宣传方式	Hypothesis	265.225	1	265.225	21.346	.019
	Error	37.275	3	12.425[b]		

续表

变异来源		平方和	自由度	均方	*F* 值	显著性
大学	Hypothesis	40.675	3	13.558	1.091	.472
	Error	37.275	3	12.425[b]		
宣传方式 * 大学	Hypothesis	37.275	3	12.425	2.290	.097
	Error	173.600	32	5.425[c]		

a.　MS(大学)　; b.　MS(宣传方式 * 大学); c.　MS(Error)

二、相关样本模型的方差分析

（一）独立样本与相关样本设计

在这之前，本书介绍的方差分析模型，每个实验处理下的被试是由不同的被试组构成的，这种设计称为被试间设计（between-subjects design，也称为独立样本设计或组间设计），模

型中的自变量称为被试间变量（between-subjects variable)。例如，假设我们要研究人们对不同色光的反应时是否存在差异，假定选择了红、黄、绿三种色光，同时将 30 名被试随机分为 3 组，分组 10 人随机接受一种色光的处理。这时，“色光”这一因素就是被试间变量。为了提高检验的灵敏性和精确度，在实际研究中，研究人员可能会让 30 名被试都对红、黄、绿三种色光进行反应。这样，各种处理条件下的被试就是同一组人了，这种设计使得处理之间存在相关，因此，

这种设计被称为被试内设计（within-subjects design，也称为相关样本设计或组内设计）。此时，“色光”这一因素就是被试内变量（within-subjects variable）。

被试内设计包括两种形式：重复测量设计（Repeated Measures design）和匹配组设计（Matched Groups design）。

1. 重复测量设计

重复测量设计是指实验各处理中的被试是同一批被试的一种设计形式。这种设计一般没有另外的控制组（Control Group），对这种设计进行分析，可以考察被试在不同实验处理条件下的行为变化。

和独立样本设计相比，重复测量设计所需的样本量较少，更为经济。由于重复测量设计中的每位被试均参加了自变量所有水平下的测试，因此，组间变异不包括个体差异的影响，这有利于提高检验处理间效应的精确度和灵敏度。但这种设计可能会产生遗留效应（carry-over effect），即一个处理条件下的效应会影响到被试在后续处理条件下的行为。例如，给被试进行系列的数学测验，被试在第一个测验中的表现，可能影响到他在后续测验中的表现。对此，常用的方法是平衡（counterbalancing）。例如，被试要接受 A、B、C 三个测验，可以让第 1 个被试测试次序为ABC，第二个被试测试次序为CBA，第三个为BCA，等等。

2. 匹配组设计（Matched Groups design）

采用匹配组设计，一般是因为研究者认为被试的某个或某些特征会对研究结果产生影响，于是，对这个或这些特性进行测量，并根据被试的得分对被试进行分组。假想我们想检测四种心理疗法对抑郁症的治疗效果。我们想到个体最初的抑郁程度会影响到心理治疗的成功，于是，采用一个标准化的抑郁症标准测验对所有被试进行测验。然后，根据被试得分高低进行排序，分数最高的四名被试被随机地分配到四种不同的治疗组，依次类推。这样做就是为了让四组被试最初的抑郁程度是相当的。这样“匹配”分配被试的方法也使得各处理组之间存在相关。

（二）相关样本模型的SPSS分析实例

【例 10-11】 适应是心理学家们研究的一个重要现象。当一个巨大响声突然出现时，我们会产生紧张反应，但当这种响声重复出现时，我们会开始忽略它并不再有紧张的反应。在一项研究中，研究者随机抽取了五位大学生，让每位大学生接受五次巨大噪声的刺激，同时测量他们的皮肤电位。前四次，噪声之间间隔 60 秒。在第四次之后，要等上 5 分钟才会出现噪声刺激。表 10-27 记录了实验的数据。问这一结果是否支持大学生出现了对噪声的适应现象。（本例题的原始数据见本书所附光盘中例 10-11.sav）

表 10-27　噪声刺激适应实验结果数据

测量时间 / *大学生*	*I*	*II*	*III*	*IV*	V
1	9.5	4.9	7.2	0.4	8.9
2	11.1	10.4	7.2	4.7	12.3
3	8.1	6.3	3.5	0.4	8.3
4	8.8	9.1	3.6	0.4	10.3
5	11.0	6.5	5.2	2.4	10.5

解：突然刺激引起紧张导致皮肤电位升高，如果一旦适应而不再有紧张反应，皮肤电位就不会升高，相对于最初紧张导致的皮肤电位升高，甚至会出现皮肤电位下降的趋势。第五次由于间隔时间发生变化，那么此时的皮肤电位又会升高。因此，通过方差分析，看各次测量之间的皮肤电位是否存在显著差异，并进行多重比较，就会知道实验结果是否支持理论假设。

SPSS 操作步骤如图 10-10 所示：

1.分析→一般线性模型→重复度量

2.“被试内因子名称”框：测量时间

3.“级别数”框：键入 5；添加

4.定义：“群体内部变量”(测量时间)：time1~time5

5.选项：（1）“显示均值”框：测量时间

（2）☑比较主效应

（3）继续

6.确定

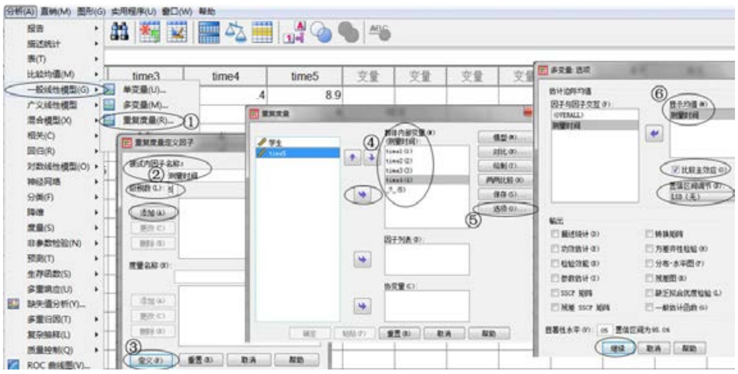


图 10-10　重复测量方差分析 SPSS 使用示意图

SPSS 分析结果中，最主要的有三张表：多元分析结果（Multivariate Tests，见表 10-28）、球形假设的检验结果（Mauchly’s Test of sphericity，见表 10-29）和被试内效应检验结果（Tests of Within-Subjects Effects，见表 10-30）。如果球形检验表明资料服从球形假设，则看被试内效应检验结果，如果不服从球形假设，则看多元分析结果。从表 10-29 可见，本例数据服从球形假设，所以，方差分析结果应看表 10-30。

表 10-28　多元检验

效果		值	F	假设 df	误差 df	显著性
测量时间	Pillai's Trace	.994	44.362[a]	4.000	1.000	.112
	Wilks' Lambda	.006	44.362[a]	4.000	1.000	.112
	Hotelling's Trace	177.449	44.362[a]	4.000	1.000	.112
	Roy's Largest Root	177.449	44.362[a]	4.000	1.000	.112

表 10-28 中给出了四个统计量：Pillai's Trace、Wilks' Lambda、Hotelling's Trace 和 Roy's Largest Root，都是多元方差分析中的统计量。这四个统计量检验结果有时会出现不一致，一般推荐以 Pillai's Trace 为最终结论。但本例中，数据服从球形假设，不以此表结论为最终结论。

表 10-29　Mauchly 的球形检验

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	Df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
测量时间	.011	10.994	9	0.383	.399	.623	.250

表 10-30　组内效应检验

来源		平方和	自由度	均方	F	显著性
测量时间	Sphericity Assumed	239.952	4	59.988	40.920	.000
	Greenhouse-Geisser	239.952	1.597	150.205	40.920	.000
	Huynh-Feldt	239.952	2.492	96.282	40.920	.000
	Lower-bound	239.952	1.000	239.952	40.920	.003
Error(测量时间)	Sphericity Assumed	23.456	16	1.466		
	Greenhouse-Geisser	23.456	6.390	3.671		
	Huynh-Feldt	23.456	9.969	2.353		
	Lower-bound	23.456	4.000	5.864		

表 10-30 对测量时间的效应进行检验的统计量也有四个：Sphericity Assumed、Greenhouse-Geisser、Huynh-Feldt 和 Lower-bound。从表中可见，不同测量时间的测量结果存在显著差异。表 10-31 的含义与前文相关内容相同，不再重复。

表 10-31　成对比较

(I) 测量时间	(J) 测量时间	均值差值 (I-J)	标准误差	显著性	95% 差异的置信区间[a]	
					下限	上限
1	2	2.260	.992	.085	-.495	5.015
	3	4.360*	.604	.002	2.684	6.036
	4	8.040*	.468	.000	6.742	9.338
	5	-.360	.430	.449	-1.553	.833
2	1	-2.260	.992	.085	-5.015	.495
	3	2.100	1.290	.179	-1.480	5.680
	4	5.780*	.806	.002	3.541	8.019
	5	-2.620*	.580	.011	-4.230	-1.010
3	1	-4.360*	.604	.002	-6.036	-2.684
	2	-2.100	1.290	.179	-5.680	1.480
	4	3.680*	.790	.010	1.488	5.872
	5	-4.720*	.822	.005	-7.003	-2.437

续表

(I) 测量时间	(J) 测量时间	均值差值 (I-J)	标准误差	显著性	95% 差异的置信区间[a]	
					下限	上限
4	1	-8.040*	.468	.000	-9.338	-6.742
	2	-5.780*	.806	.002	-8.019	-3.541
	3	-3.680*	.790	.010	-5.872	-1.488
	5	-8.400*	.402	.000	-9.517	-7.283
5	1	.360	.430	.449	-.833	1.553
	2	2.620*	.580	.011	1.010	4.230
	3	4.720*	.822	.005	2.437	7.003
	4	8.400*	.402	.000	7.283	9.517

三、混合模型的方差分析

（一）混合模型方差分析的含义

混合模型方差分析（mixed-model ANOVA）是指方差分析模型中包含被试间变量和被试内变量两种类型的自变量。

（二）混合模型的 SPSS 分析实例

【例 10-12】 假设某研究者想强化对老鼠走迷津行为的影响。为此，他从实验室的饲养的老鼠中随机抽取了 10 只老鼠，并随机分为两组。对于第一组（无强化组）的老鼠，研究者只是把它放入迷津而没有食物强化。对于第二组（有强化组），当老鼠到达目的地，则给予食物强化。研究者记录老鼠到达目的地之前改变行程方向的次数。研究者在四天内重复这一过程，所得数据结果如表 10-32 所示。问有无强化对老鼠走迷津行为有何影响？（本例题的原始数据见本书所附光盘中例 10-12.sav）

表 10-32　强化对老鼠走迷津行为的影响实验结果

分组	被试	第 1 天	第 2 天	第 3 天	第 4 天
无强化	1	10	9	10	10
	2	12	12	12	12
	3	9	9	10	12
	4	8	9	8	7
	5	11	11	10	10
有强化	6	12	3	9	3
	7	11	4	4	2
	8	8	7	3	4
	9	9	6	6	5
	10	10	8	10	4

解：SPSS 操作步骤如图 10-11 所示：

1. 分析→一般线性模型→重复度量
2. “被试内因子名称”框：天数
3. “级别数”框：键入 4，添加
4. 定义：（1）“群体内部变量”(天数)：第 1 天~第 4 天

    （2）“因子列表”：分组
5. 选项：（1）“显示均值”框：天数

    （2）☑比较主效应

    （3）继续
6. 确定

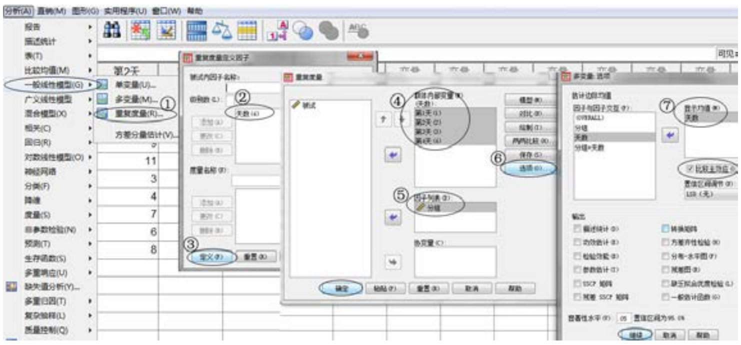


图 10-11　混合模型方差分析 SPSS 使用示意图

SPSS 分析结果中，可见数据服从球形假设（见表 10-33），所以，方差分析结果应看被试内效应检验结果（Tests of Within-Subjects Effects，见表 10-34）。结果表明，不同天数的实验结果之间存在显著差异，有无强化与实验天数之间存在交互效应。

表 10-33　Mauchly 的球形检验

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[a]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
天数	.697	2.426	5	.790	.793	1.000	.333

表 10-34　组内效应检验

来源		平方和	自由度	均方	F	显著性
天数	Sphericity Assumed	50.875	3	16.958	6.795	.002
	Greenhouse-Geisser	50.875	2.380	21.373	6.795	.004
	Huynh-Feldt	50.875	3.000	16.958	6.795	.002
	Lower-bound	50.875	1.000	50.875	6.795	.031
天数*分组	Sphericity Assumed	56.475	3	18.825	7.543	.001
	Greenhouse-Geisser	56.475	2.380	23.725	7.543	.003
	Huynh-Feldt	56.475	3.000	18.825	7.543	.001
	Lower-bound	56.475	1.000	56.475	7.543	.025
Error(天数)	Sphericity Assumed	59.900	24	2.496		
	Greenhouse-Geisser	59.900	19.043	3.146		
	Huynh-Feldt	59.900	24.000	2.496		
	Lower-bound	59.900	8.000	7.488		

本例中，还有被试间因素（有、无强化组），因此结果中还有一个重要的表格，即组间效应检验（Tests of Between-Subjects Effects，见表 10-35）。从表 10-35 可见，不同组之间的老鼠走迷津成绩存在显著差异。

表 10-35　组间效应检验

来源	平方和	自由度	均方	F	显著性
Intercept	2706.025	1	2706.025	412.347	.000
分组	133.225	1	133.225	20.301	.002
Error	52.500	8	6.563		

小　结

本章介绍了如何利用方差分析来检验多个总体或处理之间的平均数差异。本章介绍了单因素完全随机设计的方差分析模型、两因素效应可加模型和交互模型的方差分析模型。包括模型的意义、计算公式和 SPSS 操作。读者通过读详细介绍模型的学习，能够掌握方差分析的原理。另外对随机效应模型、重复测量模型和混合模型进行了简单介绍，未具体计算公式，仅介绍了枕木操作。读者通过对这些模型的学习，可以加深对方差分析思想的领会，后羿实际应用能力。

本章用单因素完全随机设计模型对方差分析的基本逻辑进行了较通俗的讲解，说明了用于方差分析的统计检验的基础是总体方差的两个独立估计量：组间估计和组内估计。这两个方差估计量之比服从 F 分布，利用 F 分布就可以作出推断，这两个方差估计量是否存在显著，从而推断出处理效应是否存在显著差异。

本章对方差分析的一般步骤提出了一个完整的框架，包括：（1）方差分析前的前提条件检验；（2）建立假设；（3）计算平方和、自由度、均方及 F 值；（4）统计决策；（5）列方差分析表；（6）方差分析效应大小和统计检验力的计算；（7）平均数的逐对比较。（8）参数估计。其中方差分析的前提条件检验的具体方法在本书前面有关章节已有介绍，方差分析效应大小和统计检验力的计算这一专题内容在本书第 11 章集中介绍，本章对其他各步骤进行了详细的介绍。

对于交互效应模型，如果因子间存在交互效应，则应进行简单效应分析。本章对简单效应分析的概念进行了讲解，并给出了 SPSS 进行简单效应分析的方法：首先，绘制简单效应分析图；然后，固定某因素做另一因素的单因素分析。

关键术语

方差分析：就是使用F检验方法来检验若干个具有相同方差的正态总体的期望（即总体平均数）是否相等的一种假设检验方法。

方差分析表：一种用来汇总方差分析计算和结果的表，包括显著变异来源、平方和、自由度、均方和F值的列。

平方和与自由度的分解：指将总平方和与总自由度分配给不同分量的过程。

多重比较方法：指在方差分析结果表明处理效应存在显著差异时，对成对的总体平均数进行统计检验的方法。

完全随机设计：指处理被随机地指派给实验单元的一种实验设计。

效应可加模型：指的是两因素模型中每种水平组合下只进行一次试验的实验设计。

交互作用：当一个因素的水平与另一个因素的水平发生作用时，对因变量产生的影响。

交互作用模型：指每种水平组合下进行多次实验，可以分析因素之间交互作用的实验设计。

固定因素：指的是因素水平的选取是出于设计者的“偏好”，而不是随机选定的。

固定效应的方差分析：指自变量是固定因素的方差分析。

随机因素：指的是因素水平是从大量的因素处理中随机选取的。

随机效应的方差分析：指的是自变量是随机因素的方差分析。

被试间设计：也称为独立样本设计或组间设计，指的是模型中的自变量各水平下的被试是由不同个体组成的实验设计。

被试内设计：也称为相关样本设计或组内设计，指的是模型中的自变量各水平下的被试是同一批个体组成的实验设计。

混合模型方差分析：是指方差分析模型中包含被试间变量和被试内变量两种类型自变量的方差分析。

重要公式

单因素完全随机设计的方差分析

总平方和：$SS_T=\sum_{j=1}^{K}\sum_{i=1}^{n}(y_{ij}-\bar{Y})^2$；

组间平方和：$SS_B=n\cdot\sum_{j=1}^{K}(\bar{Y}_j-\bar{Y})^2$；

组内平方和：$SS_w = \sum_{j=1}^{K}\sum_{i=1}^{n}(y_{ij} - \bar{Y}_j)^2$

总自由度：$df_T = n_T - 1$

组间自由度：$df_B = K - 1$

组内自由度：$df_W = n_T - K$

组间均方：$MS_B = \frac{SS_B}{df_B}$

组内均方：$MS_W = \frac{SS_W}{df_W}$

F 检验统计量：$F = \frac{MS_B}{MS_W}$

平均数间的多重比较

费希尔的 LSD 检验统计量：$t = \frac{\bar{Y}_i - \bar{Y}_j}{\sqrt{MSE(\frac{1}{n_i} + \frac{1}{n_j})}}$

Scheffé 检验法的检验统计量：$F = \frac{MS_{ij}}{MS_W}$ ， $i \neq j$

效应可加模型的方差分析

总平方和：$SST = \sum_{i=1}^{p}\sum_{j=1}^{q}(y_{ij} - \bar{Y}_{\cdot\cdot})^2$

A 因素平方和：$SSA = q \cdot \sum_{i=1}^{p}(\bar{Y}_{i\cdot} - \bar{Y}_{\cdot\cdot})^2$

B 因素平方和：$SSB = p \cdot \sum_{j=1}^{q}(\bar{Y}_{\cdot j} - \bar{Y}_{\cdot\cdot})^2$

误差平方和：$SSE = SST - SSA - SSB$

交互效应模型的方差分析

总平方和：$SST = \sum_{i=1}^{p}\sum_{j=1}^{q}\sum_{k=1}^{r}(y_{ijk} - \bar{Y}_{\cdot\cdot\cdot})^2$

A 因素平方和：$SSA = qr \cdot \sum_{i=1}^{p}(\bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$

B 因素平方和：$SSB = pr \cdot \sum_{j=1}^{q}(\bar{Y}_{\cdot j\cdot} - \bar{Y}_{\cdot\cdot\cdot})^2$

A 与 B 交互作用平方和：

$$SSAB = r \cdot \sum_{i=1}^{p}\sum_{j=1}^{q}(\bar{Y}_{ij\cdot} - \bar{Y}_{i\cdot\cdot} - \bar{Y}_{\cdot j\cdot} + \bar{Y}_{\cdot\cdot\cdot})^2$$

误差平方和：$SSE = SST - SSA - SSB - SSAB$

思考与练习

1. 下面的方差分析表是一个四水平单因素（每个处理的被试 n=10）实验数据分析结果，请将该方差分析表缺失值补充完整。

变异来源	平方和	自由度	均方	*F* 值
组间	___	___	22	*F*=___
组内	___	___	___	
总变异	138	___		

2. 某制药公司制造了一种减少饥饿感的药剂。为了检验该药剂的有效性，将一批老鼠随机每为 3 组，每组 12 只。第一组每天都喂食一次该药剂，第二组每周喂食一次该药剂，第三组从不喂食该药剂。因变量是每只老鼠一个月来的吃食总量。该研究数据的方差分析结果如下表，请将表中缺失值补充完整。

变异来源	平方和	自由度	均方	*F* 值
组间	___	___	___	*F*=6
组内	99	___	___	
总变异	___	___		

3. 一研究者研究了成就动机和任务难度对问题解决的影响。他采用了 2×4 的实验设计：成就动机包括两个水平（高成就动机、低成就动机），任务难度包括四个水平，共 8 个处理。每个处理随机分派 6 名被试。记录被试解决问题中的错误次数作为因变量。该研究数据的方差分析结果如下表，请将表中缺失值补充完整。

变异来源	平方和	自由度	均方	*F* 值
因素 *A*(成就动机)	16	___	___	*F*=___
因素 *B*(任务难度)	___	___	48	*F*=___
交互作用 *A*×*B*	120	___		*F*=___
误差 *E*	___	___		
总变异	600	___		

4. 家庭中长子（女）的语言技能，往往比后出生孩子有更快的发展。这种现象的一个可能的解释就是，长子（女）得到了父母专一的关注。如果这个解释是正确的，那么可以合乎逻辑地推理，第一胎的双胞胎孩子比单一子女的语言发展要慢，三胞胎则更慢。Davis（1937）的研究发现确实如此。下表的数据是这一关系的虚拟数据。因变量是每个孩子在三岁时的语言技能测验得分。问这一数据是否能证明上述假设。（α=.05）

单一子女	双胞胎	三胞胎
8	4	4
7	6	4
10	7	7
6	4	2
9	9	3

5. 心理学家想知道人工甜剂是否会增加老鼠按压杠杆的动机。他将老鼠随机分派到五组中的一组。老鼠在按压杠杆 10 次后，将得到混合了不同浓度甜剂的水（五组的混合浓度分别为 0%、4%、8%、16%、32%）。经过三天的训练，研究者记录了每只老鼠在 20 分钟之内按压杠杆的次数，结果如下表。问不同浓度的甜剂对老鼠按压杠杆的次数是否有影响。（α=.05）

组 1	组 2	组 3	组 4	组 5
0%	4%	8%	16%	32%
16	21	20	54	56
5	11	15	54	45
13	19	22	46	46
7	14	27	46	51
9	15	26	50	51
10				51

6. 在短文中，一些词对于理解短文的含义具有特殊重要的意义，而另外一些词则不具备这一特点。因此，我们可以假设一位好的阅读者在阅读时会首先阅读这些重要的（或关键的）词，于是在短文中，如果仅仅是这些词被大写，那么能提高阅读速度。为证明这一观点，有研究者随机抽取了 30 名被试，并随机分为 3 组，阅读同一篇短文，但印刷的形式不同：第一组阅读一般印刷的短文，第二组阅读有一些词被随机设置为大写形式的短文，第三组阅读关键词是大写形式的短文。因变量是阅读短文的时间（单位：秒）。结果如下表。问该研究结果是否支持上述假设？

	n	平均时间	标准差
第一组	10	30.2	6.21
第二组	10	38.3	7.55
第三组	10	25.6	5.75

7. 请在 α＝.05 的水平上作出判定，下列数据是否表明三个处理的平均数存在显著差异。

第 1 组	第 2 组	第 3 组
$n=4$	$n=5$	$n=6$
$\overline{X}=0.5$	$\overline{X}=2$	$\overline{X}=3$
$S^2=6$	$S^2=10$	$S^2=12$

8. 已有研究证明，当被试记忆一个单词序列时（单词依顺序呈现），对位于开始部分的单词和位于序列末尾的单词的记忆效果，要好于位于序列中间的单词，这种效应被称为序列位置效应。下表数据呈现了四个被试回忆最初识记的八个单词、中间的八个单词和最后的八个单词时的错误个数。请在 α＝.05 的水平作出判定，该数据是否支持序列位置效应。

	序列位置		
被试	最初	中间	末尾
A	1	5	0
B	3	7	2
C	5	6	1
D	3	2	1

9. 有人设计一个实验考察了不同人格类型个体的行为表现的观众效应。实验按人格类型将个体分为“高自尊”和“低自尊”两种，并随机选取“高自尊”和“低自尊”被试各 12 名，每种自尊水平的被试又被分为两组（每组 6 人），随机接受一种任务处理（单独完成任务和有观众条件下完成任务）。因变量是每个被试完成任务时的错误次数。实验结果如下表。请在 α＝.05 的水平上检验不同人格类型的个体，在有无观众条件下的行为表现是否存在显著差异？

任务情境 / 人格类型	单独完成	有观众
高自尊	3、6、2、2、4、7	9、4、5、8、4、6
低自尊	7、7、2、6、8、6	10、14、11、15、11、11

10．选择数据库管理软件的一个重要因素是需要用多长时间学会使用该系统。为了评价 3 种数据库管理系统，一家公司设计了一个有 6 位数据库管理员参加的测试。因为管理员的个体差异对学习时间有显著影响，于是让每位管理员都学习使用 3 种管理系统，学习的顺序进行了平衡。实验结果如下表（学习时间单位：小时）。请在 $\alpha=.05$ 的水平下，检验 3 种管理系统平均学会使用的时间是否存在显著差异。

	系统 A	系统 B	系统 C
管理员 1	15	15	18
管理员 2	14	14	15
管理员 3	10	11	15
管理员 4	13	12	17
管理员 5	16	13	16
管理员 6	13	13	14

11．为了考察小学生使用计算器对算术能力是否有不良影响，研究人员从小学二年级经常使用计算器的学生中随机抽取了 5 名学生，另外随机抽取了 5 名未使用计算器的学生，让每位学生接受 3 个测验（加法、减法和乘法），每个测验 10 分，要求他们尽可能快心算出结果。实验结果如下表。该实验结果是否支持孩子使用计算器。

被试		加法	减法	乘法
使用计算器的学生	1	8	5	3
	2	7	5	2
	3	9	7	3
	4	6	3	1
	5	8	3	1
不使用计算器的学生	6	10	7	6
	7	7	6	5
	8	6	5	5
	9	9	7	8
	10	9	6	9

第十一章 统计检验力和效果大小的估计

学习本章内容，将有助于你对以下问题的理解和思考：

1. 统计检验力和效果大小的含义是什么？

2. 通过 Z 检验、t 检验、方差分析等推断统计方法对实验数据进行平均数差异显著性检验后，如何对其统计检验力与效果大小进行估计？

某心理学研究者试图对某校新生在智商方面是否存在性别差异进行比较研究的过程中，从男女新生中各随机抽取 100 名学生进行韦氏智力量表测试后，测得女生平均智商为 115 分，男生平均智商为 111 分。根据常模，男女生总体智商的标准差都是 15 分。请对该校新生中男女生在智商方面进行平均数差异显著性检验后分别计算：

（1）该次差异显著性检验的效果大小（effect）是多少？

（2）在显著性水平$\alpha = 0.05$ 的条件下，该次差异显著性检验的统计检验力是多少？

（3）在显著性水平$\alpha = 0.10$ 的条件下，该次差异显著性检验的统计检验力是多少？

第一节 平均数差异显著性的统计检验力和效果大小的估计

一、统计检验力和效果大小的含义

由前面几章的内容可知，通过 Z 检验、t 检验、F 检验或 X2 检验等推断统计分析方法，我们可以得到几个样本平均数（或几个样本方差、几个样本相关系数、几个样本比率或实计数等）之间的差异是否达到显著水平的假设检验结果。但是，上述按常规心理统计方法得到的假设检验结果并不能让我们了解它们之间的差异到底有多大，其差异显著性有多重要。

以两总体平均数差异显著性 Z 检验为例，第六章例 6-7 论述了某研究者对男女新生智商差异情况的研究，检验结果是在$\alpha = 0.05$的水平下，由于实得的统计量$Z = 1.89$小于临界值 1.96，故接受虚无假设，推论该校男女新生在智商方面没有显著性差异。就常规统计而言，假设检验做到这里就算是完成了，但是，现代心理学研究还希望研究者能提供如下的信息：“当虚无假设是假的情况下，我们犯‘纳伪’错误的可能性有多大？”

第六章例 6-8 论述了某位研究者对两个省小学一年级语文成绩有无显著差异的研究，通过常规假设检验过程得到的结果是，在$\alpha = 0.01$的显著性水平上进行推断，这两个省小学一年级语文成绩存在显著差异。现代心理学研究还希望研究者能在此基础上再提供如下两种信息：（1）“做出这一推断的把握度有多大”；（2）“两总体的真实差异有多大”。

“统计检验力”和“效果大小”这两个统计指标可以为我们提供上述常规心理统计学没有要求，但现代心理学研究所需要提供的信息。而要理解这两个概念的含义，我们首先需要复习一下第五章提到的两种假设、两类错误及其相互关系的原理。

所谓两种假设是指“虚无假设（通常用符号‘H_0’表示）”和“备择假设（或称为‘研究假设’，通常用符号‘H_1’表示）”。我们对平均数进行差异显著性检验时，通常将检验的虚无假设设为$H_0 : \mu_1 = \mu_2$，而将虚无假设的反面设为$H_1 : \mu_1 \neq \mu_2$。

所谓两类错误是指我们在“虚无假设为真”或“虚无假设为假”这两种不同前提下进行统计推断时可能犯的两种不同性质的错误：当我们在“虚无假设为真”的前提下根据统计结果做出拒绝“虚无假设”的错误推断时，通常把这种错误称之为“一类错误（通常用符号‘α’表示）”；当我们在“虚无假设为假”（也就是“备择假设为真”）的前提下根据统计结果做出接受“虚无假设”的错误推断时，通常把这种错误称之为“二类错误（通常用符号‘β’表示）”。

为了让读者更好地理解两种假设和两类错误的相互关系，我们把表 5-2 和图 5-4 再列如下（表 11-1 和图 11-1）。

表 11-1（表 5-2）假设检验的各种可能结果

	接受 H_0	拒绝 H_0
H_0 为真	正确决策，概率=1-α（置信度）	第一类错误，概率=α
H_0 为假	第二类错误，概率=β	正确决策，概率=1-β（统计检验力）

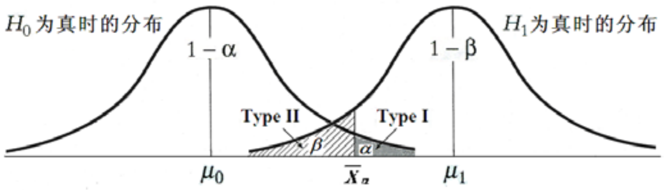


图 11-1（图 5-4） α 和 β 关系示意图

由表 11-1 和图 11-1 可知，假设检验的结果表明无论是拒绝或者不拒绝虚无假设，我们都有可能或者犯 α 错误或者犯 β 错误，并且在其他条件不变的情况下 α 和 β 不可能同时增大或减小。

前面第 5 至第 10 章论述各种平均数差异显著性检验时，都是在“虚无假设为真”的假定上进行的，因此，如果我们在抽样过程中能够把犯一类错误的可能性控制在小范围之内，例如把它控制在“100 次抽样过程中犯错误的可能性不超过 5 次”（即“α=0.05”），那么我们就可以在“一次抽样过程中‘小概率事件不可能出现’”这样的理论观点指导下，根据对抽样结果计算的统计量来进行统计推断：若该统计量低于“α=0.05”的临界值，我们就没有理由拒绝虚无假设，由此推断 μ_1 和 μ_2 相等，或者说“$\overline{X}_1$”和“$\overline{X}_2$”来自同一个总体。在“虚无假设为真”的前提下做出这种推断正确的可能性概率就是表 11-1 中所列的“1-α”。

现代心理统计学的发展要求我们在进行统计推断时，即要考虑“虚无假设 H_0 为真”的假定，也要考虑“虚无假设为假”（即“备择假设 H_1 为真”）的假定。由表 11-1 可知，当虚无假设为假时，如果我们根据对抽样结果计算的统计量做出接受 H_0 的推断就会犯 β 型错误；拒绝 H_0，则是做出了正确的决策，做出这种推断正确的可能性概率就等于 1-β。换言之， 我们是在“$1-\beta$”的概率基础上做出“拒绝虚假的虚无假设”的统计决策的，这就是现代心理统计学中“统计检验力（power of test，或称效力）”的含义。用 Cohen（1988）的话来说，“在虚无假设为假时…对一个虚无假设的统计检验力是指将引导我们拒绝这个虚无假设的概率”。根据 Cohen 的上述观点，本书把统计检验力（$1-\beta$）定义为：“在虚无假设（H_0）为假（也即备择假设（H_1）为真）时，正确拒绝这一虚假的虚无假设的概率”。

另一方面，当“虚无假设 H_0”为假时，也就意味着“备择假设 H_1” 为真。在虚无假设被设为“$H_0: \mu_1 = \mu_2$”时，备择假设通常就被设为“$H_1: \mu_1 \neq \mu_2$”，它表明 μ_1 和 μ_2 不相等，或者说这次抽取的“$\overline{X}_1$”和“$\overline{X}_2$”来自两个不同的总体。现代心理统计学用“效果大小（effect size，有时也被译为‘效应大小’或‘效应值’）”这一概念来反映“μ_1”和“μ_2”之间存在多大的差距（考虑到“效应”一词在心理学研究中的很多地方都使用，如实验设计和

方差分析中的主效应、各因素效应等，为了不使读者产生不必要的混淆,本书用“效果大小”一词来表示“effect size”的内涵）。Cohen（1988）把效果大小定义为“当虚无假设为假时，它总是在一定程度上的虚假。效果大小是指某个特定总体中的那些非零的数值,这个数值越大，就表明由研究者所处理的研究现象所造成的效果越大”;“效果大小本身可以被视为是一种参数：当虚无假设为真时，效果大小的值为零；当虚无假设为假时，效果大小为那些非零的值。因此，可以把效果大小视为某种与虚无假设分离程度的指标。”根据 Cohen 的上述观点，本书把效果大小定义为:“在虚无假设（H_0）为假时,对该次假设检验过程中所存在的备择假设（H_1）的期望值与虚无假设的期望值之间的距离的测量值”。

“效果大小”是反映心理学研究中所控制的自变量对因变量的影响效果有多大的重要指标，它表示不同处理下的总体平均数之间差异的大小，可以在不同研究之间进行比较。效果大小反映了两个总体受某种事物的影响的差异程度。两样本平均数差异显著性 Z 检验(或 t 检验)的效果大小一般用符号“d”（有时也称为“Cohen d”）表示。

我们也可以通过把d值视为两个总体分布的重叠量的途径，来理解效果大小的内涵。图 11-2 列出了四种d值在两个总体中的重叠情形。由图 11-2 可知，效果大小的d值越大，重叠程度就越小，平均数差异显著性检验的效果就会越明显；d值越小则相反。我们也可以这样来理解，即不管你取哪种样本，d值总是作为一种标准的平均数差异的估计，它与当前样本无关。

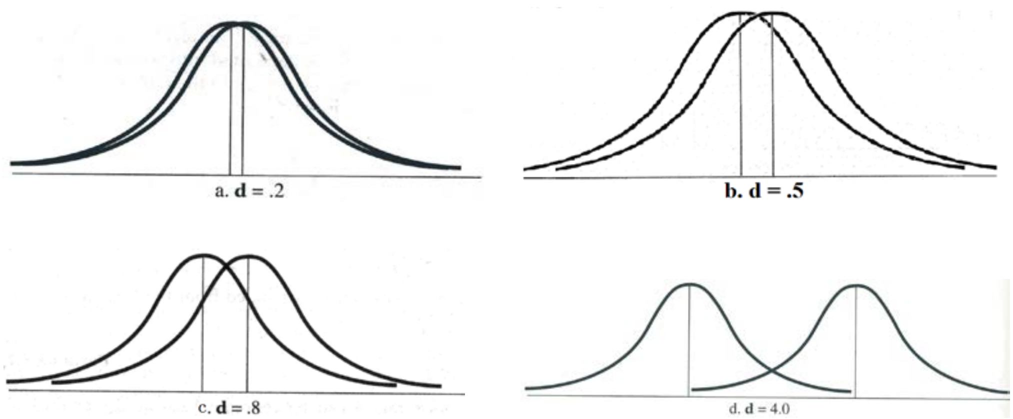


图 11-2　作为效果大小函数的两总体重叠图

二、两独立样本平均数差异显著性检验统计检验力的估计原理和估计方法

（一）统计检验力的估计原理

如前所述,当我们试图通过两总体平均数差异显著性检验过程来了解从这两个总体中所抽取的两个样本平均数之间的差异是否达到统计学意义上的显著水平的假设检验过程中,通常把“虚无假设H_0”设为“$\mu_1 = \mu_2$”时，把“备择假设H_1”设为“$\mu_1 \neq \mu_2$”。如果假设检验的结果让我们推断“虚无假设H_0”为真，意味着我们在（1-α）的置信概率上推断这次抽取的两个样本平均数是来自同一个总体；反之，如果假设检验的结果让我们推断“备择假设H_1”

为真时，意味着我们推断这次抽取的两个样本平均数是来自“μ_1”和“μ_2”相互之间有一定差距的两个不同总体，现代心理统计学告诉我们，做这种推断的把握度（即概率值）是通过（1-β）这一指标来反映的。要了解如何计算（1-β）的概率值，我们需要在理解第五章所论述的“总体分布”、“样本分布”、“抽样分布”以及第六章所论述的“虚无假设分布”这四种分布的基础上，进一步理解“备择假设分布”这一概念。

第六章曾指出：虚无假设分布是指在对两总体平均数差异显著性进行假设检验过程中，通常将检验的虚无假设设为H_0：$\mu_1=\mu_2$，这时，所谓虚无假设分布就是指当虚无假设H_0为真时，我们在这两个总体中每次各抽取一定数量的样本进行差异显著性检验，在进行无数多次抽样后，每次抽取的两个样本平均数之差（即“$\overline{X}_1-\overline{X}_2$”之差）的分布。它是一种以零为中心的正态分布。

与虚无假设分布相对应，所谓备择假设分布（alternative hypothesis distribution, AHD），就是指当虚无假设H_0为假而备择假设H_1为真时，我们在这两个总体中每次各抽取一定数量的样本进行差异显著性检验，在进行无数多次抽样后，每次抽取的两个样本平均数之差(即“$\overline{X}_1-\overline{X}_2$”之差）的分布。

例如，我们可以用“μ_1”来代表男性人类个体的平均身高，用“μ_2”来代表女性人类个体的平均身高，那么，根据常识，男人和女人在身高方面总是有一定差距的，虽然我们不知道差距有多大，但总的趋势是男性高于女性；此外，常识还告诉我们两性的身高差异在不同年龄段中是不一样的。

如果我们要在某年龄段男生（总体平均身高等于“μ_1”）和女生（总体平均身高等于“μ_2”）这两个总体中通过随机抽样各抽取一定数量的样本（例如 n=100）来进行假设检验，又假定这两个总体的平均身高相差为 3 厘米，那么，“虚无假设H_0”通常被设为“$\mu_1=\mu_2$”，“备择假设H_1”通常被设为“$\mu_1\neq\mu_2$”。我们的做法通常是：

在男生中抽取 100 人进行身高测试后求得平均身高“$\overline{X}_{1-1}$”，在女生中抽取 100 人进行身高测试后求得平均身高“$\overline{X}_{1-2}$”，由此可以求得这次抽取的两个样本平均数之差（用符号“D”表示）为：“$D_1=\overline{X}_{1-1}-\overline{X}_{1-2}$”（注：下标中第一个数字是指第几次抽样，第二个数字表示是从哪个总体抽取的样本）；

我们还可以继续在这一年龄段的男女生中进行抽样，即：在男生中抽取 100 人进行身高测试后求得平均身高“$\overline{X}_{2-1}$”，在女生中抽取 100 人进行身高测试后求得平均身高“$\overline{X}_{2-2}$”，由此可以求得第二次抽取的两个样本平均数之差为：“$D_2=\overline{X}_{2-1}-\overline{X}_{2-2}$”（注：下标中第一个数字是指第几次抽样，第二个数字表示是从哪个总体抽取的样本）；

理论上，我们可以无数次地重复这样的抽样过程，如果我们这样做了 n 次，那么就可以得到 n 个两样本平均身高之差，即D_1，D_2，……D_n，如果我们把这 n 个两样本平均身高之差相加，结果会是怎样的呢？心理统计学家告诉我们，每次抽样过程中，所抽取的两样本平均身高之差实际上包括两个部分：一部分是“系统误差”，即这两个总体平均数之间实际存在的身高差距；另一部分是随机误差，即由“每次随机抽样过程中所抽取的个体是不一样”而造成的

误差。相关的概率理论告诉我们，当抽样次数足够多时，随机误差会趋向于零，因而从这两个总体中各抽取一定样本的平均数之差会稳定在“系统误差”上。

上述就是“备择假设分布”的含义。

遗憾的是，由于与虚无假设相对立的备择假设H_1 ：$\mu_1 \neq \mu_0$是虚无假设H_0 ：$\mu_1 = \mu_0$的补集，两个平均数之间不相等的值几乎是无限多的，因此，它们之间的差值到底是多少是不确定的。在理论上，备择假设H_1 ：$\mu_1 \neq \mu_0$的分布不是一个以零为中心的分布（a noncentral Z or t distribution），它的中心值也有着无数多个选择，因此，一般情况下，我们是难以对β或（$1-\beta$）值作出准确的估计的。如果我们需要对β或（$1-\beta$）的值作出估计，我们就还需要一些统计理论支持。现代心理统计学指出，我们可以在以下两个假定基础上来对某次假设检验过程中可能犯二类错误β的概率值或统计检验力（$1-\beta$）的概率值进行估计：

假定一：假定该次假设检验中的“备择假设分布”服从正态分布，并且用符号“δ”来表示这一“备择假设正态分布”的期望值；

假定二：假定该次“备择假设正态分布”的期望值与这次假设检验的“Z 统计量（或 t 统计量）”相等。

上述二个假定可以做这样的解读：虽然在一般情况下，由于不知道备择假设分布的期望值是多少因而无法精确计算β或（$1-\beta$）的值，但是，我们在进行以“两个平均数差异是否显著”为目的的假设检验过程中，总会得到一个 Z（或 t）统计量。如果该次检验的备择假设服从正态分布，在没有其他可用信息的情况下，我们有理由认为“通过在两个总体中进行随机抽样得到的假设检验统计量 Z 值，应该最有资格来代表这次假设检验过程中备择假设分布的中心点”。换言之，我们可以用本次假设检验所得到的 Z（或 t）值作为本次假设检验过程中备择假设分布的期望值（the expected Z value，或称为备择假设分布的平均数，通常用希腊字母“δ”表示备择假设分布的期望 Z 值）。如果本次假设检验的备择假设分布确实服从正态分布，那么在同样的“μ_1”和“μ_2”两个总体中各抽取与本次假设检验过程中相同数量的样本继续进行多次，那么，对每次抽取的两个样本平均数之差（即“$\overline{X}_1 - \overline{X}_2$”之差）进行假设检验后得到的 Z 统计量的值，大于这次 Z 统计量的数量与小于这个 Z 统计量的数量会各占 50%。

（二）统计检验力或犯二型错误的概率值的主要估计方法

借助上述两个假定，我们就可以通过对统计检验力（$1-\beta$）或犯二型错误β的概率值进行近似的估计了。心理统计学家提供了多种估计方法，本书主要介绍其中两种：（1）利用正态分布表求取；（2）利用 δ 和 α 值直接查表求取。下面我们分述之。

方法一：利用正态分布表求统计检验力或犯二型错误的概率值

根据上述统计检验力的估计原理，我们可以根据以下几个步骤，来估计两个独立样本平均数差异显著性检验之后该检验可能存在的“二类错误β值”（当假设检验的结果是接受虚无假设时）或“统计检验力 1-β的值”（当假设检验的结果是拒绝虚无假设时）（注：t检验及其

他条件的 Z 检验可以参照本节内容计算其统计检验力或效果大小）：

（1）根据已知条件建立需要检验的假设；

（2）用相应的公式计算 Z 统计量；

（3）确定做出统计决策的α水平及相应的临界值；

（4）计算实际得到的 Z 值与α水平临界值的差；

（5）根据 Z 值与α水平临界值的差查正态分布表,确定可能犯的β型错误或统计检验力 1-β的概率。

【例 11-1】根据【例 6-7】的数据，按上述方法 1 的五个步骤求$\alpha = 0.05$时平均数假设检验（双侧）的“二类错误”或“统计检验力”的概率值。

解：在例 6-7 中，已知的条件是：

$$n_1 = 100, \overline{X}_1 = 115, \sigma_1 = 15$$
$$n_2 = 100, \overline{X}_2 = 111, \sigma_2 = 15$$

求解过程重复如下：

（1）建立假设：
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

（2）用公式 11-1（即公式 6-10）计算检验统计量

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{115 - 111}{\sqrt{\frac{15^2}{100} + \frac{15^2}{100}}} = \frac{4}{2.12} = 1.89$$

（3）确定显著性水平并做出决策

通常我们是在显著性水平$\alpha = 0.05$的基础上进行双侧检验，这时作出接受或拒绝虚无假设的临界值为$\left|Z_{\frac{\alpha}{2}}\right| = 1.96$，以此为分界点，当实际得到的 Z 值小于 1.96 时，就认为没有充分理由拒绝虚无假设，反之则拒绝虚无假设。由于本次假设检验结果为Z=1.89 小于$\left|Z_{\frac{\alpha}{2}}\right|$，据此做出的统计推论是“没有充分理由去拒绝虚无假设，因而接受虚无假设。由此推断该校男女生在智商方面不存在显著差异,男女生两个样本平均数之间所存在的 4 分差异是由于抽样过程中的偶然因素造成的”。

上述三步就是在本书第六章中**【例 6-7】**所介绍的两独立样本平均数差异显著性检验的全过程。一般情况下，假设检验的过程至此已经完成。但是，上述推断只能在“虚无假设为真”的前提下成立，而在假设检验过程中我们是不知道“虚无假设是真还是假”，如果这次假设检验是在“虚无假设为假”的前提下进行的话，根据假设检验结果做出“该校男女生在智商方面不存在显著差异”的统计推断就可能犯“二类错误即β型错误（也称为“纳伪”错误）”，如果我们想知道犯β型错误的概率值是多少，则还需要做以下两步工作：

（4）计算实际得到的 Z 值与α水平临界值的差，得到

1.89-1.96=-0.07；

（5）根据 Z 值与α水平临界值的差查正态分布表，确定可能犯β型错误的概率

查正态分布表可知，从中心点为零到右边 0.07 个标准差所占的面积是 0.0279，加上中心点左边的 0.50 的面积，共有曲线下面积 0.50 + 0.0279＝0.5279，约等于 53%（如图 11-3 所示）。这就是本例情况下估计出来的犯β型错误的概率值。

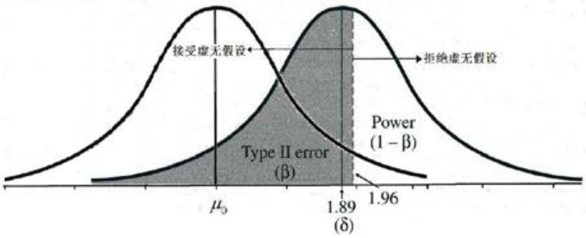


图 11-3 本例中的备择假设分布图

答：本例中$\alpha = 0.05$时平均数假设检验的结果是“没有充分理由去拒绝虚无假设因而接受虚无假设。由此推断该校男女新生在智商方面不存在显著差异”。如果这种推论是错误的，那么犯“二类错误”的概率是β=0.53。

【例 11-2】根据【例 6-8】的数据，按上述方法 1 的五个步骤求$\alpha = 0.05$时平均数假设检验（双侧）的“二类错误”或“统计检验力”的概率值。

解：在例 6-8 中，已知的条件是：

$$n_1 = 180, \overline{X}_1 = 82, \sigma_1 = 11.5$$
$$n_2 = 160, \overline{X}_2 = 78.42, \sigma_2 = 10.5$$

求解过程重复如下：

（1）建立假设：
$$H_0 : \mu_1 = \mu_2$$
$$H_1 : \mu_1 \neq \mu_2$$

（2）用公式 11-1（即公式 6-10）计算检验统计量

$$Z = \frac{(\overline{X_1} - \overline{X_2}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(82 - 78.42) - 0}{\sqrt{\frac{11.5^2}{180} + \frac{10.5^2}{160}}} = \frac{3.58}{1.1932} = 3.0$$

（3）确定显著性水平并做出决策

就像在例 11-1 中所述的那样，通常我们是在显著性水平$\alpha = 0.05$的基础上进行双侧检验， 这时作出接受或拒绝虚无假设的临界值为$\left|Z_{\frac{\alpha}{2}}\right| = 1.96$（当然我们也可以在显著性水平$\alpha$=0.01 的基础上进行双侧检验，这时作出接受或拒绝虚无假设的临界值则是$\left|Z_{\frac{\alpha}{2}}\right| = 2.58$），以此为分界点，当实际得到的 Z 值小于 1.96 时，就认为没有充分理由拒绝虚无假设，反之则拒绝虚无假设。由于本次假设检验结果为Z=3 大于$\left|Z_{\frac{\alpha}{2}}\right|$，据此做出的统计推论是“拒绝虚无假设$H_0$，推断这两个省小学一年级语文成绩存在显著差异”。

上述三步就是在本书第六章中**【例 6-8】**所介绍的两独立样本平均数差异显著性检验的全过程。一般情况下，假设检验的过程至此已经完成。但是，就像在例 11-1 所述的推论过程那

样，在假设检验过程中我们并不清楚我们预设的“虚无假设”是真还是假：如果这次假设检验是在“虚无假设为真”的前提下进行的话，那么我们根据这次假设检验结果做出“拒绝虚无假设 H_0，推断这两个省小学一年级语文成绩存在显著差异”的统计推断就犯了“一类错误即 α 型错误（也称为“拒真”错误）”，可能犯此类错误的概率值就是预设的 α 值，在本例中是指“在这两个省的小学一年级学生中用同样方法抽样进行平均数差异检验，得到本次假设检验结果的次数不会超过 5%”；相反，如果这次假设检验是在“虚无假设为假”的前提下进行的话，那么我们根据这次假设检验结果做出“拒绝虚无假设 H_0，推断这两个省小学一年级语文成绩存在显著差异”的统计推断就是正确的推断。如果我们想知道我们有多大把握度来推断“这次假设检验的虚无假设为假（通过统计检验力 $1-\beta$ 的概率值反映）”，则还需要做以下两步工作：

（4）计算实际得到的 Z 值与 α 水平临界值的差，得到

$$3-1.96=1.04;$$

（5）根据 Z 值与 α 水平临界值的差查正态分布表，确定统计检验力 1-β 的概率

查正态分布表（附表 1）可知，从中心点为零到左边 1.04 个标准差所占的面积约为 0.35，加上中心点左边的 0.50 的面积，共有曲线下面积 0.50 + 0.35＝0.85，（如图 11-4 白色区域所示）。这就是本例情况下估计出来的“在虚无假设为假时正确拒绝虚无假设的概率值”，即统计检验力 $1-\beta$ 的概率值。

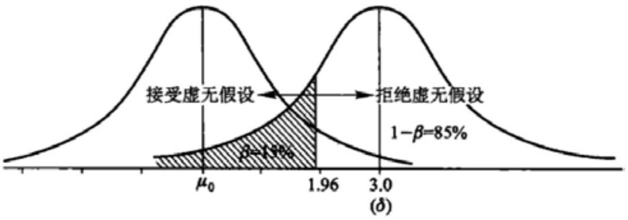


图 11-4 本例中的备择假设分布图

答：两省小学一年级语文成绩存在显著差异，做出这一推断的统计检验力为：$1-\beta=0.85$。

方法二：利用 δ 和 α 值直接查表求统计检验力或犯二型错误的概率值

第二种评估两个独立样本平均数差异显著性检验之后该检验可能存在的“二类错误 β 值”或“统计检验力 1-β 的值” 的概率值的方法是表格换算法，换算的根据之一是备择假设分布的期望值 δ，根据之二是显著性水平 α **值**。

如前所述，通常情况下我们是不知道 δ 值是多少的，但是根据前述统计检验力估计原理的第二个假定，即令“δ”=“Z”，因此 δ 值的计算方法就与 Z 值的计算方法有密切关联了，实际上，两个独立样本平均数差异显著性检验后其备择假设分布的期望值 δ 的计算方法就是以 Z 统计量的计算公式为基础的（参见公式 6-10）。为更好地理解 Z 值与 δ 值的关系，现将公式 10-6 再列如下：

$$Z=\frac{(\overline{X}_1-\overline{X}_2)-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$ （公式 11-1，即公式 6-10）

请读者注意：虽然统计检验力估计原理的第二个假定是令“δ”=“Z”，但求“δ”的计算公式却与公式 11-1 所示的求“Z”值的公式不完全一样，这是因为我们在进行两个独立样本平均数差异显著性检验 Z 统计量的过程中，通常是在把虚无假设设为“$\mu_1 = \mu_2$”并假定虚无假设为真的前提下进行，虚无假设真就意味着“$\mu_1 - \mu_2 = 0$”，因此，实际计算 Z 值的公式就成为：

$$Z=\frac{(\overline{X}_1-\overline{X}_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$ （公式 11-2）

但是，当我们在虚无假设为假的前提下进行两个独立样本平均数差异显著性检验时，就意味着存在一个与虚无假设有一定距离的“备择假设”并且我们假定该备择假设的分布服从期望值为“δ”的正态分布，在公式 11-1 中，$\overline{X}_1$是指从其所属的总体μ_1中所抽取一个样本的平均数，同样，$\overline{X}_2$是指从其所属的总体μ_2中所抽取一个样本的平均数，而备择假设分布的期望值实质上是指“μ_1”和“μ_2”之间的距离，因此，计算两个独立样本平均数差异显著性检验的δ值的公式就有如公式 11-2 所示：

$$\delta=\frac{(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$$

（公式 11-3）

“δ”只是理论上存在的“备择假设分布”的期望值，在实际操作过程中，我们只需要通过常规的假设检验过程求出“Z”统计量的值，令“δ”=“Z”就可以了。

有了“δ”值，我们就可以根据附表 10 通过δ值和做出决策所定的α水平直接求出统计检验力的值或可能犯二类错误的值了。

附表 10 的结构有如表 11-2 所示（附表 11 则是作为统计检验力与显著性水平的函数的δ值表）。

表 11-2 作为δ与α水平函数的统计检验力表

δ	单侧检验（α）			
	0.05	0.25	0.01	0.05
	双侧检验（α）			
	0.10	0.05	0.02	0.01
0.5	.14	.08	.03	.02
0.6	.16	.09	.04	.02
……	……	……	……	……
1.8	.56	.44	.30	.22
1.9	.60	.48	.33	.25
……	……	……	……	……
4.0	.99	.98	.95	.92
……	……	……	……	……

有了附表 10，对平均数差异显著性的统计检验力的求法就更为简单了，只要将在 Z 检验过程中实际计算获得的 Z 值（或 t 检验中的 t 值）作为该次假设检验过程中备择假设分布的期望值 δ 值，就可以直接查附表 10 获得相应的统计检验力的值。

在例 11-1 中，对某校男女新生在智商方面是否存在显著差异进行 Z 检验后所得到的 Z 统计量是 $Z=1.89$，以此作为该次假设检验备择假设分布的期望 δ 值，查附表 10，最接近 1.89 的 δ 值是 1.9，表中第一列 1.9 横向与双侧检验 $\alpha=0.05$ 相交的值是 0.48，数值“0.48”就是本次假设检验的统计检验力。由于本次的假设检验结果是接受虚无假设，如果虚无假设是假就犯了“纳伪”性质的“二类错误”，犯二类错误的概率值就是“β=1-0.48=0.52”，这与前面第一种方法通过正态分布表计算求得的 53%的 β 值非常接近。

在例 11-2 中，对两个省小学一年级学生语文成绩是否存在显著差异进行 Z 检验后所得到的 Z 统计量是 $Z = 3.00$，以此作为该次假设检验备择假设分布的期望 δ 值，查附表 10，对应于表中第一列 δ=3.0 横向与双侧检验 $\alpha=0.05$ 相交的值是 0.85，这就是本次假设检验的统计检验力。这与前面第一种方法通过正态分布表计算求得的 $1-\beta$=0.85 的概率值完全相同。

（三）统计检验力或犯二型错误的概率值与正态分布中曲线下面积的关系

与前面所述例 11-1 和例 11-2 相关的问题是，在什么情况下我们需要求取犯二类错误的 β 值（图 11-3 中阴影部分），什么情况下我们需要求取统计检验力 $1-\beta$ 值（图 11-4 中白色区域）呢？我们应该如何解读图 11-3 和图 11-4 所包含的信息呢？

解答这一问题的关键点在于如何正确地理解“真假虚无假设”与“接受或拒绝虚无假设”之间的相互关系。

以平均数差异显著性检验为例，在第六章中，无论是例 6-7 还是例 6-8，关心的都是在“虚无假设为真”的前提下，如何通过假设检验过程得到的 Z 统计量做出“接受”或“拒绝”这一虚无假设的决策，在“小概率事件一般不会在一次抽样中出现”的思想指导下，我们预设一个较小的小概率 α 值，如果实际得到的 Z 统计量小于 α 值需要的临界值就做出“接受”该虚无假设的决策；如果实际得到的 Z 统计量大于 α 值需要的临界值就做出“拒绝”该虚无假设决策。第六章（以及随后的 7-10 章）介绍的统计决策过程不考虑“虚无假设为假”的问题。

本章当我们考虑我们的假设检验过程有可能是在“虚无假设为假”的前提下如何进行统计决策时，那就意味着存在一个与“虚无假设”有一定距离的“备择假设”，根据前面“统计检验力的基本原理”中所述的两个假定，我们以实得的 Z 统计量作为本次假设检验“备择假设分布”的中心点，并以此为基础根据统计推断结果来决定是求取 β 值还是 $1-\beta$ 值。这时我们再根据是接受还是拒绝虚无假设的假设检验结果来确定“通过 Z 统计量与显著性水平临界值之差查 Z 分布表后所得到的概率值”具有什么样的内涵。

在例 11-1（即例 6-7）中，根据已知数据对平均数差异显著性进行检验后，得到的 Z 统计量是 Z=1.89，当显著性水平为 $\alpha=0.05$ 时，$Z=1.89<Z_{\frac{\alpha}{2}}=1.96$，因此，我们的统计推论是“没有充分理由去拒绝虚无假设因而接受虚无假设。由此推断该校男女新生在智商方面不存

在显著差异”。在“虚无假设”为真时，做出这一正确推论的概率值为95%；但是，如果“虚无假设”本身为假，换言之，在本例中如果男女新生在智商方面事实上是存在显著差异，那么我们根据这一次抽样得到的数据进行平均数差异显著性检验的结果所做出的推论就是错误的，在心理统计学中，这种错误属于假设检验过程中的“二类错误β”或“纳伪错误”(即接受了错误的虚无假设)。心理统计学告诉我们，如果“虚无假设”为假，则存在一个与虚无假设有一定差距的备择假设，一般情况下我们并不知道备择假设的期望值离虚无假设的期望值有多远，但根据前面“统计检验力的基本原理”中所述的两个假定，我们以本次所得到的假设检验结果“Z=1.89”为中心点建构本次假设检验的备择假设分布，其期望值“δ”与虚无假设曲的期望值的距离是 1.89 个标准差，同时，仍然以$\alpha = 0.05$的显著性水平上（相应的临界点为 Z=1.96）来进行统计决策，由于得到的 Z 统计量即 Z＝1.89 小于显著性水平$\alpha = 0.05$双侧检验的临界值$Z_{\frac{\alpha}{2}} = 1.96$，并因此做出“接受这一虚假的虚无假设”的统计决策是错误的决策，就犯了“纳伪型二类错误”，在备择假设分布的曲线下面积中，距离虚无假设分布 1.96 个标准差左边阴影部分的面积都属于犯二类错误的概率值，通过 Z 统计量（在例 11-1 中即 Z＝1.89）与显著性水平α临界值（Z=1.96）之差（1.89-1.96=-0.07），再通过 Z 分布表所求得的曲线下面积（如图 11-3 所示阴影部分的面积）是 53%，其含义是“如果虚无假设是假，接受它做出错误决策的概率为$\beta = 0.53$”。

同样，当我们考虑我们的假设检验过程有可能是在“虚无假设为假”的前提下如何对例 11-2（即例 6-8）中的数据进行统计决策时，也意味着存在一个与“虚无假设”有一定距离的“备择假设”，根据本例已知数据对平均数差异显著性进行检验后，得到的 Z 统计量是Z=3.00，当显著性水平为$\alpha = 0.05$时，Z=3.00 $> Z_{\frac{\alpha}{2}}$（临界值为 Z=1.96），因此，我们的统计推论是“拒绝虚无假设，由此推断这两个省小学一年级学生的评议成绩存在显著差异”。在“虚无假设”为真时，做出这一错误推论（即犯一类错误）的概率值不超过 5%；但是，如果“虚无假设”本身为假，换言之，在本例中如果这两个省小学一年级学生的评议成绩事实上确实存在显著差异，那么我们的推论就是正确的（即正确拒绝了虚假的虚无假设）。如果“虚无假设”为假则存在一个与虚无假设有一定差距的备择假设，根据前面“统计检验力的基本原理”中所述的两个假定，我们以本次所得到的假设检验结果“Z=3.00”为中心点建构本次假设检验的备择假设分布，其期望值“δ”与虚无假设曲的期望值的距离是 3.00 个标准差，同时，仍然以$\alpha = 0.05$的显著性水平上（相应的临界点为 Z=1.96）来进行统计决策，由于得到的 Z 统计量即 Z＝3.00 大于显著性水平$\alpha = 0.05$双侧检验的临界值$Z_{\frac{\alpha}{2}} = 1.96$，并因此做出 “拒绝这一虚假的虚无假设” 的正确统计决策，在备择假设分布的曲线下面积中，距离虚无假设分布 1.96 个标准差右边白色的那部分面积都属于正确决策的概率值，通过 Z 统计量（在例 11-2 中即 Z＝3.00）与显著性水平α临界值（Z=1.96）之差（3.00-1.96=1.04），再通过 Z 分布表所求得的曲线下面积（如图 11-4 所示白色区域的面积）是 85%，其含义是“如果虚无假设是假，拒绝它做出正确决策的概率为$1 - \beta = 0.85$”。

总之，在进行假设检验的过程中，当我们的统计决策是“拒绝虚无假设”时（如例 11-2），如果“虚无假设”本身是虚假的则做出了正确决策，这时需要求取“备择假设分布”曲线下面积中显著性水平临界点右边白色区域的那部分面积就是表示正确决策的统计检验力$1-\beta$的概率值；当我们的统计决策是“接受虚无假设”时（如例 11-1），如果“虚无假设”是假的，则需要求取犯“二类错误”β的概率值，在“备择假设分布”的曲线下面积中就是显著性水平临界点左边的阴影部分面积。需要提醒读者注意的是，虽然“β”和“$1-\beta$”是数值互补的两个指标，但从表 11-1 中所列的各种关系以及“统计检验力”这一概念的定义出发，在不拒绝虚无假设的情况下，更多学者此时关注的是二类错误概率“β”而不是统计检验力“$1-\beta$”。

三、两独立样本平均数差异显著性检验效果大小的估计方法

本书第一部分曾指出：根据 Cohen（1988）的观点，本书把效果大小定义为：“在虚无假设（H_0）为假时，对该次假设检验过程中所存在的备择假设（H_1）的期望值与虚无假设的期望值之间的距离的测量值”。下面将以两总体平均数差异显著性检验为例来论述效果大小的计算公式和主要求取方法。

（一）效果大小的计算公式

公式 11-3 是用于计算备择假设期望值“δ”的公式，在两个独立样本的方差和样本容量都相等的条件下，可作如下推导：

$$\delta=\frac{(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2+\sigma_2^2}{n}}}=\frac{(\mu_1-\mu_2)}{\sqrt{\frac{2\sigma^2}{n}}}=\frac{(\mu_1-\mu_2)}{\sqrt{\frac{2}{n}}\sqrt{\sigma^2}}=\frac{(\mu_1-\mu_2)}{\sigma\sqrt{\frac{2}{n}}}=\frac{(\mu_1-\mu_2)\sqrt{\frac{n}{2}}}{\sigma}$$

由此可得：

$$\delta=\frac{(\mu_1-\mu_2)}{\sigma}\sqrt{\frac{n}{2}}$$
（公式 11-4）

令
$$d=\frac{(\mu_1-\mu_2)}{\sigma}$$
（公式 11-5）

公式 11-5 就是求取两总体平均数差异显著性检验后衡量该检验结果效果大小的计算公式，由于该公式最早是由心理统计学家 Cohen 推导出来，因此，通常也将通过该公式计算得到的结果称之为“Cohen d”

当两个独立样本的方差和样本容量都不相等时，则公式 11-5 方程右侧的分母“σ”需要用“联合方差（pooled-variance）S_p”来取代（见公式 11-6）

$$d=\frac{(\mu_1-\mu_2)}{S_p}\quad\left(S_p=\sqrt{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\right)$$
（公式 11-6）

请读者注意：（1）在实际计算时，通常是用$\overline{X}_1$取代μ_1，用$\overline{X}_2$取代μ_2；（2）Cohen（1988）建议，在评估某个假设检验的效果大小时，d = 0.2 为小的检验效果， d=0.5 为中等的检验效果， d=0.8 及以上为大的检验效果。

（二）效果大小的主要求取方法

有了公式 11-4、公式 11-5 和公式 11-6，我们就可以根据不同的数据情况分别通过这几个不同的公式来求取两总体平均数差异显著性检验效果大小的值（即 Cohen d）了。

【例 11-3】根据例 11-1（即例 6-7）的数据，求该假设检验的效果大小。

方法 1：通过公式 11-5 或公式 11-6 直接求取

解：在例 11-1 中，已知的条件是：

$$n_1 = 100, \overline{X}_1 = 115, \sigma_1 = 15$$
$$n_2 = 100, \overline{X}_2 = 111, \sigma_2 = 15$$

由于本次假设检验过程中所抽取的样本容量是一样的，并且已经知道两个独立样本所属总体的标准差（或方差）也相等（即$\sigma_1 = \sigma_2 = 15$），因此，可以将相关数据直接代入公式 11-5 求取 Cohen d：

$$d = \frac{(\mu_1 - \mu_2)}{\sigma} = \frac{115 - 111}{15} = 0.27$$

答：本次假设检验的效果大小是：d = 0.27，根据 J. Cohen 的观点，这属于较小程度的效果大小。

【例 11-4】根据例 11-2（即例 6-8）的数据，求该假设检验的效果大小。

解：在例 11-2 中，已知的条件是：

$$n_1 = 180, \overline{X}_1 = 82, \sigma_1 = 11.5$$
$$n_2 = 160, \overline{X}_2 = 78.42, \sigma_2 = 10.5$$

由于本次假设检验过程中所抽取的样本容量和两个独立样本所属总体方差都不相等，如果用上述方法 1 求解，则需要使用公式 11-6，将相关数据代入公式 11-6 后可得：

$$d = \frac{(\mu_1 - \mu_2)}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}}$$

$$= \frac{(82 - 78)}{\sqrt{\dfrac{(180 - 1) \times 11.5^2 + (160 - 1) \times 10.5^2}{180 + 160 - 2}}}$$

$$= 0.36$$

答：本次假设检验的效果大小是：d = 0.36，

方法 2：通过假设检验过程中所求取的 Z 统计量来求取

将前述公式 11-5 代入公式 11-4，整理后可得：

$$\delta = d\sqrt{\frac{n}{2}} \qquad \text{（公式 11-7）}$$

$$d = \frac{\delta}{\sqrt{\frac{n}{2}}} \qquad \text{（公式 11-8）}$$

根据前述统计检验力计算原理中的假定二：令“$\delta = Z$”，则有

$$d = \frac{Z}{\sqrt{\frac{n}{2}}} \qquad \text{（公式 11-9）}$$

公式 11-9 中的“n”是指两个样本容量相等时各自的样本容量。在两个样本容量不相等的情况下,可以通过公式 11-10 先求这两个样本容量的调和平均数,然后再用公式 11-9 来计算 d 值。

$$\bar{n}_H = \frac{2n_1n_2}{n_1 + n_2} \qquad \text{（公式 11-10）}$$

在例 11-1 中,各样本容量相等且都等于“100”，通过假设检验过程得到的 Z 统计量是“1.89”，将相关数据代入公式 11-9 就可求得本次假设检验的效果大小为：

$$d = \frac{Z}{\sqrt{\frac{n}{2}}} = \frac{1.89}{\sqrt{\frac{100}{2}}} = \frac{1.89}{7.071} = 0.27$$

结果表明，用这种方法求取的效果大小 d 值结果与前面方法一所得到的结果完全一样。

再如，在例 11-2 中，两个样本容量并不相等，这时我们可以先利用公式 11-10 求这两个样本的调和平均数：

$$\bar{n}_H = \frac{2n_1n_2}{n_1 + n_2} = \frac{2 \times 180 \times 160}{180 \times 160} = 169.41$$

然后再将相关数据代入公式 11-9 求本次假设检验结果的效果大小：

$$d = \frac{Z}{\sqrt{\frac{n}{2}}} = \frac{3}{\sqrt{\frac{169.41}{2}}} = \frac{3}{9.20} = 0.33$$

用这种方法求取的效果大小 d 值与前面方法一所得到的结果非常接近。

四、统计检验力与效果大小、显著性水平、样本容量等因素之间的相互关系

本书第五章第五节（P104）论述第一类（α）错误和第二类（β）错误的相互关系时，已经提到了在统计决策过程中影响“二类错误（β）”的因素主要有三个：（1）α 水平；（2）两个总体平均数 μ_1 与 μ_2 之间的距离，也就是效果大小；（3）样本容量 n 的大小。由于统计检验力 $1-\beta$ 是“二类错误（β）”的补数，因此，前述影响“二类错误（β）”的三个因素也就是统计检验力 $1-\beta$ 的影响因素，正如著名统计学家 Cohen（1992）所说：“显著性检验的统计检验力是在给定总体效果大小、显著性水平 α 和样本容量 N 的条件下拒绝 H_0 的概率”。下面我们分述之。

（一）统计检验力与显著性水平的相互关系

对表 11-1（详见附表 10）中所列各数据进行分析后可知：

一方面，统计检验力 $1-\beta$ 的值在显著性水平 α 不变的前提下，将随着 δ 值的增大而增大，例如，在 $\alpha=0.05$ 的前提下进行双侧检验，δ=0.5 时，$1-\beta$=0.08；δ=1.8 时，$1-\beta$=0.44。

另一方面，统计检验力 $1-\beta$ 的值在 δ 不变的前提下，将随着显著性水平 α 值的降低而减小，例如，在 δ=1.90 的情况下，当我们在 α=0.10 的显著性水平下进行双侧检验，检验结果的统计检验力为 $1-\beta$=0.60；当我们在 α=0.05 的显著性水平下进行双侧检验，检验结果的统计检验力则降为 $1-\beta$=0.48；如果我们在 α=0.01 的显著性水平下进行双侧检验，检验结果的统计检验力 $1-\beta$ 则只有“0.25”了。

统计检验力与 α 水平的关系还可以通过图 11-5 来进一步理解。

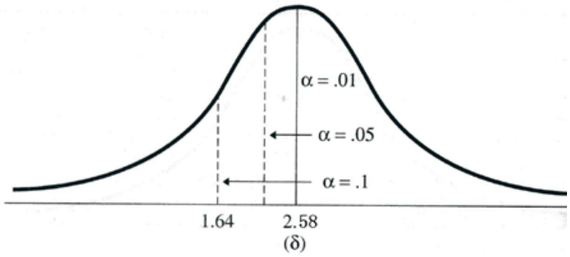


图 11-5 作为固定 δ 值的 α 的函数的统计检验力示意图

由图 11-5 可知，对于任何的 δ 值，以及由横轴表示的相应于该 δ 值建构的正态分布曲线下面积，进行假设检验之后该检验的统计检验力值的变化都将是 α 的函数。例如，在 $\delta = 2.58$ 的情况下，当 $\alpha=0.10$ 时，其统计检验力为 0.83；当把统计决策的显著性水平提高到 $\alpha=0.05$ 时，统计检验力就降低到 0.74；当 α 水平进一步提高到 $\alpha=0.01$ 时，统计检验力就降低到只有 0.50 了。

上述表明：统计决策过程中，预设的显著性水平越是严格，统计检验力的概率值就越小。

下面，我们通过对同一批数据进行假设检验时在不同 α 水平上做出决策后所具有的不同统计检验力值，来进一步说明统计检验力与显著性水平 α 之间的相互关系。

【例 11-5】如果我们在对【例 11-1】中所列数据进行平均数差异显著性假设检验的过程中，做统计决策时把显著性水平从 $\alpha=0.05$ 改为 $\alpha=0.10$ 的水平来进行双侧检验，求该检验结果的统计检验力。

利用方法 1 的五步骤方法求解：

前两个步骤与例 11-1 的求解过程一样，得到的检验统计量为$Z=1.89$。

（3）确定做出统计决策的α水平并做出决策

查正态分布表可知，在显著性水平$\alpha=0.10$（双侧）时，其临界值为$Z_{\frac{\alpha}{2}}=1.645$，由于$Z=1.89>Z_{\frac{\alpha}{2}}=1.645$，因此，正常情况下做出的决策是拒绝虚无假设而接受备择假设，认为这次随机抽样的结果表明某校男女新生在智商方面存在差异，女生的智商要高于男生的智商。

（4）计算实际得到的 Z 值与α水平临界值的差，得到

1.89-1.645=0.245；

（5）根据 Z 值与α水平临界值的差查正态分布表并确定统计检验力 1-β的概率

查正态分布表可知，从中心点为零到左边 0.245 个标准差所占的面积是 0.0968，约等于 0.10，如图 11-6 白色区域所示，这 10%加上该正态分布右边的 50%就是本次检验的统计检验力，其概率值为“1-β=0.60”。

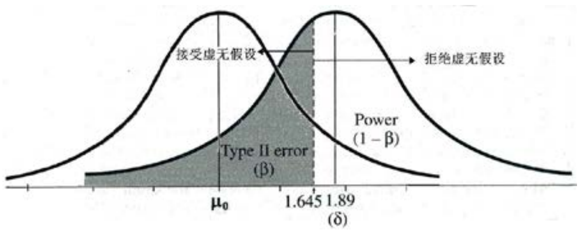


图 11-6 备择假设分布图（$\alpha=0.10$双侧检验）

像图 11-3 那样，图 11-5 的含义是说，我们以这次抽样中得到的 Z=1.89 的检验统计量作为估计备择假设分布的中心点（或期望值δ），建构一个正态分布（1.89 左右两边各占 50%的面积），以 Z=1.645 作为接受或拒绝假设的分界点，如果实际获得的 Z 值小于 Z=1.645（图 11-4 左边阴影部分）就接受虚无假设,这时如果虚无假设为假则犯β型错误的概率值约为 40%；如果实际获得的 Z 值大于 Z=1.645（图 11-5 右边部分，如本例）就拒绝虚无假设，这时如果虚无假设为假则是作出了正确的决策，其正确决策的概率值是（1-β）=1-0.40=60%。

答：本例中$\alpha=0.10$时平均数假设检验后的统计检验力是$1-\beta=0.60$。

例 11-1 和例 11-7 这两个例题所述的是对同一批数据在不同显著性α水平上进行平均数差异显著性检验，得到的 Z 统计量都是 Z=1.89，但是，当我们在显著性水平$\alpha=0.05$进行统计决策时（例 11-1），Z＝1.89 小于显著性水平$\alpha=0.05$双侧检验的临界值$Z_{\frac{\alpha}{2}}=1.96$，因此我们的统计决策结果是接受虚无假设，这时，“如果虚无假设是假，那么我们接受它做出错误决策的概率为$\beta=0.53$（这种决策不存在‘统计检验力’）”；而当我们在显著性水平$\alpha=0.10$进行统计决策时（例 11-7），Z＝1.89 大于显著性水平$\alpha=0.10$双侧检验的临界值$Z_{\frac{\alpha}{2}}$=1.645，因此我们的统计决策结果是接受虚无假设，我们的统计决策结果是拒绝虚无假

设，这时，“如果虚无假设是假，那么我们有 60%的把握度（也就是统计检验力$1-\beta=0.60$）来拒绝这一虚假的虚无假设”（这种决策不存在‘纳伪错误’）。上述两个例子中，同样的数据在进行平均数差异显著性检验过程中，由于在不同显著性水平上进行决策会得出不同的推论结果，由此也就会得出不同的统计检验力$1-\beta$值或犯二类错误的β值，表明统计检验力和显著性水平α两者之间确实具有非常密切的关系。

（二）统计检验力与样本容量的相互关系

由公式 11-7 即$\delta=d\sqrt{\frac{n}{2}}$可知，在实际计算过程中，$\delta$值是$d$值与样本容量除以 2 之后的算术平方根两部分的乘积。这表明，在其他条件不变的情况下，样本容量除越大，δ值也会越大。进而由表 11-2（附表 10）可知，在固定显著性水平α的前提下，δ值越大，则统计检验力$1-\beta$也越大（附表 10），因此，样本容量除与统计检验力$1-\beta$之间存在正相关的关系。

下面我们通过例 11-6 来进一步理解这种关系。

【例 11-6】以【例 6-7】的数据为基础，如果我们在该校男女新生中各抽取 200 名学生进行韦氏智力测验，如果也得到女生平均智商为“$\overline{X}_1$=115 分”，男生平均智商为“$\overline{X}_2$=111 分”，求在显著性水平为$\alpha=0.05$时进行平均数差异显著性检验（双侧）的“二类错误”或“统计检验力”的概率值。

解：已知条件为：

$$n_1=200,\overline{X}_1=115,\sigma_1=15$$
$$n_2=200,\overline{X}_2=111,\sigma_2=15$$

按前述方法一所介绍的求解该检验统计检验力的过程如下：

（1）建立假设：
$$H_0:\mu_1=\mu_2$$
$$H_1:\mu_1\neq\mu_2$$

（2）用公式 11-1（即公式 6-10）计算检验统计量

$$Z=\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{115-111}{\sqrt{\frac{15^2}{200}+\frac{15^2}{200}}}=\frac{4}{1.5}=2.67$$

（3）确定显著性水平并做出决策

在显著性水平$\alpha=0.05$进行双侧检验时，由于本次假设检验结果为 $Z=2.67>Z_{0.05}=1.96$，因此，就像在**【例 11-2】**时所做决策那样，正常情况下做出的决策是拒绝虚无假设而接受备择假设，认为这次随机抽样的结果表明某校男女新生在智商方面存在差异，女生的智商要高于男生的智商。

（4）计算实际得到的 Z 值与α水平临界值的差

2.67-1.96=0.71；

（5）根据 Z 值与α水平临界值的差查正态分布表并确定统计检验力 1-β的概率

查正态分布表可知，从中心点为零到左边 0.71 个标准差所占的面积是 0.26，这 26%加上该正态分布右边的 50%就是本次检验的统计检验力，其概率值为“1-β=0.76”。

答：本次随机抽样进行平均数差异检验的结果表明，该校男女新生在智商方面存在差异，女生的智商要高于男生的智商。做出这一推论的把握度为“1-β=0.76”。

对**【例 11-1】**和**【例 11-6】**所列数据的假设检验结果进行比较可知：在其他条件不变的条件下，样本容量越大，假设检验结果的 Z 统计量就会越大，与此相关联的统计检验力的概率值也就越大。

请读者注意：虽然增加样本容量 n 可以提高 Z 统计量，但例 11-8 中的效果大小并没有提高。例如我们可以利用公式 6-7 来计算本例的效果大小：

$$d = \frac{Z}{\sqrt{\frac{n}{2}}} = \frac{2.67}{\sqrt{\frac{200}{2}}} = \frac{1.89}{7.071} = 0.27$$

这表明与例 11-1 相比较，例 11-8 通过增加样本容量增加了 Z 统计量的值，也因此增加了统计检验力（由例 11-1 的 0.47 增加到本例的 0.77），但两例中的效果大小都是同样的 0.27。

公式 11-7 即 $\delta = d\sqrt{\frac{n}{2}}$ 还可以让我们求取与特定效果大小 d 值相关的样本容量 n。

换言之，在确定显著性水平和d值的条件下，我们可以通过公式 11-11 推导出要达到某种程度的统计检验力时所需要的样本容量：

$$n = 2\left(\frac{\delta}{d}\right)^2 \qquad \text{（公式 11-11）}$$

【例 11-7】在$\alpha = .05$，$d = .50$的条件下，求对实验结果进行双侧假设检验后达到 80%的统计检验力所需要的样本容量。

解：

（1）先从附表 10 查双尾检验$\alpha = .05$那一列（即最中间一列），对应于 80%的δ值是 2.8；

（2）将这些数据代入公式 11-11 可得：

$$n = 2\left(\frac{\delta}{d}\right)^2 = 2\left(\frac{2.8}{0.5}\right)^2 = 2\times(5.6)^2 = 62.72$$

答：在上述条件下要达到 80%的统计检验力，每组被试应不少于 63 名，实施实验约需要 130 名被试。

（三）统计检验力与效果大小之间的相互关系

1.两者的联系

就像样本容量与统计检验力的相互关系那样，公式 11-7 也清楚地表明了效果大小也统计检验力之间的相互关系：在其他条件不变的情况下，d 值越大，δ 值也会越大。进而由表 11-1（附表 10）可知，δ 值越大，则统计检验力$1-\beta$也越大（附表 10），因此，效果大小d 值与统计检验力$1-\beta$之间也存在正相关的关系。

【例 11-8】根据【例 6-7】的数据，按上述方法 1 的五个步骤求$\alpha=0.05$时平均数假设检验（双侧）的“二类错误”或“统计检验力”的概率值。

解：在例 6-7 中，已知的条件是：

$$n_1=100, \overline{X}_1=116.66, \sigma_1=15$$
$$n_2=100, \overline{X}_2=111, \sigma_2=15$$

求解过程重复如下：

（1）建立假设：
$$H_0:\mu_1=\mu_2$$
$$H_1:\mu_1\neq\mu_2$$

（2）用公式 11-1（即公式 6-10）计算检验统计量

$$Z=\frac{(\overline{X_1}-\overline{X_2})-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}=\frac{116.66-111}{\sqrt{\frac{15^2}{100}+\frac{15^2}{100}}}=\frac{5.66}{2.12}=2.67$$

（3）确定做出统计决策的α水平并做出决策

就像例 11-8 那样，当我们在显著性水平$\alpha=0.05$进行双侧检验时，由于本次假设检验结果为 $Z=2.67>Z_{0.05}=1.96$， 因此，正常情况下做出的决策是拒绝虚无假设而接受备择假设，认为这次随机抽样的结果表明某校男女新生在智商方面存在差异，女生的智商要高于男生的智商。

（4）计算实际得到的 Z 值与α水平临界值的差，得到

$$2.67-1.96=0.71;$$

（5）根据 Z 值与α水平临界值的差查正态分布表并确定统计检验力 1-β的概率

查正态分布表可知，从中心点为零到左边 0.71 个标准差所占的面积是 0.26，这 26%加上该正态分布右边的 50%就是本次检验的统计检验力，其概率值为“1-β=0.76”。

答：本次随机抽样进行平均数差异检验的结果表明，该校男女新生在智商方面存在差异，女生的智商要高于男生的智商。做出这一推论的把握度为“1-β=0.76”。

对**【例 11-1】**和**【例 11-8】**所列数据的假设检验结果进行比较可知：在例 11-1 中，两个样本平均数之间的差距是 115-111=4 分；而在例 11-8 中，两个样本平均数之间的差距则是

116.66-111=5.66 分，因为例 11-8 中这两个样本平均数之间的差距比例 11-1 中的更大，因此，不仅增大了 Z 统计量（从例 11-1 中的“Z=1.89”增加到例 11-8 中的“Z=2.67”）和统计检验力（从例 11-1 中的“1-β=0.53”增加到例 11-8 中的“1-β=0.76”），而且该假设检验的效果大小也得到相应的增大：

$$d = \frac{(\mu_1 - \mu_2)}{\sigma} = \frac{115.66 - 111}{15} = 0.38$$

或：

$$d = \frac{Z}{\sqrt{\frac{n}{2}}} = \frac{2.67}{\sqrt{\frac{100}{2}}} = \frac{2.67}{7.07} = 0.38$$

由此可知：在其他条件不变的条件下，两个样本的平均数之间的差距越大，假设检验结果的 Z 统计量就会越大，与此相关联的统计检验力的概率值也就越大；此外，该假设检验的效果大小也会相应的增大。

2.两者的区别

虽然效果大小d值与统计检验力$1-\beta$之间有着密切关系，但我们也要注意这两者之间的区别：统计检验力$1-\beta$的大小会受到样本容量大小的影响，而效果大小d值则不受样本容量影响：我们可以通过公式 11-5，在不考虑样本容量的情况下直接计算某个假设检验结果的d值。

我们可以通过对**【例 11-1】**、**【例 11-6】**和**【例 11-8】**这三个例题的假设检验结果进行综合比较来理解上述观点：

对**【例 11-1】**和**【例 11-6】**的假设检验结果进行比较可知：虽然例 11-6 在其他条件与例 11-1 相同的情况下，由于提高样本容量（从例 11-1 中的“=100”增加到例 11-6 中的“=200”）而增大了 Z 统计量（从例 11-1 中的“Z=1.89”增加到例 11-6 中的“Z=2.67”）并因此增加了统计检验力的值（从例 11-1 中的“1-β=0.53”增加到例 11-8 中的“1-β=0.76”），但这两个例题假设检验结果的效果大小是一样的，都是“d=0.27”；

对**【例 11-6】**和**【例 11-8】**的假设检验结果进行比较可知：虽然这两个例题与例 11-1 的假设检验结果相比较，Z 统计量都从例 11-1 中的“Z=1.89”增加到例 11-6 中的“Z=2.67”，因而有着相同的统计检验力（“1-β=0.76”），但这两个例题假设检验结果的效果大小是不同的：例 11-1 由于增加样本容量而使 Z 统计量和统计检验力得以增大，结果效果大小与例 11-1 一样是“d=0.27”；但例 11-8 由于增加了两个样本平均数之间的差距，不仅使 Z 统计量和统计检验力得以增大，结果效果大小“d=0.38”也比与例 11-1 的“d=0.27”更大。

综合本节内容可知：当假设检验的结果是接受虚无假设时，我们关心的是犯“纳伪”错误可能性即β值的大小：β值越小，那么犯纳伪错误的可能性就越小（这种情况下不存在“统计检验力”的概念）；

当假设检验的结果是拒绝虚无假设时（这种情况下不存在“纳伪错误”的概念），我们关

心以下两个指标：

（1）统计检验力$1-\beta$，它反映的是我们通过这次检验的结果做出“拒绝虚无假设”这一推断的把握度的大小：“$1-\beta$”值越大，意味着我们有较大的把握认为“虚无假设是虚假的”；反之，“$1-\beta$”值越小，那么，虽然统计学意义上这次假设检验结果表明两个平均数之间的差异具有显著意义，但我们做出“虚无假设是虚假的”这一推断的自信度并不高。

（2）效果大小 d 值，它反映的是通过抽样所得到的两个平均数各自所属的两个总体之间的实际差异有多大：d 值反映这次假设检验过程中，备择假设与虚无假设所存在的实际差异越大。

如前所述，“统计检验力$1-\beta$”值的大小不仅受效果大小的影响，也受到“样本容量 N”的大小的影响到，如果某次假设检验过程中所得到的较大的“统计检验力$1-\beta$”主要是由样本容量大所造成的，那么，它所反映的只是“在统计学意义上能在较小的显著性水平上拒绝虚无假设”，而通常我们使用假设检验过程是为了检验我们在实验研究中所控制的自变量对因变量的影响有多大，这一问题正好就是“效果大小”这一统计指标的内涵。因此，在实际应用中，我们通常更关注效果大小的值，而不怎么关注“$1-\beta$”。

第二节　方差分析的统计检验力和效果大小的估计

一、单因素方差分析统计检验力和效果大小的含义

在心理统计学中，平均数差异显著性检验、方差分析（F 检验）、$\overline{X}^2$ 检验和相关系数差异显著性检验等所有检验统计量的统计检验力和效果大小的含义实质上都是一样的：统计检验力都是“指在虚无假设H_0为假（备择假设H_1为真）时，正确拒绝H_0的概率”；效果大小则都是“指如果存在备择假设，它与虚无假设之间的距离有多大”。

第一节我们以两个独立样本平均数的差异显著性检验为例论述了统计检验力的原理，并且曾指出理解统计检验力和效果大小的含义的关键在于正确理解“虚无假设分布”和“备择假设分布”这两个概念的内涵。

在对两个独立样本平均数差异显著性进行假设检验的过程中，通常虚无假设（NHD）是$H_0: \mu_1 = \mu_2$，其分布以及对其进行 Z 检验或 t 检验后 Z（或 t）统计量的分布都是一种“以零为中心的分布”；备择假设是$H_1: \mu_1 \neq \mu_2$，其分布的期望值δ以及所有绝对值大于Z_α的 Z 统计量（在 t 检验中则是指绝对值大于$t_{\alpha,(n-1)}$的t统计量）都属于该次假设检验的备择假设，在理论上，所有这些备择假设的分布值以及对其进行检验后相应的 Z 或 t 统计量的分布值都是一种“不以零为中心的分布”。

就单因素方差分析而言，从某种意义上说，这种统计方法要处理的数据是 Z（或 t）检验的延伸：从 Z 检验对两个抽样样本数据的处理延伸到对三个或以上抽样样本数据的处理，例如，在单因素方差分析的假设检验中，通常虚无假设是“$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$”，对其进行

假设检验的统计量计算公式是：$F=\frac{MS_B}{MS_W}$，当虚无假设为真，对其进行多次实验和用单因素方差分析方法进行统计检验后，所得到的所有F值将围绕$F=1$波动，这是一种“以 1 为中心的分布”，这时 F 统计量分布的平均值F_0等于$\frac{df_W}{df_W-2}$；备择假设是指“H_1：$\mu_1,\mu_2,\cdots\mu_k$”不完全相等，当虚无假设为假时，所有大于或等于$F_{\alpha,(df_1,df_2)}$的 F 统计量都属于该次假设检验的备择假设，从理论上说，这是一种“非中心F分布（the noncentral F distribution)”。备择假设的期望分布值通常用希腊字母Φ表示。

单因素方差分析的效果大小与两个独立样本平均数差异显著性检验的效果大小一样：当虚无假设为假时，那么就意味着在多组平均数中，至少有两个组之间存在备择假设（也可能不止两个组之间存在备择假设），对此，我们自然想知道这次假设检验备择假设的期望值离虚无假设期望值有多远，以此来作为自变量对因变量有影响作用的指标。

在 SPSS 软件包的“分析”一栏下拉菜单中所包含的“一般线性模型的多因素方差分析”和“重复测量方差分析”等计算模块都提供了输出方差分析“效果大小”和“统计检验力”的选项，我们会在后面适当加以介绍。

二、单因素方差分析的效果大小和统计检验力的计算

（一）单因素方差分析效果大小的估计

J. Cohen（1988）提出用小写英文字母“f”来表示方差分析效果大小这一统计量的值。至于方差分析效果大小的“f值”如何进行评估，不同学者有不同的看法，目前至少存在以下四种不同的观点，我们也可以把这四种方法分为两大类：第一类（下面介绍的前两种方法）评估方法以方差分析后得到的 F 统计量为基础；第二类（下面介绍的后两种方法）评估方法则以“实验处理之后各组间平方和在总体平方和中所占的比重”的计算方法为基础，下面我们分述之。

（1） 有些学者（如甘怡群，2005）认为，可以用公式 11-12 来计算方差分析的效果大小f值：

$$f=\sqrt{\frac{F}{n}} \qquad \text{（公式 11-12）}$$

式中，根号内大写字母F是该次方差分析后得到的检验统计量，n是指各组人数相等时每个组的人数。

（2） 美国纽约大学的学者 B. H. Cohen 认为，用$f=\sqrt{\frac{F}{n}}$计算得到的f值是效果大小的有偏估计，因此提出通过公式 11-13 来对方差分析效果大小f值加以校正，他提议用粗体的小写字母“**f**”来表示，以示与f区别。

$$\mathrm{f} = \sqrt{\left(\frac{k-1}{k}\right)\frac{F}{n}}$$ （公式 11-13）

式中，根号内大写字母 F 是该次方差分析后得到的检验统计量，k 是分组数，n 是各实验分组的人数。

上述两种计算方差分析效果大小的公式中的“n”ftjb 是指各组人数相等时每个组的人数，如果各组人数不相等，则需要用公式 11-14 来计算各组人数的调和平均数 n_h。

$$n_h = \frac{k}{\sum \frac{1}{n_i}}$$ （公式 11-14）

（3） Aron 等（2006）认为，可以使用 η^2 作为评估方差分析效果大小的指标，η^2 的含义是指经过实验处理之后各组间平方和在总体平方和中所占的比重，其计算公式为：

$$\eta^2 = \frac{SS_{组间}}{SS_{总体}}$$ **（公式 11-15）**

（4）舒华等（2008）认为，可以使用 η^2 的平方根即 η 作为方差分析效果大小的指标，其计算公式为：

$$\eta = \sqrt{\frac{SS_{组间}}{SS_{总体}}}$$ **（公式 11-16）**

J. Cohen（1992）认为，当用 f 值做方差分析效果大小的指标时，$f = 0.10$ 时属于小的效果，$f = 0.25$ 时属于中等效果，$f = 0.40$ 时属于大的效果。

当用 η^2 或其平方根 η 做方差分析效果大小的指标时，通常认为 $\eta^2 = .01$ 时属于小的效果，$\eta^2 = .06$ 时属于中等效果，$\eta^2 = .14$ 时属于大的效果。

【例 11-9】以例 10-1 有关反馈类型影响自尊水平的研究资料为例，求方差分析结果的统计检验力。

解：

传统的心理统计方法对例 10-1 的数据进行方差分析后得到的结果有如表 11-3 所示（即表 10-6）。

表 11-3（表 10-6）不同反馈类型是否影响被试自尊水平方差分析表

变异来源	平方和	自由度	均方	F
组间	448	2	224	7.814**
组内	430	15	28.67	
总变异	878	17		

**表示在 α =.01 水平上拒绝 H_0

假如该假设检验中“虚无假设为假”，就需要我们提供该检验“效果大小”指标，我们可以通过上述四种方法来分别计算，由此也就会得到不同的效果大小值：

用方法（1）即公式 11-12 计算时，F 统计量为“F=7.814”，每组人数为 6 人（即“n=6”），将相应数据代入公式后可得：

$$f=\sqrt{\frac{F}{n}}=\sqrt{\frac{7.814}{6}}=\sqrt{1.30}=1.14$$

用方法（2）即公式 11-13 计算时可得：

$$\mathbf{f}=\sqrt{\frac{(k-1)}{k}\cdot\frac{F}{n}}=\sqrt{\frac{(3-1)}{3}\cdot\frac{7.814}{6}}=\sqrt{0.67\times1.30}=0.93$$

用方法（3）即公式 11-15 计算时，本例中的组间平方和为 448，总体平方和为 878，代入公式后可得：

$$\eta^2=\frac{SS_{组间}}{SS_{总体}}=\frac{448}{878}=0.51$$

用方法（4）即公式 11-16 计算时可得：

$$\eta=\sqrt{\frac{SS_{组间}}{SS_{总体}}}=\sqrt{.51}=0.714$$

由上面的计算结果可知，四种计算方法得出的方差分析效果大小的结果是不一样的。要注意的是，第一种、第二种与第三种、第四种方法之间由于计算方法所依赖的理论基础是不一样的，因此所用指标的量纲也不一样。

在上述四种计算方法的计算结果中，虽然用方法（3）所得出的方差分析效果大小值最小（η^2=0.51），但仍然有统计学家（如 Howell）认为，用η^2作为实验处理在总变异中所占比重的估计方法仍然有一定的偏差，一般会高估实验处理的效果，因此主张用另外一个反映实验处理效果大小的指标ω^2来对η^2加以校正（计算这两个指标的理论基础大致是相同的），其计算公式为：

$$\varpi^2=\frac{SS_{组间}-(k-1)MS_e}{SS_{总体}+MS_e} \qquad （公式 11-17）$$

还是以例 10-1 为例，将已知数据代入公式后可得：

$$\varpi^2=\frac{SS_{组间}-(k-1)MS_e}{SS_{总体}+MS_e}=\frac{448-(3-1)\times28.67}{878+28.67}=0.43$$

Howell 认为，用“$\omega^2=0.43$”比用“$\eta^2=0.51$”能更准确地反映本研究中反馈类型对被试自尊水平的实验效果。

使用不同的方法就可以得出不同的方差分析效果大小值，那么，在众多的指标中我们应该选择哪个更好呢？本书编者认为可以用 SPSS 软件包给出的结果作为参考。

【例 11-10】用 SPSS 软件包求例 10-1 中方差分析过程中的效果大小和统计检验力。

解：

将数据正确输入到 SPSS（中文版 18.0）数据框。其中在“反馈类型”变量中分别用“1”代表“积极反馈组”，“2”代表“对照组”， “3”代表“消极反馈组”；在“自尊水平得分”变量中按序录入每组 6 名共 18 名被试的测试成绩。然后可根据图 11－7 所示的操作步骤求解本例的效果大小和统计检验力。

1. 分析→一般线性模型→单变量
2. 因变量框：自尊水平得分
3. 固定因子框：反馈类型
4. 选项：（1）输出框中选择功效估计、检验效能
   （2）$\boxed{\text{继续}}$
5. $\boxed{\text{确定}}$

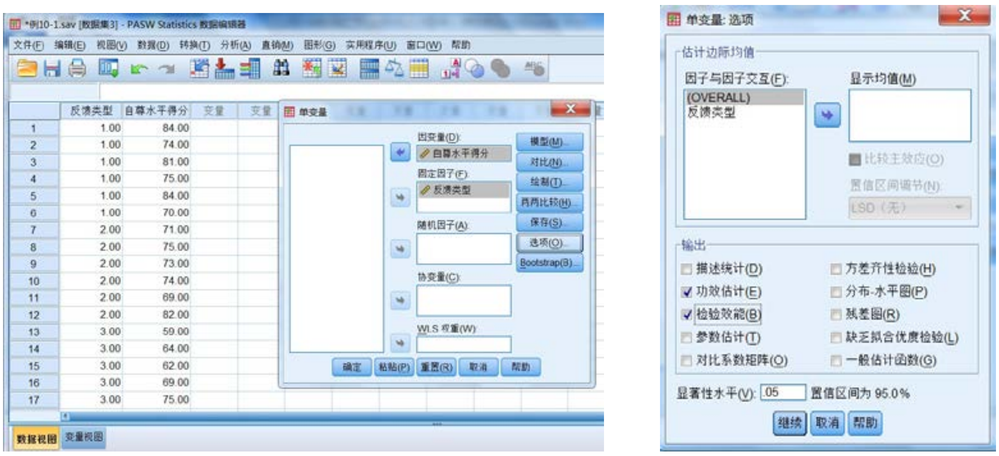


图 11-7 在 SPSS 中求解单因素方差分析功效估计、检验效能各步骤示意图

将 SPSS（中文版 18.0）的计算结果中的相关信息整理后有如成表 11-4 所示。

表 11-4 主体间效应检验表

变异源	平方和	自由度	均方	F	Sig.	η^2	统计检验力
反馈类型	448.00	2	224.000	7.814	.005	.510	.901
误差	430.00	15	28.667				
总体	878.00	18					

a. R 方 = .510（调整 R 方 = .445）

b. 使用 alpha 的计算结果 = .05

由表 11-4 中数据可知，与表 11-3 所列传统心理统计方法的方差分析表相比较，表 11-4 增加了最后两列内容，表明本例中的效果大小为“η^2=0.51”；统计检验力为“$1-\beta$=0. 90”。

对上述 SPSS 软件包给出的计算结果进行分析后可知，该软件包使用前述方法（3）所述的“η^2”作为方差分析效果大小的指标。Cohen 认为，“η^2 的含义是指在实验处理所造成的

变异在总变异中所占的比重”，就本例而言，“η^2=0.51”这一结果表明实验的总体变异中约有51%是来自“反馈类型“这一自变量的影响。

答：本次方差分析结果的效果大小为“η^2=0.51”。

（二） 单因素方差分析的统计检验力的计算方法

通常认为需要通过以下三个步骤来求单因素方差分析的统计检验力：

（1）用公式 11-13 求小写粗体“**f** 值”的值（如前所述也是一种效果大小的指标）；

（2）用公式 11-19 求备择假设分布的期望值“Φ”：

$$\Phi = \mathrm{f}\sqrt{\mathrm{n}} \qquad \text{（公式 11-18）}$$

（3） 根据Φ值、分组数 k，组内自由度 df_w 查附表 12 求统计检验力的值

附表 12 包括因分组数 k 不同而不同的几个分表，不同分表之间的结构是一样的。以 k=3 为例，其结构有如表 11-5 所示。

表 11-5 方差显著性检验的统计检验力表

K=3	（Φ）								
df_w	1.0	1.2	1.4	1.6	1.8	2.0	2.2	2.6	3.0
4	.18	.23	.30	.38	.46	.54	.62	.76	.86
8	.23	.32	.42	.52	.63	.72	.80	.92	.97
……									
16	.27	.38	.49	.61	.72	.81	.88	.96	.99
……									
∞	.32	.44	.57	.70	.80	.88	.94	99	*

【例 11-11】 以例 10-1 为例，求该次方差分析结果的统计检验力。

解：计算单因素方差分析统计检验力的过程有如下述：

（1）根据实际得到的方差分析统计量 F 值，用公式 11-13 计算 **f** 值：

$$\mathrm{f} = \sqrt{\left(\frac{k-1}{k}\right)\frac{F}{n}} = \sqrt{\left(\frac{3-1}{3}\right)\frac{7.814}{6}} = \sqrt{0.67\times1.30} = 0.93$$

（2） 根据公式 11-12 计算Φ值：

$$\Phi = \mathrm{f}\sqrt{n} = 0.93\times\sqrt{6} = 2.28$$

（3）查附表 12 中对应于 k=3 的分表，应该查 $df_W = 15$（最接近的自由度值是 $df_W = 16$），$\Phi = 2.28$（介于 $\Phi = 2.2$ 和 $\Phi = 2.6$ 之间）的交叉点，因为没有准确的值，因此，例 10-1 方差分析的统计检验力 $1-\beta$ 应该介于 0.88-0.96 之间。这与表 **11-4** 所示通过 SPSS 软件包计算给出的“$1-\beta$=0.91”的结果基本一致。

答：本次方差分析结果的统计检验力为“$1-\beta$=0.91”。

J.Cohen（1988）指出，方差分析的统计检验力与效果大小之间具有密切的联系，这种关系可以通过公式 11-20 来表示：

$$f=\sqrt{\frac{\eta^2}{1-\eta^2}} \qquad \text{（公式 11-19）}$$

二、两因素方差分析的效果大小和统计检验力

两因素有交互作用方差分析的效果大小计算方法，与单因素完全随机设计方差分析求取效果大小和统计检验力的过程基本相同。下面我们以例 10-8 所示方差分析结果的数据为例分别介绍使用两因素方差分析进行假设检验后其效果大小和统计检验力的计算过程。传统的心理统计方法对例 10-8 的数据进行方差分析后得到的结果有如表 11-6（即表 10-22）所示。

表 11-6（表 10-22）　方差分析表

变异来源	平方和	自由度	均方	F 值
因素 A(人员)	64.8	1	64.8	18.65**
因素 B(棋局)	51.2	1	51.2	14.73**
交互作用 $A\times B$	39.2	1	39.2	11.28**
误差 E	55.6	16	3.475	
总变异 T	210.8	19		

（一）两因素方差分析效果大小的计算方法

【例 11-12】根据表 11 - 6 中所列方差分析结果的数据计算其因素 A、因素 B 和两因素交互作用的效果大小的值。

解：

在论述单因素方差分析效果大小的求解方法时，虽然不同心理统计学家提出了不同的求解方法，但目前更多的学者认同使用“η^2”来做效果大小的指标，因此，在讨论两因素方差分析效果大小的计算方法时，“η^2”仍然是我们用于反映其效果大小的首选统计指标。

不过，如何求取“η^2”的值，则主要有以下两种不同的方法。

解法 1：直接用公式 11-15 即 $\eta^2=\frac{SS_{组间}}{SS_{总体}}$ 计算各因素的“η^2”做相应因素效果大小的指标（Howell，2011；Runyon，2004；张敏强，2010）。

使用这种方法求解表 11-8（表 10-22）所列数据各因素效果大小的计算过程如下：

$$\eta_A^2=\frac{SS_A}{SS_T}=\frac{64.8}{210.8}=0.31$$

$$\eta_B^2=\frac{SS_B}{SS_T}=\frac{51.2}{210.8}=0.24$$

$$\eta_{A\times B}^2=\frac{SS_{AB}}{SS_T}=\frac{39.2}{210.8}=0.19$$

答：以η^2作为本例方差分析后各因素效果大小的指标，结果分别为：因素 A 的效果大小为$\eta_A^2 = 0.31$，、因素 B 的效果大小为$\eta_B^2 = 0.24$，两因素交互作用的效果大小为$\eta_{A\times B}^2 = 0.19$。

解法 2：B. Cohen 等心理统计学家认为用η^2作为两（或多）因素方差分析后各因素效果大小的指标，其值偏小，因此主张使用偏η^2（partial_η^2，或η_p^2）作为评估各因素效果大小的指标，其计算公式为：

$$\eta_{effect}^2 = \frac{SS_{effect}}{SS_{effect} + SS_W} \quad \text{（公式 11-20）}$$

或

$$\eta_{effect}^2 = \frac{df_{effect}F_{effect}}{df_{effect}F_{effect} + df_W} \quad \text{（公式 11-21）}$$

式中，df_{effect}、SS_{effect}和F_{effect}分别是与各因素对应的自由度、平方和与 F 值。例如，如果是对 A 因素的效果大小进行计算，则这三者就分别是df_A、SSA和F_A；SS_W是误差项的平方和；df_W是误差项的自由度。

使用这种方法求解表 11-8（表 10-22）所列数据各因素效果大小的计算过程如下：

$$\eta_A^2 = \frac{SS_A}{SS_A + SS_W} = \frac{64.8}{64.8 + 55.6} = 0.5382$$

$$\eta_B^2 = \frac{SS_B}{SS_B + SS_W} = \frac{51.2}{51.2 + 55.6} = 0.4794$$

$$\eta_{A\times B}^2 = \frac{SS_{A\times B}}{SS_{A\times B} + SS_W} = \frac{39.2}{39.2 + 55.6} = 0.4135$$

答：以偏η^2作为本例方差分析后各因素效果大小的指标，结果分别为：因素 A 的效果大小为$partial_\eta_A^2 = 0.5382$，、因素 B 的效果大小为$partial_\eta_B^2 = 0.4794$，两因素交互作用的效果大小为$partial_\eta_{A\times B}^2 = 0.4135$。

使用上述介绍的两种不同方法，对表 11-8 所列方差分析结果中各因素效果大小的计算得到了不同的计算结果。这两种方法都有其存在的数学基础，因此，两种方法计算的结果都可以作为衡量方差分析结果中各因素效果大小的指标。不过，我们要注意这两种方法所得结果的名称和内涵的不同：前者计算结果称为"η^2"，其内涵是指"某个因素的平方和在总平方和中所占的比值"；后者计算结果称为"偏η^2"，其内涵是指"某因素平方和与'该因素平方和与误差项平方和之和'相比占多大的比值"，其分母部分不包含其他因素（或交互作用）的平方和。

为使读者对多因素方差分析结果中各因素效果大小的求取方法有更为全面的认识，下面我们介绍如何使用 SPSS 软件包求解本例效果大小值的操作过程（如图 11-8 所示，与图 10-5 所

示求解例 10-8 的操作过程相比较，其操作过程除增加了第五步之外，其他操作过程都是一样的)。将 SPSS 给出的计算结果进行简单整理后有如表 11-7 所示。

1. 分析→一般线性模型→单变量
2. 因变量框：记忆成绩
3. 固定因子框：人员、棋局
4. $\boxed{\text{模型}}$:（1）选择“全因子”　（这是 SPSS 默认选项）
   （2）$\boxed{\text{继续}}$
5. 选项：（1）输出框中选择功效估计、检验效能
   （2）$\boxed{\text{继续}}$
6. $\boxed{\text{确定}}$

图 11-8 在 SPSS 中求解两因素方差分析功效估计、检验效能各步骤示意图

表 11-7　主体间效应检验表

变异源	平方和	自由度	均方	F	Sig.	偏 η^2	检验力
人员（因素 A）	64.80	1	64.80	18.65	.001	.538	.982
棋局（因素 B）	51.20	1	51.20	14.73	.001	.479	.949
人员 * 棋局	39.20	1	39.20	11.28	.004	.414	.883
误差	55.60	16	3.48				
总体	794.0	20					

a. R 方 = .736（调整 R 方 = .687）

b. 使用 alpha 的计算结果 = .05

将前面所述两种方法计算多因素方差分析各因素效果大小的结果相比较，似乎可以推断 SPSS 软件包给出的结果是按前述第二种方法即求取偏 η^2 的方法所求取。

（二）两因素方差分析统计检验力的计算方法

两因素方差分析统计检验力的计算方法与单因素方差分析统计检验力的计算方法基本相同，也是按三个步骤来求取。

【例 11-13】以第六章中例 6-9 中的数据用 SPSS 求其统计检验力。

解法 1：是以各因素的备择假设期望值 Φ 和 **f** 代表的效果大小作指标的计算方法。具体计算过程如下：

（1）根据方差分析实际得到的各因素 F 统计量的值，用公式 11-13 计算小写粗体“f 值”

$$\mathbf{f_A}=\sqrt{\frac{(k-1)}{k}\cdot\frac{F}{n}}=\sqrt{\frac{(2-1)}{2}\cdot\frac{18.65}{10}}=\sqrt{0.5\times1.865}=0.965$$

$$\mathbf{f_B}=\sqrt{\frac{(k-1)}{k}\cdot\frac{F}{n}}=\sqrt{\frac{(2-1)}{2}\cdot\frac{14.73}{10}}=\sqrt{0.5\times1.473}=0.858$$

$$\mathbf{f_{A\times B}}=\sqrt{\frac{(k-1)}{k}\cdot\frac{F}{n}}=\sqrt{\frac{(2-1)}{2}\cdot\frac{11.28}{10}}=\sqrt{0.5\times1.128}=0.751$$

（2）用公式 11-19 即$\Phi = f\sqrt{n}$求各因素的“Φ值”：

$$\Phi_{A} = f\sqrt{n} = 0.965\times\sqrt{10} = 3.05$$

$$\Phi_{B} = f\sqrt{n} = 0.858\times\sqrt{10} = 2.71$$

$$\Phi_{A\times B} = f\sqrt{n} = 0.751\times\sqrt{10} = 2.37$$

（3） 根据Φ值、分组数 k，组内自由度df_{w}查附表 12，求得统计检验力

本例中，$df_{W} = 16$，因素 A 的$\Phi_{A} = 3.05$（表中Φ值最高为$\Phi = 3.0$），因此，因素 A 的统计检验力$1-\beta$比 0.98 更高；因素 B 的$\Phi_{B} = 2.71$（介于$\Phi = 2.6$和$\Phi = 3.0$之间），因此因素 B 的统计检验力$1-\beta$应该介于 0.93-0.98 之间；两因素交互作用的$\Phi_{A\times B} = 2.375$（介于$\Phi = 2.2$和$\Phi = 2.6$之间），因此，两因素交互作用方差分析的统计检验力$1-\beta$应该介于 0.83-0.93 之间。

将这一结果与表 11-9 通过 SPSS 软件包计算后给出的统计检验力的结果基本相符。

答：本例方差分析后各因素的统计检验力分别为：因素 A 的统计检验力要高于 0.98；因素 B 的统计检验力应该介于 0.93-0.98 之间；两因素交互作用的方差分析的统计检验力$1-\beta$应该介于 0.78-0.92 之间。

小　结

本章先是以两独立样本平均数差异显著性检验为例介绍了假设检验后估计统计检验力和效果大小的基本原理，然后分别介绍了平均数差异显著性检验、方差分析的统计检验力和效果大小的估计方法。

估计统计检验力的基本原理是以对两种假设分布的分析为基础的。

虚无假设（$H_0 : \mu_1 = \mu_2$）分布是以零为中心的正态分布，由于我们可以通过预先设定α水平的方式来控制当虚无假设为真时拒绝它可能会犯错误的概率。因此，在此基础上得到的虚无假设差异显著性检验的 Z 分布（或 t 分布）在置信度范围内也是以零为中心的分布。

由于两个平均数之间不相等的值几乎是无限多的，因此，备择假设（$H_1 : \mu_1 \neq \mu_2$）的分布是一个不以零为中心的分布，它以什么值为中心也有着无数多个选择，因此，一般情况下，我们是难以对β或（$1-\beta$）值作出准确的估计的。

对某次假设检验统计检验力的估计原理以下述两条假定为基础：

假定一：假定“备择假设的分布”服从正态分布，并且用符号“δ”来表示这一“备择假设”正态分布的期望值；

假定二：假定“备择假设”正态分布的期望值与这次假设检验的“Z 统计量（或‘t’统计量”相等

在上述两个假定基础上，两样本平均数差异显著性检验通常通过以下五个步骤来估计来对某次假设检验结果的统计检验力或犯二型错误的概率值进行估计：

（1）根据已知条件建立需要检验的假设；

（2）用相应的公式计算 Z 统计量；

（3）确定做出统计决策的α水平及相应的临界值；

（4）计算实际得到的 Z 值与α水平临界值的差；

（5）根据 Z 值与α水平临界值的差查正态分布表，确定可能犯的β型错误或统计检验力 1-β的概率。

方差分析的统计检验力则通常需要通过以下三个步骤来求取：

（1）用公式 11-13 求小写粗体“**f** 值”的值（如前所述也是一种效果大小的指标）；

（2）用公式 11-19 求备择假设分布的期望值“Φ”：

（3） 根据Φ值、分组数 k，组内自由度df_w查附表 12，求得统计检验力

统计检验力与效果大小、显著性水平、样本容量等因素之间都有密切关系。

（1）统计检验力与显著性水平的关系是：做统计决策时选用的显著性水平越严格，则相应的统计检验力 1-β的概率值越小；

（2）统计检验力与样本容量 n 的关系是：在其他条件不变的前提下，样本容量 n 越大，由此得出的统计检验值（如 Z 值或 t 值等）就会越大，同时相应的统计检验力 1-β的概率值也会越大；

（3）统计检验力与效果大小的关系是：在其他条件不变的前提下，效果大小的值越大，由此得出的统计检验值（如 Z 值或 t 值等）也会越大，同时相应的统计检验力 1-β的概率值也会越大；

当假设检验的结果是接受虚无假设时，我们关心的是犯“纳伪”错误可能性即β值的大小：β值越小，那么犯纳伪错误的可能性就越小（这种情况下不存在“统计检验力”的概念）；

当假设检验的结果是拒绝虚无假设时，我们关心以下两个指标：

（1）统计检验力$1-\beta$，它反映的是我们通过这次检验的结果做出“拒绝虚无假设”这一推断的把握度的大小：“$1-\beta$”值越大，意味着我们有较大的把握认为“虚无假设是虚假的”；反之，“$1-\beta$”值越小，那么，虽然统计学意义上这次假设检验结果表明两个平均数之间的差异具有显著意义，但我们做出“虚无假设是虚假的”这一推断的自信度并不高。

（2）效果大小 d 值，它反映的是通过抽样所得到的两个平均数各自所属的两个总体之间的实际差异有多大：d 值反映这次假设检验过程中，备择假设与虚无假设所存在的实际差异越大。

由于心理学研究中使用心理统计方法主要是要了解研究者在实验中所控制的自变量对因变量是否造成有意义的影响，而效果大小正好能达到这一要求，因此在实际应用中我们通常更关注效果大小的值，而不怎么关注统计检验力“$1-\beta$”值的大小。

关键术语

统计检验力：在虚无假设（H_0）为假（也即备择假设（H_1）为真）时，正确拒绝这一虚假的虚无假设的概率，通常以符号“$1-\beta$”表示。

效果大小：在虚无假设（H_0）为假时，对该次假设检验过程中所存在的备择假设（H_1）的期望值与虚无假设的期望值之间的距离的测量值。对两样本进行平均数差异显著性检验，无论是用 Z 检验还是用 t 检验，通常都使用“Cohen d”作为该检验效果大小的指标；单因素方差分析的效果大小通常使用“η^2” 作为该检验效果大小的指标；两（或多）因素方差分析后则通常使用“偏η^2（$partial\_\eta^2$，或η_p^2）” 作为各因素效果大小的指标。

备择假设分布：是指当虚无假设H_0为假而备择假设H_1为真时，我们在这两个总体中每次各抽取一定数量的样本进行差异显著性检验，在进行无数多次抽样后，每次抽取的两个样本平均数之差（即“$\overline{X}_1 - \overline{X}_2$”之差）的分布。它是一个不以零为中心的分布。Z 检验或 t 检验中的备择假设H_1的分布通常是指“$\mu_1 \neq \mu_2$”的分布；Z 检验或 t 检验中备择假设分布的期望值通常用符号“δ”表示。方差分析中备择假设H_1的分布通常是指“$\mu_1, \mu_2, \cdots \mu_k$不完全相等”的分布。该分布的期望值通常用符号“$\Phi$”表示。

重要公式

两个独立样本平均数差异显著性检验的备择假设期望值δ值：

$$\delta = \frac{(\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$\delta = d\sqrt{\frac{n}{2}}$$

两个独立样本平均数差异显著性检验的效果大小d值：

$$d = \frac{(\mu_1 - \mu_2)}{\sigma}$$

$$d = \frac{\delta}{\sqrt{\frac{n}{2}}}$$

样本容量的调和平均数：

$$\overline{n}_H = \frac{2n_1 n_2}{n_1 + n_2}$$

在不同d值条件下，要达到某种程度的统计检验力所需要的样本容量：

$$n = 2\left(\frac{\delta}{d}\right)^2$$

方差分析的假设检验中备择假设的期望分布值：

$$\Phi = \mathrm{f}\sqrt{\mathrm{n}},$$

其中：

$$\mathbf{f} = \sqrt{\frac{(k-1)}{k} \cdot \frac{F}{n}}$$

样本容量的倒数平均数：$n_h = \dfrac{k}{\sum \frac{1}{n_i}}$

方差分析效果大小的η^2值：

$$\eta^2 = \frac{SS_B}{SS_T}$$

反映实验处理效果大小的ϖ^2值：

$$\varpi^2 = \frac{SS_B - (k-1)MS_e}{SS_T + MS_e}$$

两（或多）因素方差分析效果大小的偏η^2值：

$$\eta^2_{effect} = \frac{SS_{effect}}{SS_{effect} + SS_W}$$

或：

$$\eta^2_{effect} = \frac{df_{effect}F_{effect}}{df_{effect}F_{effect} + df_W}$$

思考与练习

1. 有研究者在研究中得到如下一组数据，$n_1=30, \overline{X}_1=114, \sigma_1=5$，$n_2=27, \overline{X}_2=112.5, \sigma_2=6.5$，请根据这组数据进行平均数差异显著性假设检验并计算检验后的统计检验力。如果实验结果需要达到 80%的统计检验力，每组需要多少被试？

2. 有人研究了不同类型的指导语对测验焦虑的影响。让 15 名被试参加一项知识测验，被试分为 3 组，每组 5 名被试。三组的指导语不同，在正式测验前，让被试参加一项焦虑测验，测验总分为 10 分，得到的分数越高，表明焦虑水平越高。实验结果如下表所示，问不同指导语对被试测验焦虑水平是否存在显著影响？假设检验后的效果大小的值和统计检验力的值各是多少？如果实验结果需要达到 80%的统计检验力，那么每组需要多少被试？

不同指导语下的测验焦虑得分表

	指导语 1	指导语 2	指导语 3
	8	5	2
	7	6	4
	9	7	5
	10	4	3
	6	3	6
$\overline{Y}_j$	8	5	4
S_j^2	2.5	2.5	2.5

第十二章　非参数统计

学习本章内容，将有助于你对以下问题的理解与思考：

1. 什么叫非参数统计？它有什么特点？
2. 参数检验方法与非参数检验方法有何区别？
3. 在什么情况下使用二项式检验比较合适？
4. 两独立样本的非参数检验应该注意哪些问题？
5. 两相关样本的非参数检验应该注意哪些问题？

某企业 A 准备正式投产一种桔子汁饮料，为此调查了 12 名消费者，了解他们对自己的产品和另一企业 B 的同类产品的喜好。提供给 12 名消费者的产品均没有贴出品牌标签，每名消费者第一次品尝的品牌是随机选择的。在下表中所示的数据中，1 表示喜好 A 企业的产品，0 表示喜好 B 企业的产品（表 12-1）。这个结果是否表明消费者对两个企业所生产的桔子汁的喜好态度存在差异？

表 12-1 消费者对不同企业生产的桔子汁类产品的偏好数据表

消费者	偏好产品的生产企业	记录的数据
1	A	1
2	A	1
3	B	0
4	A	1
5	A	1
6	A	1
7	A	1
8	A	1
9	B	0
10	A	1
11	A	1
12	A	1

第一节　非参数统计概述

一、参数统计与非参数统计

本书之前介绍的诸如 t 检验、F 检验等统计方法都属于参数统计方法，这些方法往往假设总体分布形式是已知的，根据样本资料来检验各自来自的总体是否具有相同的分布时，就转变为对分布的有关参数是否相等的检验。例如“两个正态总体的检验问题”，设有两个相互独立的 X 和 Y，具有相同方差的正态分布，即

$$X \sim N(\mu_1,\ \sigma^2)，Y \sim N(\mu_2, \sigma^2)$$

其中 μ_1, μ_2, σ^2 是未知参数，μ_1，μ_2 的范围均在（$-\infty, +\infty$），$\sigma^2 > 0$。这时检验两个总体是否具有相同分布，就转化为参数的检验问题：

$$H_0: \mu_1 = \mu_2，\ H_1: \mu_1 \neq \mu_2。$$

因此，所谓参数检验（parametric test）是指对总体分布服从正态分布或总体分布

已知条件下对诸如总体平均数、总体方差、总体标准差、总体相关系数等总体参数所进行的统计检验。研究这一问题的统计分支，也就是在假设总体分布形式是已知的情况下根据样本资料来进行参数估计或假设检验的统计分析方法统称为参数统计（parametric statistics）。

绝大多数的参数统计方法要求使用的数据是等距变量或是比率变量，在这些变量水平上进行数学运算是有意义的，例如可以计算均值、标准差、方差，并用于分析。对于称名变量或顺序变量，计算均值、标准差和方差是不恰当的，因此，通常情况下不能使用参数统计方法。另外，在许多的实际场合，常常因为缺乏足够信息，人们无法确定一个总体究竟服从哪种分布形式，所以无法使用参数统计方法。如果为了使用参数统计而牵强附会地假设总体具有某种分布形式，这样得出的结果是不可靠的，或者说是不合理的。

这时，就需要一种统计检验方法，能够适用于称名变量或顺序变量的统计检验，或者虽然是等距变量或比率变量，但检验的正确性和有效性并不依赖于总体的特定分布。而适用于这些条件的统计检验方法就是非参数检验方法。换言之，所谓非参数检验（nonparametric test）是指能够适用于称名变量或顺序变量，或总体分布未知条件下的等距变量或比率变量的统计检验方法。研究这一问题的统计分支，也就是在总体分布未知或知之甚少的条件下根据样本资料下对总体性质和特征进行推论的统计分析的方法统称为非参数统计（nonparametric statistics）。具体而言，一种统计方法被归为非参数方法，它至少应该满足下列条件之一：

（1）这种方法可用于称名变量。

（2）这种方法可用于顺序变量。

（3）当总体的概率分布不遵从正态分布，或无法对总体的概率分布做出假设时，这种方法可用于等距变量或比率变量。

二、非参数检验方法的特点

非参数检验方法是相对于参数检验方法而出现的，其特点也常常是与参数检验方法对比而言的。

1.非参数检验方法的优点

（1）适用面广。非参数检验方法不仅可以用于等距变量、比率变量的统计检验，还可用于称名变量、顺序变量的统计检验，其适用面较参数检验方法更广。

（2）假设条件少。这是它与参数检验方法相比的最大优点。几乎每种参数检验都有一些严格的假设前提条件，例如总体遵从正态分布等，若不满足这些条件仍使用参数检验方法，很可能会得出错误结论，而非参数检验方法的不必过多考虑这些条件。

（3）具有稳健性。稳健性（Robustness）指的是，当真实模型与假定的理论模型有不大的偏离时，统计方法仍能维持较好的性质，至少不会变得很坏。参数检

验方法建立在严格的假设条件基础上，一旦条件不满足，其推断的正确性就会受到严重影响。非参数检验方法，由于假设条件弱，对模型的限制少，因而天然地具有稳健性，这是非参数检验方法常被使用的一个原因。

（4）非参数检验方法适用于小样本，计算简便。在心理学研究领域中，常常在正式实验之前需要做一些预备性的实验，这时被试人数较少且要求结果尽快处理，使用非参数检验方法非常方便。

2.非参数检验方法的不足

（1）未能充分利用资料的全部信息。非参数检验的方法常常对测量数据进行等级转换，或只考虑符号而忽视其大小，这都导致一些信息的丢失。但当数据不符合参数检验的假设条件时，虽然非参数检验方法不能充分利用资料的信息，但比不顾条件地采用参数检验方法要好。

（2）检验功效比参数检验方法低。当假设条件满足时，非参数统计方法虽然也可以使用，但检验功效不如参数检验方法。这是研究人员愿意采用参数检验的原因之一。但假设条件不满足时，参数检验的结果是缺乏解释意义的，此时，宁可采用检验功效稍低的非参数检验方法。

（3）非参数检验方法目前还不能处理交互作用。很多研究，希望考察因子之间是否存在交互作用，例如多因素方差分析，但目前的非参数检验方法还不能处理交互作用。

三、非参数检验的预备知识

1.秩和秩统计量

对于样本 x_1，x_2，…，x_n，假定它们的值各不相同，将 x_1，x_2，…，x_n 由小到大排成一列。若 x_i 在这一列中占据第 R_i 位，则称 x_i 的**秩**（Rank）为 R_i（i=1,2,…,n)，称（R_1，R_2，…，$_i$）是原样本 x_1，x_2，…，x_n 的**秩统计量**（rank statistic）。例如，成绩排名（因为成绩变量的特殊性，应该采用倒数排名，即从大到小排名），每个学生的成绩对应的倒数排名就是秩。基于秩统计量所产生的方法，它在非参数统计方法中占有极其重要的地位。

2.结

在许多时候，样本 x_1，x_2，…，x_n 会存在相同的值，此时排序的结果就会出现同秩的现象，这种情况被称为数据中的结（ties）。例如，考试成绩排名中会出现并列第3、并列第 10 等并列名次，这就是结。结中数据的秩为它们按大小顺序排列后所处位置的平均值。例如，数据 25，26，26，28，28，28，30，36，那么数据 26 对应的秩应该为(2+3)/2＝2.5，因为 26 占据了第 2 名和第 3 名两个位置；同样，28 对应的秩应该为(4+5+6)/3=5。因此，相应的秩统计量为（1，2，5，2，5，5，5，6，7）。

四、参数检验方法与非参数检验方法的对照

在统计检验时，常常根据数据单样本、两独立样本、两相关样本和多个独立样本、多个相关样本等情况，而采用相应的统计检验方法。当条件许可时，例如，来自正态总体、方差齐性等，我们可以采用参数检验的方法。但如果条件不许可，例如，参数检验的假设前提不满足或是变量类型不满足参数检验的要求，我们可以采用非参数检验的方法。表 12-2 列出了参数检验与非参数检验常用方法的对照情况。

表 12-2　非参数方法与参数方法对应表

样本情况	非参数方法	参数方法
单样本	二项式检验（含符号检验、中数检验） 游程检验 K-S 检验	单总体均值假设检验
两独立样本	Mann-Whitney U 检验 K-S 检验 Wald-Wolfowitz 游程检验 Moses extreme reactions 检验	独立样本 t 检验
多个独立样本	Kruskal-Wallis H 检验 中数检验 Jonckheere-Terpstra 检验	单因素完全随机设计方差分析
两相关样本	符号检验 Wilcoxon 符号秩检验 McNemar 检验 边际齐性检验	配对样本 t 检验 相关系数
多个相关样本	Friedman 检验 Kendall W 检验 Cochrans Q 检验	随机区组设计方差分析

第二节　单样本非参数检验

单样本非参数检验通常属于拟合优度检验，通过考察观察次数和特定分布下的期望次数是否存在显著差异，回答样本来自某特定分布的总体的假定是否合理。

一、二项式检验的含义与方法

二项式检验（binomial test）是对二分变量的两个类别的观察频率与指定概率参数的二项分布下的期望频率是否存在显著差异进行检验的一种非参数检验方法。例如，抛一枚硬币试验。正面朝上的概率等于 1/2，假设抛掷 40 次得到 25 次下面朝上的结果，那么，我们可以根据二项分布的规律，计算在服从二项分布的前提下得到 25 次正面朝上结果的可能性有多大，从而做出抛硬币试验结果是否有偏的结论。

在实际问题中，如果所研究的问题，每次试验只有两种结果：“成功”或“失败”，“赞成”或“反对”，等等，并且这两种结果的出现被假定服从二项分布，那么，当我们希望判定，一个“成功”概率为 p（“成功”的次数为 n_p），“失败”概率为 q（“失败”的次数为 n_q）的随机样本，是否来自某特定参数的二项分布时，就可采用二项式检验。试验的总次数 $n_T = n_p + n_q$。实际研究的数据不一定都是二分类的数据，但如果可以依据某一规定将其转化为二分数数据时，也可采用二项式检验。

记某二项分布总体参数“成功”概率为 π， 根据一随机样本的资料来检验其是否来自成功概率为的 π_0总体，则建立假设为：

$$\begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi \neq \pi_0 \end{cases}$$

然后根据二项分布的公式或查附表 13，计算在 H_0 为真的前提下，得到本样本试验数据的概率。根据这一概率是否为小概率，来作出是否拒绝 H_0 或是无充分理由拒绝 H_0的统计决策。

以上为 $n_T \leqslant 20$ 的情况。当 $n_T > 20$，可以采用正态分布近似解决，先用公式 12-1 计算 $Z_{+,R}$，$Z_{-,R}$统计量，再查附表 1（正态分布表），作出判定。

$$Z_{+,R} = \frac{n_p - 0.5 - n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}} \quad , \quad Z_{-,R} = \frac{n_q - 0.5 - n(1-\pi_0)}{\sqrt{n\pi_0(1-\pi_0)}} \qquad \text{（公式 0.1）}$$

【例 12-1】　以本章开头的问题为例：这个结果是否表明消费者对两个企业所生产的橘子汁的喜好态度存在差异？（本例题原始数据见本书所附光盘中例 12-1.sav。）

解：

由于每次试验结果是非此即彼的两种结果“喜好 A”或“喜好 B”，因此可以考虑用二项式检验来进行。如果消费者对两个企业的产品喜好态度不存在显著差异，那么各自喜好的人数应该相等，即总体上 π＝0.50。因此，检验假设为：

$$\begin{cases} H_0 : \pi = 0.50 \\ H_1 : \pi \neq 0.50 \end{cases}$$

本例中，$n_T=12, n_p=10$，$p=10/12$；$n_q=2$，$q=2/12$。

查附表 13 得，在 $\pi=0.50$（表中为 p=0.50），n=12 时，x=10（即 $n_p=10$）的概率为 0.0161。但由于双侧检验的 p 值是二项抽样分布尾部概率的 2 倍，见图 12-1。因此，本例假设检验的 p 值应该是(0.0161+0.0029+0.0002)×2=0.0384。由于 p 值小于 α=0.05，所以拒绝 H_0，即认为消费者对两个企业所产的橘子汁类饮料的偏好存在显著差异。

答：这 12 名消费者对两个企业所产的橘子汁类饮料的偏好存在显著差异。

当然，二项式检验也可以利用二项分布展开式计算得到相应的 p 值。如本例中，$p\text{值}=(C_{12}^{10}0.5^{10}0.5^{2}+C_{12}^{11}0.5^{11}0.5^{1}+C_{12}^{12}0.5^{12}0.5^{0})\times 2$，但由于计算麻烦，一般还是查表得到 p 值。

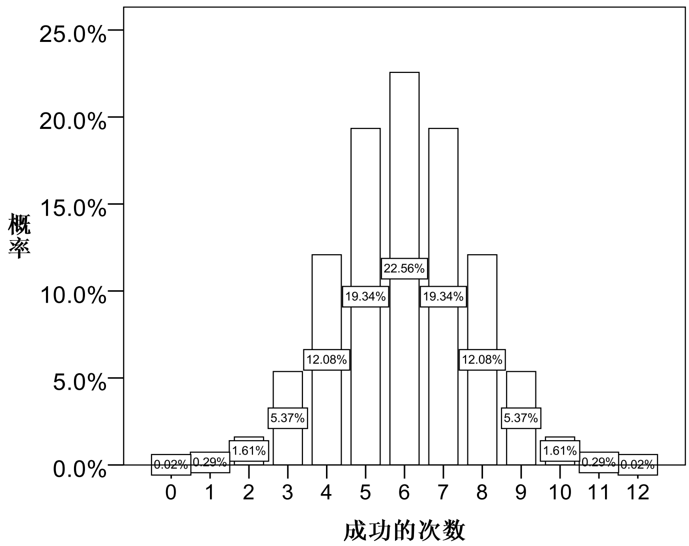


图 12-1　n=12，$\pi=0.50$ 的成功次数的二项抽样分布

【例 12-2】 某企业生产一种钢管，规定长度的中数是 10 米。某天随机抽取了 10 根进行测量，结果见表 12-3。问该天生产的钢管长度是否符合质量标准。（本例题原始数据见本书所附光盘中例 12-2.sav。）

表 12-3　抽取的钢管的长度测量结果

样本	1	2	3	4	5	6	7	8	9	10
长度	9.8	10.1	9.7	9.9	9.8	10.0	9.7	10.0	9.9	9.8

解：

如果抽取的 10 根钢管的长度中数为 10 米，则表明该天生产的钢管长度符合质量标准，那么，10 根钢管就会有一半的长度小于 10 米，一半的长度大于 10 米。因此，对这 10 根钢管的长度中数为 10 米，也可采用二项式检验。因此，对表 12-3 的数据进行整理，大于 10 米的记为 1，小于 10 米的记为 0，等于 10 米的不记录。通过这种方法对表 12-3 中的数据处理后，所得结果有如表 12-4 所示：

表 12-4 抽取的钢管的长度与 10 米相比的二项式比较结果表

样本	1	2	3	4	5	6	7	8	9	10
长度	9.8	10.1	9.7	9.9	9.8	10.0	9.7	10.0	9.9	9.8
与 10 米相比	0	1	0	0	0		0		0	0

$$\begin{cases} H_0: \pi = 0.50 \\ H_1: \pi \neq 0.50 \end{cases}$$

此时，n_T＝8，n_p＝1，p＝1/8；n_q＝7，q＝7/8。

查附表 13 得，在 π＝0.50，n=8 时，x=1（即 n_p＝1）的概率为 0.0312。与例 12-1 相同，本例假设检验的 p 值应该是(0.0312+0.0039)×2=0.0702。由于 p 值大于 α=0.05，所以无充分理由拒绝 H_0，即认为该天生产的钢管长度是符合质量标准的。

答：该天生产的钢管长度是符合质量标准的。

【例 12-3】 某大学原来晚上就寝熄灯时间为 23:00，后来改为不统一熄灯。现在有人提出还是统一就寝熄灯较好，学校为此进行了预调查，并决定如果有 25%以上的同学赞成统一熄灯就寝，就改回原来的规定。随机抽取的 18 个同学中，有 7 人赞成统一熄灯，11 人反对。学校应该如何决策？

解：

这个问题实质上是一个二项式检验的问题，需要调查数据支持这样一个结论：成功的概率大于 25%。因此，相应的假设为：

$$\begin{cases} H_0: \pi = 0.25 \\ H_1: \pi > 0.25 \end{cases}$$

此时，n_T＝18，n_p＝7，p＝7/18；n_q＝11，q＝11/18。

查附表 13 得，在 π＝0.25，n=18 时，x=7（即 n_p＝7）的概率为 0.0820，这表明在 H_0：π＝0.25 成立的前提下，得到 n_p＝7 这个结果的概率是 0.0820。本例是一个单尾检验的例子，应该将 n_p＝7 至 18 的概率相加，即 p 值应该是 0.0820+0.0376+⋯+0.0000=0.1389，也就是说，H_0：π＝0.25 成立的前提下，得到 n_p≥7 的概率为 0.1389。显然，本例假设检验的 p 值应该大于 α=0.05，所以无充分理由拒绝 H_0，即没有充分证据表明学生赞成统一熄灯的人数比例超过 25%。

答：学校可以继续实施不统一熄灯的规定。

二、游程检验的含义与方法

游程检验（run Test）是检验某一变量的两个值的出现顺序是否是随机的一种非参数检验方法，也称为连贯检验或串检验。游程指的是，当样本按某种顺序排列（如抽样时间的先后）时，一串不间断的 0（或 1）所组成的序列。游程中数字 0 或 1 的个数，称为游程的长度。这里的 0 或 1 表示观察值的两种类型。例如序列

1　1　0　0　0　1　1　0　0　1　1　1　0

中有 6 个游程，用 $R=6$ 表示。其中 3 个由 0 组成的游程，长度分别为 3,2,1；记数字 0 的个数为 $m=6$。3 个由数据 1 组成的游程，长度分别为 2,2,3；记数字 1 的个数为 $n=7$。

如果一个序列游程个数 R 太小，说明 0 和 1 过分集中，例如序列

0　0　0　0　0　0　1　1　1　1　1　1　1

$R=2$，游程总数过少，表明某一游程的长度过长，意味着有较多的相同观察值相连，序列存在成群的倾向。

如果一个序列游程个数 R 太大，说明 0 和 1 过分分散，例如序列

1　0　1　0　1　0　1　0　1　0　1　0　1

$R=13$，游程总数过多，表明游程长度很短，意味着两个观察值频繁交替，出现一种周期性的变化，序列具有混合的倾向。

因此，游程总数过多或过少，都表明序列不是随机的。至于何为过多或过少，可以查附表 14，拒绝域为 $\{R\leqslant c_1\}\cup\{R\geqslant c_2\}$。在上面的 3 个序列中，$m=6$，$n=7$，在给定检验水平 $\alpha=0.05$，查附表 14 得 $c_1=3$，$c_2=12$。对于第一个数列，$R=6$，$3<R<12$，认为数据符合随机性假设；对于第二个数列，$R=2$，$R<3$，则认为序列中的数据不符合随机性假设；对于第三个数列，$R=13$，$R>12$，则认为序列中的数据也不符合随机性假设。

实际中的数据未必都是二分类数据，此时我们也可把数据转化为二分类数据。例如，指定大于中数的值为 1，小于中数的值为 0。

以上方法仅适用于 $m\leqslant10$，$n\leqslant20$ 的情形，当 m、n 较大时（$m+n=N>20$ 或 $m>12$，$n>12$），游程总数 R 近似均值为 $1+2mn/N$，标准差为 $\sqrt{2mn(2mn-N)/N^2(N-1)}$ 的正态分布，此时，可利用(公式 12.2)求得 Z 值，再查正态分布表（即附表 1）作出判定。

$$Z=\frac{(R+0.5)-(1+2mn/N)}{\sqrt{2mn(2mn-N)/N^2(N-1)}} \qquad \text{（公式 0.2）}$$

【例 12-4】 某城市坚持记录每天的气温，并与历史上的天气记录作比较，其中某月 10 天的最高气温与历史上同期最高气温平均值比较的结果如下数列：

1　1　0　1　0　0　1　1　1　0

1 表示高于历史同期最高气温平均值，0 表示低于历史平均值。请在 α=0.05 的水平，检验高温的偏差是否随机。

解：

$$\begin{cases} H_0:序列是随机的 \\ H_1:序列非随机的 \end{cases}$$

此时，m＝4，n＝6，R＝6，查附表 14 得 c_1＝2，c_2＝9。2<R<9，所以无充分理由拒绝 H_0，即认为该 10 天的气温与历史同期相比，其高温偏差是随机的。

【例 12-5】 生产流水线上要监控产品质量，就必须抽取一些产品样本。若不合格产品是随机出现的，那么每天可以间隔较长地等距抽样或随机抽样，即可得到一个较好的估计；如果不合格产品常常是“连续”出现，似乎存在成群的倾向，那么抽样则应以多次抽取小样本，以保证估计的可靠性。现从某生产流水线上随机抽取了 30 个产品，测量结果用 1 表示合格，用 0 表示不合格，按从生产线抽取的顺序排列如下

0 0 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 1 1 1 1 1 1

请在 α=0.05 的水平，检验出现不合格产品是否是随机的。

解：

$$\begin{cases} H_0:序列是随机的 \\ H_1:序列非随机的 \end{cases}$$

由于 m=7，n=23，m+n=30>20，所以采用(公式 12.2)，得

$$Z=\frac{(R+0.5)-(1+2mn/N)}{\sqrt{2mn(2mn-N)/N^2(N-1)}}=\frac{(4+0.5)-(1+2\times7\times23/30)}{\sqrt{2\times7\times23(2\times7\times23-30)/30^2(30-1)}}=-3.81$$

因为 $Z=-3.81<-Z_{\frac{0.05}{2}}=-1.96$，因此拒绝 H_0，即认为出现不合格产品不是随机的。

答：就本次抽样而言，出现不合格产品不是随机的。

三、单样本柯尔莫哥诺夫—斯米尔诺夫检验的含义与方法

单样本柯尔莫哥洛夫-斯米尔诺夫检验（Kolmogorov-Smirnov test）是将变量的观察累积分布函数与指定的理论分布进行比较，以检验观察值是否合理来自指定的理论分布的一种非参数检验方法。这里的理论分布可以是正态分布、均匀分布、泊松分布或指数分布等。检验常简写为 K-S 检验。

K-S 检验的检验统计量由观察累积分布函数和理论累积分布函数之间的最大差分（取绝对值）计算而得。若 $S_n(x)$ 表示一个 n 次的随机样本观察值的累积分布函数 $S_n(x)=i/n$，i 是等于或小于 x 的所有观察结果的数目，i＝1, 2, …, n。$F_0(x)$ 表示一个特定的累积分布函数，也就是说，对于作任一给定的 x 值，$F_0(x)$

值代表等于或小于 x 值的理论上的概率值。于是，$K-S$ 检验的检验统计量为：

$$D = \max|S_n(x) - F_0(x)| \quad \text{（公式 0.3）}$$

然后，查附表 15（单样本 $K-S$ 检验统计量）作出统计决策，拒绝域为 $D>d_{\alpha/2,n}$，其中 $d_{\alpha/2,n}$ 是根据 α 和 n 查表而得到的临界值。如果 $n>40$，则根据表中提供的公式计算出临界值。

【例 12-6】 按规定，某路公交车每 15 分钟通过一公交车站。实际上，由于多种因素的影响，公交车通过时间会有些差异。某天测量的 20 辆公交车到达时间情况如表 12-5 所示（提前到计为负值，比计划晚到为正值），问今天公交车到达时间的分布是否服从正态分布？

表 12-5　汽车到达时间统计表

到达时间（x）	-5	-3	-1	0	1	2	4	7	8
观测频率（f）	1	1	2	1	5	5	3	1	1

解：

采用 $K-S$ 检验关键要得到观察值的累积概率分布 $S_n(x)$ 和指定理论分布的累积概率 $F_0(x)$。其中 $S_n(x)$ 可以通过计算相对累积次数获得，见表 12-7 中的第 4 列。$F_0(x)$ 可将到达时间 x 转换为标准分数，然后查正态分布表获得。标准分数计算中的均值和标准差采用观测值的均值与标准差，本例中观测值的平均数为 1.6，标准差为 2.998（为计算方便采用标准差为 3）。

表 12-7　$S_n(x)$、$F_0(x)$ 和 $|S_n(x)-F_0(x)|$ 计算表

到达时间（x）	观测频率（f）	累积次数	$S_n(x)$	Z＝(x-1.6)/3	$F_0(x)$	$\|S_n(x)-F_0(x)\|$
-5	1	1	.05	-2.20	.0139	.0361
-3	1	2	.10	-1.53	.0630	.0370
-1	2	4	.20	-.87	.1922	.0078
0	1	5	.25	-.53	.2981	.0481
1	5	10	.50	-.20	.4207	.0793
2	5	15	.75	.13	.5517	.1983
4	3	18	.90	.80	.7881	.1119
7	1	19	.95	1.80	.9641	.0141
8	1	20	1.00	2.13	.9834	.0166

$$\begin{cases} H_0: S_n(x) = F_0(x) \text{ 对所有}x\text{而言} \\ H_1: \text{不是所有的}S_n(x)\text{与}F_0(x)\text{都相等} \end{cases}$$

$$D = \max|S_n(x) - F_0(x)| = 0.1983$$

查附表 15 得，$d_{\alpha/2,n}$＝0.294。因为 D＝0.1983＜ $d_{\alpha/2,n}$＝0.294，所以无充分理由拒绝 H_0，即认今天公交车到达时间服从正态分布。

【例 12-7】 某校一周内学生考勤情况统计如表 12-7 所示，问该校学生考勤情况是否正常。（指每天考勤人数应基本相等，维持一平均水平，即是否服从平均分布。）

表 12-7　某校一周内学生考勤情况统计表

星期 / 缺勤人数	一	二	三	四	五	六	合计
f_o	13	11	7	10	8	11	60

解：

$S_n(x)$、$F_0(x)$和$|S_n(x)-F_0(x)|$的计算见表 12-8。

表 12-8　$S_n(x)$、$F_0(x)$和$|S_n(x)-F_0(x)|$计算表

星期	缺勤人数	累积次数	$S_n(x)$	理论次数	理论累积次数	$F_0(x)$	$\|S_n(x)-F_0(x)\|$
一	13	13	.22	10	10	.17	.05
二	11	24	.40	10	20	.33	.07
三	7	31	.52	10	30	.50	.02
四	10	41	.68	10	40	.67	.01
五	8	49	.82	10	50	.83	.01
六	11	60	1.00	10	60	1.00	.00

$$\begin{cases} H_0: S_n(x)=F_0(x) \ \ \text{对所有}x\text{而言} \\ H_1: \text{不是所有的}S_n(x)\text{与}F_0(x)\text{都相等} \end{cases}$$

$$D=\max|S_n(x)-F_0(x)|=0.07$$

因为 n=60>40，根据附表 15 提供的 α =.05 临界值计算公式得，$d_{\alpha/2,n}$= $1.36/\sqrt{n}=1.36/\sqrt{60}=.1756$。因为 D=0.07< $d_{\alpha/2,n}$=0.1756，所以无充分理由拒绝 H_0，即认该校学生这一周内考勤情况服从均匀分布。

答：该校学生考勤情况属于正常。

四 单样本非参数检验的 SPSS 操作

（一）二项式检验

1. 以例 12-1 的资料为例（原始数据见本书所附光盘中“例 12-1.sav”。)，二项式检验的 SPSS 操作过程如下：

1.分析→非参数检验→旧对话框→二项式

2.“检验变量列表”框：记录的数据

3.“定义二分法”：选择“从数据中获取”　（因为本例中的数据是二值数据）

4.“检验比例”: 0.50　（因为本例中的要检验的是 π=0.50）

5.$\boxed{\text{确定}}$

2. 以例 12-2 的资料为例（原始数据见本书所附光盘中“例 12-2.sav”），数据中删除了恰好等于 10 的数据），二项式检验的 SPSS 操作过程如下：

1.分析→非参数检验→旧对话框→二项式

2.“检验变量列表”框：长度

3.“定义二分法”：选择“割点”，并输入 10　（因为本例中的数据不是二值数据，因此要输入分割点的值，将数据分为两组）

4.“检验比例”: 0.50　（因为本例中的要检验的是 π=0.50）

5.$\boxed{\text{确定}}$

3. 以例 12-3 的资料为例（（原始数据见本书所附光盘中“例 12-3.sav”），构建 SPSS 数据时要注意，第一条记录应该是属于 n_p 的记录），二项式检验的 SPSS 操作过程如下：

1.分析→非参数检验→旧对话框→二项式

2.“检验变量列表”框：态度

3.“定义二分法”：选择“从数据中获取”　（因为本例中的数据是二值数据）

4.“检验比例”: 0.25　（因为本例中的要检验的是 π=0.25）

5.$\boxed{\text{确定}}$

4. 二项式检验 SPSS 分析结果的解释

二项式检验 SPSS 分析结果十分清楚，以上三个命题分析结果的结构十分相似，此处以例 12-3 的 SPSS 分析结果（见表 12-9）为例予以解释。

表 12-9　例 12-3 的 SPSS 分析结果

		类别	N	观察比例	检验比例	精确显著性（单尾）
态度	群组 1	赞成	7	.39	.25	.139
	群组 2	反对	11	.61		
	总计		18	1.00		

表 12-8 显示 n_T＝18，n_p＝7，p＝7/18＝.39；n_q＝11，q＝11/18＝.61；π=0.25；p 值＝0.139。因为 p 值＝0.139>.05，所以无充分理由拒绝 H_0。

（二）游程检验

1. 以例 12-5 的资料为例（原始数据见本书所附光盘中 “例 12-5.sav”），游程检验的 SPSS 操作过程如下：

1.分析→非参数检验→旧对话框→游程

2.“检验变量列表”框：质量检验

3.“割点”：选择“设定”，并输入 0.5 （因为本例中的数据是二值数据,我们指定 0.5 只是为了区分 1 和 0；如果是连续数据，另外可采用中数或平均数等为分割点。）

4.$\boxed{\text{确定}}$

2. 游程检验 SPSS 分析结果的解释

以例 12-5 的分析结果为例，SPSS 分析结果见表 12-11。

表 12-11 例 12-5 的 SPSS 分析结果

	质量检验
Test Value[a]	.50
Total Cases	30
Number of Runs	4
Z	-3.811
Asymp. Sig. (2-tailed)	.000

表中“Test Value”是我们指定的分割点的值，在本例中，只要指定一个介于 0、1 之间的任一数值均可。“Total Cases”表明共有 30 个观测值；“Number of Runs”表示共有 4 个游程。最末一行表示检验的 p 值。

（三）K-S 检验

1. 以例 12-6 的资料为例（原始数据见本书所附光盘中 “例 12-6.sav”），*K-S* 检验的 SPSS 操作过程如下：

1.数据→加权个案：选择“加权个案”，在“频率变量”框中加入“观测频率” (本例中各个到达时间出现次数不等，所以要加权)

2.分析→非参数检验→旧对话框→1-样本 K-S(1)

3.“检验变量列表”框：到达时间

4.“检验分布”：选择“常规” （本例要检验的是到达时间是否服从正态分布，所以选择“常规”根据需要还可选择“相等”(均匀分布)、“泊松”、“指数分布”等。）

5.$\boxed{\text{确定}}$

2. K-S 检验 SPSS 分析结果的解释

以例 12-6 的资料为例，*K-S* 检验的 SPSS 分析结果如表 12-11。

表 12-11　例 12-6 的 SPSS 分析结果

		到达时间
N		20
Normal Parameters[a,b]	Mean	1.60
	Std. Deviation	2.998
Most Extreme Differences	Absolute	.197
	Positive	.197
	Negative	-.171
Kolmogorov-Smirnov Z		.881
Asymp. Sig. (2-tailed)		.420

表 12-12 表明正态分布的均值为 1.60，标准差为 2.998；$D=\max\left|S_n(x)-F_0(x)\right|=.0197$，最后的 p 值为.420，p 值 > 0.05，所以无充分理由拒绝 H_0。

第三节　独立样本的非参数检验

一、两独立样本的非参数检验

两独立样本的非参数检验主要有曼－惠特尼检验、柯尔莫哥诺夫－斯米尔诺夫 Z 检验、沃尔德－沃尔福威茨游程检验和 Moses 极端反应检验等方法。接下来我们介绍前三种方法。

（一）曼-惠特尼 U 检验

1. 基本含义与方法

曼-惠特尼 U 检验（Mann-Whitney U test）是最常用的两个独立样本非参数检验方法，用于检验两
个总体分布各自的中心位置是否相同。该方法与曼-惠特尼-维尔克松（Mann-Whitney-Wilcoxon）检验（简写为 M-W-W 检验）的原理和检验结果完全等价，仅仅是统计量的构造略有不同。

设有 x_1，$x_2, \cdots, x_m$ 和 y_1，$y_2, \cdots, y_n$ 是分别来自具有连续分布总体的两个样本，数据类型至少为等距变量，若是顺序变量，则每个观察值的大小能够被确定。若 M_X、M_Y 分别是 X、Y 总体的中心位置，M_X、M_Y 一般是总体的中数，则建立检验假设为：

$$\begin{cases} H_0: M_X = M_Y \\ H_a: M_X \neq M_Y \end{cases}$$

若 X、Y 存在算术平均数时，也可以建立两个均值相等的虚无假设。

将 m 个 x、n 个 y 的数据，一共 $N=m+n$ 个数据一起按从小到大排序，如果 H_0 为真，那么这 N 个数据可以被看作来自于共同总体的一个单一的随机样本。如果大部分的 y 大于 x，或大部分的 x 大于 y，将不能证实这个有序的序列是一个随机的混合，那么将拒绝 x 和 y 来自相同总体的假设。

为了证实 x 和 y 的样本数据是否为随机“混合”，可通过考察 N 个数值在“混合”样本中的位置次序，即对应的秩 R。在有结的情况下，每个结得到平均秩（见本章第一节有关内容）。分别计算样本 x 和 y 各数据对应的秩和，即

$$T_x=\sum_{i=1}^{m}R_i\text{，}T_y=\sum_{j=1}^{n}R_i \qquad \text{（公式 0.4）}$$

如果 H_0 为真，那么两个样本中各数据相应的秩就应当围绕着平均秩次 $(N+1)/2$ 均匀分布，样本 x 的秩和应当接近于 $m(N+1)/2$，样本 y 的秩和应当接近于 $n(N+1)/2$，如果实际得到的秩和与理论值差别较大时，则可推断两总体的中心位置存在显著差异。同时，可计算曼-惠特民 U 检验的检验统计量：

$$U_{xy}=mn+m(m+1)/2-\sum_{i=1}^{m}R_i=T_y-n(n+1)/2 \qquad \text{（公式 0.5）}$$

$$U_{yx}=mn+n(n+1)/2-\sum_{j=1}^{n}R_j=T_x-m(m+1)/2 \qquad \text{（公式 0.6）}$$

式中，U_{xy} 表示 y 的观察值大于 x 观察值的个数，U_{yx} 表示 x 的观察值大于 y 观察值的个数。

一般情况下，两组样本数据个数不等时，较少数据的组记为 x，即 $m\leqslant n$。在 $m\leqslant n\leqslant 10$ 时，可以根据 m、n 及 T_x 的值查附表 16（Mann-Whitney-Wilcoxon 分布表），得到相应的 p 值并作出统计决策。但表中给出的是单侧检验的 p 值，如果是双侧检验则应乘以 2。

在 $m\leqslant n\leqslant 20$ 时，可依据 m、n 查附表 17（Mann-Whitney U 的临界值表）。在 m 与 n 交界处有两个数字，上面一个是 $\alpha=.10$（双尾检验）和 $\alpha=.05$（单尾检验）的临界值，下面一个是 $\alpha=.05$（双尾检验）和 $\alpha=.025$（单尾检验）的临界值。如果检验统计量 U_{xy} 和 U_{yx} 中较小的者小于临界值，则拒绝 H_0。

当 m、n 都大于 20 时，检验统计量 U 近似服从正态分布，从而可将 U 值变换成相应的 Z 值

$$Z=\frac{U-mn/2}{\sqrt{\dfrac{mn(m+n+1)}{12}}} \qquad \text{（公式 0.7）}$$

式中，U 是 U_{xy} 和 U_{yx} 中的较小者。由于检验统计量 U 是两个样本统计量中较小

者，所以决策规则为：对于左侧检验，若 $Z<-Z_{\alpha}$，则拒绝 H_0；对于双侧检验，若 $Z<-Z_{\alpha/2}$，则拒绝 H_0。

2. 实例

【例 12-8】 某人提出一种指导学习的新方法，为了证明新方法的有效性，他从某班中随机抽取了 9 名同学，随机对其中 6 名同学采用新方法指导，对另 3 名同学采用传统方法指导。经一个月后，两组同学的一次统一考试成绩如表 12-12 所示。问新方法的效果是否好于传统方法。

表 12-12 两组同学统一考试的成绩表

传统方法组（X）	94	156	115			
新方法组（Y）	120	163	226	187	220	130

解：

$$\begin{cases} H_0: M_X = M_Y \\ H_a: M_X \neq M_Y \end{cases}$$

将考试成绩从低到高排列为如表 12-13 所示。

表 12-13 两组同学统一考试成绩合并后的排序表

成绩	94	115	120	130	156	163	187	220	226
秩（排名）	1	2	3	4	5	6	7	8	9
所属组别	X	X	Y	Y	X	Y	Y	Y	Y

此时，m=3，n=6，T_x=1+2+5=8，查附表 16 得 p 值＝0.048，因为这是单侧检验概率，本例为双侧检验，所以 p 值＝0.048×2＝.096>α＝0.05，所以无充分理由拒绝 H_0，即在.05 的显著性水平上，认为新方法的效果与传统方法没有差别。

$$U_{xy} = T_y - n(n+1)/2 = 37 - 6(6+1)/2 = 16$$

$$U_{yx} = T_x - m(m+1)/2 = 8 - 3(3+1)/2 = 2$$

此时，m=3，n=6，查附表 17 得临界值为 1，因为 U_{yx}＝2>1，所以无充分理由拒绝 H_0。

答：没有充分根据表明新方法的效果好于传统方法。

（二） 柯尔莫哥诺夫-斯米尔诺夫 Z 检验

（1）基本含义与方法

柯尔莫哥诺夫-斯米尔诺夫Z检验（Kolmogorov-Smirnov Z test）是两个独立样本是否来自同一总体的

一种非参数检验方法。简称为两独立样本的 $K-S$ 检验。它单样本的 $K-S$ 检验相似，可以对连续性资料的分布情况予以考察。不同的是，单样本的 $K-S$ 检验的检验统计量是观察累积分布函数和理论累积分布之间的最大绝对差值，而两独立样本 $K-S$ 检验是以两个样本的观察累积分布函数之间的最大绝对差值为基础的，当这个差很大时，就将这两个分布视为不同的分布。

假设从两个总体中各随机抽取一个样本，样本容量分别为 m、n，两个样本的观察累积分布函数分别记作 $S_1(x)$和 $S_2(x)$，其中

$$S_1(x) = \text{第一个样本观察值小于等于}x\text{的数目}/m \qquad \text{（公式 0.8）}$$

$$S_2(x) = \text{第二个样本观察值小于等于}x\text{的数目}/n \qquad \text{（公式 0.9）}$$

那么，两独立样本 $K-S$ 检验的检验统计量为：

$$D = \max|S_1(x) - S_2(x)| \qquad \text{（公式 0.10）}$$

根据 D 值进行统计决定需要查附表 18（两样本 $K-S$ 检验统计量）。当 $2 \leqslant m \leqslant n \leqslant 12$ 或 $m+n \leqslant 16$ 时，表中第一部分给出了相应的 p 值。当 $9 \leqslant m = n \leqslant 20$ 时，附表 18 中第二部分给出了相应的 p 值。表的最后给出了大样本的近似临界值。

2.实例

【例 12-9】 有人调查了某大学来自大城市和中小城市学生的家长学历背景，大城市共 274 名学生，中小城市 236 名学生，结果如表 12-15 所示，问不同生源地学生家长学历分布是否存在显著差异。

表 12-15　不同生源地学生家长学历分组表

学历分组	大城市	中小城市
小学及以下	58	31
初中	51	46
高中及中专	47	53
大专	44	73
本科	22	51
研究生以上	14	20
合计	236	274

解：

$$\begin{cases} H_0: \text{两总体分布相同} \\ H_a: \text{两总体分布不同} \end{cases}$$

$S_1(x)$、$S_2(x)$和$|S_1(x) - S_2(x)|$计算的计算见表 12-16。

表 12-15　$S_1(x)$、$S_2(x)$和$|S_1(x)-S_2(x)|$计算表

学历分组	次数		累积次数		$S_1(x)=$	$S_2(x)=$	$\|S_1(x)-S_2(x)\|$
	f_1	f_2	Σf_1	Σf_2	$\Sigma f_1/m$	$\Sigma f_2/n$	
小学及以下	58	31	58	31	.2458	.1131	.1327
初中	51	46	109	77	.4619	.2810	.1809
高中及中专	47	53	156	130	.6610	.4745	.1865
大专	44	73	200	203	.8475	.7409	.1066
本科	22	51	222	254	.9408	.9270	.0138
研究生以上	14	20	236	274	1.0000	1.0000	.0000
合计	236	274					

$D=\max|S_1(x)-S_2(x)|=.1865$

由于 $m=236$，$n=274$，$N=m+n=510$，为大样本，查附表 18 得 $p=.01$ 的临界值为

$$1.63\sqrt{N/mn}=1.63\sqrt{510/(236\times 274)}=.1448$$

因为 $D=.1865$，大于这个临界值，所以拒绝 H_0，这表明，不同生源地学生家长的学历水平分布存在显著差异。

答：不同生源地学生家长学历分布存在显著差异。

（三）沃尔德-沃尔福威茨游程检验的含义与方法

沃尔德-沃尔福威茨游程检验(Wald-Wolfowitz run test)又称串检验，简写为 W-W 游程检验。与两独立样本 K-S 检验一样，W-W 游程检验主要用于考察两总体的分布是否相同，或或者说主要是考察两样本资料是否来自相同分布的总体。

具体方法为，将从 X 中随机抽取的 m 个数据 x_1，$x_2,\cdots,x_m$，和从 Y 总体中随机抽取的 n 个数据 y_1，$y_2,\cdots,y_n$，共 $N=m+n$ 个数据一起从小到大排列，以确定这个序列的游程数。一个游程指的是取自同一样本的一串相连的数据。例如，观察两组学生的成绩如下

X：70，　76，　61

Y：63，　77，　80，　83

将 7 个成绩由小到大排列后有如表 12-16 所示。

表 12-16　7 个成绩由小到大排列表

分数序列	61	63	70	76	77	80	83
所属组别	X	Y	X	X	Y	Y	Y

那么，该序列中有 4 个游程：一个由来自 X 的 61 分的游程，随后是一个由来自 Y 的 63 分构成的游程，再后是来自 X 的 70 和 76 构成的游程，最后是来自 Y 的三个分数构成的游程。

记游程总数目为 R，当 $m<10, n<20$ 时，则依据 m、n 和 U 查附表 14（游程检验表），得到相应的 c_1 和 c_2 值，拒绝域为 $\{R \leqslant c_1\} \cup \{R \geqslant c_2\}$。当 $m+n=N>20$ 或 $m>12$，$n>12$ 时，则 R 的抽样分布近似正态分布，可采用单样本相似的方法计算 Z 值并作出决策。

【例 12-10】有研究人员想考察试卷中试题难度编排对学生成绩的影响。他设计了两份试卷，试题内容完全一样，但其中一份试卷试题由易到难来呈现，另一份试卷试题由难到易来呈现。并随机向两组学生施测，结果如表 12-17 所示。问试题难度编排方式对学生成绩是否产生影响。

表 12-17　两种难度编排方式下学生的测验成绩

由难到易（X）	40	50	59	62	63	73	76	76	76	79
由易到难（Y）	67	69	70	73	80	81	82	88	89	94

解：

将学生考试成绩由小到大排成序列后有如表 12-18 所示。

表 12-18　两组学生测验成绩合并后的排序表

成绩序列	40	50	59	62	63	67	69	70	73	73	76	76	76	79	80	81	82	88	89	94
所属组别	X	X	X	X	X	Y	Y	Y	Y	X	X	X	X	X	Y	Y	Y	Y	Y	Y

由表 12-18 可知，序列的游程总数目 $R=4$。根据 $m=10$，$n=10$ 查附表 14 得，$c_1=6$，$c_2=16$，因为 $R=4<c_1=6$，所以拒绝 H_0，即认为试题难度编排方式对学生成绩有显著影响。

答：试题难度编排方式对学生成绩是有影响作用。

（四）Moses 极端检验

Moses 极端反应检验假定实验变量在一个方向影响某些主体，而在相反方向影响其他主体。例如，对电信资费下调的态度调查，一般群众可能都持支持的态度，而电信行业的工作人员可能持反对的态度。对于这样的数据资料，可以采用该方法。

由于篇幅有限，该方法在此不详述，在 SPSS 中可以很方便地加以应用。

（五）两独立样本非参数检验的 SPSS 操作

（1）以例 12-8 的资料为例，在 SPSS 中进行两独立样本曼-惠特尼 U 检验

的操作过程如下：

1.分析→非参数检验→旧对话框→2 个独立样本(2)
2.“检验变量列表”框：成绩
3.“分组变量”框：组别
4.定义组：（1）组 1(1):1
（2）组 1(2):2
（3）继续
5.“检验类型”：☑ Mann-Whitney U
6.确定

结果如表 12-20：

表 12-20 例 12-8 的 SPSS 分析结果

	成绩
Mann-Whitney U	2.000
Wilcoxon W	8.000
Z	-1.807
Asymp. Sig. (2-tailed)	.071
Exact Sig. [2*(1-tailed Sig.)]	.095[a]

表中第一行的结果是例 12-8 中的两个 U 值中较小的一个；第二行是例 12-8 中 T_x 值，第三行是常用的 Z 值，第四、五行是给出检验 p 值。由此可见，没有充分理由拒绝 H_0。

（2）以例 12-9 为例，两样本的 K-S 检验的 SPSS 操作过程如下：

1.数据→加权个案：选择“加权个案”，在“频率变量”框中加入“人数” (本例中各学历人数不等，所以要加权)
2.分析→非参数检验→旧对话框→2 个独立样本(2)
3.“检验变量列表”框：学历
4.“分组变量”框：组别
5.定义组：（1）组 1(1):1
（2）组 1(2):2
（3）继续
6.“检验类型”：☑ Kolmogorov-Smirnov Z
7.确定

结果如表 12-21：

表 12-21　例 12-9 的 SPSS 分析结果

		学历
Most Extreme Differences	Absolute	.187
	Positive	.000
	Negative	-.187
Kolmogorov-Smirnov Z		2.101
Asymp. Sig. (2-tailed)		.000

表中的第二行是最大绝对差值，第二行是最大正差，第三行是最大负差，第四行是 Kolmogorov-Smirnov Z 值，最后一行是我们统计决策的依据，是检验的 p 值。由此看来，应拒绝 H_0。

（3）以例 12-10 为例，两样本 Wald-Wolfowitz 游程检验的 SPSS 操作过程如下：

1.分析→非参数检验→旧对话框→2 个独立样本(2)
2.“检验变量列表”框：成绩
3.“分组变量”框：组别
4.定义组：（1）组 1(1):1
（2）组 1(2):2
（3）继续
5.“检验类型”：☑ Wald-Wolfowitz runs
6.确定

结果如表 12-22。

表 12-22　例 12-10 的 SPSS 分析结果

		Number of Runs	Z	Exact Sig. (1-tailed)
成绩	Minimum Possible	4	-2.987	.001
	Maximum Possible	6	-2.068	.019

由于两组中有相同的成绩，所有在排序时有不同的形式，表 12-22 给出了不同排序时的游程数，最小的游程数为 4，最大的游程数为 6，两种情况下的概率（最后一列）都小于.05，所以拒绝 H_0。

二、多个独立样本的非参数检验

多个独立样本的非参数检验是用于确定多个（大于 2）独立样本是否来自同一总体的一类检验方法。此类方法主要包括克拉夏尔-瓦里斯 H 检验（简称克氏 H 检验）、中数检验和 Jonckheere-Terpstra 检验等方法。这些方法在 SPSS 中可以十分方便地进行，此处仅对其中之克氏 H 检验进行介绍。

（一）、克氏 H检验的基本含义与方法

克氏 H检验（Kruskal-Wallis H test）是两个独立样本曼-惠特尼 U检验的一种推广。

如方差分析一样，在克氏 H 检验中，独立变量称为因子，因子的各个取值称为水平或处理。例如，检验不同年级的学生在图书馆的时间是否有差异，年级的各个取值（一、二、三、四）就是水平或处理。克氏 H 检验的虚无假设为：样本容量为 n_1，n_2，…，n_k 的 k 个独立处理来自于同一总体，或者说所有抽样总体有相同分布；备择假设是至少有一个总体分布不同于其他的总体分布。

克氏 H检验步骤如下：

（1） 将来自 k 个独立处理的所有观测值合并，然后由小到大排序，并记录下各观测值对应的秩；存在结的情况，求其平均秩。

（2）计算每个处理的秩和 R_i。

（3）使用(公式 12.11)计算检验统计量 H值。

$$H=\frac{12}{N(N+1)}\sum_{i=1}^{k}(\frac{R_i^2}{n_i})-3(N+1) \qquad \text{（公式 0.11）}$$

式中，N 是样本观测总数，n_i 是第 i 个处理的样本观测值个数，R_i^2 是第 i 个处理的秩的平方和。

当得分中相等观测值的个数超过 35%时，需要采用(公式 12.12)对 H值进行修正。

$$H=\frac{\frac{12}{N(N+1)}\sum_{i=1}^{k}(\frac{R_i^2}{n_i})-3(N+1)}{1-\frac{\sum(t_i^3-t_i)}{N^3-N}} \qquad \text{（公式 0.12）}$$

式中，t_i表示对于某个秩 i 相等的观测值的个数。

（4）统计决策

对于小样本（n≤5），且只有三个处理时，H 分布的临界值可由附表 19（Kruskal-Wallis H 临界值表）得到，如果 H 值大于临界值，则拒绝 H_0。对于大样

本（n>5）或者处理数多于三个，检验统计量 H 近似服从自由度为 k-1 的 x^2 分布（k 为处理数）。如果 H 值大于 x^2 临界值，则拒绝 H_0。

（二）克氏 H 检验的实例

【例 12-11】 某公司想购买 A、B、C 三种机器之一，要考虑的重要因素就是各机器的产出效率。为此，将 12 名熟练工人分为四组，每组随机使用一种机器进行生产，表 12-23 列出了在相同时间内不同机器条件下生产出来的产品数。问三种机器生产的产品数量分布是否存在显著差异。

表 12-23　三种机器条件下生产的产品数

A	50	8	40	45
B	10	5	30	10
C	2	3	5	14

解：

$$\begin{cases} H_0: \text{三种机器生产的产品数分布相同} \\ H_a: \text{至少一个总体分布与其他总体不同} \end{cases}$$

将所有生产的产品数排序有如表 12-24 所示。

表 12-24　三种机器条件下所有生产产品数的排序表

产品数排序	2	3	5	5	8	10	10	14	30	40	45	50
秩	1	2	3.5	3.5	5	6.5	6.5	8	9	10	11	12
所属组别	C	C	C	B	A	B	B	C	B	A	A	A

$$H=\frac{12}{N(N+1)}\sum_{i=1}^{k}(\frac{R_i^2}{n_i})-3(N+1)=\frac{12}{12(12+1)}(\frac{38^2}{4}+\frac{25.5^2}{4}+\frac{14.5^2}{4})-3(12+1)=5.3173$$

查附表 19 得，α＝.05 的临界值为 5.692，因为 H＝5.3173<5.692，所以无充分理由拒绝 H_0。

答：没有充分理由表明三种机器生产的产品数量分布存在显著差异。

【例 12-12】为了解年级对大学生在图书馆的时间是否有影响，某研究人员从每个年级（大一至大四年级）随机抽取 7 名学生，并记录下每个学生一周内在图馆的时间，结果如表 12-25 所示。问四个年级的学生在图书馆的时间分布是否相同。

表 12-25　四个年级大学生一周内在图书馆的时间情况表

一年级	11.5	15.0	12.0	6.0	7.0	9.0	10.5
二年级	11.0	19.0	18.0	14.0	8.5	6.5	5.0
三年级	12.5	8.5	4.0	13.0	19.5	20.0	10.0
四年级	16.0	10.0	7.5	8.0	17.0	4.5	21.0

解：

$$\begin{cases} H_0: \text{四个年级学生在图书馆的时间分布相同} \\ H_a: \text{至少一个总体分布与其他总体不同} \end{cases}$$

将所有生产的产品数排序（略），可发现相等的观测值只有 4 个（占 14.3%），故不必采用修正公式。

$$H = \frac{12}{28(28+1)}\left(\frac{89.0^2}{7}+\frac{101.5^2}{7}+\frac{113.0^2}{7}+\frac{102.5^2}{7}\right)-3(28+1)=0.6112$$

查 x^2 分布表得，$\chi^2_{.05,df=4-1}=7.81$。因为 H=0.6112<7.81，所以无充分理由拒绝 H_0。

答：没有充分理由认为四个年级的学生在图书馆的时间分布不相同。

（三）多个独立样本非参数检验的 SPSS 操作

（1）以例 12-11 的资料为例，在 SPSS 中进行克氏 H 检验的操作过程如下：

1.分析→非参数检验→旧对话框→*K* 个独立样本
2.“检验变量列表”框：产品数
3.“分组变量”框：机器
4. 定义范围：（1）最小:1
（2）最大:3
（3）继续
5.“检验类型”：☑ Kruskal-Wallis H
6.确定

结果如表 12-26。

表 12-26　例 12-11 的 SPSS 分析结果

	产品数
卡方	5.355
df	2
渐近显著性	.069

从表 12-26 中可见，SPSS 进行克氏 H 检验，是采用 x^2 分布进行统计决策的。从表中最末一行可见，p 值大于.05，所以无充分理由拒绝 H_0。

2. 以例 12-12 的资料为例，在 SPSS 中进行克氏 H 检验的操作过程如下：

1.分析→非参数检验→旧对话框→*K* 个独立样本
2.“检验变量列表”框：在图书馆的时间
3.“分组变量”框：年级
4. 定义范围：（1）最小: 1
（2）最大: 4
（3）继续
5.“检验类型”：☑ Kruskal-Wallis H
6.确定

结果如表 12-27，其含义与表 12-25 相同，不再赘述。

表 12-27　例 12-12 的 SPSS 分析结果

	在图书馆的时间
卡方	.612
df	3
渐近显著性	.894

第四节　相关样本的非参数检验

一、两相关样本的非参数检验

（一）概述

两相关样本（或称配对样本）的非参数检验方法主要有符号检验、维尔克松符号秩检验、McNemar 检验和边际齐性检验。

1. 符号检验

符号检验（sign test）是指通过比较两个相关样本每对数据之差的正负符号数目之间差异是否显著，进而比较两个样本数据之间是否有显著差异的非参数检验方法。其原理是，如果两个相关样本没有差别，则样本数据相关所得的差
值应当大致有一半为正、一半为负，数量基本平衡。用 S^+ 表示差值为正的个数，S^- 表示差值为负的个数，当 S^+、S^- 过大或过小，则说明两样本之间存在显著差别。该方法简便易行，但检验效能较低精度较差。

2.维尔克松符号秩检验

维尔克松符号秩检验（Wilcoxon signed-rank test）是对符号检验的改进，既考虑了样本差数的符号，又考虑到了差数的顺序。不同的符号代表了在中心位置的哪一边，而差的绝对值代表了距离中心的远近，二者结合会更有效。该检验方法使用较广，下文会着重予以介绍。

3.McNemar 检验

McNemar 检验(McNemartest)仅适用于二分类变量，它考察的重点是两组间分类的差异，对相同的分类则忽略不计。该检验特别适合于重复测量情况，用于分析每个主体在事件前后反应的变化情况。

4.边际齐性检验

边际齐性检验（Marginal Homogeneity test）是 McNemar 检验向多分类情形下的扩展，适用于资料为有序
分类的情况。与 McNemar 检验相似，在前后对比设计中检测因实验干预所导致的因变量变化很有用。与 McNemar 检验不同的是，在 McNemar 检验中只采用二分法，例如成绩分为“及格”与“不及格”，而边际齐性检验中的因变量是多个有序分类，例如成绩分为“优”、“良”、“及格”和“不及格”等。

（二）相关样本的 Wilcoxon 符号秩检验

1.基本含义与方法

两个相关样本的秩次检验是研究成对抽取的观测值之间的差，该检验与单样本的检验十分相似，不同之处在于：这里考虑的差是两个样本中数的差，而不是样本中数与假设的中数的差。

维尔克松符号秩检验步骤如下：

（1） 随机抽取 n 对观测值，计算出每对观测值的差，将这些差值按绝对值从小到大排列，并记录各自对应的秩。若有相同差值，则取其平均秩。差为正，则在秩的前面加上“+”号；差为负，则在秩的前面加上“-”号；若差为 0，则在计算过程中删除此对观测值。

（2）分别计算正的秩和（W_+）和负的秩和（W_-）。

（3）根据(公式 12.13)计算检验统计量 W。

$$W = \min\{|W_+|, |W_-|\} \qquad \text{（公式 0.13）}$$

（4）统计决策。

对于小样本（$n<30$），附表 20（Wilcoxon W 的临界值表）给出了各种显著性水平和样本容量 n 下的的 W 临界值，如果检验统计量 W 小于临界值，则拒绝 H_0；对于大样本（$n\geqslant30$），利用(公式 12.14)将 W 值转换为 Z 值，然后根据正态分布的规律作出判断。

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$ （公式 0.14）

2.实例

【例 12-13】一研究者想知道某项新型锻炼方式对 75 至 80 岁之间的妇女的脉搏速率是否有影响，随机抽取该年龄段的妇女 12 人并测量她们的脉搏速率，随后请她们按照新型锻炼方式进行 2 个月的锻炼，在锻炼结束以后再次测量她们的脉搏速率，结果如表 12-27 所示。问锻炼前后的脉搏速率是否存在显著差异？

表 12-27　12 名妇女锻炼前后的脉搏速率

被试	1	2	3	4	5	6	7	8	9	10	11	12
锻炼前	75	81	74	75	70	74	82	64	79	83	73	82
锻炼后	71	83	70	60	75	67	85	65	69	71	65	76

解：

$$\begin{cases} H_0: \text{锻炼前后的中位数没有差异} \\ H\ : \text{锻炼前后的中位数存在差异} \end{cases}$$

中数的差、差的绝对值、秩和符号秩的计算见表 12-28。

表 12-28 中数的差、差的绝对值、秩和符号秩的计算表

被试	锻炼前	锻炼后	锻炼前减去锻炼后	差的绝对值	秩	符号秩
1	75	71	+4	4	4.5	+4.5
2	81	83	−2	2	2.0	−2.0
3	74	70	+4	4	4.5	+4.5
4	75	60	+15	15	12.0	+12.0
5	70	75	−5	5	6.0	−6.0
6	74	67	+7	7	8.0	+8.0
7	82	85	−3	3	3.0	−3.0
8	64	65	−1	1	1.0	−1.0
9	79	69	+10	10	10.0	+10.0
10	83	71	+12	12	11.0	+11.0
11	73	65	+8	8	9.0	+9.0
12	82	76	+6	6	7.0	+7.0

$$|W_+|=66.0,\quad |W_-|=12.0$$

$$W=\min\{|W_+|,|W_-|\}=12.0$$

根据 n=12，$\alpha=.05$ 查附表 20 得检验统计量 W 的临界值为 14，因为 W=12<14，所以拒绝 H_0，说明锻炼前后脉搏速率的中数存在显著差异。

答：锻炼前后的脉搏速率存在显著差异。

（三）相关样本的 Wilcoxon 符号秩检验的 SPSS 操作

以例 12-13 的资料为例，在 SPSS 中进行 Wilcoxon 符号秩检验的操作过程为：

1.分析→非参数检验→旧对话框→2 个相关样本

2.“检验对”框：(1)Variable1:锻炼前

(2)Variable2:锻炼后

3.“检验类型”：☑ Wilcoxon

4.$\boxed{\text{确定}}$

分析结果主要包括两个表，见表 12-29 和表 12-30。

表 12-29　Wilcoxon 符号秩检验秩和汇总表

		N	平均等级	等级总和
锻炼后 - 锻炼前	负等级	8	8.25	66.00
	正等级	4	3.00	12.00
	结	0		
	总计	12		

表 12-30　例 12-13 的 SPSS 分析结果

	锻炼后 - 锻炼前
Z	-2.119^a
渐进显著性（双尾）	.034

表 12-30 汇总了符号秩的和；表 12-31 给出了假设检验的 p 值（表中最后一行），由此可见，应该拒绝 H_0。

二、多个相关样本的非参数检验

多个相关样本的非参数检验用于确定多个（大于 2）相关样本是否来自同一总体的一类检验方法，主要包括弗里德曼（Friedman）检验、肯德尔（Ken-Dall） W 检验和 Cochrans Q 检验等方法。这些方法在 SPSS 中都可以方便地进行，

此处仅对其中之弗里德曼检验进行介绍。

（一）弗里德曼检验的基本含义与方法

弗里德曼检验(Friedman test)也称为弗里德曼双向评秩方差分析(Friedman’s Two-way analy-

sis of Variance by Ranks 或 Friedman’s Two-way analysis with Ranks)。该方法的基本思想是：由于区组间的差异是各式各样的，只有同区组的处理值的比较才有意义，一个观察值的秩是在某一区组中的秩，而不是对所有数据而言的。因此应当独立地在每个区组内对数据进行排序，这样就可以消除区组间的差异以检验各种处理之间是否存在差异。该检验的假设是：

$$\begin{cases} H_0: \mu_{R_1} = \mu_{R_2} = \cdots = \mu_{R_n} \\ H_a: \text{至少一对}\mu_{R_j}\text{不等，}(j = 1, 2, \cdots, n) \end{cases}$$

这里，μ_{R_j}（j=1, 2, …, n)是第 j 个处理的平均效果，即平均秩。为对假设进行检验，所需要的数据至少是顺序变量，要能将数据排成 k 行 n 列的双向表，如表12-31。

表 12-31　Friedman 检验的数据模式

	处理						行的秩和
区组	R_{11}	R_{12}	⋯	R_{1j}	⋯	R_{1n}	$n(n+1)/2$
	R_{21}	R_{22}	⋯	R_{2j}	⋯	R_{2n}	$n(n+1)/2$
	⋮	⋮		⋮		⋮	⋮
	R_{i1}	R_{i2}	⋯	R_{ij}	⋯	R_{in}	$n(n+1)/2$
	⋮	⋮		⋮		⋮	⋮
	R_{k1}	R_{k2}	⋯	R_{kj}	⋯	R_{kn}	$n(n+1)/2$
列的秩和	R_1	R_2	⋯	R_j	⋯	R_n	$k\ n(n+1)/2$

表 12-26 中，$R_j = \sum_{i=1}^{k} R_{ij}$ 。对于 H0 是否成立的判定，各个 R_j(j=1, 2, ⋯,

n)与平均秩和 $k\ n(n+1)/2$ 之差的大小来作出定义：

$$S = \sum_{j=1}^{n} [R_j - \frac{k(n+1)}{2}]^2 \qquad \text{（公式 0.15）}$$

那么，检验统计量 Q 就可通过计算而得到。

$$Q = \frac{12S}{kn(n+1)} = 12\sum_{j=1}^{n} R_j^2 / kn(n+1) - 3k(n+1) \qquad \text{（公式 0.16）}$$

统计量 Q 近似为自由度 df=n−1 的 x^2 分布，通过查 x^2 分布得到临界值。如果

Q大于等于临界值，则拒绝 H_0。

（二）弗里德曼检验的实例

例 12-14　四种药物可以治疗同一疾病，为检验其效果是否有显著差异，选取了 32 名患该种病的病人接受治疗。每 4 人一组，共 8 组。同一组的 4 个病人符合配对条件，随机地指定某人使用药物 A、B、C、D。经过一个周期的治疗后，根据病情好转情况评分，结果如表 12-32。问四种药物的疗效是否存在显著差异。

表 12-32　病情好转评分结果表

病人组	药物 A	药物 B	药物 C	药物 D
1	14	23	26	30
2	19	25	25	33
3	17	22	29	28
4	17	21	28	27
5	16	24	28	32
6	15	26	27	26
7	18	26	27	36
8	16	22	30	32

解：

$$\begin{cases} H_0: \mu_{R_1} = \mu_{R_2} = \cdots = \mu_{R_n} \\ H_a: \text{至少一对}\mu_{R_j}\text{不等，}(j = 1, 2, \cdots, n) \end{cases}$$

将病情好转得分转换为秩得分到表 12-33。

表 12-33　病情好**转**得分的秩

病人组	药物 A	药物 B	药物 C	药物 D
1	1	2	3	4
2	1	2.5	2.5	4
3	1	2	4	3
4	1	2	4	3
5	1	2	3	4
6	1	2.5	4	2.5
7	1	2	3	4
8	1	2	3	4
合计（R_j）	8	17	26.5	28.5

$$Q = 12(8^2 + 17^2 + 26.5^2 + 28.5^2)/4(4+1) - 3\times 8\times(4+1) = 20.0625$$

查 x^2 分布表得，$\chi^2_{.05,df=3} = 7.82$，因为 $Q = 20.0625 > \chi^2_{.05,df=3} = 7.82$，所以拒绝 H_0，说明不同药物的疗效存在显著差异。

（三）弗里德曼检验的 SPSS 操作

以例 12-14 的资料为例，在 SPSS 进行 Friedman 检验的操作过程如下：

1.分析→非参数检验→旧对话框→*K* 个相关样本

2.“检验变量”框：药物 A、药物 B、药物 C、药物 D

3.检验类型：☑ Friedman

4.确定

SPSS 分析的结果十分简洁地呈现在两个表中，见表 12-35 和表 12-36。

表 12-35　例 12-14 的 SPSS 分析结果 1

	平均等级
药物 A	1.00
药物 B	2.13
药物 C	3.31
药物 D	3.56

表 12-36　例 12-14 的 SPSS 分析结果 1

N	8
卡方	20.577
df	3
渐进显著性	.000

表 12-34 呈现了四个处理的平均秩；表 12-35 呈现了统计检验的 x^2 值和 p 值，从表中可见，p 值<.05，所以拒绝 H_0。

小　结

非参数检验是一类适用于称名变量、顺序变量，或者虽然是等距变量和比率变量，但进行参数检验的前提不能满足条件，或者是样本容量较小时的一类统计检验方法。它们扩展了统计分析研究对象问题的范围。

本章对不同样本情况下（单样本与多样本；独立样本与相关样本）的非参数检验方法进行了介绍。这些条件下的具体方法众多，本章仅对统计软件 SPSS 中提

供的非参数检验方法中，部分常用的方法进行了介绍，以利于读者对照使用 SPSS 软件。

单样本非参数检验主要介绍了二项式检验、游程检验和 K-S 检验。独立样本包括两独立样本和多个独立样本两种情形。两独立样本非参数检验主要介绍了曼-惠特尼 U 检验、K-S 检验和沃尔德-沃尔福威茨游程检验，多个独立样本非参数检验主要介绍了克氏 H 检验。相关样本包括两相关样本和多个样本两种情形。两相关样本非参数检验主要介绍了氏维尔克松符号检验，多个相关样本主要介绍了弗里德曼检验。

这些非参数检验方法的具体运用时，多数方法根据样本容量的大小给出了不同的办法，本章对于小样本情形给出了临界值表，对于大样本情形则给出了正态近似的公式。

关键术语

参数统计：通常是指那些假设总体分布形式是已知的情况下根据样本资料来进行参数估计或假设检验的统计分析方法。

参数检验：对总体分布服从正态分布或总体分布已知条件下对诸如总体平均数、总体方差、总体标准差、总体相关系数等总体参数所进行的假设检验过程。

非参数检验：是指能够适用于称名变量或顺序变量，或总体分布未知条件下的等距变量或比率变量的统计检验方法。

非参数统计：在总体分布未知或知之甚少的条件下根据样本资料下对总体性质和特征进行推论的统计分析的方法。

秩：根据一组数据的值从小到大（或从大到小）排成一列后，数据对应的排名等级。

秩统计量：对于样本 x_1，x_2，…，x_n，假定它们的值各不相同，将 x_1，x_2，…，x_n 由小到大（或由大到小）排成一列。若 x_i 在这一列中占据第 R_i 位，则称 x_i 的**秩**为 R_i（i=1,2,…,n），称（R_1，R_2，…，$_i$）是原样本 x_1，x_2，…，x_n 的秩统计量。

结：在许多时候，样本 x_1，x_2，…，x_n 会存在相同的值，此时排序的结果就会出现同秩的现象，这种情况被称为数据中的结。

二项式检验：是对二分变量的两个类别的观察频率与指定概率参数的二项分布下的期望频率是否存在显著差异的一种非参数检验方法。

游程检验：是检验某一变量的两个值的出现顺序是否是随机的一种非参数检验方法，也称为连贯检验或串检验。

游程：指的是当样本按某种顺序排列（如抽样时间的先后）时，一串不间断的 0（或 1）所组成的序列。

柯尔莫哥诺夫-斯米尔诺夫检验（K-S 检验）：是一种分布拟合检验的方法。单样本时，K-S 检验用于检验观察值是否合理来自指定的理论分布。两独立样本时，K-S 检验用于检验两独立样本是否来自相同分布的同一总体。

曼-惠特尼 U 检验：是最常用的两个独立样本非参数检验方法，用于检验两个总体分布各自的中心位置是否相同。

沃尔德-沃尔福威茨游程检验：是考察两独立样本资料是否来自相同分布的总体的非参数检验方法。

克氏 H 检验：是两个独立样本曼-惠特尼 U 检验的一种推广，是一种用于确定三个或三个以上总体间差异的非参数检验方法。

符号检验：是指通过比较两个相关样本每对数据之差的正负符号数目之间差异是否显著，进而比较两个样本数据之间是否有显著差异的非参数检验方法。

维尔克松符号秩检验：是以对两相关样本的分析为基础来判断两个总体间的差异的非参数检验方法。

弗里德曼**检验**：是以三个或三个以上相关样本为基础，用于检验三个或三个以上处理间差异的非参数统计方法。

重要公式

二项式检验 $Z_{+,R}$，$Z_{-,R}$ 统计量

$$Z_{+,R}=\frac{n_p-0.5-n\pi_0}{\sqrt{n\pi_0(1-\pi_0)}}\quad,\quad Z_{-,R}=\frac{n_q-0.5-n(1-\pi_0)}{\sqrt{n\pi_0(1-\pi_0)}}$$

游程检验的 Z 统计量：$Z=\dfrac{(R+0.5)-(1+2mn/N)}{\sqrt{2mn(2mn-N)/N^2(N-1)}}$

单样本 K-S 检验的检验统计量：

$$D=\max|S_n(x)-F_0(x)|$$

曼-惠特尼 U 检验中两样本各数据对应的秩和：

$$T_x=\sum_{i=1}^{m}R_i\ ,\ \ T_y=\sum_{j=1}^{n}R_i$$

曼-惠特尼 U 检验的检验统计量：

$$U_{xy}=mn+m(m+1)/2-\sum_{i=1}^{m}R_i=T_y-n(n+1)/2$$

$$U_{yx}=mn+n(n+1)/2-\sum_{j=1}^{n}R_j=T_x-m(m+1)/2$$

$$Z=\frac{U-mn/2}{\sqrt{\dfrac{mn(m+n+1)}{12}}}$$

两独立样本 K-S 检验的检验统计量：

$$D = \max|S_1(x) - S_2(x)|$$

克氏 H 检验的检验统计量：

$$H = \frac{12}{N(N+1)}\sum_{i=1}^{k}(\frac{R_i^2}{n_i}) - 3(N+1)$$

两相关样本的维尔克松符号秩检验检验统计量：

$$W = \min\{|W_+|,|W_-|\}$$

$$Z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

弗里德曼检验的检验统计量：

$$S = \sum_{j=1}^{n}[R_j - \frac{k(n+1)}{2}]^2$$

$$Q = \frac{12S}{kn(n+1)} = 12\sum_{j=1}^{n}R_j^2 / kn(n+1) - 3k(n+1)$$

思考与练习

1. 下表是评价一临床治疗效果的研究数据。对每个病人都要进行治疗前后的症状评估，并记录治疗后症状是加重了还是减轻了。该研究数据是否表明临床治疗有效果。

病人	A	B	C	D	E	F	G	H
治疗效果（-表示症状减轻，+表示症状加重）	－	－	－	+	－	－	－	－

2. 某大型考试的评卷工作中，全体评卷员对一主观题评分的中数是 15 分。有人声称 A 评卷员评卷偏严，为此随机抽取了 15 份 A 评卷员所评试卷，15 份试卷得分如下：

6　6　7　8　8　9　10　12　13　14　16　16　17　18　19

问该数据是否表明 A 评卷员评卷偏严。

3. 某商场每晚 18:30 打烊。有人建议延长营业时间至 22:00。为此，商场对老顾客进行了电话调查，随机抽取了 20 个家庭，有 9 个家庭表示如果延长营业时间将可能去购买。商场管理人员认为 30%的老顾客支持延长营业时间就实行，现在的调查结果是否支持商场管理层作出延长营业时间的决定。

4. 某地区记录了 16 年来的“年度最高水位“（单位：英尺）：7.4，7.8，6.9，8.1，8.0，7.1，7.4，6.8，6.9，7.4，7.7，8.0，8.3，7.6，7.8，7.1。问数据表明“年度最高水位”的变化是随机的，还是表现出越来越低或是越来越高的趋势。（提示：可将小于中数的记为 0，大于中数的记为 1）

5. 某班男女生人数比例相当，现在为了调查男生和女生对学校实行某项政策的态度，从该班抽取了 25 名学生，抽取的学生性别按抽取的顺序依次为（0 表示女生，1 表示男生）：

0 0 1 0 1 0 1 1 1 0 0 0 0 1 1 0 0 0 1 1 1 0 1 1 0

问该数据是否是随机抽取的。

6. 通过对上一年数千辆汽车排放的氮氧化物的测量，发现它大约服从平均数为 5.6，标准差为 1.2 的正态分布。今年测量的 12 辆汽车的氮氧化物排放量为：

4.8　6.2　6.0　5.9　6.6　5.5　5.8　5.9　6.3　6.6　6.2　5.0

问今年汽车的氮氧排放量的分布模型与去年的是否相同。

7. 有人想研究在 1 小时内，不同时间段（每 10 分钟为一时间段）的解题效率是否会发生变化。为此精心设计了一批难度相当的试题，让学生回答并记录下每 10 分钟内完成的题数，所有学生在不同时间段平均完成的试题数见下表。问不同时间段学生解题效率是否发生变化。

时间段	1	2	3	4	5	6
平均解题数	14	11	10	7	7	11

8. 有人发现较小年龄的孩子组织故事时倾向于采用简单的描述模式（例如，“然后……”，等），而更年长一些的孩子会加上因果关系的陈述和自己的推论等内容。假设我们两组不同年龄的孩子，要求他们对刚阅读完的故事做一概述，记录概述内容中可归为推论内容的句子个数。结果如下表。该数据是否支持上述发现？

较小的孩子	0	1	0	3	2	5	2
较大的孩子	4	7	6	4	5	7	

9. 有人对来自农村和城市的大学生的家庭月收入情况进行了调查，结果如下表。问来自不同地方的大学生月家庭收入分布是否存在显著差异？

家庭月收入分组	生源地 城市	生源地 农村
500 元以下	144	235
500-1000 元	643	753
1000-3000 元	1266	483
3000-5000 元	348	30
5000-8000 元	66	17
8000 元以上	62	5

10. 厂家研究出一种新型牙箍，该牙箍更舒适美观，但如果使用寿命明显不如老的牙箍则不予投产。为此随机抽取了 20 名儿童，随机分为两组，分别戴新、老牙箍，记录牙箍使用寿命如下表。问新牙箍使用寿命是否不如老牙箍长？

牙箍使用寿命	老牙箍	14	16	18	20	15	13	10	12	19	24
（单位：月）	新牙箍	13	15	20	18	17	19	20	22	24	21

11. 有位教师提出了三种个别辅导的教学方案，为了比较教学方案的效果，他随机抽取了 15 名学生，并随机分为 3 组，每组 5 名学生，每组学生随机接受一种教学方案的训练。一段时间后，这 15 名学生期末考试的该科成绩如下表。假定其他因素得到了较好的控制，问该数据是否表明三种教学方案的效果存在显著差异。

方案 A	68	72	77	42	53
方案 B	72	53	63	53	48
方案 C	66	82	74	75	72

12. 三位竞争教授的英语教师都声称自己所教的学生成绩最好，为了澄清问题，从他们任教的班中各自随机抽取 8 名学生，让抽取的学生授受相同的测验，并让另一位中立的教授（他不知道学生来自哪个班）评分。结果如下表。问三位老师所教学生的成绩是否存在显著差异？

教师 A	82	81	56	58	63	64	62	53
教师 B	55	88	85	83	71	70	68	72
教师 C	65	54	66	68	72	78	65	73

13. 有人提出进行诊断技能训练，将使临床医生产生并检验更多的假设来确诊病患的情况。为此，研究者随机选择了 10 名刚开始担任实习医生的精神病治疗医生，让他们观看病患门诊的录像，并要求他们将自己的治疗意见和观看录像时的想法写下来。研究者对每位实习医生的笔记进行分析，并记录每位实习医生提出的假设个数。在实习期间，对这 10 名实习医生进行诊断技能训练。在实习期结束，采用了可比的录像重复上述研究。所得数据如下表。问该数据是否支持研究者的假设？

实习医生	A	B	C	D	E	F	G	H	I	J
实习期开始时	8	4	2	2	4	8	3	1	3	9
实习期结束时	7	9	3	6	3	10	6	7	8	7

14. 为了考察视觉教具的使用对教学效果的影响，随机抽取了 17 名教师，让他们在三种条件下对同一内容进行讲解。这三种条件是：没有视觉教具、采用较少视觉教具用于说明主要观点、采用视觉教具和挂图对每个观点进行说明。然后让学生对教学效果作出评定，并将学生评定的平均分作为教学效果的得分。假定其他条件得到了较好的控制，实验的最后结果如下表。请在 $\alpha=.05$ 的水平下检验不同条件下的教学效果是否存在显著差异。

	教学条件		
教师	无视觉教具	较少视觉教具	较多视觉教具
1	50	58	54
2	32	37	25
3	60	70	63
4	58	60	55
5	41	66	59
6	36	40	28
7	26	25	20
8	49	60	50
9	72	73	75
10	49	54	42
11	52	57	47
12	36	42	29
13	37	34	31
14	58	50	56
15	39	48	44
16	25	29	18
17	51	63	68

第十三章　线性回归

学习本章内容，将有助于你对以下问题的理解和思考：

1. 回归分析与相关分析有何区别与联系？
2. 回归分析的主要内容是什么？
3. 如何建立回归方程以及如何对回归方程的有效性进行检验？
4. 回归方程的有效性检验与回归方程有效程度高低有何区别与联系？
5. 如何在实际工作中应用回归分析？

客观事物间普遍存在联系。相关是描述事物数量间的相互关系，能反应出事物间数量变化的方向及密切程度。但一个事物数量变化了，另一个事物数量如何变化？变化的范围有多大？以及如何通过一个事物的数量来预测或控制另一个事物的数量？在回答这些问题时，相关分析就无能为力了。本章将对这些问题做出回答。

第一节 回归分析的基本概念

一、回归分析的概念

现实生活中，人们经常会根据当前已知的一些变量（称之为“自变量”），来预测或控制其它变量（称之为“因变量”）的变化发展趋势，以达到认识世界并改造世界的目的。比如，根据当前的气象指标来预测未来气象发展及变化、根据学生平时考试成绩来预测高考成绩等。

心理学研究中经常要考察几个变量之间的相互关系。有些变量会因为自身的变化而引起另外一个或一些变量也同时发生变化，这样的变量被称为**自变量** (independent variable) ，通常用英文字母 A，B，C…等来加以表示。有些变量则会在自变量的影响下，随着自变量的变化而发生变化，这样的变量被称为**因变量** (dependent variable) ，或反应变量，常用英文字母 Y 表示。因变量通常是实验研究的结果或观察指标。

第三章中我们用相关系数来描写两个变量间的相互关系，它可以对变量间是正相关还是负相关作出定性描述，也能对两变量的相关密切程度作出定量描述。但对于自变量变化了因变量如何变化的问题，以及需要通过自变量的值去预测、估计因变量的值的问题，相关系数就显得无能为力了。这项任务只能由回归分析来完成。当然，要想真正通过自变量来预测或控制因变量，那么自变量与因变量间必须存在一定的相关。因此回归分析是相关分析的延伸和推广。

回归分析是统计分析中一种有效的工具。回归（regression）现象最早由英国生物学家高尔顿和皮尔逊在遗传学研究中发现。他们发现，高个子父母的子代平均身高比较高，矮个子父母的子代平均身高比较矮。进一步研究发现，高个子子代的平均身高要比父代的平均身高矮，而矮个子子代的平均身高要比父代的平均身高高，形成向种族平均身高靠拢的趋势，高尔顿将这种现象称为“回归”。研究发现，这种回归现象普遍存在于相关关系之中，故“回归”一词沿用至今，其含义也更为丰富了。回归分析（regression analysis, 简称回归）是研究自变量和因变量之间的共变关系，研究在自变量已经发生变化的情况下因变量会如何变化、变化的范围有多大的一整套统计分析方法。通过回归分析而建立的描述因变量的均值与自变量之间回归关系的函数或方程称为回归方程（regression equation）。

二、回归分析的基本类型及其主要内容

根据自变量的个数，可将回归分析分为一元回归分析和多元回归分析。一元回归分析（simple regression analysis）是指只有一个自变量的回归分析，多元回归分析（multiple regression analysis）是指自变量超过一个的回归分析。还可以根据自变量和因变量间的函数关系，将回归分析分为线性回归和非线性（或曲线）回归。线

性回归（linear regression analysis）是指自变量与因变量间存在线性函数关系的回归，非线性（或曲线）回归（curvilinear regression analysis）是指
自变量与因变量之间存在非线性函数关系的回归。将这两种分类方法结合起来，回归分析可分为一元线性回归、一元曲线回归、多元线性回归和多元曲线回归，等等。在实际数据分析时，采用何种回归分析不是分析者主观确定的，而是必须依据客观事实来确定。分析者选用的函数形式和变量个数如果不符合客观事实，那么所作的回归分析必定达不到预期效果。

回归分析主要有三大内容：建立回归方程，对回归方程的有效性进行检验，根据回归方程进行预测和解释。

第二节 一元线性回归方程的建立

一、一元线性回归的数学模型

一元线性回归（simple linear regression analysis）是指自变量只有一个，且自变量与因变量之间存在线性函数关系的回归分析方法。其数学模型为

$$y = \alpha + \beta x + \varepsilon \qquad \text{（公式 13-1）}$$

该模型表明，因变量 y 受已知的自变量 x 的影响，同时还受一些未知因素或随机因素 ε 的影响。对于自变量 x 的影响，我们可以通过函数 $\alpha + \beta x$ 来表示，而对于未知因素或随机因素的影响则用 ε 来表示。

由于随机因素 ε 的不确定性，上述模型也是不确定的。但在回归分析中，一般假设 ε 服从均数为 0 的正态分布，因此可通过计算因变量 y 的平均数（记为 $\tilde{y}$）来剔除随机因素 ε。

$$\tilde{y} = \alpha + \beta x \qquad \text{（公式 13-2）}$$

$\tilde{y}$ 又称为回归的主值，即在自变量给定的情况下，因变量的平均值。因此在回归分析中，主要是根据自变量的值，来预测因变量的平均值。例如体重对身高的回归，主要是预测或解释特定身高（如 $x = 1.7$ 米）的人的平均体重（$\tilde{y}$）。

上述模型建立在总体数据的情况下，但在实际中往往只能得到样本数据，在样本数据下建立的一元线性回归模型记为

$$\hat{y} = a + bx \qquad \text{（公式 13-3）}$$

其中，α 为方程在 y 轴上截距，b 为回归系数（regression coefficient）即斜率。

二、一元线性回归方程的建立

要建立一元线性回归方程，实际上只要求出回归方程系数 a 和 b 即可。

（一）回归分析的最小二乘法原理

一元线性回归的基本思想是通过一条直线来反应或拟合因变量与自变量之间的关系。但因变量与自变量之间的关系到底如何，是否真的存在线性关系，我们可以通过实际调查，采用散点图的方法来直观了解。图 13-1 是 Wagnger 等人(1988)调查的 107 名大学一年级学生的心理压力水平与心理健康水平之间关系的散点图

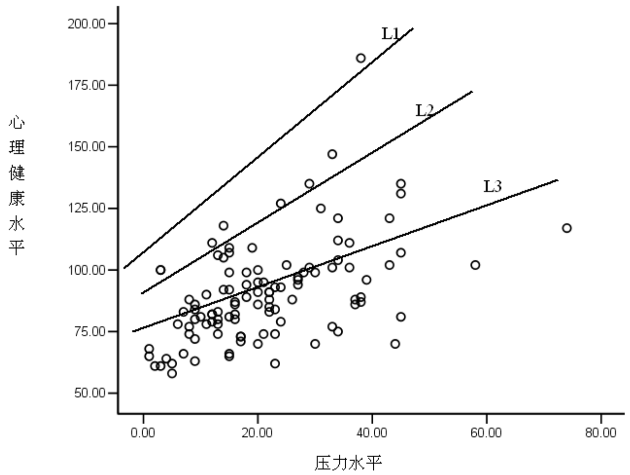


图 13-1　心理压力与心理健康水平之间关系的散点图

从图中我们可以看出，心理压力水平与心理健康水平基本呈线性相关，这时的回归分析可以通过建立一个线性数学函数（即一元线性方程）来反映心理压力水平与心理健康水平间的关系，即用一条直线来拟合它们间的关系。根据图 13-1 可以看出，直线 L_2 较 L_1 更能拟合心理压力水平与心理健康水平间的关系，同样 L_3 较 L_2 更能拟合这种关系。但 L_3 是所有可能的直线中拟合最佳的吗？拟合最佳是指方程预测的因变量值（ $\hat{y}$ ）与实际调查因变量值（ y ）间差异最小，统计学中常用指标 $Q = \sum_{i=1}(\hat{y}_i - y_i)^2$ 来反应这种差异，也就是 Q 最小的直线即为拟合最佳直线（参见图 13-2）。

建立回归方程其实就是求解最佳拟合直线的系数 a 和 b 。统计学上，把 Q 最小的条件下估计方程系数 a 和 b 的方法称为最小二乘法（method of least squares)。即在

$$Q = \sum_{i=1}(\hat{y}_i - y_i)^2 = \sum_{i=1}(a + bx_i - y_i)^2 \qquad \text{（公式 13-4）}$$

最小的情况下求解系数 a 和 b 。

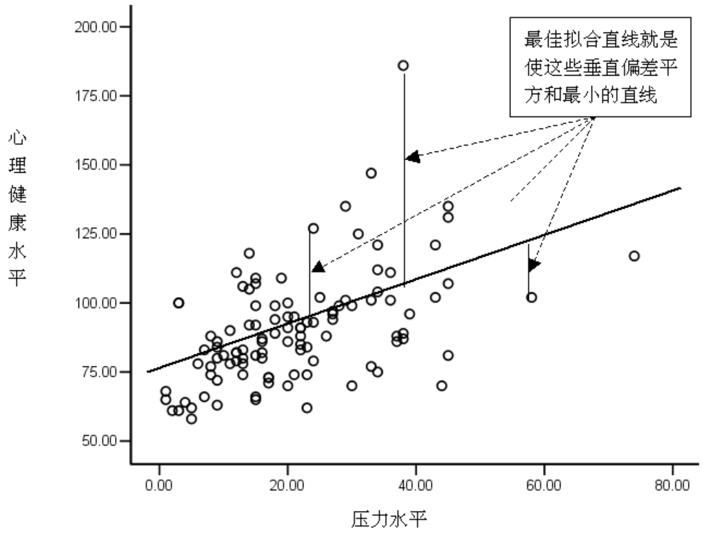


图 13-2　心理压力水平与心理健康水平的散点图

（二）一元线性回归方程的建立步骤

根据最小二乘法原理，建立一元线性回归方程（即求解 a 和 b ）一般有以下几个步骤：

（1）　由自变量 x，求 $\overline{x}$ 和 l_{xx}.

$$l_{xx} = \sum (x - \overline{x})^2 = N \cdot \sigma_x^2 \quad (13\text{-}5)$$

（2）　由因变量 y，求 $\overline{y}$ 和 l_{yy}.

$$l_{yy} = \sum (y - \overline{y})^2 = N \cdot \sigma_y^2 \quad (13\text{-}6)$$

（3）　由自变量 x 与因变量 y，求 l_{xy}.

$$l_{xy} = \sum (x - \overline{x})(y - \overline{y}) = \sum xy - N \cdot \overline{x} \cdot \overline{y} \quad (13\text{-}7)$$

（4）　求系数 α 和 b：

$$b = l_{xy} / l_{xx} \quad (\text{公式 } 13\text{-}8)$$

$$a = \overline{y} - b\overline{x} \quad (\text{公式 } 13\text{-}9)$$

（5）　写出方程：

$$\hat{y} = a + bx$$

【例 13-1】，表 13-1 是 15 名大学一年级学生的心理压力水平 (x) 与心理健康水平 (y) 得分数据，试建立心理健康水平对心理压力水平的一元线性回归方程。

表 13-1　15 名学生的心理压力水平与心理健康水平

序号	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
(x)	20	8	15	14	13	12	18	29	37	13	23	11	8	4	13
(y)	100	88	66	92	74	79	99	735	86	83	84	90	77	64	78

注：表中心理压力得分越高说明心理压力越大；
心理健康得分越高说明心理越倾向于不健康

解：

（1）由自变量 x，求 $\bar{x}$ 和 l_{xx}

$$\bar{x} = \frac{\sum x}{n} = 15.867$$

$$l_{xx} = N \cdot \sigma_x^{\ 2} = 15 \times 68.249 = 1023.736$$

（2）由因变量 y，求 $\bar{y}$ 和 l_{yy}

$$\bar{y} = \frac{\sum y}{n} = 86.333$$

$$l_{yy} = N \cdot \sigma_y^{\ 2} = 15 \times 270.356 = 4055.34$$

（3）由自变量 x 与因变量 y，求 l_{xy}

$$l_{xy} = \sum xy - N \cdot \bar{x} \cdot \bar{y} == 21658 - 15 \times 15.867 \times 86.333 = 1110.314$$

（4）求系数 a 和 b

$$b = l_{xy} / l_{xx} = \frac{1110.314}{1023.736} = 1.085$$

$$a = \bar{y} - b\bar{x} = 86.333 - 1.085 \times 15.867 = 69.117$$

（5）写出方程

$$\hat{y} = 69.117 + 1.085x$$

因此心理压力水平对心理健康水平的一元线性回归方程为 $\hat{y}$ ＝ 69.117 + 1.085x。由于回归系数为正，说明心理压力越大者，心理越倾向于不健康。

第三节　一元线性回归方程的检验

一、一元线性回归方程有效性检验

理论上，只要给出一组数据，都能找到一条最佳的拟合直线，即都能求出一个回归方程。但最佳拟合只是一个相比较而言的最佳，是一个相对的最佳，不一定是真正有效的回归方程。这类似于“小明在班上是最高分，但他的分数只有

40 分，仍然是不及格”。因此，须对所建立的回归方程的有效性进行检验，经检验不具备有效性的回归方程是不能使用的。

检验的基本思想是考察因变量 y 的变异中，有多少是由自变量 x 所引起的，如果因变量 y 的变异中大部分是由自变量 x 所引起，则说明用自变量来预测或解释因变量是有效的，即所求的回归方程有效。

回归分析中，因变量 y 的变异是因变量的离均差平方和 SS_T（即 l_{yy}），它可以分解成两个部分，一部分是由自变量所决定的变异即回归平方和（U），另一部分是残差平方和即误差平方和（Q）：

$$U = b^2\sum(x_i - \bar{x})^2 = b^2 l_{xx} \qquad （公式 13-10）$$

$$SS_T = l_{yy} = U + Q \qquad （公式 13-11）$$

$$Q = \sum(\hat{y}_i - y_i)^2 \qquad （公式 13-12）$$

或 $$Q = l_{yy} - U \qquad （公式 13-13）$$

离均差平方和是所有样本观测值与平均值离差的平方和，样本容量越大时，离均差平方和的值也会越大，因此它的大小会受样本容量的影响。不同研究者若调查的被试容量不同，则所计算的离均差平方和也不尽相同，因此离均差平方和需除以相应的自由度（df）以消除样本容量的影响。平均化后的离均差平方和称为均方，平均化后的回归平方和称为回归均方（记为 MS_U），而平均化后的残差平方和称残差均方（记为 MS_Q）。

一元线性回归方程有效性检验的原假设为所求的回归方程无效，即回归方程中来自总体自变量的系数为 0（$\beta = 0$）。研究发现在这一假设成立的前提下，$\frac{MS_U}{MS_Q}$ 这一统计量服从自由度为（df_U, df_Q）的 F 分布。因此检验一元线性回归方程的有效性可以通过构建 $\frac{MS_U}{MS_Q}$ 统计量，并在 F 分布的基础上，计算统计量 $\frac{MS_U}{MS_Q}$ 服从 F 分布的概率。若 $\frac{MS_U}{MS_Q}$ 服从 F 分布的概率属小概率事件，则认为 $\frac{MS_U}{MS_Q}$ 不服从 F 分布，与“原假设成立下，$\frac{MS_U}{MS_Q}$ 服从 F 分布”相矛盾，从而可以推断原假设错误。反之，则无充分理由拒绝原假设。

一般来讲，$\frac{MS_U}{MS_Q}$ 的值越大于 1，说明回归的均方 MS_U 越大于残差均方 MS_Q，也就是说在因变量总变异中，自变量所决定的变异占总变异的比重越大于误差因素所决定的变异，这时统计量 $F = \frac{MS_U}{MS_Q}$ 越容易进入拒绝域，从而越容易拒绝原假

设，即方程有效。

在完成F检验时，需列出如下方差分析表（表 13-2）

表 13-2 方差分析表(一元回归)

方差来源	平方和	自由度	均方	F 值
回归	$U=b^2 l_{xx}$	1	$MS_U=\frac{U}{1}$	$F=\frac{MS_U}{MS_Q}$
残差	$Q=l_{yy}-U$	$n-2$	$MSQ=Q/(n-2)$	
总体	l_{yy}	$n-1$		

【例 13-2】 请对例 13-1 建立的回归方程有效性进行检验$(\alpha=0.05)$。

解：

（1）建立假设 $\begin{cases} H_0：所建的回归方程无效 \\ H_1：所建的回归方程有效 \end{cases}$

（2）根据b, l_{xx}, l_{yy}，计算回归平方和（U）和残差平方和（Q）

$$U=b^2 l_{xx}=1.085^2\times 1023.736=1205.168$$

$$Q=l_{yy}-U=4055.34-1205.168=2850.172$$

（3）列方差分析表

表 13-3 **例 13-2** 数据的一元回归方差分析表

方差来源	平方和	自由度	均方	F 值
回归	1205.168	1	1205.168	5.497
残差	2850.172	13	219.244	—
总体	4055.34	14	—	—

（4）查表求临界值并做出决策

$$F_{\alpha=0.05}(1,14)=4.67$$

因为$F=5.497>F_{\alpha=0.05,(1,14)}=4.67$，因此拒绝原假设，认为所求的回归方程是有效的，这样下结论犯错误的概率不超过 5%。

二、一元线性回归方程有效性程度指标

一元线性回归方程的有效性检验，只是检验自变量系数即回归系数是否为 0，若检验结果为自变量系数不为 0（即方程有效），也只能说明这个回归方程有别于无使用价值的方程，但无法指出该方程的有效性程度到底有多高。因此需找

到衡量回归方程有效性高低的指标。

回归分析中，总平方和（l_{yy}）可分解为回归平方和（U）和残差平方和（Q），而回归平方和是由自变量 x 所决定的部分，若回归平方和越大，则说明因变量的变异中由自变量决定的部分越大，也就说明自变量越能有效地预测或解释因变量，因此可用指标 $\frac{U}{l_{yy}}$ 来表明一元线性回归方程的有效性程度，并称指标 $\frac{U}{l_{yy}}$ 为决定系数，记为 R^2，即

$$R^2 = \frac{U}{l_{yy}} \qquad \text{（公式 13-14）}$$

由公式 13-14 可知，决定系数（coefficient of determination）是指因变量 y 的总变异（即 l_{yy}）中可以由自变量的变异来决定或解释的比率。从上式可知，若 $U = l_{yy}$，那么因变量就完全由自变量所决定，两者就成了确定的线性关系。我们还可以对上式进行如下变换：

$$R^2 = \frac{U}{l_{yy}} = b^2\frac{lxx}{lyy} = \frac{l_{xy}^2}{(l_{xx}\cdot l_{yy})} = \left(\frac{l_{xy}}{\sqrt{(l_{xx}\cdot l_{yy})}}\right)^2$$

展开得

$$R^2 = \left(\frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2}\sqrt{\sum(y-\bar{y})^2}}\right)^2$$

可以发现上式括号中所求的为变量 x 与 y 的积差相关系数 r_{xy}，因此

$$R^2 = r_{xy}{}^2 \qquad \text{（13-15）}$$

也就是说，在一元线性回归中决定系数 R^2 是因变量与自变量积差相关系数的平方。回归分析是相关分析的继续与发展，数学上可以证明，回归分析对所建回归方程进行有效性检验，其本质是对变量间相关关系的显著性检验。因此如果两个变量间不存在显著线性相关，则没有必要进行线性回归分析，只有当两个变量存在显著相关，回归分析才有意义。

【例 13-3】 计算例 13-1 所建立回归方程有效性程度指标

解：

$$R^2 = \frac{U}{l_{yy}} = \frac{1205.168}{4055.34} = 0.297$$

说明因变量的变异中由自变量所决定的部分占 29.7%。

第四节 一元线性回归方程的预测

一 回归分析的误差

建立回归方程不是我们的终极目标，回归分析的最终目标是应用所建立的回

归方程对因变量作估计和预测。任何估计和预测都有误差的，因此首先需计算出估计误差。

在本章第三节曾指出，Q 是指除了自变量 x 的影响之后的剩余方差，MS_Q 是指残差均方，回归分析中将 $\sqrt{MS_Q}$ 称为回归估计标准误差，简记为 $S_{y\cdot x}$，即

$$S_{y\cdot x} = \sqrt{MS_Q} \tag{13-16}$$

$S_{y\cdot x}$ 实际上可以作为回归模型中对于确定自变量值 x，随机误差分布的标准差的估计值，因此可用做观察值变异情况的指标。

对于给定的自变量 x_0，回归方程的预测有两种情况，一种是 x_0 对应的因变量平均值（即主值 $\hat{y}$）的预测，另一种是对单个因变量值的预测。如，预测所有心理压力水平为 x_0 的学生的平均心理健康水平，这属于对因变量主值的预测；已知张三的心理压力水平为 x_0 而预测张三的心理健康水平，这属于对单个因变量值的预测。

数学上可以证明，用 $\hat{y}$ 作为 $\tilde{y}$ 的点估计量，其估计标准误 $S_{\hat{y}i}$ 为：

$$S_{\hat{y}i} = S_{y\cdot x} \cdot \sqrt{\frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{13-17}$$

而用 $\hat{y}$ 作为单个因变量进行估计或预测时，其估计的标准误 S_{y0} 为：

$$S_{yi} = S_{y\cdot x} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}} \tag{13-18}$$

二、回归方程的预测

（一）对因变量主值的估计及预测

点估计：$\hat{y}_i = \alpha + bx_i$

区间估计：$\hat{y} \in \left[\hat{y} - \left|t_{\alpha/2}\right| \cdot S_{\hat{y}_i}, \hat{y} + \left|t_{\alpha/2}\right| \cdot S_{\hat{y}_i}\right]$ （13-19）

其中 $t_{\alpha/2}$ 是自由度为 $n-2$，夹中间概率面积为 $1-\alpha$ 的 t 分布双侧分位数值。在样本容量较大时（$n \geq 30$），可以用正态分布 $Z_{\alpha/2}$ 代替 $t_{\alpha/2}$。

（二）对单个因变量的估计及预测

点估计：$\hat{y}_i = \alpha + bx_i$

区间估计：$\hat{y} \in \left[\hat{y} - \left|t_{\alpha/2}\right| \cdot S_{\hat{y}_i}, \hat{y} + \left|t_{\alpha/2}\right| \cdot S_{\hat{y}_i}\right]$ （13-20）

【例 13-4】 请根据例 13-1 所建的回归方程，预测表 13-1 中所有自变量所对应的因变量主值（$\alpha = 0.05$）

解：

根据回归方程 $\hat{y}=69.117+1.085x$、表 13-1 中自变量的值以及对主值区间估计的误差公式，预测结果如下表（表 13-4）。

表 13-4　对表 13-1 的预测结果

序号	心理压力水平（预测源）	心理健康水平（实际观测值）	预测的因变量主值（点估计）	预测的因变量主值（区间估计）
1	20	100	90.82	[81.58, 100.05]
2	8	88	77.80	[66.39, 89.20]
3	15	66	85.39	[77.09, 93.70]
4	14	92	84.31	[75.84, 92.77]
5	13	74	83.22	[74.48, 91.96]
6	12	79	82.14	[73.02, 91.25]
7	18	99	88.65	[80.12, 97.18]
8	29	135	100.58	[85.07, 116.09]
9	37	86	109.26	[86.58, 131.94]
10	13	83	83.22	[74.48, 91.96]
11	23	84	94.07	[83.16, 104.98]
12	11	90	81.05	[71.47, 90.64]
13	8	77	77.80	[66.39, 89.20]
14	4	64	73.46	[59.00, 87.91]
15	13	78	83.22	[74.48, 91.96]

第五节　多元线性回归的初步应用

一 多元线性回归的数学模型

前面讲的一元线性回归主要是讨论一个自变量对因变量的影响，实现生活中，尤其是心理现象中影响因变量的因素可能是多个。比如影响心理健康的因素不仅只有心理压力一个，还有诸如遗传、人际关系、家庭环境等等。这时根据多个自变量来解释或预测因变量，就会比只用一个自变量更加准确和有效。多元线性回归分析就是讨论多个自变量对因变量的影响。一般认为自变量的个数多于 2 个并且变量之间呈线性关系的回归分析方法称为多元线性回归分析（multiple linear regression analysis）。其分析的基本思想和原理与一元线性回归基本相似。

多元线性回归的数学模型为

$$y=\beta_0+\beta_1 x_1+\beta_2 x_2+\cdots+\beta_K x_K+\varepsilon \tag{13-21}$$

或简写为：

$$y=\beta_0+\sum_{j=1}^{K}\beta_j x_j+\varepsilon \tag{13-22}$$

由公式 13-22 式可得总体数据基础上的多元回归模型为：

$$\tilde{y} = \beta_0 + \sum_{j=1}^{K} \beta_j x_j \tag{13-23}$$

其样本数据的多元线性回归记为

$$\hat{y} = b_0 + \sum_{j=1}^{K} b_j x_j \tag{13-24}$$

有几个自变量我们就相应的称为几元线性回归。在多元线性回归方程中，β_0 依然是方程在 y 轴上的截距，$\beta_j(j = 1,2,\cdots,K)$ 是第 j 个自变量 x_j 的回归系数。如果所有自变量的量纲（即单位）都一致，则 β_j 的大小就反应了自变量 x_j 对因变量的影响大小，这时 $|\beta_j|$ 的值越大那么自变量 x_j 对因变量的影响越大，因此可以通过比较经量纲化后求得的 $|\beta_j|$ 的大小，来判断各自变量对因变量的影响大小。统计上，使自变量有相同等量纲的常用方法，是将所有自变量均独自标准化（即 Z 分数化）。

二、多元线性回归方程的初步应用

（一）多元线性回归方程的建立

多元线性回归方程建立的原理也是最小二乘法，相对于一元回归而言，由于多元回归的多变量，因此用最小二乘法求解回归方程要复杂的多。在高等数学中，比较常用的是用矩阵算法进行推演，在这我们不作详细推演。在实际应用时，研究者可以借助相应的统计软件（如 SPSS，SAS）等，由电脑来完成所有计算。

（二）多元线性回归方程有效性检验

多元线性回归方程建立后，也应对方程是否有效进行检验。多元线性回归方程的总离差平方和（l_{yy}）还是分解为回归平方和（U）和残差平方和（Q），其方差分析表见表 13-5。

表 13-5 方差分析表(多元回归)

方差来源	平方和	自由度	方差	值
回归	$U = b^2 l_{xx}$	K	$MSU = U/K$	$F = \frac{MSU}{MSQ}$
残差	$Q = l_{yy} - U$	$n-K-1$	$MSQ = Q/(n-K-1)$	—
总体	l_{yy}	$n-1$	—	—

在原假设成立的前提下，$F = \frac{MSU}{MSQ}$ 服从自由度为 $(K, n-K-1)$ 的 F 分布，其中 K 为自变量的个数。

多元线性回归方程有效性高低的指标仍然用决定系数 R^2 表示，$R^2 = U/l_{yy}$，

其意义与一元线性回归一样，但 R^2 这时表示的是因变量与 K 个自变量间复相关系数平方。

（三）多元线性回归各自变量有效性检验

多元线性回归有效，并不代表其中的每个自变量均有效，因此还应对每个自变量的显著性进行检验，以便剔除不显著的变量，从而简化回归方程。检验的原假设为 $\beta_j = 0$。其检验过程是建立在回归方程建立的基础上，在这不作相关推演，研究者可借助 SPSS 统计软件来完成。这项工作在一元线性回归分析中是不必要的，因为一元回归中方程的有效性检验与一个自变量的显著性检验是完全等价的。

（四）多元线性回归方程的应用

下面以一个例子来探讨多元线性回归的应用。

【例 13-5】 为了解居民住房价格的影响因素，随机调查了 30 户居民，并收集了居民当时购买住房的一些基本情况：住房价格、住房居住面积、住房基本配置和住房地理位置。下表为调查结果。请通过多元回归分析的方法，分析住房居住面积、住房基本配置和住房地理位置是否会影响住房价格？

表 13-6 30 户居民住房价格等情况调查

编号	住房价格	住房居住面积	住房基本配置	住房地理位置	编号	住房价格	住房居住面积	住房基本配置	住房地理位置
1	4350	80	4	1	16	3005	81	2	0
2	4100	147	2	1	17	2929	110	1	0
3	3760	86	3	1	18	3175	93	2	0
4	3575	96	3	1	19	3139	89	1	0
5	3470	70	3	1	20	2780	130	1	0
6	3355	106	2	1	21	2899	156	3	0
7	3160	97	3	0	22	3329	103	2	1
8	3220	92	2	0	23	3245	100	2	0
9	3220	104	2	1	24	3200	95	3	0
10	3050	110	1	1	25	3230	95	1	1
11	2920	74	2	0	26	3140	83	2	0
12	2949	105	3	0	27	3120	89	2	0
13	4350	90	4	1	28	3072	79	2	0
14	3250	102	2	1	29	3069	130	2	0
15	3249	113	2	1	30	2966	118	2	0

注：住房价格单位为元/每平方米；住房居住面积单位为平方米；住房基本配置有四种水平：4 代表优、3 代表良、2 代表中、1 代表差；住房地理位置有二种水平：1 代表好，0 代表差。

解：

运用 SPSS18.0 软件包解本题时，首先将以上数据录入数据文件窗口中，根据图 13－3 所示的操作步骤可以求解本例的多元回归方程（图 13-4）。主要分析结果有如表 13-7 至表 13-9 所示。

1. 分析→回归→线性
2. 因变量框：住房价格
3. 自变量框：住房居住面积、住房基本配置、住房地理位置
4. 确定

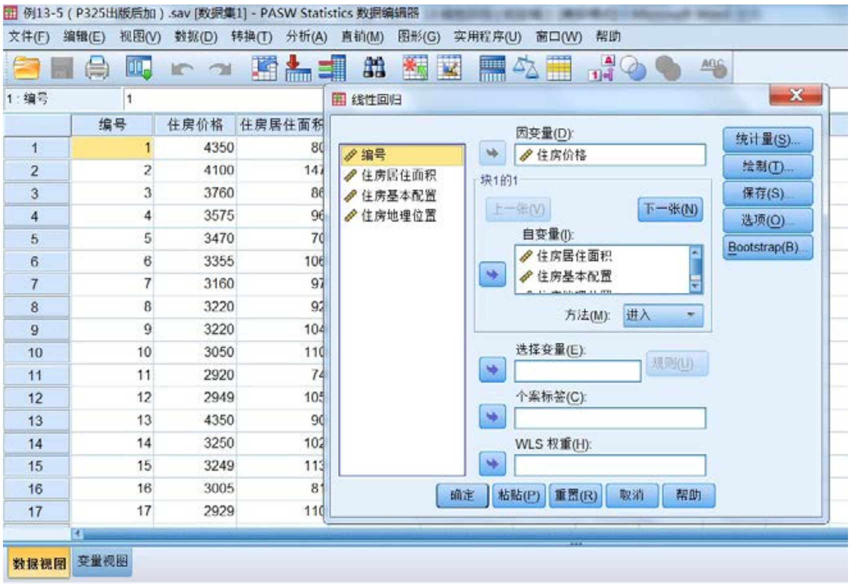


图 13-3 在 SPSS 中设置多元回归分析程序各步骤示意图

表 13-7　方程有效性检验

模型		*平方和*	*df*	*均方*	*F*	*Sig.*
1	回归	2954686.571	3	984895.524	16.514	.000[a]
	残差	1550672.895	26	59641.265		
	总计	4505359.467	29			

a. 预测变量：(常量)，住房地理位置，住房居住面积，住房基本配置。

b. 因变量：住房价格

表 13-8　方程有效性程度指标

模型	*R*	*R 方*	*调整 R 方*	*标准 估计的误差*
1	.810[a]	.656	.616	244.216

a. 预测变量：(常量)，住房地理位置，住房居住面积，住房基本配置。

表13-8 第三列所示的“ $R^2 = 0.656$ ”就是本次回归方程的决定系数。

表 13-9　各自变量回归系数有效性检验

模型		非标准化系数		标准系数		
		B	标准 误差	试用版	t	Sig.
1	(常量)	2649.396	289.299		9.158	.000
	住房居住面积	-.867	2.304	-.044	-.376	.710
	住房基本配置	241.142	58.698	.493	4.108	.000
	住房地理位置	422.998	91.919	.541	4.602	.000

a. 因变量：住房价格

表13-8第二列第一行数据“2805.76”是截距，其他三项数据分别是三个自变量未标准化的回归系数；第四列的三项数据则是这三个自变量标准化后的回归系数；对各项进行t检验后结果表明，除“住房居住面积”外，其他各项检验结果都达显著水平。由于自变量“居住面积”的回归系数不显著性，因此为简化模型，可将该变量从回归模型中剔除，从而构成二元线性回归。具体做法是：在图15-3所示的第3步自变量框中只选入住房基本配置和住房地理位置两个变量，然后让系统再做一次多元回归分析。结果有如表13-10至表13-12）所示。其结果的解释与前面的分析相同。

表 13-10　方程有效性检验

模型		*平方和*	*df*	*均方*	*F*	*Sig.*
1	回归	2946245.734	2	1473122.867	25.511	.000[a]
	残差	1559113.733	27	57744.953		
	总计	4505359.467	29			

a. 预测变量：(常量)，住房地理位置，住房基本配置。

b. 因变量：住房价格

13-11　方程有效性程度指标

模型	*R*	*R 方*	*调整 R 方*	*标准 估计的误差*
1	.809[a]	.654	.628	240.302

a. 预测变量：(常量)，住房地理位置，住房基本配置。

表 13-12　各自变量回归系数有效性检验

模型		非标准化系数		标准系数		
		B	标准 误差	试用版	t	Sig.
1	(常量)	2552.635	130.312		19.589	.000
	住房基本配置	245.520	56.611	.502	4.337	.000
	住房地理位置	422.509	90.437	.540	4.672	.000

a. 因变量：住房价格

综上，通过对调查结果分析可知：自变量“住房基本配置”和“住房地理位置”两个自变量能显著地解释或预测因变量“住房价格”，其解释率高达65.4%，它们是影响住房价格的主要因素。而自变量“住房面积”不影响住房价格。

小　结

回归分析主要是描述事物间的不确定关系的统计分析方法，主要用于探明变量间在数量上的因果关系，其目的在于为不确定现象的研究提供科学、精细的方法，并对相关变量或事物进行估计、预测和控制。回归分析主要包括回归方程的建立、回归方程有效性检验及回归方程的应用等几项内容。

关键术语

自变量：因为自身的变化而能引起另外一个或一些变量发生变化的变量

因变量：在自变量的影响下，随着自变量的变化而发生变化的变量。

回归分析：研究在自变量已经发生变化的情况下因变量会如何变化、变化的范围有多大的一整套统计分析方法。

回归方程：通过回归分析而建立的描述因变量的均值与自变量之间回归关系的函数或方程。

一元线性回归：指自变量只有一个，且自变量与因变量之间存在线性函数关系的回归分析方法

回归系数：样本数据下建立的一元线性回归模型记为 $\hat{y}=a+bx$，其中，α 为方程在 y 轴上截距，b 为回归系数即斜率。

最小二乘法：建立回归方程其实就是求解最佳拟合直线的系数 a 和 b。统计学上，把 Q 最小的条件下估计方程系数 a 和 b 的方法称为最小二乘法。

方程有效性检验：回归方程建立后需要对其有效性进行检验，检验的基本思想是考察因变量 y 的变异中，有多少是由自变量 x 所引起的，如果因变量 y 的变异中大部分是由自变量 x 所引起，则说明用自变量来预测或解释因变量是有效的，即所求的回归方程有效。

决定系数：指因变量 y 的总变异中可以由自变量的变异来决定或解释的比率。回归分析中，总平方和（l_{yy}）可分解为回归平方和（U）和残差平方和（Q），而回归平方和是由自变量 x 所决定的部分。回归平方和越大，说明因变量的变异中由自变量决定的部分越大，自变量越能有效地预测或解释因变量，因此可用指标 $\frac{U}{l_{yy}}$ 来表明一元线性回归方程的有效性程度，并称指标 $\frac{U}{l_{yy}}$ 为决定系数，记为 R^2，即 $R^2=\frac{U}{l_{yy}}$。

多元线性回归分析：自变量的个数多于 2 个并且变量之间呈线性关系的回归分析方法。

重要公式

样本数据基础上建立的一元线性回归模型：

$$\hat{y}=a+bx$$

根据最小二乘法原理建立的一元线性回归模型：

$$l_{xx}=\sum(x-\overline{x})^2=N\cdot\sigma_x{}^2\text{；}$$

$$l_{yy} = \sum(y-\bar{y})^2 = N \cdot \sigma_y{}^2\text{；}$$

$$l_{xy} = \sum(x-\bar{x})(y-\bar{y}) = \sum xy - N \cdot \bar{x} \cdot \bar{y}\text{；}$$

$$b = l_{xy} / l_{xx}$$

$$a = \bar{y} - b\bar{x}\text{；}$$

回归平方和： $U = b^2 \sum(x_i - \bar{x})^2 = b^2 l_{xx}$

决定系数： $R^2 = U / l_{yy}$

回归分析误差： $S_{y \cdot x} = \sqrt{MSQ}$， $\hat{y} = b_0 + \sum_{j=1}^{K} b_j x_j$

用 $\hat{y}$ 作为 $\tilde{y}$ 的点估计量，其估计标准误

$$S_{\tilde{y}0} = S_{y \cdot x} \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

用 $\hat{y}$ 作为单个因变量的点估计量，其估计标准误

$$S_{y0} = S_{y \cdot x} \cdot \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}}$$

基于样本数据的多元线性回归模型：

$$\hat{y} = b_0 + \sum_{j=1}^{K} b_j x_j$$

统计软件使用提示

回归分析可借助 SPSS 统计软件完成，具体为“分析（Analyze）”命令功能中的“回归（Regression）”，并选择“线性（回归，Linear）”，具体可参考本章相关 SPSS 操作。

思考与练习

1. 回归分析的基本原理是什么？
2. 回归方程有效性检验的原假设为什么是“方程无效”而不是“方程有效”？
3. 多元回归分析中，为什么要对各自变量的显著性进行检验

第十四章　一般线性模型的方差与协方差分析

学习本章内容，将有助于你对以下问题的理解与思考：

1.什么是一般线性模型？

2.一般线性模型的单因素方差分析与传统单因素方差分析的异同？

3.什么是协方差分析？

4.一般线性模型中方差分析与协方差分析的区别及联系是什么？

5.如何在 SPSS 软件中进行一般线性模型分析？

统计分析有多种方法，人们很容易认为各种方法之间没什么关系，事实上并不是这样，很多方法从某个角度看是相同的。在本章中，读者将发现，t 检验、方差分析和多元回归原来可以统一到一个模型下，这个模型就是一般线性模型。本章首先介绍一般线性模型，再用多个例子对比详细说明 t 检验、方差分析和多元回归都属于一般线性模型，最后介绍一般线性模型下的协方差分析的方法。

第一节 一般线性模型简介

一、一般线性模型的概念

一个实际的统计问题，在进行数学的抽象之后，往往可归结为一个统计模型。一般而言，这个模型无非是联系着变量（包括随机变量和一般变量）和参数（已知和未知的）的一组数学关系式。假设有一个可观察的随机变量 Y，和一组可观察的一般变量 x_1，…，x_m；β_1，…，β_p 是未知参数，ε 是不可观察的随机变量，如果它们满足

$$Y=\sum_{i=1}^{p} f_i(x_1,\cdots,x_m)\beta_i+\varepsilon,\ f_i\text{是已知函数} \qquad \text{（公式 14－1）}$$

则称公式（14－1）是线性统计模型，也称为一般线性模型，简称为线性模型；式中 ε 被称为随机误差，一般总是假定为正态独立随机向量，其均值 $E(\varepsilon)=0$。可观察的随机变量，指的是在一次观察或试验中可以测得它的取值的随机变量；可观察的一般变量，指的是在一次观察或试验中可以精确测定或严格控制的，不具有随机性的变量。

如果将公式（14－1）中的 $f_i(x_1,\cdots,x_m)$ 记为 $\tilde{x}_i$，$i=1,\cdots,p$，从而公式（14－1）可记为

$$Y=\sum_{i=1}^{p}\tilde{x}_i\beta_i+\varepsilon \qquad \text{（公式 14－2）}$$

于是，不失一般性可记线性模型为

$$Y=\sum_{i=1}^{p}x_i\beta_i+\varepsilon \qquad \text{（公式 14－3）}$$

或用均值的形式来表达，记公式（14－3）为

$$EY=\sum_{i=1}^{p}x_i\beta_i \qquad \text{（公式 14－4）}$$

应注意的是，记为公式（14－3）形式是一种一般的表达方式，其本质仍是公式（14－1），是 $f_i(x_1,\cdots,x_m)\beta_i$，也就是说是 x_1，…，x_m 的函数乘以 β_i。例如，可以是 $\beta_1x_1+\beta_2x_2$，也可以是 $\beta_1x_1+\beta_2x_2^2$，还可以是 $\beta_1x_1+\beta_2x_2^2+\beta_3x_1x_2$，等等。显然，作为 x_1 和 x_2 的函数，后面两个例子是非线性的。在此着重指出一点，一般线性模型的线性是指对参数 β_1，…，β_p 而言的，而不是 x_i 在模型中出现的方式。

在公式（14－4）中，EY 是 x_1，…，x_p 的函数，故称 x_1，…，x_p 为自变量，称 Y 为因变量。当然，我们不应把这里的 Y 和 x_1，…，x_p 的统计依赖关系和函数关系相混淆，这里不存在决定性的因素关系。

在对公式（14—3）的模型进行统计推断时，由于它含有 p 个未知参数和误差，必须获取有 n 次试验的数据（如表 14-1 所示），这里的试验次数 n 一般应大于 p。

表 14-1　公式（14—3）模型推断时的数据结构

试验序次	y	x_1	x_2	…	x_p
1	y_1	x_{11}	x_{12}	…	x_{1p}
2	y_2	x_{21}	x_{22}	…	x_{2p}
⋮	⋮	⋮	⋮	…	⋮
n	y_n	x_{n1}	x_{n2}	…	x_{np}

设第 α 次试验中自变量的取值为 $x_\alpha=(x_{\alpha 1},\cdots,x_{\alpha p})'$，则 x_α 是一个试验点。注意，在此用 α 这个下标表示试验的序次，避免与变量的序次相混淆。相应的因变量 Y 的观察值记为 y_α。从而有样本满足的模型为

$$y_\alpha=\sum_{i=1}^{p}x_{\alpha i}\beta_i+\varepsilon_\alpha,\quad E\varepsilon_\alpha=0,\quad \alpha=1,\cdots,n.\qquad \text{（公式 14—5）}$$

这里 y_α 为观察值。由于 ε 是不可观察的随机变量，人们对只能根据 $x_{\alpha 1},\cdots,x_{\alpha p}$ 对 y_α 进行预测，预测值记为 $\hat{y}_\alpha$。为了精确地对 y_α 进行预测，即 $\hat{y}$ 与 y 的误差最小，模型中会给出 y 的预测初始值 β_0，即为

$$\hat{y}_a=\beta_0+\beta_1x_{\alpha 1}+\beta_2x_{\alpha 2}+\cdots+\beta_px_{\alpha p},\quad (\alpha=1,\cdots,n)\qquad \text{（公式 14—6）}$$

在回归分析中，β_0 就是截距；在方差分析中，β_0 就是观察数据的总均值。若引进矩阵记号

$$y=\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix},\quad X=\begin{pmatrix}1&x_{11}&\cdots&x_{1p}\\1&x_{21}&\cdots&x_{2p}\\\vdots&\vdots&\vdots&\vdots\\1&x_{n1}&\cdots&x_{np}\end{pmatrix},\quad \beta=\begin{pmatrix}\beta_0\\\beta_1\\\vdots\\\beta_p\end{pmatrix},\quad \varepsilon=\begin{pmatrix}\varepsilon_1\\\varepsilon_2\\\vdots\\\varepsilon_n\end{pmatrix}$$

y 为 $(n\times 1)$ 的观测向量，X 为由自变量构造的 $n\times(p+1)$ 的设计矩阵，β 为 $(p+1)\times 1$ 的系数向量，ε 为 $(n\times 1)$ 的误差向量。则公式（14-6）式就可以记为如下简洁形式

$$\hat{Y}=X\beta\qquad \text{（公式 14-7）}$$

则 y 为观察值可记为

$$Y=X\beta+\varepsilon\qquad \text{（公式 14-8）}$$

二、一般线性模型的分类

一般线性模型的分类，谈不上有很严格的标准，只是可以从不同角度进行分类。实验工作者常从模型的来源和要解决的实际问题的角度进行分类，而统计学者常从模型的数学特征和它所侧重的统计问题的角度进行分类。本书将从后者的角度进行分类。

前文指出，自变量是可观察的一般变量，是在一次观察或试验中可以精确测定或严格控制的，不具有随机性的变量。从这里可以看出，自变量可以是连续变量，例如考察个体智力

水平对学业成绩的影响，自变量智力水平从理论上讲就是一个连续变量；另外，自变量也可以是分类变量，例如考察不同干预方法对抑郁症恢复效果的影响，自变量干预方法就是一个分类变量，有几个类别就是几个水平。

根据自变量的变量性质，可以将一般线性模型分为三类：如果自变量均为连续变量，则该模型为回归分析模型；如果自变量均为分类变量，则该模型为方差分析模型；倘若自变量中，既有分类变量，也有连续变量，则称之为协方差分析模型。

另外还有一种依据参数 β_1，…，β_p 的性质不同进行的分类。从某种意义上说 β_i 反映了自变量的第 i 个因子对观察值 y 的影响大小，被称为第 i 个因子的效应。这个效应在实际问题里可以是随机的，于是可依据 β_1，…，β_p 是否是随机的来分类。如果 β_1，…，β_p 均为非随机变量而是固定的未知参数，则称模型为固定效应模型；如果 β_1，…，β_p 均为不可观察的随机变量，则称模型为随机效应模型；如果 β_1，…，β_p 当中既有未知参数又有不可观察的随机变量，就称模型为混合效应模型。读者可对照第十章中随机效应模型方差分析的内容加以理解。

三、一般线性模型的重要性

一般线性模型是一个极为重要的统计模型，它之所以重要，在于它是最简便地描述因变量与自变量之间的依存关系的模型，t 检验、方差分析、多元回归、协方差分析、描述性判别分析、多元方差分析、典型相关分析、结构方程模型等都可以统一到一般线性模型框架中。并且，这些统计方法是有层次区分的，某些方法是其他方法的特例。例如，t 检验是单变量方差分析的特例；所有单变量的统计分析技术，都是多元回归分析的特例；所有单变量和多元分析的方法又是典型相关的特例；而上述各种方法又是结构方程模型的特例。

上述各种统计方法都是基于一般线性模型的方法，有着一些共同的特征，本质上都是基于最小二乘法的思想，使观察值与预测值之间的差异最小。所以，一般线性模型提供了一个学习与理解上述各种统计方法的整体框架，可以加深学习者对各种统计方法的理解，还可以为学习者学习更高级统计方法（如多层线性模型等）奠定更好的知识基础，拓广统计方法的应用范围。在本章，就是要让学习者能够理解：包括 t 检验、方差分析、多元回归、协方差分析等常见的统计方法都可以统一到一般线性模型的理论框架上来。回归分析是最典型的（或称为最“标准”的）一般线性模型，其求解过程与结果解释在第十三章已有陈述，本章不再赘述。本章在第二节将对 t 检验、方差分析采用一般线性模型的方法进行求解，并与传统的方法进行对照，让学习者理解作为一般线性模型的 t 检验和方差分析；在第三节将对协方差分析采用一般线性模型的方法进行求解。作为一般线性模型的 t 检验、方差分析和协方差分析方法的求解，都与回归分析的求解方法相同。

另外，使用一般线性模型在计算上是有优势的，如，当各实验处理的样本容量不同时，一般线性模型处理数据更直接；再如，传统上，协方差分析用的是比较笨的方法，一般线性模型使得协方差分析运行简单，并容易理解（详细介绍见本章第三节）。

第二节 一般线性模型与“t 检验”和“方差分析”

本节主要是用几个实例来说明一般线性模型这一统计分析方法与 *t* 检验和方差分析这两种常见的统计方法的相互关系，并将这三种统计不同方法所得到的统计结果进行比较，由此加深读者对一般线性模型的理解。

一、t 检验与两个水平的单因素方差分析

【例 14-1】假设有 24 名被试，其中 12 人被临床诊断为抑郁症患者，另 12 人为非抑郁症患者，这 24 名被试在某抑郁症测试量表得分见表 14-2。问诊断为抑郁症的患者与非抑郁症的患者在该量表上的得分是否存在显著差异？

表 14-2　24 名被试的抑郁测试得分

被试	抑郁量表得分	临床诊断结果	诊断结果编码
1	3	非抑郁症	-1
2	5	非抑郁症	-1
3	6	非抑郁症	-1
4	8	非抑郁症	-1
5	9	非抑郁症	-1
6	11	非抑郁症	-1
7	19	抑郁症	1
8	15	抑郁症	1
9	16	抑郁症	1
10	16	抑郁症	1
11	19	抑郁症	1
12	17	抑郁症	1
13	3	非抑郁症	-1
14	5	非抑郁症	-1
15	4	非抑郁症	-1
16	6	非抑郁症	-1
17	8	非抑郁症	-1
18	10	非抑郁症	-1
19	24	抑郁症	1
20	24	抑郁症	1
21	22	抑郁症	1
22	23	抑郁症	1
23	19	抑郁症	1
24	20	抑郁症	1

（1）用传统的 *t* 检验方法求解

由于本例题的求解目的是诊断为“抑郁症”与“非抑郁症”这两个不同组的患者在某个量表上的得分是否存在显著差异，因此，我们可以使用第六章所介绍的“*t* 检验分析方法”来达到对这些数据进行平均数差异显著性检验的目的。有关 *t* 检验分析方法在第六章有详细讲解，这里直接给出 SPSS 分析的结果如表 14-3 和表 14-4。

表 14-3　分组统计量

	临床诊断结果	N	均值	标准差	均值的标准误
抑郁量表得分	非抑郁症	12	6.50	2.680	.774
	抑郁症	12	19.50	3.177	.917

表 14-4　独立样本检验（SPSS 分析部分结果）

		方差方程的 Levene 检验		均值方程的 t 检验			
		F	Sig.	t	df	Sig.(双侧)	均值差值
抑郁量表得分	假设方差相等	.298	.590	-10.836	22	.000	-13.000
	假设方差不相等			-10.836	21.393	.000	-13.000

答：由于$|t| = -10.836 > t_{\frac{\alpha}{2}(46)} = 3.551$，因此我们在$\alpha = 0.01$的显著性水平上推断诊断为抑郁症的患者与非抑郁症的患者在该量表上的得分具有显著差异。

（2）用传统的方差分析方法求解

方差分析方法通常用于求解多于两个分组的平均数差异显著性检验问题，但作为特例，也可以用它来求解类似于上述"'抑郁症'与'非抑郁症'这两个不同组的患者在某个量表上的得分是否存在显著差异"这样的假设检验问题。有关方差分析方法在第十章有详细讲解，这里直接给出 SPSS 分析的结果如表 14-5（用"ANOVA"程序和用"一般线性模型中的单变量"程序计算结果是一样的）。

表 14-5　方差分析 SPSS 检验结果

变异来源	III型平方和	*df*	均方	*F*	*Sig.*	偏*Eta*方
临床诊断结果	1014.000	1	1014.000	117.411	.000	.842
误差	190.000	22	8.636			
校正的总计	1204.000	23				

答：由于 $F=117.41 > F_{0.01(1,22)}=7.95$，因此我们在$\alpha = 0.01$的显著性水平上推断诊断为抑郁症的患者与非抑郁症的患者在该量表上的得分具有显著差异。

（3）用一般线性模型分析方法求解

本章第一节曾提到，根据数理统计学家的观点："t 检验、方差分析、多元回归、协方差分析等，都可以统一到一般线性模型的理论框架上来"。换言之，我们也可以通过一般线性模型来求解类似于上述"'抑郁症'与'非抑郁症'这两个不同组的患者在某个量表上的得分是否存在显著差异"这样的假设检验问题。不过，求解的目的不只是在于做出两个组的平均数之间是否存在显著差异，更要通过一般线性模型分析方法估计出适合表达表 11-1 中相关数据的"抑郁量表得分=$\beta_0 + \beta_1$×临床诊断结果+ε"这一方程中的β_0和β_1参数。

在对“β_0”和“β_1”这两个参数进行求解的过程中，“临床诊断结果”作为自变量进入方程，必须是一个量化的指标，这正是用一般线性模型来解决 t 检验和方差分析问题的关键所在。在 t 检验和方差分析时，临床诊断结果仅起到分类的作用，给予“任何”编码都是可以的，如“1 和 2”或“0 和 1”；但在一般线性模型中的编码是很关键的问题，特别是三个水平以上的方差分析时，在两个水平的方差分析时，采用哪个编码都可以得到相似的但本质上相同的结果。本例中采用“－1”和“1”的编码（读者也可试试“0 和 1”编码进行分析，以进行对照分析，加深理解），这样编码有其特殊优势，这一点在后文有详细解释。

当对“临床诊断结果”这列数据作好编码后，求解 β_0 和 β_1 参数，本质上就是进行回归分析。这里直接给出 SPSS 的回归分析结果如表 14-6、表 14-7 和表 14-8。

表 14-6　回归分析模型概要表

模型	R	R 方	调整 R 方	标准估计的误差
1	.918[a]	.842	.835	2.939

表 14-7　回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	1014.000	1	1014.000	117.411	.000
	残差	190.000	22	8.636		
	总计	1204.000	23			

表 14-8　回归系数及其显著性检验结果

模型		非标准化系数		标准系数	t	*Sig.*
		B	标准误差	β		
1	(常量)	13.000	.600		21.671	.000
	诊断结果编码	6.500	.600	.918	10.836	.000

答：由一般线性模型方法对例 14-1 所列数据求解结果为：$\beta_0 = 13$，$\beta_1 = 6.5$，用方程表示就是：抑郁量表得分=13 + 6.5 × 临床诊断结果+ε。

（4）传统 t 检验、方差分析与一般线性模型分析结果的对比

通过对照传统 t 检验、方差分析与一般线性模型分析的结果可以看到，无论是哪种方法，我们都得到了“诊断结果为非抑郁症组和抑郁症组的两组被试的抑郁量表得分存在显著差异”这样一个相同的统计推论：t 检验结果通过“得到的统计量 t =-10.84 在显著性水平 $\alpha = 0.001$ 上大于其临界值（即 $t_{\alpha/2,(46)} = 3.551$）”而得出这一推论；方差分析结果通过“得到的统计量 F=117.41 在显著性水平 $\alpha = 0.01$ 上大于其临界值（即 $F_{0.01(1,22)}$ =7.95”也得出同样的推论。实际上，数理统计学家告诉我们，虽然“t 检验”和“方差分析”这两种统计分析方法是从不同角度来解决诸如例 11-1 中所求解的“两组被试的抑郁量表得分存在显著差异”这样的

假设检验问题，但这两种假设检验方法之间也是有密切联系的：表现为“$t^2=F$”。

通过一般线性模型的分析，我们得到方程“抑郁量表得分=13.0+6.5×临床诊断结果+ε”，在对这一方程的求解过程中，表 11-7 所列有关“回归方程有效性检验”的结果与表 11-5 所列的“方差分析”结果几乎完全一致，这说明通过一般线性模型的分析方法我们也能达到“求解‘两组被试的抑郁量表得分存在显著差异’”这样的假设检验问题。此外，通过一般线性模型的分析所得到的“抑郁量表得分=13.0+6.5×临床诊断结果+ε”这一方程，还能告诉我们“当把‘诊断结果’作为自变量时如何有效预测被试抑郁量表得分”这样的问题。

总之，诊断结果的回归系数检验显著（t=10.836**，与 t 检验结果相同），也说明诊断结果能有效预测被试抑郁量表得分；这说明，抑郁症患者与非抑郁症患者的抑郁量表得分存在显著差异，与传统 t 检验、方差分析的结果完全相同。另外，决定系数 R^2=0. 842,说明诊断结果能解释抑郁量表得分的 84. 2%的变异，这与方差分析的η^2相同。

另外，回归方程的截距就是两组被试得分的总均值。当临床诊断结果为“-1”（即为非抑郁症组）时，抑郁量表得分＝13.0－6.5＝6.5，这就是非抑郁症组的平均得分；当临床诊断结果为“1”（即为抑郁症组）时，抑郁量表得分＝13.0+6.5＝19.5，这就是抑郁症组的平均得分。

我们还可以通过表 11-1 中每位被试的得分与 “抑郁量表得分=13.0+6.5×临床诊断结果+ε”这一方程的相互关系来理解一般线性模型对表中数据的分析结果。数理统计学家指出，通过一般线性模型还可以将每个被试的成绩用下列通式来表达：

$$Y_{ij}=\mu+\tau_j+e_{ij} \qquad \text{（公式 14-9）}$$

μ表示总平均效应；τ表示处理效应，τ_j是$\tau_1\tau_2\tau_3\ldots\tau_p$中的某一个。对于某个被试来讲，他只会受某个处理的影响；e表示误差。

将公式 11-9 与“抑郁量表得分=13.0+6.5×临床诊断结果+ε”相联系，则有“$\mu=13$”，共有“非抑郁症组（τ_1，编码为（-1））”和“抑郁症组（τ_2，编码为（1））”两个组，因此$\tau_1=6.5\times(-1)=-6.5$；$\tau_2=6.5\times(1)=6.5$。

表 11-1 数据中被诊断为“非抑郁症组（τ_1）”的 12 名被试中，最低得分为 3 分，最高得分为 11 分；被诊断为“抑郁症组”的 12 名被试中，最低得分为 15 分，最高得分为 24 分。这 24 人中每个人的得分都可以通过“抑郁量表得分=13.0+6.5×临床诊断结果+ε”这一方程来表示，例如，表中所列第 1 位被试属于“非抑郁症组（τ_1=-6.5’）”者，其得分是“3 分”，若用上述方程表示就是：“$3=13-6.5-3.5$”；表中所列第 20 位被试属于“抑郁症组（τ_2=6.5）”者，其得分是“24 分”，若用上述方程表示就是：“$24=13+6.5+4.5$”；其他以此类推。由此可以得到“抑郁量表得分=13.0+6.5×临床诊断结果+ε”这一方程对 24 名被试得分情况的表达式为：

$$
\begin{pmatrix} 3 \\ \vdots \\ 11 \\ 19 \\ \vdots \\ 17 \\ 3 \\ \vdots \\ 10 \\ 24 \\ \vdots \\ 20 \end{pmatrix} = 13 + \begin{pmatrix} -6.5 \\ \vdots \\ -6.5 \\ 6.5 \\ \vdots \\ 6.5 \\ -6.5 \\ \vdots \\ -6.5 \\ 6.5 \\ \vdots \\ 6.5 \end{pmatrix} + \begin{pmatrix} -3.5 \\ \vdots \\ 4.5 \\ -0.5 \\ \vdots \\ -2.5 \\ -3.5 \\ \vdots \\ 3.5 \\ 4.5 \\ \vdots \\ 0.5 \end{pmatrix}
$$

上述方程式中，将第二列每个数据加以平方后求和，即得到“$(-6.5)^2 \times 12 + (6.5)^2 \times 12 = 1014$”，这就是方差分析中的“组间平方和”；将第三列每个数据加以平方后求和，即得到“$(-3.5)^2 + (-1.5)^2 + \cdots + (-0.5)^2 + (0.5)^2 = 190.00$”，这就是方差分析中的“误差平方和”。

通过上面的论述可知，一般线性模型的分析结果与传统 t 检验、方差分析的结果相比较，除了能得到与前两种统计方法相比较完全相同的假设检验结果之外，还可以提供前两种统计方法所没有提供的其他有用的统计信息。当自变量增多时，使用一般线性模型的统计分析方法进行矩阵运算，可以使得计算过程更加简洁。

二、2×2 设计的方差分析

【例 14-2】假定**【例 14-1】**的数据中，其中第 1－12 名被试为男性，第 13－24 名被试为女性，数据见表（14-9）问性别、诊断结果的主效应是否显著？二者的交互作用是否显著？

表 14-9　24 名被试的抑郁测试得分

被试	抑郁量表得分	性别（男=1，女=-1）	诊断结果（抑郁=1，非抑郁=-1）	性别×诊断结果
1	3	1	-1	-1
2	5	1	-1	-1
3	6	1	-1	-1
4	8	1	-1	-1
5	9	1	-1	-1
6	11	1	-1	-1
7	19	1	1	1
8	15	1	1	1
9	16	1	1	1
10	16	1	1	1
11	19	1	1	1
12	17	1	1	1
13	3	-1	-1	1
14	5	-1	-1	1
15	4	-1	-1	1
16	6	-1	-1	1
17	8	-1	-1	1
18	10	-1	-1	1

续表

被试	抑郁量表得分	性别（男=1，女=-1）	诊断结果（抑郁=1，非抑郁=-1）	性别×诊断结果
19	24	-1	1	-1
20	24	-1	1	-1
21	22	-1	1	-1
22	23	-1	1	-1
23	19	-1	1	-1
24	20	-1	1	-1

（1）传统的 2×2 设计的方差分析结果

直接给出SPSS分析的结果如表14-10。

表14-10　2×2设计的方差分析SPSS计算结果

变异来源	III 型平方和	*df*	均方	*F*	*Sig.*	偏 *Eta* 方
校正模型	1092.000[a]	3	364.000	65.000	.000	.907
截距	4056.000	1	4056.000	724.286	.000	.973
性别	24.000	1	24.000	4.286	.052	.176
诊断结果	1014.000	1	1014.000	181.071	.000	.901
性别×诊断结果	54.000	1	54.000	9.643	.006	.325
误差	112.000	20	5.600			
总计	5260.000	24				
校正的总计	1204.000	23				

a. *R* 方 = .907（调整 *R* 方 = .893）

（2）一般线性模型的分析结果

如前文所述，要构建如下模型：

$$抑郁量表得分=\beta_0+\beta_1\times性别+\beta_2\times临床诊断+\beta_3\times（性别\times临床诊断）+\varepsilon$$

首先就必须解决给性别、临床诊断以及二者交互作用的数量编码问题。对此，常用的有两个方法：一是对照编码（contrast codes），另一个哑变量编码（dummy codes）。两种编码结果的比较见表（14-11）。

表14-11　两种编码方法结果的比较

编码方法	性别		诊断结果	
	男	女	抑郁症	非抑郁症
对照编码（Contrast）	1	-1	1	-1
哑变量编码（Dummy）	1	0	1	0

虽然采用两种方法进行编码都是允许的，但是建议采用对照编码，特别是要进行交互作用检验时。因为采用对照编码方法时，如果各组样本容量相等，那么各自变量的和将等于 0。例如，12 位男性编码为“+1”、12 位女性编码为“-1”，那么 12+（-12）=0。显然，性别的平均分将等于 0，这等于是将变量作了中心化处理。同时，“(+1)×男+(-1)×女=0”，就是假设男性的得分等于女性，在检验 β_1 是否等于 0 就等于是在检验男、女性在抑郁量表上的得分是否存在显著差异。另外，采用对照编码，交互作用的编码将会是男性抑郁组和女性非抑郁组被编码为“+1”，而男性非抑郁组和女性抑郁组被编码为“-1”，那么在检验交互作用是否显

著时，等于是检验这两个“重新组合的组”的均值是否存在显著差异，考察β_3是否等于 0 就合乎逻辑了。如果采用哑变量编码，那么在检验交互作用时就不是“男性抑郁组和女性非抑郁组”与“男性非抑郁组和女性抑郁组”之间的比较，而变成了“男性抑郁组”与其他三组之间的比较了。

同样采用 SPSS 进行多元回归分析，得到模型的有关结果见表 14-12、表 14-13 和表 14-14。

表 14-12　回归分析模型概要表

模型	*R*	*R* 方	调整 *R* 方	标准估计的误差
1	.952	.907	.893	2.366

表 14-13　回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	1092.000	3	364.000	65.000	.000
	残差	112.000	20	5.600		
	总计	1204.000	23			

表 14-14　回归系数及其显著性检验结果

模型		非标准化系数		标准系数	*t*	*Sig.*
		B	标准误差	β		
1	(常量)	13.000	.483		26.913	.000
	性别	-1.000	.483	-.141	-2.070	.052
	诊断结果	6.500	.483	.918	13.456	.000
	性别×诊断结果	-1.500	.483	-.212	-3.105	.006

（3）一般线性模型方程系数的解释

对比表 14-10 和表 14-12、表 14-13、表 14-14，大家可以很容易看出两种分析方法得出了相同的结论，对此不再赘述。既然两种方法分析结果一样，为什么要两种方法呢？除了前文所说的一般线性模型提供了一个理解各种统计方法的整体框架、更加便于计算（矩阵运算）之外，还有一点就是，一般线性模型的结果能提供更加丰富的信息，而传统的方差分析不能提供。例如，在本例中，回归方程如下：

抑郁量表得分=13-1(性别)+6.5(诊断结果)-1.5(性别×诊断结果)+ε　　（方程 14-1）

方程中的回归系数描述了抑郁量表得分与每个自变量和交互变量之间的关系。对此进行逐一解释如下：

描述性别主效应的系数是“−1”。这表明，当性别的“取值”从较小值向较大值变化时，抑郁量表得分会降低。本例中，女性编码为“−1”，男性编码为“+1”，这也就是说，男性比女性有更低的抑郁量表得分。

描述诊断结果主效应的系数是“+6. 5”。这表明，当诊断结果的“取值”从较小值向较大值变化时，抑郁量表得分会提高。本例中，非抑郁症患者编码为“−1”，抑郁症患者编码

为“+1”，这也就是说，抑郁症患者比非抑郁症患才有更高的抑郁量表得分。

对交互变量系数的解释相对复杂一些。在方差分析一章，我们已经知道交互作用就是一个自变量与因变量的关系，在另一个自变量的不同水平上的表现不一样；为了考察交互作用，常常固定一个自变量的水平，然后来考察另一个自变量与因变量的关系，也就是简单效应分析。在此，为了考察性别对诊断结果与抑郁量表得分之间关系的影响，我们对方程 14-1 作一点小的变化：

抑郁量表得分=[13-1(性别)]+[6.5(诊断结果)-1.5(性别×诊断结果)] （方程 14-2）

对于重新安排的回归方程的各项，我们可以把诊断结果作为公因子提取出来：

抑郁量表得分=[13-1(性别)]+[6.5-1.5(性别)]诊断结果 （方程 14-3）

在方程(14-3)中，“**[13-1(性别)]**”描述的是方程的截距项——抑郁量表得分的初始预测值。在此我们称这为“合成截距”。为了解释这个“合成截距”，请注意这个“13”是总平均。如果把女性编码“-1”代入到“合成截距”，我们将得到女性组的抑郁量表的平均得分（13－1×(-1)=13+1=14)；代入男性的编码“+1”，我们将得到男性组的抑郁量表的平均得分（13－1×(+1)=13-1=12)。

“**[6.5-1.5(性别)]**”描述了诊断结果与抑郁量表得分之间的关系。如果把性别固定为其均值（由于我们使用了对照编码，其均值为 0)，我们会发现随着诊断结果“取值”的增大，抑郁量表得分也会增大。如果把诊断结果的可能取值代入方程，模型预测的抑郁组均值为（13+6.5×(+1)=13+6.5=19.5)，非抑郁组均值为（13+6.5×(-1)=13-6.5=6.5)。

现在“独立地”看看“**[6.5-1.5(性别)]**”这一项。该项表明诊断结果与抑郁量表得分的关系以 6.5 作为基础，并随着性别的变化而变化：男性组为 6.5-1.5×(+1)=5；女性组为 6.5+1.5×(-1)=8。这样，我们就得到了两个不同的回归系数去描述诊断结果与抑郁量表得分的关系：一个是男性组的系数 5；另一个是女性组的系数 8。从而，我们就得到了两个不同的回归模型：一个男性组的，另一个是女性组的。通过男性组的回归方程（方程 14-4）可见，截距为 12（男性组的总均值)，非抑郁组（编码为-1）的预测得分为 7（男性非抑郁组均值)，抑郁组（编码为+1）的预测得分为 17（男性抑郁组均值)；同样地通过女性组的回归方程（方程 14-5)，女性组均值为 14，女性非抑郁组均值为 6，女性抑郁组均值为 22。事实上，这也体现了采用对照编码方式的优势。

抑郁量表得分=12+5(男性组的诊断结果) （方程 14-4）

抑郁量表得分=14+8(女性组的诊断结果) （方程 14-5）

三、三个水平以上的单因素方差分析

【例 14-3】假定**【例 14-1】**的数据中，有 16 人有药物滥用情况（其中 8 人为酗酒，8 人为其他药物滥用)，另外 8 人无药物滥用情况，数据见表（14-9)。问酗酒组与其他药物滥用组的得分是否存在显著差异？无药物滥用组的得分与另两组均值是否存在显著差异？

表 14-15　24 名被试的抑郁测试得分

被试	抑郁量表得分	药物滥用情况	λ_1	λ_2
1	3	无（对照组）	0	-2
2	5	无（对照组）	0	-2
3	6	无（对照组）	0	-2
4	8	无（对照组）	0	-2
5	9	酗酒	-1	1
6	11	酗酒	-1	1
7	19	酗酒	-1	1
8	15	酗酒	-1	1
9	16	药物	1	1
10	16	药物	1	1
11	19	药物	1	1
12	17	药物	1	1
13	3	无（对照组）	0	-2
14	5	无（对照组）	0	-2
15	4	无（对照组）	0	-2
16	6	无（对照组）	0	-2
17	8	酗酒	-1	1
18	10	酗酒	-1	1
19	24	酗酒	-1	1
20	24	酗酒	-1	1
21	22	药物	1	1
22	23	药物	1	1
23	19	药物	1	1
24	20	药物	1	1

（1）传统的三水平单因素方差分析结果

直接给出SPSS分析的结果如表14-16和表14-17。

表 14-16　SPSS 方差分析结果

变异来源	III 型平方和	*df*	均方	*F*	*Sig.*	偏 *Eta* 方
校正模型	832.000[a]	2	416.000	23.484	.000	.691
截距	4056.000	1	4056.000	228.968	.000	.916
组别	832.000	2	416.000	23.484	.000	.691
误差	372.000	21	17.714			
总计	5260.000	24				
校正的总计	1204.000	23				

a. R 方 = .691（调整 R 方 = .662）

表 14-17　方差分析后逐对比较结果（S-N-K 法）

组别	N	子集	
		1	2
无	8	5.00	
酗酒	8		15.00
药物	8		19.00
Sig.		1.000	.071

（2）一般线性模型的分析结果

运用一般线性模型解决这个问题，关键问题仍在于给自变量进行编码，也就是构建设计矩阵。当自变量有三个或三个以上水平时，给自变量进行“编码”会稍微复杂一点。主要原因在于，给各组指定一个数值编码等于是强制规定了这些组之间的数值关系。如果只是两组，这是很简单的，我们给一组编码为“+1”，另一组为“-1”即可，因为总是一组均值更高，而另一组更低，这不存在次序出错的问题。但是当水平数在三个以上时，这个问题麻烦了。例如，使用一个简单编码给三个组（例如-1、0、+1），就等于强制一个组均值比另一组更低（或更高），第三组的均值则必须在这两组中间。这就可能出现次序的错误，比如，编码为“0”的中间组，事实上均值比另两组要高。

当自变量水平在三个以上时，需要创建 K-1 个预测变量（K 为水平数），如果是三个水平，则应创建 2 个预测变量；如果是四个水平，则需要 3 个预测变量。构建 K-1 个预测变量时，如果采用对照编码（推荐采用），编码必须符合两个条件：一是每个预测变量内的编码和等于 0；另一个是多个预测变量对应相乘得到的结果之和等于 0。对此，以本例的数据解释如下：

第一，本例数据有三个水平，需构建 2 个预测变量，每个变量内给“无”、“酗酒”、“药物”都要予以编码，这些编码加起来的和应等于 0。例如，我们设置第一个预测变量（记为 λ_1）为“酗酒=-1，无=0，药物=+1”，这时它们的和为 0；设置第二个预测变量（记为 λ_2）为“酗酒=+1，无=－2，药物=+1”，这时它们的和为 0。这样的编码有其“特殊作用”，如 λ_1 的编码实质上就是检验酗酒组与药物滥用组的均值是否存在显著差异，因为“(-1)×酗酒+0×对照+1×药物=0”，化简后即为“酗酒=药物”；同样的，λ_2 的编码是要检验“无滥用组”均值与“酗酒组”和“药物组”共同均值是否存在显著差异，因为“(+1)×酗酒+(-2)×无+(+1)×药物=0”，化简后为“2 无=酗酒+药物”，即“无=(酗酒+药物）/2”。因此，设置预测变量的编码，事实上充当了事先计划的组间比较，我们在工作中可以根据理论上或实践中的需要确定要进行哪些比较，在编码时予以考虑，事后比较的方法事实已不需要了。

第二，当我们“确定”好了两个预测变量的编码后，那么每个实验处理组就都有两个编码了：$\lambda_{1(无)}$ 和 $\lambda_{2(无)}$、$\lambda_{1(酗酒)}$ 和 $\lambda_{2(酗酒)}$、$\lambda_{1(药物滥用)}$ 和 $\lambda_{2(药物滥用)}$。如果把每个处理的两个预测变量代码相乘，就会得到每个组的第三个结果（见方程 14-6）。对照编码方法的第二个条件就是这第三个结果的和为 0。

$$\lambda_{1(无)}\lambda_{2(无)}+\lambda_{1(酗酒)}\lambda_{2(酗酒)}+\lambda_{1(药物滥用)}\lambda_{2(药物滥用)}=0 \quad （方程 14-6）$$

本例中，采用对照编码的结果见表 14-15；以 λ_1 和 λ_2 为自变量，以抑郁量表得分为因变量进行回归分析结果见表 14-18、表 14-19 和表 14-20。

表 14-18　回归分析模型概要表

模型	R	R 方	调整 R 方	标准估计的误差
1	.831	.691	.662	4.209

表 14-19　回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	832.000	2	416.000	23.484	.000
	残差	372.000	21	17.714		
	总计	1204.000	23			

表 14-20　回归系数及其显著性检验结果

模型		非标准化系数		标准系数	t	*Sig.*
		B	标准误差	β		
1	(常量)	13.000	.859		15.132	.000
	λ_1	2.000	1.052	.231	1.901	.071
	λ_2	4.000	.607	.799	6.584	.000

（3）传统方差分析与一般线性模型分析结果的对比

从表 14-16 至表 14-20 可见，两种分析方法结果也是一致的。与前文相似的结果不重复作比较说明，在此仅对逐对比较结果进行说明。方差分析的事后检验表明（见表 14-17），“无药物滥用”组的平均得分显著低于另外两组，而另外两组之间不存在显著差异。回归分析的结果与此是完全一致的（见表 14-20），λ_1的回归系数不显著（t = 1.901），说明“酗酒”组和“药物”组均值差异不显著；λ_2的回归系数显著（t = 6.584^{**}），说明“无滥用”组与另外两组的均值存在显著差异。

四、2×3 设计的方差分析

【例 14-4】假定得到 24 名被试的抑郁测试得分如表 14-21 所示。问性别、药物滥用情况的主效应是否显著？二者的交互作用是否显著？

表 14-21　24 名被试的抑郁测试得分

被试	抑郁量表得分	性别（男=1，女=-1）	药物滥用情况	λ_1	λ_2	性别×λ_1	性别×λ_2
1	3	1	无	0	-2	0	-2
2	5	1	无	0	-2	0	-2
3	6	1	无	0	-2	0	-2
4	8	1	无	0	-2	0	-2
5	9	1	酗酒	-1	1	-1	1
6	11	1	酗酒	-1	1	-1	1
7	19	1	酗酒	-1	1	-1	1
8	15	1	酗酒	-1	1	-1	1
9	16	1	药物	1	1	1	1
10	16	1	药物	1	1	1	1
11	19	1	药物	1	1	1	1
12	17	1	药物	1	1	1	1
13	3	-1	无	0	-2	0	2
14	5	-1	无	0	-2	0	2
15	4	-1	无	0	-2	0	2
16	6	-1	无	0	-2	0	2
17	8	-1	酗酒	-1	1	1	-1
18	10	-1	酗酒	-1	1	1	-1
19	24	-1	酗酒	-1	1	1	-1
20	24	-1	酗酒	-1	1	1	-1
21	22	-1	药物	1	1	-1	-1
22	23	-1	药物	1	1	-1	-1
23	19	-1	药物	1	1	-1	-1
24	20	-1	药物	1	1	-1	-1

（1）传统的 2×3 设计的方差分析结果

直接给出SPSS分析的结果如表14-22。方差分析后逐对比较的结果与**【例14-3】**一样，见表14-17。

表 14-22　SPSS 方差分析结果

变异来源	III 型平方和	*df*	均方	*F*	*Sig.*	偏 *Eta* 方
校正模型	884.000a	5	176.800	9.945	.000	.734
截距	4056.000	1	4056.000	228.150	.000	.927
性别	24.000	1	24.000	1.350	.260	.070
组别	832.000	2	416.000	23.400	.000	.722
性别×组别	28.000	2	14.000	.788	.470	.080
误差	320.000	18	17.778			
总计	5260.000	24				
校正的总计	1204.000	23				

a. R 方 = .734（调整 R 方 = .660）

（2）一般线性模型的分析结果

一般线性模型的 2×3 方差分析与**【例 14-2】**的 2×2 方差分析十分相似，唯一不同的是，“药物滥用情况”有三个水平，需要编码为两个预测变量（参见**【例 14-3】**），因此，考察性别与药物滥用情况的交互作用也将有两个交互作用，即性别×λ_1和性别×λ_2。各预测变量的编码结果见表 14-21，此时只要建立抑郁量表得分对这些预测变量的回归方程即可。

表 14-23　回归分析模型概要表

模型	*R*	*R* 方	调整 *R* 方	标准估计的误差
1	.857	.734	.660	4.216

表 14-24　回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	884.000	5	176.800	9.945	.000
	残差	320.000	18	17.778		
	总计	1204.000	23			

表 14-25　回归系数及其显著性检验结果

模型		非标准化系数		标准系数	*t*	*Sig.*
		B	标准误差	β		
1	(常量)	13.000	.861		15.105	.000
	性别	-1.000	.861	-.141	-1.162	.260
	λ_1	2.000	1.054	.231	1.897	.074
	λ_2	4.000	.609	.799	6.573	.000
	性别×λ_1	-.250	1.054	-.029	-.237	.815
	性别×λ_2	-.750	.609	-.150	-1.232	.234

（3）传统方差分析与一般线性模型分析结果的对比

方差分析结果表明，性别主效应不显著（F =1.350，p=0.260），组别（药物滥用情况分组）主效应显著（F =23.400***，p<0.001），二者的交互作用不显著（F =0.788，p=0.470）；组别主效应显著，说明三组的均值中至少有一对不等，究竟是哪些之间不等需要作事后检验的逐对均值比较，结果表明（见表 14-17），“酗酒”与“药物”组之间差异不显著，但二者与“无滥用药物”组存在显著差异。对照表 14-23、表 14-24、表 14-25 的结果，可以发现一般线性模型的分析结果与传统的方差分析结果完全一致。并且由于“编码”是事先设置好的，并不需要事后检验，λ_1系数不显著就说明了“酗酒”与“药物”组之间差异不显著，λ_2系数显著就说明“酗酒”与“药物”与“无滥用药物”组存在显著差异。交互作用不显著也体现在性别×λ_1和性别×λ_2的系数不显著。

第三节　一般线性模型的单因素协方差分析

本节首先介绍协方差分析的有关知识，接着结合实例介绍一般线性模型的单因素协方差分析的具体过程。

一、协方差分析概述

协方差分析(analysis of covariance)是将回归分析与方差分析结合在一起，对试验数据进行分析的方法。在协方差分析中，涉及到协变量（covariate）。

为了提高实验的精确性和准确性，对处理以外的一切条件都需要采取有效措施严加控制，使它们在各处理间尽量一致，这叫实验控制。但在有些情况下，即使做出很大努力也难以使实验控制达到预期目的。比如研究几种教学方法对学生学业成绩的影响，希望参与实验的学生的初始水平都相同。如果学生的初始水平不同，教学方法对学生的影响也可能会不同，这会影响到对因教学方法的改变而改变的学业成绩的评价。在这样的实验中，学生的初始水平会影响到实验结果，但不是实验关注的变量，又无法通过实验设计加以控制，只能以统计分析方法加以控制，此时，学生的初始水平就是协变量。

当实验中存在协变量时，如果不排除协变量的影响，将使研究的误差增大，影响研究结果。如果协变量（记为 X）与因变量(记为 Y)之间存在线性回归关系，又简单地假设不同组的协变量的平均值不相同，协变量和因变量的关系如图 14-1 所示（以三个处理为例）：

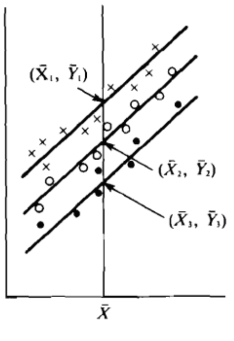


图 14-1 不同处理组的因变量对协变量（平均值相同）回归直线

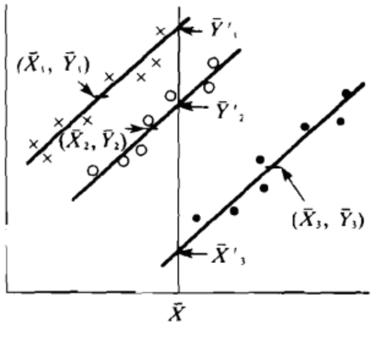


图 14-2 不同处理组的因变量对协变量（平均值不同）回归直线

从图 14-1 中可以看出，协变量的变异很大，会影响到对因变量的变异来源的分析。协变量并不是实验关注的内容，如果我们能排除因变量的变异中协变量的影响，能减少统计分析的误差，所得到的结果会更符合研究目的。这也正是协方差分析的重要任务。

当各处理组在协变量平均值上没有差异时，协方差分析只要简单地减少统计分析的误差；当各处理组在协变量平均值上有差异时，不仅要减少统计分析的误差，还要校正因变量的平平均值。如果协变量与因变量之间存在线性回归关系，又假设不同组的协变量的平均值不相同，协变量和因变量的关系如下图所示（以三个处理为例）：

图 14-2 将图 14-1 中两个处理沿 x 轴平移。三条回归线和 $X=\overline{X}$ 交汇的地方就是 $\overline{Y}_1^{'}, \overline{Y}_2^{'}$ 和 $\overline{Y}_3^{'}$。这些都是经校正的 Y 的平均数，即各处理在协变量上没有差异时，Y 的平均数。

若 Y 的变异主要由 X 的不同造成（处理没有显著效应），则各校正后的 Y' 间将没有显著差异（但原 Y 间的差异可能是显著的)。若 Y 的变异除掉 X 不同的影响外，还存在不同处理的显著效应，则可期望各 Y' 间将有显著差异（但原 Y 间差异可能是不显著的）。此外，校正后的 Y' 和原 Y 的大小次序也可能不一致。所以，在协方差分析中要进行处理平均数的回归校正和校正平均数的显著性检验。这些工作能够提高实验的准确性和精确性，从而更真实地反映实验实际。

协方差分析是实验数据分析中一个非常有用的工具。然而，用传统的方差分析方法进行协方差分析，计算非常繁琐，尤其协变量不只一个的时候。如果用一般线性模型的方式，协方差分析就变得非常简单，所用到的数据结构、统计过程和一般线性模型的方差分析差不多。

二、协方差分析模型的假设

除了方差分析中通常要求的总体正态和方差一致性的假设外，在协方差分析中还有另外两个假设：（1）因变量 Y 和协变量成线性相关（非线性关系也能处理，但本章不讨论这个）；（2）回归具有同质性。也就是说不同处理的回归系数相等（$\beta_1 = \beta_2 = \ldots = \beta_p$）。只有不同处理的回归系数相等，才能将不同处理的数据合并起来求一条回归线。不同处理的回归系数是否相等是可以检验的，具体的检验方法后文将会详细谈到。对于协变量的性质没有要求，协变量可以是固定变量，也可以是随机变量。

三、单因素协方差分析实例

【例 14-5】假定在**【例 14-2】**的数据中，加上了被试的年龄，数据见表（14-9）问性别对抑郁量表得分是否有显著影响？

表 14-26　24 名被试的抑郁测试得分

被试	抑郁量表得分	性别（男=1，女=-1）	年龄（*Covariate*）	性别×年龄
1	3	1	21	21
2	5	1	22	22
3	6	1	22	22
4	8	1	22	22
5	9	1	21	21
6	11	1	19	19
7	19	1	20	20

续表

被试	抑郁量表得分	性别（男=1，女=-1）	年龄（*Covariate*）	性别×年龄
8	15	1	19	19
9	16	1	18	18
10	16	1	17	17
11	19	1	21	21
12	17	1	22	22
13	3	-1	23	-23
14	5	-1	24	-24
15	4	-1	25	-25
16	6	-1	23	-23
17	8	-1	22	-22
18	10	-1	21	-21
19	24	-1	18	-18
20	24	-1	18	-18
21	22	-1	19	-19
22	23	-1	20	-20
23	19	-1	20	-20
24	20	-1	21	-21

（一）一般线性模型下的分析方法

使用一般线性模型进行协方差分析一般包括如下几个步骤：

（1）检验不同处理组的协变量与因变量的关系是否是线性的，以及各处理组的因变量对协变量的回归系数是否同质（也就是各组的回归斜率是否相等，亦即各组的回归线是否平行）。这个步骤也是采用传统的协方差分析时必须做的，是能否进行协方差分析的基本前提条件的检验。

（2）依据不同的自变量建立对因变量预测的不同回归模型，再通过不同回归模型对总变异解释比例的差异来分解各自变量对因变量的影响效应（参见表 14-27）。

表 14-27　分解协变量与自变量处理效应的模型

模型类别	**【例 14-5】**示例	备注
模型 A：包含所有自变量和协变量的回归模型	抑郁得分= $\beta_0+\beta_1\times$性别+ $Z_1\times$年龄+ ε	以“*Z*”作为年龄的系数，表示“年龄”是协变量
模型 B：包含所有自变量的回归模型	抑郁得分= $\beta_0+\beta_1\times$性别+ ε	用模型 B 减去模型 A 不能解释的平方和，就可得到单纯的年龄能解释的平方和
模型 C：包含所有自变量和协变量的回归模型	抑郁得分= $\beta_0+Z_1\times$年龄+ ε	用模型 C 减去模型 A 不能解释的平方和，就可得到单纯的性别能解释的平方和

（3）计算各处理校正均值及其差异显著性检验。也就是，计算排除协变量影响之后的各处理的平均数，以及重新计算的平均数（校正均值）是否存在显著差异。

下面就以【**例 14-5**】数据为例，对上述三个步骤过程进行演示：

1. 检验回归系数是否同质

对于不同处理组的协变量与因变量的关系是否是线性的，可以采用散点图等方式进行检验，具体参加本书前面的章节，在此不再赘述。

对每个处理的数据单独求回归方程，如果这些回归线都是平行的，就说明这些回归线的斜率都是同质的，那么在统计过程中不考虑交互效应数据可解释的变异的百分比只会减少一点点。如果这些回归线不平行，也就是说这些回归线的斜率不同质的，那么在统计过程中不考虑交互效应则数据可解释的变异的百分比会显著减少，此时就不宜采用协方差分析，而必须对资料做一定处理以满足回归系数同质的条件，或者选用其他方法分析了。

检验回归系数是否同质，可以通过检验考虑交互效应和不考虑交互效应两种模型下，数据可解释变异百分比的差异是否显著来进行，也可以直接考察交互效应预测变量的回归系数是否显著。第一种方法需要手工计算，后一种方法用 SPSS 可直接给出结果，本文采用第二种方法（见表 14-28）。从表中可见，交互效应项“性别×年龄”的回归系数不显著（t =1.965，p=0.063），符合协方差分析的模型假设要求。

表 14-28　回归系数及其显著性检验结果

模型		非标准化系数		标准系数	*t*	*Sig.*
		B	标准误差	*β*		
1	(常量)	65.761	10.144		6.483	.000
	性别	-22.016	10.144	-3.108	-2.170	.042
	年龄	-2.523	.490	-.707	-5.155	.000
	性别×年龄	.962	.490	2.831	1.965	.063

2. 建立回归方程并计算自变量和协变量对总变异解释的比例

采用回归分析方法建立“三类”模型如下：

（1）模型 A（包含所有自变量和协变量）：抑郁得分=70.886-2.162×性别-2.790×年龄+ε

表 14-29　模型 A 回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	726.996	2	363.498	16.003	.000
	残差	477.004	21	22.714		
	总计	1204.000	23			

（2）模型 B（仅含所有自变量，不含协变量）：抑郁得分=13-1×性别+ε

表 14-30　模型 B 回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	24.000	1	24.000	.447	.511
	残差	1180.000	22	53.636		
	总计	1204.000	23			

（3）模型 C（不含自变量，仅含协变量）：抑郁得分=66.138-2.561×年龄+ε

表 14-31　模型 C 回归方程有效性检验的方差分析表

模型		平方和	*df*	均方	*F*	*Sig.*
1	回归	619.725	1	619.725	23.335	.000
	残差	584.275	22	26.558		
	总计	1204.000	23			

模型 A 与模型 C 对比，模型 A 能解释的平方和为 726.996；模型 C 少了性别因素，能解释的平方和为降 619.725，这意味着性别因素能有效解释的平方和为 726.996-619.725=107.271；同理，可得年龄能有效解释的平方和为 726.996-24.000=702.996。至此，我们就有了完成协方差分析表的所有信息。协方差分析表的一般格式如下（表 14-32）：

表 14-32　协方差分析表的一般格式

变异来源	平方和	自由度	均方	*F* 值
协变量	$SS_{回归(模型A)}-SS_{回归(模型B)}$	c	$SS_{协变量}/c$	$MS_{协变量}/MS_{残差}$
校正后处理	$SS_{回归(模型A)}-SS_{回归(模型C)}$	k-1	$SS_{校正后处理}/(k-1)$	$MS_{校正后处理}/MS_{残差}$
残差	$SS_{残差(模型A)}$	N-k-c	$SS_{残差}/(N-k-c)$	
总体	SST	N-1		

表中 c 指的是协变量的个数，k 为自变量水平数，N 为被试数。将已获得的信息填入下表，即可完成协方差分析（表 14-33）。经查 F 分布表可知，协变量（年龄）对抑郁得分的影响显著（F=30.949，p<0.001）；在排除年龄影响之后，性别对抑郁得分的影响也显著（F=4.723，p<0.05）。

表 14-33　性别对抑郁得分影响的协方差分析表

变异来源	平方和	自由度	均方	*F* 值
协变量（年龄）	702.996	1	702.996	30.949***
校正后处理（性别）	107.271	1	107.271	4.723*
残差	477.004	21	22.714	
总体	1204.000	23		

3. 计算各处理校正均值及其差异显著性检验

把协变量均值与自变量各处理的取值代入模型 A（抑郁得分=70.886-2.162×性别-2.790×年龄+ε），即可得到男、女性的抑郁得分校正均值：

男性抑郁得分校正均值=70.886-2.162×(+1)-2.790×20.75=10.83

女性抑郁得分校正均值=70.886-2.162×(-1)-2.790×20.75=15.16

得到两组的校正均值后，即可用公式（14-9）计算得到 F 值：

$$F_{(1,N-a-1)} = \frac{(\overline{Y}_j^{'} - \overline{Y}_k^{'})^2}{MS_{残差}^{'}\left[(\frac{1}{n_j}+\frac{1}{n_k})+\frac{(\overline{C}_j-\overline{C}_k)^2}{SS_{e(c)}}\right]} \qquad 公式 14\text{-}9$$

式中，N 指的是被试数，a 为自变量水平数；$\overline{Y}_j^{'}$ 和 $\overline{Y}_k^{'}$ 是要进行差异显著性检验的两组校正均值，n_j 和 n_k 分别为两组的样本容量，$\overline{C}_j$ 和 $\overline{C}_k$ 分别为两组对应的协变量均值；$MS_{残差}^{'}$ 指的是协方差分析中的残差均方；$SS_{e(c)}$ 指的是把协变量作为因变量进行方差分析时的误差平方和（在本例中，就是以年龄为因变量，性别为自变量进行方差分析，所得的误差平方和）。把**【例 14-5】**相关数据代入公式（14-9），得

$$F_{(1,24-2-1)} = \frac{(10.83-15.16)^2}{22.714\left[(\frac{1}{12}+\frac{1}{12})+\frac{(20.33-21.17)^2}{90.333}\right]} = \frac{18.7489}{3.9631} = 4.73$$

本附表 5-A（F 分布的临界值表），得 $F_{0.05,(1,24-2-1)}$ =4.32，因为 F =4.74> $F_{0.05,(1,24-2-1)}$ =4.32，因此认为两个处理的校正均值存在显著差异。

（二）应用 SPSS 进行协方差分析的方法

如果用传统的方差分析的方法，进行平方和分解计算，协方差分析的计算过程将会很复杂，特别是当协变量个数增多时。好在通过一般线性模型的讲解，大家对协方差分析的原理有了认识，此时，我们只要通过软件得到结果即可。用 SPSS 进行协方差分析十分简便，下文作一简介。

在 SPSS 中，通过“一般线性模型”命令，可以完成方差分析、多元方差分析（指的是多个因变量，参见第 15 章）、重复测量方差分析等分析过程，见图 14-3。

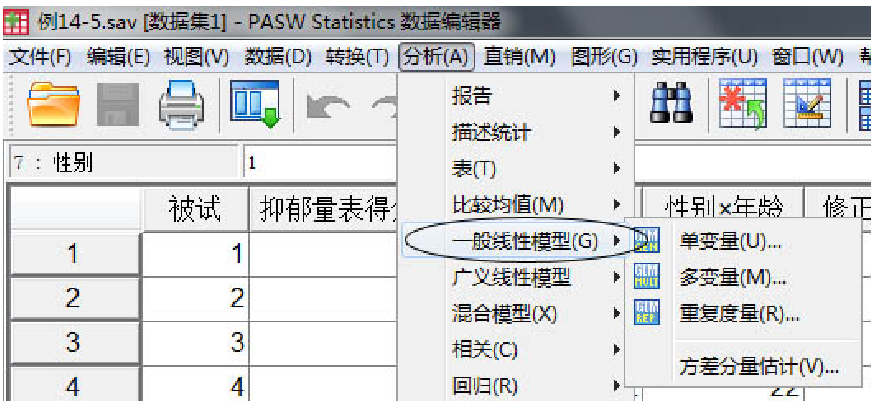


图 14-3　SPSS 一般线性模型分析命令图示

1. 检验回归系数是否同质

在 SPSS 中检验回归系数是否同质，只需在进行正式分析之前，采用“设定”模型规定模型包括性别、年龄的主效应，以及二者的交互作用，进行一次“预分析”即可（如图 14-4 所示，请注意椭圆框住的部分）。如果二者交互作用不显著，则可认为回归系数是同质的，可以进行正式分析了。如果交互作用显著，则要另想办法了。在本例中，结果表明回归系数是同质的，因为性别与年龄的交互作用不显著（F=3.863，p<0.063；与一般线性模型的分析结果一样，见表 14-28）。

图 14-4　SPSS 协方差分析全模型“预分析”图示

2. SPSS 进行协方差分析的方法

SPSS 进行协方差分析与刚才讲的“预分析”过程基本相同，唯一需要注意的是在模型设定时不要包含自变量与协变量交互效应项。在 SPSS18.0 中，可以指定模型为“全因子”，因为软件默认的是没有交互效应项的；也可以“设定”模型仅包含主效应即可（见图 14-5）。分析结果见表 14-34，从表中可见，性别、年龄对抑郁量表得分都有显著影响，注意，此时性别的影响是排除了年龄因素后的“影响”。至于校正均值的比较与一般的方差分析过程中均值比较的操作方法相似，在此不重复叙述。

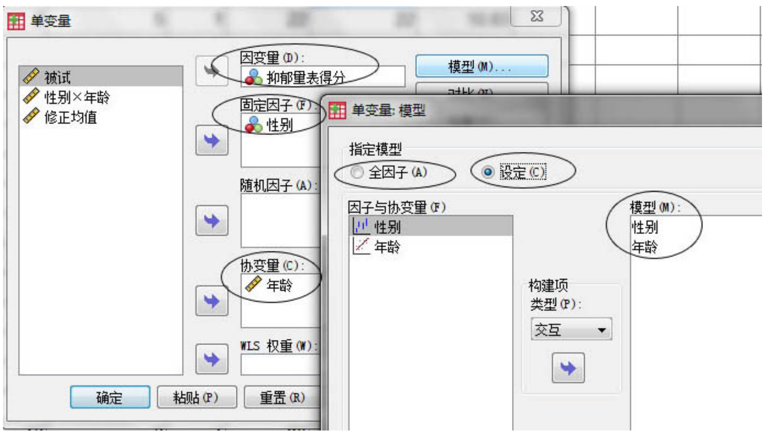


图 14-5　SPSS 协方差分析图示

表 14-34　性别对抑郁得分影响的协方差分析表

变异来源	平方和	*df*	均方	*F*	*Sig.*
校正模型	726.996	2	363.498	16.003	.000
截距	1045.076	1	1045.076	46.009	.000
性别	107.271	1	107.271	4.723	.041
年龄	702.996	1	702.996	30.949	.000
误差	477.004	21	22.714		
总计	5260.000	24			
校正的总计	1204.000	23			

四、协方差分析与方差分析的比较

对于同一批数据，做方差分析和做协方差分析结果是不一样的。对照同一批数据的方差分析的结果和协方差分析的结果能清楚看出这种差异。对**【例 14-5】**的数据，表 14-30 是对数据仅做方差分析的结果，表 14-33 是对数据做协方差分析的结果。将表 14-30 和表 14-33 对照一下可以发现：首先，两种分析的残差项有变化。协方差分析的残差项减少很多，这是因为由协变量导致的变异被排除了；其次，两种分析的处理效应不一样。这是因为在协方差分析中，由协变量的平均差异导致的因变量的差异得到了校正。而在方差分析中并没有这种校正。

当研究中涉及到多个协变量时，如果在方差分析的框架下进行协方差分析计算量就非常大，如果在一般线性模型的框架中进行协方差分析计算就很简单。如，用 $R^2{}_{c,\alpha,\beta,\alpha\beta,\beta}$ 来表示一批的预测变量和协变量所能解释的变异。β 用来表示一系列的预测变量（$B_1, B_2, \cdots B_b$），c 可以用来代表一系列的协变量（$c_1, c_2, \cdots c_k$）。例如，研究不同教学方法的教学效果是否存在显著差异，可以将学生的智商（IQ）、年龄（Age）和学校类型（如保守的和激进的）(School) 作为协变量，那么 $R^2{}_{c,\alpha,\beta,\alpha\beta,\beta}$ 就可以写为：

$$R^2{}_{IQ,Age,School,A_1,B_1,B_2,AB_{11},AB_{12}}$$

这样来看的话，多元协方差分析和单元协方差分析没什么区别，c 代表学生的智商、年龄和学校类型，则 $SS_{AB}(adj)$ 为：

$$SS_{\gamma(R^2_{c,\alpha,\beta,\alpha\beta}-R^2_{c,\alpha,\beta})} = SS_{reg\,(IQ,Age,School,A_1,B_1,B_2,AB_{11},AB_{12})} - SS_{reg\,(IQ,Age,School,A_1,B_1,B_2)}$$

尽管如果在一般线性模型的框架中进行协方差分析计算很简单，但是有一个问题，就是关于协方差分析结果的解释。其实在协方差分析中，只有一个协变量时，协方差分析结果的解释已经很难，当协变量增加时，这种难度迅速变大。当要评价不同教学方法的优劣时，被试组在智商、年龄、学习类别等条件上都是匹配的是最好，但这样的实验条件很难获得。

小　结

t 检验、方差分析和多元线性回归等分析方法本质上是相同的，都可以统一到一般线性模型的框架下。将方差分析和多元线性回归等众多统计方法统一到一般线性模型框架下的一个重要工具是设计矩阵，也就是要把一些变量重新编码构建成预测变量。

一般线性模型下的分差分析的过程包括两个步骤，一是构造设计矩阵，二是进行回归分析。本文对构造设计矩阵中关键的“对照编码”技术进行了详细讲解。

进行协方差分析的目的是消除协变量对研究的影响。一般线性模型的单因素协方差分析要注意设计矩阵的构造。和一般线性模型下的单向分差分析的设计矩阵相比，一般线性模型的单向协方差分析的设计矩阵多了几列用来检验不同处理组的回归系数是否同质的数据。并要进行多次回归，以获得完成协方差分析表的数据。协方差分析中可以计算出各实验处理校正后平均数并进行校正后平均数的差异显著性检验。

一般线性模型下的协方差分析和方差分析相比，处理效应和误差项都不一样。当研究中涉及到多个协变量时，如果在方差分析的框架下进行协方差分析计算量就非常大，如果在一般线性模型的框架中进行协方差分析计算就很简单。协变量多时，协方差分析结果难以解释。

关键术语

一般线性模型：统计分析的对象是统计资料，当资料中包含着自变量 X 和连续变化的因变量 Y 时，可用最简便的方式即一般线性模型 $Y = X\beta + e$ 来描述因变量与自变量之间的依存关系。其中，Y 为因变量的观测值向量，X 为由自变量构造的设计矩阵，β 为回归参数向量，e 为正态独立随机误差向量，并假定其平均数 E（e）=0，协方差矩阵为 Ω＝Cov（e）。当设计矩阵 X 和误差的协方差矩阵 Ω 具有各种不同结构时，一般线性模型就会有各种不同的变形，包括线性回归分析模型、方差分析模型、协方差分析模型等。

方差分析模型：若定义向量 τ' 为 $[\mu \quad \tau_1 \quad \tau_2 \quad \tau_3]$，方差模型可以记为：$X_\tau + e$。

设计矩阵：为了将方差分析模型和一般线性模型联系起来，需要构造一个设计矩阵。设计矩阵 X 有 $(p+1)$ 列，代表着总平均效应 μ 和 p 个处理效应。设计矩阵 X 中的每一列，用 1 和 0 代表某被试的成绩有没受到该列的影响，1 表示受到影响，0 表示没受到影响。在设计矩阵 X 中，每个被试在总平均效应（μ）所在的列都为 1，因为 μ 是所有观测值的一部分。每个被试只在他所受到影响的列上（或处理上）值为 1，在其他列上都是 0。

协方差分析：协方差分析是将回归分析与方差分析结合在一起，对试验数据进行分析的方法。

协变量：为了提高实验的精确性和准确性，对处理以外的一切条件都需要采取有效措施严加控制，使它们在各处理间尽量一致，这叫实验控制。但在有些情况下，即使做出很大努力也难以使试验控制达到预期目的。比如研究几种教学方法对学生学业成绩的影响，希望参与实验的学生的初始水平都相同。如果学生的初始水平不同，教学方法对学生的影响也可能会不同，这会影响到对因教学方法的改变而改变的学业成绩的评价。在这样的实验中，学生的初始水平会影响到实验结果，但又不是实验关注的变量，学生的初始水平就是协变量。

重要公式

一般线性模型：$Y = X\beta + e$

一般线性模型下协方差分析检验不同处理组的回归是否同质的检验统计量：

$$F_{(f-r,N-f-1)} = \frac{(N-f-1)(R^2_{\tau,c,\tau c} - R^2_{\tau,c})}{(f-r)(1-R^2_{\tau,c,\tau c})}$$

一般线性模型下协方差分析中，各处理变量能解释的因变量的变异：

$$SS_{\text{treat(adj)}} = SS_{\text{regression}(\tau,\text{c})} - SS_{\text{regression(c)}}$$

一般线性模型下协方差分析中，各处理变量能解释的因变量的变异：

$$SS_{\text{covariate}} = SS_{\text{regression}(\tau,\text{c})} - SS_{\text{regression}(\tau)}$$

校正后各处理平均数差异显著性检验的检验统计量：

$$F_{(1,N-a-1)} = \frac{(\overline{Y}'_j - \overline{Y}'_k)^2}{MS'_{error}\left[(\frac{1}{n_j} + \frac{1}{n_k}) + \frac{(C_j - C_k)^2}{SS_{e(c)}}\right]}$$

思考与练习

1. 如何理解一般线性模型和回归分析、方差分析以及协方差分析的关系。

2. 一般线性模型下方差分析的过程

3. 什么是协变量，为什么要进行协方差分析。

4. 一般线性模型下协方差分析的过程。

5. 协方差分析和方差分析有何不同。

6．某实验有三个实验处理（B1、B2、B3），实验中的某因素（C）可能会对实验结果有影响，实验的因变量为（Y），实验数据列入下表中：

B1		B2		B3	
C	Y	C	Y	C	Y
8	5	10	15	11	28
10	12	13	10	8	26
12	16	5	26	6	13
15	28	17	23	5	12
8	8	7	7	11	32
10	16	10	23	7	30
13	18	12	25	5	17
16	27	14	36	4	17
11.5	16.25	12.25	20.625	7.125	21.875

请对这批数据完成以下任务：

(1) 进行一般线性模型下的方差分析；

(2) 进行一般线性模型下的协方差分析。

第十五章 实用多元统计分析方法简介

学习本章内容，将有助于你对以下问题的理解与思考：

1．应如何使用多元方差分析方法来考察不同类别的多个自变量同时对多个因变量是否存在显著影响？

2. 如何使用典型相关方法来考察一批变量与一批变量之间的相关？

3．如何用聚类分析方法来按照描述分类对象所用的变量中空间关系的亲疏程度对其进行分类？

4．如何通过因素分析方法对变量或个体之间相关系数矩阵的考察，寻找几个新的变量（潜变量或者因子）以再现这种变量之间或个体之间的内部关系？

参考文献

[1] 张厚粲， ， 伯 .心理与教育统计学[M].北京：北京师范大学出版社，1988.

[2] 张敏强.教育与心理统计学[M].（第 3 版）.北京：人民教育出版社，2010.

[3] 甘怡群,等.心理与行为科学统计[M].北京：北京大学出版社，2005.

[4] 舒华，张亚 .心理学研究方法[M].北京：人民教育出版社，2008.

[5] 盛骤,谢式千， 承 . 概率论与数理统计[M].第 3 版. 北京：高等教育出版社，2001.

[6] 范大 ，陈永华. 概率论与数理统计[M].修订版. ： 江大学出版社，2003.

[7] 华中科技大学数学系. 概率论与数理统计[M].第 2 版.北京：高等教育出版社，2003.

[8] 胡竹菁.平均数差异显著性检验统计检验力和效果大小的估计原理与方法.心理学探新, 2010,30(1),68-73.

[9] 胡竹菁,戴海琦.方差分析的统计检验力和效果大小的常用方法比较.心理学探新,2011,31(3),254-259.

[10] 胡竹菁,董圣鸿，张阔.《心理统计学》教学内容的新探索.心理学探新, 2013,33(5),402-408.

[11] 美国心理协会.美国心理协会写作手 [M]. 第 5 版. 陈 玲，等译. 重 ：重 大学出版社，2008.

[12] Cohen, B. H. Explaining Psychological Statistics[M].4rd Edition. John Wiley & Sons, Inc., 2013.

[13] Aron, A. Coups, E. J.Aron, E. N. Statistics for Psychology[M]. 6rd Edition. Pearson Education, Inc.，2013.

[14] Howell, D. C. Fundamental statistics for the behavioral science[M]. 7rd Edition. Wadsworth, Cengage Learning. 2011.

[15] Runyon, R. P. Fundamentals of behavioral statistics[M]. 北京：人民 电出版社，2004.

[16] Devore, J. L. Probability and statistics[M]. 北京：高等教育出版社，2004.

[17] Cohen, J. A power primer[J]. Psychological Bulletin, 1992, 112: 155–159.

[18] Cohen, J. The earth is round (p < .05) [J] . American Psychologist, 1994, 49: 997–1003.

[19] American Psychological Association. Publication Manual of the American Psychological Association[M]. 6rd Edition. American Psychological Association, 2010.

当我们在进行某项社会调查后，得到了有关编号（id）、性别（gender）、出生日期（bdate）、受教育水平（educ）、工作类别（jobcat）、现在薪金（salary）、起薪（salbegin）、工作时间（jobtime）、以前的工作经验（prevexp）和是否少数族裔（minority）等变量的信息，我们应如何通过多元统计分析方法来分析其中某些自变量对另外一些因变量的影响作用？

第一节 多元方差分析

前面所讲过的大多数统计方法，都是用一个变量来描述个体，尽管有一些统计方法如相关和回归能够探讨两个或多个变量之间的关系，但是因变量还是只有一个。实际上在日常生活或者研究领域中，往往用多个变量来描述一个个体，例如人的运动能力包括跳高、跳远、引体向上、推铅球、百米跑等。要对这些个体或者变量之间的关系进行描述和推断，就涉及到同时操作多个变量的方法，在心理统计学中，研究多个自变量与因变量相互关系的统计理论和分析方法被称为多元统计分析（multivariate statistical analysis，又译为“多变量统计分析”，简称为多元统计或多元分析）。

与前面初等统计相似，针对变量的不同水平，以及要解决不同类型的统计问题（描述变量之间的关系强弱，考察组间差异的显著性，探索变量之间的内部结构，根据已知的信息对其它变量或个体观测值的情况进行预测等等），可以选用不同的多元统计方法。多元统计方法有很多而且还在不断发展，本章只选用多元方差分析，典型相关分析，聚类分析以及因子分析等几个比较常用的方法作简单介绍，以便读者初步了解这些方法。与前面所讲的大多数统计方法不同，多元统计方法涉及的变量较多，往往都是需要用向量和矩阵来描述的，而且其计算量很大，无法通过简单的计算理解，需要统计软件帮助求解。为了帮助大家尽快熟悉和了解多元统计方法的应用，本章尽量少使用数学公式，而主要介绍各种方法所能够解决的问题，以及如何利用统计软件 SPSS 实现并分析其结果。为了方便，本章中的所有例子都是采用 SPSS 安装之后自带的数据文件“employee data.sav”中的同一批数据，选用其中一部分变量进行分析，同时为了使大家能够对这些统计方法和 SPSS 的实现有更深的认识，在文中还对同一批数据提出了和例子中相差不大的问题要求读者自己动手。

本章各节所介绍的内容都是采用 SPSS 中文版 18.0 运行的结果，不同版本的 SPSS 运行结果的形式可能会有细微的不同。在 SPSS 中文版 18.0 中，与本文开头所述各种数据对应的编码有如图 15-1 所示（括号中的英文词组是 SPSS 英文版中的相应编码）：

	员工代码	性别	出生日期	教育水平	雇佣类别	当前薪金	起始薪金	雇佣时间	经验	少数民族
1	1	m	02/03/1952	15	3	$57,000	$27,000	98	144	0
2	2	m	05/23/1958	16	1	$40,200	$18,750	98	36	0
3	3	f	07/26/1929	12	1	$21,450	$12,000	98	381	0

图15-1 本章例题各变量编码图

一、多元方差分析的含义与要求

多元方差分析（Multivariate Analysis of Variance，MANOVA）与前面的多因素方差分析不同，它是处理多个因变量的方差分析方法，是同时考察不同类别的自变量在多个因变量上是否存在差异的统计方法。比如要研究在课本上所花的时间对学习成绩的影响，而学生的阅读以及数学都可以作为学习成绩的合适测量。这种方法是由 Wilk 于 1932 年在《Biometrika》中发表了《Certain Generalization in the Analysis of Variance》一文首先提出的。多元方差分析是一般的方差分析在多因变量情况下的拓展。尽管我们可以对于阅读和数学分别作一般的方差分析，但是在因变量之间有相关时，单因变量的结果可能无法推广到多因变量的情况。

多元方差分析能够解决的问题是自变量对因变量的影响是否显著，也可以解决因变量之间的交互作用以及自变量之间的交互作用（J.P.Stevens，2002).

多元方差分析对于数据的要求与一元的方差分析相类似：独立性，总体正态分布以及方差齐性。每个组的个体观察值都是从相应总体中随机抽样得到的，因此各观测值之间是相互独立的。一般来说，独立性假设在横贯性研究中都是可以满足的，但是这也只是在实验设计过程中凭经验判断，而很难通过统计方法来检验。多元方差分析的要求与一元方差分析不同的是，不仅针对各个因变量的总体分布都必须是正态的，而且要求因变量之间的联合分布也是正态的。在多元方差分析中方差齐性是指对应自变量的每个不同类别，因变量分布的总体方差相等，对应每个自变量，任何两个因变量之间的协方差也同质。多元方差分析中方差齐性假设有多种统计方法进行检验，一般来说，多元方差分析的结果对于该假设有一定的稳健性，尤其是在样本容量均衡时。除了要满足这几个理论假设之外，多元方差分析的数据还必须满足自变量是一个或多个分类变量，有两个或多个连续的自变量，而且因变量之间有一定程度的相关。在样本上也有一定的要求，不仅要求样本总量有一定的规模，而且各个分组的样本规模差别最好也不大。

多元方差分析的许多概念与一元方差分析基本一样，只是同时考察一个或多个自变量在多个因变量上的差异。最简单的一个自变量两个因变量的单因素二元模型，相当于 t 检验在多维中的推广，实际上就是 Hotelling T^2 检验，其在多元方差分析中的操作与多因素的二元模型并无区别。本节选用两个自变量两个因变量的双因素二元模型（自变量间有交互作用的饱和模型）作为例子，以方便读者了解多元方差分析的一般操作步骤，多因素（有交互作用的和无交互作用的）多元模型的例子可参见其它统计书籍。

二、利用 SPSS 软件包求解多元方差分析的实例

【例 15-1】 以 SPSS 自带数据 employee data.sav 为例，本例要求解的问题是：“起始薪金和受教育年限”在“不同的性别”与“工作类别”中是否有差异？”（文件中受教育年限是定义为等级变量的，但因为其取值较多，且可以当成等距变量，因此我们在本例中将其作为连续变量处理）。

运用 SPSS18.0 软件包，在已打开的数据文件窗口中，根据下列操作步骤（如图 15－2 所示）可以求解本例的多元方差分析结果（请注意：因为“起始薪金”是严重的正偏态分布，将其转换成自然对数后与正态分布的偏差更小。下面的分析用的是原始数据）。

1.分析→一般线性模型→多变量
2.因变量框：起始薪金、教育水平
3.固定因子框：性别、雇佣类别
4.选项：（1）方差齐性检验
（2）继续
5.确定

图 15-2 在 SPSS 中设置多元方差分析程序各步骤示意图

三、多元方差分析的结果解释

通过 SPSS 软件包进行分析后，得到的结果摘选分析如下（表 15-1）：

（一）各因变量之间的方差协方差齐性分析

表15-1 协方差矩阵齐性检验表

Box’s M	433.615
F	47.462
df1	9
df2	66185.172
Sig.	.000

Box’s M 用 F 分布检验各因变量之间的方差协方差矩阵齐性。其虚无假设是协方差矩阵无显著差异，当伴随概率（sig.）大于 0.001 时多元方差分析的前提假设满足。不过需要指出的是，Box’s M 对于是否正态分布前提非常敏感。但是在样本容量均匀时，MANOVA 对于该假设具有稳健性。

（二）多元方差分析结果

表15-2 多元方差分析表（Multivariate Tests(c)）

效应		值	F	假设 df	误差 df	Sig.
截距	Pillai 的跟踪	.923	2812.269[a]	2.000	468.000	.000
	Wilks 的 Lambda	.077	2812.269[a]	2.000	468.000	.000
	Hotelling 的跟踪	12.018	2812.269[a]	2.000	468.000	.000
	Roy 的最大根	12.018	2812.269[a]	2.000	468.000	.000
性别	Pillai 的跟踪	.158	44.028[a]	2.000	468.000	.000
	Wilks 的 Lambda	.842	44.028[a]	2.000	468.000	.000
	Hotelling 的跟踪	.188	44.028[a]	2.000	468.000	.000
	Roy 的最大根	.188	44.028[a]	2.000	468.000	.000
雇佣类别	Pillai 的跟踪	.509	80.073	4.000	938.000	.000
	Wilks 的 Lambda	.511	93.477[a]	4.000	936.000	.000
	Hotelling 的跟踪	.920	107.409	4.000	934.000	.000
	Roy 的最大根	.876	205.420[b]	2.000	469.000	.000
性别 * 雇佣类别	Pillai 的跟踪	.063	15.614[a]	2.000	468.000	.000
	Wilks 的 Lambda	.937	15.614[a]	2.000	468.000	.000
	Hotelling 的跟踪	.067	15.614[a]	2.000	468.000	.000
	Roy 的最大根	.067	15.614[a]	2.000	468.000	.000

a. 精确统计量

b. 该统计量是 F 的上限，它产生了一个关于显著性级别的下限。

c. 设计 : 截距 + 性别 + 雇佣类别 + 性别 * 雇佣类别

这部分是多元方差分析的结果。SPSS 的运行结果提供了四个多元统计量：Pillai’s Trace，Wilks’ Lambda，Hotelling’s Trace，Roy’s Largest Root。这些不同的统计量都可以检验自变量“性别”，“工作类别”以及二者交互作用“性别*工作类别”在两个因变量“起始薪金”和“受教育水平”上的效应。这几个统计量尽管采用不同的公式计算，但是都服从 *F* 分布，因此都给出了 *F* 值以及相应的伴随概率。虽然这四种方法的值可能有差异，但是其结论的显著性一般来说都是一致的，即要么都显著，要么都不显著。

在实际的结果报告中，Hotelling’s Trace 在自变量只有两类时较为常用，Wilk’s Lambda 方法的 *F* 检验结果是精确值，且具有针对假设违背时的稳健性，也是个常用的统计量。而Pillai’s Trace检验相对更为保守，更容易保留无差异假设，且在样本规模小，各组样本规模不等以及各组方差不齐时稳健性更好。故而在方差不齐的情况下 Pillai’s Trace 方法用的较多。Roy’s Largest Root 检验在因变量可以用一个维度很好地表示时的统计检验力最强。

从表 15-2 中的结果可以看出：性别，工作类型以及两者的交互作用对于薪水和受教育水平两个因变量都是具有显著的影响的。或者说，不同性别，不同工作类型以及两者的不同结合时，薪水与受教育水平的某个线性组合新变量上的取值不全相等。但是和一元方差分析一样，得到 F 检验显著的结果只是说明至少两种自变量的不同分类所对应的因变量取值不同，具体是哪些分类不同还需要做成对比较。性别变量只有两种取值不需要再做成对比较。

表15-3 多元方差分析的方差齐性表（Levene's Test of Equality of Error Variances）

	F	*df1*	*df2*	*Sig.*
起始薪金	38.694	4	469	.000
教育水平（年）	6.085	4	469	.000

Levene 也是多元方差分析的假设检验。它的虚无假设是每个因变量在各个自变量的分组中是否方差齐性。由表 15-3 可以看到，对于受教育水平和起始薪金两个因变量，其伴随概率都小于 0.05，说明多元方差分析的方差齐性假设被拒绝。

（三）多元方差分析表

表15-4 组间效应检验

源	因变量	III 型平方和	*df*	均方	*F*	*Sig.*
校正模型	起始薪金	1.995E10	4	4.988E9	250.307	.000
	教育水平（年）	1856.834[b]	4	464.208	104.689	.000
截距	起始薪金	3.964E10	1	3.964E10	1989.067	.000
	教育水平（年）	21924.287	1	21924.287	4944.378	.000
性别	起始薪金	1.713E9	1	1.713E9	85.947	.000
	教育水平（年）	71.018	1	71.018	16.016	.000
雇佣类别	起始薪金	5.820E9	2	2.910E9	146.022	.000
	教育水平（年）	1089.533	2	544.766	122.856	.000
性别 * 雇佣类别	起始薪金	5.652E8	1	5.652E8	28.358	.000
	教育水平（年）	.154	1	.154	.035	.852
误差	起始薪金	9.347E9	469	1.993E7		
	教育水平（年）	2079.633	469	4.434		
总计	起始薪金	1.665E11	474			
	教育水平（年）	90215.000	474			
校正的总计	起始薪金	2.930E10	473			
	教育水平（年）	3936.466	473			

a. R 方 = .681（调整 R 方 = .678）

b. R 方 = .472（调整 R 方 = .467）

以上部分是对每个因变量分别进行一元的方差分析，将两个二因素方差分析的结果合并成一个表（表 15-4）。表格中“校正模型”和“截距”是采用一般线性模型计算时所得的结果，主要关注的还是因素的主效应、因素间的交互效应以及残差项。由表中“性别”，“工作类别”，“性别*工作类别”以及误差行中对应“受教育水平”的 F 和 sig.值可知，对于“受教育水平”这个因变量来说，性别、工作类别不同，受教育水平有差异，但是两者的交互作用效应不显著（性别×工作类别的 $F(1,469)=0.035$，$p=0.852$）。同样可以看出，对于起始薪金这个因变量来说，性别，工作类别以及两者的交互作用都有显著影响。分析中“性别*工作类别”的自由度本应为 2，此处为 1，是因为数据中无人属于“性别=f，工作类别=2”这种分类。

结合多元方差分析和两个一元方差分析的结果可知，本例中性别和工作类别不同，受教育程度和起始薪金都有显著不同，且对于起始薪金来说两者交互作用显著，而对于受教育程度来说两者交互作用不显著。将受教育程度和起始薪金联合起来看（多元方差分析），性别，工作类别以及交互作用都效应显著。由此可见，多元方差分析并不仅仅是几个一元方差分析的结果合成，甚至有些情况下，对几个因变量分别进行的一元方差分析中自变量的效应都不显著，但是将几个因变量联合起来同时做多元方差分析时，自变量的效应显著。

若是不考虑自变量之间的交互作用，对于受教育程度和起始薪金的二元方差分析如何进行？结果如何？请同学们自行练习。

第二节 典型相关分析

一、典型相关分析的含义与数学模型

在相关分析中，考察一个变量与另一个变量之间的关系，可以用简单相关系数；考察一个变量与一批变量之间的关系，可以用复相关系数。Hotelling 在 1936 年提出典型相关的方法考察一批变量与一批变量之间的相关。比如想研究人格与工作绩效之间的关系，人格有外向性，情绪稳定性，宜人性，尽责性，开放性等，而工作绩效又可以有多种指标，此时要考察人格与工作绩效的相关，简单相关系数和复相关系数就不够了。

通常情况下，为了研究两组变量之间的相关关系，可以用最原始的方法，分别计算两组变量之间的全部相关系数，一共有 $p \times q$ 个简单相关系数，这样又烦琐又不能抓住问题的本质。如果能够采用类似于主成分的思想，分别找出两组变量各自的某个线性组合，讨论线性组合之间的相关关系，则更简捷。首先分别在

每组变量中找出第一对线性组合，使其具有最大相关性；然后，再在每组变量中找出第二对线性组合，使其分别与本组内的第一线性组合不相关，第二对本身具有次大的相关性。如此继续下去，直至进行到 r 步，两组变量的相关性被提取完为止。r≤min(p,q)，可以得到 r 组变量。这样造出来的成对的线性组合称为典型变量，典型变量之间的相关系数称为典型相关系数。心理统计学把这种先将较多变量转化为少数几个典型变量，再通过其间的典型相关系数来综合描述两组多元随机变量之间关系的统计分析方法称为“典型相关分析（canonical correlation analysis）”。其数学模型可以表示为：

已知一组变量（X_1，X_2，…X_P）和另一组变量（Y_1，Y_2，…Y_q）是相互关联的两批变量。

$$\begin{cases} u_1 = a_{11}x_1 + a_{21}x_2 + ... + a_{p1}x_p \\ v_1 = b_{11}y_1 + b_{21}y_2 + ... + b_{q1}y_q \end{cases}$$

$$\begin{cases} u_2 = a_{12}x_1 + a_{22}x_2 + ... + a_{p2}x_p \\ v_2 = b_{12}y_1 + b_{22}y_2 + ... + b_{q2}y_q \end{cases}$$

…

求 a_{11}，a_{21}…a_{p1}，b_{11}，b_{21}，…b_{q1}；a_{12}，a_{22}…a_{p2}，b_{12}，b_{22}，…b_{q2} 等系数，使得 u_1v_1 之间的相关最高，u_2v_2 之间的相关次高，且 u_1u_2，v_1v_2 之间互相独立。

SPSS 中典型相关的分析可以通过调用专门的宏程序来执行，这种调用使用上十分简单，结果报告也比较详细。下面将详细介绍这种方法的调用及其结果。

二、利用 SPSS 软件包求解典型相关分析的实例

【例 15-2】以 SPSS 自带数据 employee data.sav 为例，分析薪水与工作阅历之间的关系。第一组变量薪水用起始薪金(salbegin)和当前薪金(salary)为指标，第二组变量工作阅历用受教育水平(educ)，工作时间(jobtime)和以往工作经验(prevexp)为指标。

在进行 SPSS 分析之前，先将文件“canonical correlation.sps”拷贝到“PASWStatistics18”文件夹中。然后，在数据文件窗口中，点击“文件（File）”菜单中的“新建（New）”子菜单中的“语法（Syntax）”，新建一个句法窗口。在新的窗口中输入以下两行：

Include ‘canonical correlation.sps’.

Cancorr set1=第一组变量 / set2=第二组变量.

在本例中第二句替换成“Cancorr set1=起始薪金 当前薪金 /set2=教育水平 雇佣时间 经验”即可。

最后在句法窗口中选择“运行（run）”菜单中的“全部（Al l）”，运行程序，典型相关分析的结果将出现在结果窗口。以下是部分结果的解释。

三、典型相关分析的主要结果解释

表15-5 两组变量内部相关系数矩阵表

Correlations for Set-1

	起始薪金	当前薪金
起始薪金	1.0000	.8801
当前薪金	.8801	1.0000

Correlations for Set-2

	教育水平	雇佣时间	经验
教育水平	1.0000	.0474	-.2524
工作时间	.0474	1.0000	.0030
经　验	-.2524	.0030	1.0000

表 15-5 给出的是两组变量内部各自的相关矩阵。可见薪水之间的相关是较高的。

表 15-6 两组变量之间相关系数矩阵表

Correlations Between Set-1 and Set-2

	教育水平	雇佣时间	经验
起始薪金	.6332	-.0198	.0451
当前薪金	.6606	.0841	-.0975

表15-6是两组变量之间各变量的两两相关系数矩阵。典型相关就是要找到典型变量来描述它们之间的这种相关性。

表 15-7 典型相关系数表

Canonical	Correlations
1	.683
2	.362

表 15-7 是提取的两个典型相关系数的大小。第一典型相关系数为 0.683，第二典型相关系数为 0.362。

表 15-8 典型相关系数的显著性检验表

Test that remaining correlations are zero:

	Wilk's	Chi-SQ	DF	Sig.
1	.464	361.293	6.000	.000
2	.869	65.953	2.000	.000

表 15-8 是典型相关系数的显著性检验，可见两个典型相关系数都有统计显著性。

表 15-9 两个典型变量与第一个变量组中各个变量之间
的标准化系数以及非标准化系数表

Standardized Canonical Coefficients for Set-1

	1	2
起始薪金	-.544	-2.035
当前薪金	-.487	2.049

Raw Canonical Coefficients for Set-1

	1	2
起始薪金	.000	.000
当前薪金	.000	.000

表 15-9 是两个典型变量与第一个变量组中各个变量之间的标准化系数以及非标准化系数。可见标准化的典型变量转换公式为：

u_1=0.554*“起始薪金” + 0.487*“当前薪金”

u_2=2.035*“起始薪金” – 2.049*“当前薪金”

非标准化的转换公式因为“当前薪金”和“起始薪金”的原始单位很大，所以非标准化的原始系数很小，显示位数不够显示为 0。要比较在典型变量的转换中哪个变量的作用大，还是要比较标准化系数。

表 15-10 典型变量与第二组变量之间的转换关系表

Standardized Canonical Coefficients for Set-2

	1	2
教育水平	-1.033	-.055
工作时间	.005	.592
经　验	-.227	-.821

Raw Canonical Coefficients for Set-2

	1	2
教育水平	-.358	-.019
工作时间	.001	.059
经　验	-.002	-.008

表 15-10 是典型变量与第二组变量之间的转换关系。同样可知标准化的典型变量为：

v_1 = 1.033 * 教育水平 – 0.005 * 雇佣时间 + 0.227 * 经验

v_2 = 0.055 * 教育水平 – 0.592 * 雇佣时间 + 0.821 * 经验

表 15-11 两组变量与自身所组成的典型变量 u_1，u_2 之间的相关表

Canonical Loadings for Set-1

	1	2
起始薪金	-.973	-.231
当前薪金	-.966	.258

Cross Loadings for Set-1

	1	2
起始薪金	-.665	-.084
当前薪金	-.660	.094

表 15-11 中上半部分是第一组的变量 x_1，x_2 与自身所组成的典型变量 u_1，u_2 之间的相关，称为典型负载系数。表中下半部分是第一组的变量 x_1，x_2 与第二组变量 y_1，y_2，y_3 所组成的典型变量 v_1，v_2 之间的相关，称为交叉负载系数。可见它们与第一典型变量 u_1，v_1 之间的相关更高。交叉负载系数的大小体现在后面的冗余分析中。

表 15-12 第二组变量与 v_1，v_2 和 u_1，u_2 之间的相关表

Canonical Loadings for Set-2

	1	2
教育水平	-.976	.180
雇佣时间	-.044	.587
经　验	.034	-.806

Cross Loadings for Set-2

	1	2
教育水平	-.666	.065
雇佣时间	-.030	.213
经　验	.023	-.292

表 15-12 中上半部分是第二组变量 y_1，y_2，y_3 与自身所组成的典型变量 v_1，v_2 之间的相关，下半部分是 y_1，y_2，y_3 与 u_1，u_2 之间的相关。可见受教育程度与第一典型变量相关高，但是雇佣时间和先前的工作经验两个变量与第二典型变量的相关更高，而且先前的工作经验与典型变量之间的关系方向与雇佣时间不同。

表 15-13 冗余分析的结果汇总表

Redundancy Analysis:	
Proportion of Variance of Set-1 Explained by Its Own Can. Var.	
	Prop Var
CV1-1	.940
CV1-2	.060
Proportion of Variance of Set-1 Explained by Opposite Can.Var.	
	Prop Var
CV2-1	.438
CV2-2	.008
Proportion of Variance of Set-2 Explained by Its Own Can. Var.	
	Prop Var
CV2-1	.318
CV2-2	.342
Proportion of Variance of Set-2 Explained by Opposite Can. Var.	
	Prop Var
CV1-1	.148
CV1-2	.045

表 5-13 是冗余分析的结果汇总，共由四个分表所组成。第一分表是第一个组的变量的总方差由其典型变量所解释的部分所占比率，其中 CV1-1 表示第一个组的第一个典型变量，CV1-2 表示第一个组的第二个典型变量。本例结果中，第一组变量的第一个典型变量解释了所有方差的 94%（其实这就是变量与自身典型变量相关系数或者说典型负载系数的平方和的平均，$[0.973^2+0.966^2]/2=0.94$），第二个典型变量只解释了总方差的 6%，可以看出因为只有第一组只有两个变量，其全部方差由两个典型变量全部解释。第三分表与第一分表类似，描述的是第二个组的变量总方差由其自身典型变量所解释的比率，其中 CV2-1 表示的是第二组变量的第一个典型变量，CV2-2 表示的是第二个组的第二个典型变量。由结果可知，第二组变量的第一典型变量只解释了总方差的 31.8%，而第二组变量的第二个典型变量则解释了总方差的 34.2%，但是因为第二组有三个变量，这两个典型变量并不能全部解释三个原始变量的总方差，其总和不等于 100%。第二分表和第四分表给出的是冗余指数，第二分表是第一个变量组的总方差由第二个变量的典型变量所解释的部分占总方差的比率，第二组的第一个典型变量（CV2-1）对第一组总方差的解释率为 43.8%（这是变量与另一组的典型变量相关系数或者说交叉负载系数的平方和的平均，$[0.665^2+0.660^2]/2=0.438$，这可能是由于受教育程度与薪水之间的相关造成），而第二组的第二个

典型变量（CV2-2）对第一组变量的总方差几乎没有解释力（0.8%）。同样，第四分表可以看出第二组变量的总方差可以有 14.8%由第一组的第一个典型变量解释，4.5%由第一组的第二个典型变量解释。冗余指数在想要考察一群自变量和一群因变量之间的关系时有用，它可以反映自变量的各种典型变量对于因变量所有变量的解释能力。

为了熟悉典型相关的操作和解释，请读者试用同一批数据分析第一组（当前薪金，起始薪金和受教育水平）与第二组（工作时间，以前的工作经历）之间的相关。

第三节 聚类分析

一、聚类分析的思路与要求

聚类分析（cluster analysis）是一种采用数值的方法对个体观察值或者变量进行分类的多元统计方法。人类认识事物往往从分类开始，但是早期的分类学中，人们主要依靠的是经验和专业知识。直到二十世纪，生物学、心理学等学科的发展衍生出一门新的学科——数量分类学（numerical taxonomy）。1939 年美国心理学家特雷恩（R. C. Tryon）出版了 Cluster Analysis:Correlation Profile and Orthometric(factor) Analysis for the Isolation of Unities in Mind and Personality 一书，使得用数学方法解决聚类问题的多元理论得到发展，分化出一个新的分支——聚类分析，并成为数量分类学的数学基础。现在聚类分析已经成为生物学，地质学，社会学，人口学等研究领域中常用的数值分类方法。聚类分析的内容非常丰富，包括系统聚类法，动态聚类法，模糊聚类法，有序个体观察值聚类法，图论聚类法等，本节重点介绍 SPSS 中可以实现的系统聚类法和动态聚类法两种。

聚类分析的基本思想是按照描述分类对象所用的变量，将每个对象相当于多维空间中的一个点，按照它们空间关系的亲疏程度进行分类，其原则是分类后同一类中的个体观察值间有较大的相似性，不同类中的个体观察值差异很大。聚类分析时有两种，一种是对个体观察值进行聚类，我们称为 Q 型聚类（Q-type cluster），另一种是对变量进行聚类，我们称为 R 型聚类（R-type cluster）。

进行聚类对于变量有一定的要求。在对个体观察值进行聚类时，根据实际问题选择能够反应个体观察值特性的聚类变量是很关键的，因为聚类分析的结果仅仅反映的是按照所选择的变量定义的数据结构。很显然，同一个样本根据不同的聚类变量所得到的聚类结果可能相差很大。一般来说，聚类变量应该尽可能符合以下特征：

（1）与聚类分析的目标密切相关，能够反应分类对象的特征；

（2）应该只引入在不同类间有显著差别的变量，变量不是越多越好，有时无关变量的引入反而会引起严重的错分；

（3）变量之间不存在高度的相关，引入高度相关的变量相当于在计算相似性时进行了加权；

（4）在聚类分析中，尽量只使用相同类型的变量进行分析。

二、相似性测度

将个体观察值或者变量进行聚类，首先需要了解它们之间的相似性。相似性测度有两类方法，一类是用距离描述相似性，另一类是用相似系数描述相似性。

（一）距离

在聚类分析中采用距离来描述数据点之间相似程度，这里需要区分两个概念：点与点之间的距离以及类与类之间的距离。

1. 点与点之间的距离

用来描述点与点之间距离的方法非常多，一般来说只要所定义的计算方法满足非负性，对称性以及三角不等式，都可以称为距离。常用的三种距离计算方法有：

（1）欧氏距离（Euclidean distance）

$$d_{ij}=\sqrt{\sum_{k=1}^{m}(x_{ik}-x_{jk})^2}$$

其中 d_{ij} 表示第 i 个点与第 j 个点之间的距离，x_{ik} 表示第 i 个点在第 k 个变量上的取值，m 为变量的个数。在 SPSS 中为了计算的简便还提供了平方欧氏距离（Squared Euclidean distance）。

（2） 兰氏距离（Lanberra distance）

$$d_{ij}=\frac{1}{m}\sum_{k=1}^{m}\frac{\left|x_{ik}-x_{jk}\right|}{x_{ik}+x_{jk}}$$

式中各符号含义同上。

（3）马氏距离（Mahalanobis distance）

$$d_{ij}=(X_i-X_j)'\Sigma^{-1}(X_i-X_j)$$

其中 X_i 表示第 i 个点在各个变量上的值所组成的向量，Σ^{-1} 为变量的协方差矩阵的逆矩阵。

这三种计算距离的方法中欧氏距离是用得最多，也最容易理解的。但是欧氏距离有两个缺陷：一是欧氏距离的值与各指标的量纲有关，而各指标计量单位的选择有一定的人为性和随意性，各变量计量单位的不同不仅使此距离的实际意义难以说清，而且，任何一个变量计量单位的改变都会使此距离的数值改变，从而

使该距离的数值依赖于各变量计量单位的选择。二是没有考虑各个变量之间的相关性和重要性。实际上，欧氏距离将各个变量都同等看待，而没有考虑变量的离差大小的不同。如果变量之间的高相关，则所计算的距离相当于多次重复加权。兰氏距离对自身进行了标准化，去除了单位的影响，因此对于大的奇异值不敏感，适用于高度偏倚的数据。尽管对于欧氏距离的第一个缺点提出了改善，但兰氏距离还是没有考虑变量之间的相关。马氏距离称为广义的欧氏距离，它主要的不同就是考虑到变量之间的相关性。可以看到，当变量之间互相独立时，即变量的协方差矩阵为对角矩阵，马氏距离就就退化为用各个观测指标的标准差的倒数作为权数进行加权的欧氏距离。因此，马氏距离既考虑到变量取值的差异程度，又考虑到了变量之间的相关。

2. 类与类之间的距离

类与类之间的距离实际上包含了点与类之间的距离。如果将只有一个点的类看成是最基本的类，此时点与类的距离就可以通过类与类的距离来计算。常用的类与类之间距离的计算方法有：

（1） 最短距离法（Nearest neighbor）： 将 A 类中每一个点分别与 B 类中所有点求出距离，这些距离中最短的就定义为 A 类与 B 类之间的距离，这种方法称为最短距离法。最短距离法容易聚成一个较大的类，使大多数个体观察值被聚在一起。

（2） 最长距离法（Furthest neighbor)：将 A 类中每一个点分别与 B 类中所有点求出距离，这些距离中最长的就定义为 A 类与 B 类之间的距离，这种方法称为最长距离法。最长距离法容易受到极端值的影响。

（3）平均联结法：最短距离法和最长距离法都只采用了两个类的点之间距离中极端的一个，没有充分利用其它点之间的距离信息。平均联结法是将两个类所有点的距离的平均作为类与类之间的信息。它有两种形式，一种称为组间联结法（Between-groups linkage)，另一种称为组内联结法（within-groups linkage)。组间联结法是计算在 A 类的每个点与 B 类所有点之间的距离，汇总之后求平均数作为类和类之间的距离。组内联结法则是计算 A 类和 B 类合并之后所有点两两之间的距离（包括 A 类内部的点两两之间的距离以及 B 类内部的点两两之间的距离）的平均作为类和类之间的距离。

（4）重心法（centroid clustering)：与平均联结法求所有距离的平均不同，重心法是先计算出 A 类和 B 类各自的平均（或重心)，然后计算两个重心之间的距离作为类和类之间的距离。重心法要求计算点与点的距离使用欧氏距离法。

（5）离差平方和法（Ward 法)：其方法是使每个个体观察值自成一类，每一步使得合成之后的离差平方和增量最小。即 A 类与 B 类的距离定义为 A 与

B 合并之后的类内离差平方和与原有的 A 类离差平方和以及 B 类离差平方和相比较的增量。

平均链接法和重心法都较少受到奇异值的影响，但是重心法在聚类的过程中合并的类之间的距离不能保证单调递增，即合并之后的两类之间的距离可能小于上次合并的两类之间的距离。离差平方和法对奇异值很敏感，但是对较大的类倾向产生较大的距离，从而不易合并，因此容易产生规模大致相同的类。平均联结法和离差平方和法是聚类中应用较为广泛的方法。

（二）相似系数

在对变量进行聚类时，常用相似系数来描述变量之间的相似性。相似系数常用的算法有两种：夹角余弦法以及皮尔逊相关法。

1. 夹角余弦法

$$r_{ij}=\cos\theta_{ij}=\frac{\sum_{k=1}^{n}x_{ki}x_{kj}}{\sqrt{\sum_{k=1}^{n}x_{ki}^2\sum_{k=1}^{n}x_{kj}^2}}$$

其中 r_{ij} 表示第 i 个变量与第 j 个变量之间的相似系数，x_{ki} 表示第 k 个样品在第 i 个变量上的取值，n 表示样品的个数。

2. 皮尔逊相关法

$$r_{ij}=\frac{\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)(x_{kj}-\bar{x}_j)}{\sqrt{\sum_{k=1}^{n}(x_{ki}-\bar{x}_i)^2\sum_{k=1}^{n}(x_{kj}-\bar{x}_j)^2}}$$

采用相似系数描述变量相似性时，相似系数绝对值越大，说明变量之间的相似程度越高。因此，相似系数大的变量优先聚在一起。

三、聚类方法

常用的聚类方法有系统聚类和动态聚类两种，下面我们分别加以介绍：

（一）系统聚类

系统聚类（Hierarchical Cluster）又称作谱系聚类或层次聚类，其做法是先将所有 n 个变量 / 观察值看成不同的 n 类，然后将性质最接近（相似性程度最高或者距离最近）的两类合并为一类，再从这剩下的（n-1）类中找到最接近的两类加以合并，依此类推，直到所有的变量 / 样品被合为一类。最后使用者再根据具体的问题和聚类结果来决定应当分为几类。其特点是可以用于对变量或观察值进行聚类，但是一旦变量或者样品被划到某个特定类别，其分类结果就不会再进行更改。相对来说，系统聚类

方法运行速度较慢，易受奇异值的影响。

（二）动态聚类

动态聚类（K-Means cluster）又称作快速聚类，其做法是先指定要形成的聚类数，按某种原则选择或人为指定某些观察值作为凝聚点，然后计算每个观察值到这些凝聚点的距离，将它们归入到距离最近的类中，得到初始分类结果，之后重新计算每类的重心作为新的凝聚中心，再次计算每个观察值到新的凝聚中心的距离，将它们归入最近的类，如此循环，直到新一次的归类过程中没有样品点的归类需要调整为止。这种聚类方法克服了系统聚类不能调整归类结果的缺陷，因此受奇异值或不合适的聚类变量的影响较小，而且相较于系统聚类，需要存储的信息少，运算时间更短，往往适用于大型数据的情况。但是采用这种方法需要预先知道类别数，而且只适用于样品聚类，要求变量是连续型变量，而且对于初始分类敏感。

一般来说，对样品进行聚类时可以将两种方法联合起来使用。首先采用系统聚类法确定分类数，检查奇异值，并以系统聚类的结果作为迭代聚类时的初始凝聚中心。

四、聚类方法的效果

前面我们提到过，聚类变量的选用对于聚类的结果会有影响，除此之外，还有许多其它因素也会影响聚类的结果。数据中的奇异值对于聚类的结果会有比较大的影响，通常我们要剔除奇异值然后进行聚类分析。数据内是否有区分比较明显的层或者类也会影响聚类方法的效果，无论用哪种聚类方法，要想将类与类之间有高度重叠的样品中准确地区分出来都是很困难的。另外，我们知道，对于同样的数据，采用的不同的相似性测度（距离或者相似系数）所得到的聚类结果会不尽相同，而选用不同的聚类方法所得到的聚类结果也可能不一致。但是我们没有特别行之有效的方法来判断哪种聚类方法所得到的结果更值得信赖。通常的做法是将一个样本随机分成两组，分别进行聚类分析，将结果进行比较；或者在样本容量较小时将同一个样本采用不同的方法重复聚类，然后比较结果。

五、利用 SPSS 软件包求解聚类分析的实例

【例 15-3】 以 SPSS 自带数据 employee data.sav 为例，试对年龄(age)，受教育水平(educ)，当前薪金(salary)，起始薪金(salbegin)，工作时间(jobtime)，以前的工作经验(prevexp 等变量进行分类研究（注意：原始数据中关于员工年龄的变量是“出生年月”，如果把该日期型变量转换成以月为单位的“年龄”新变量，在年龄的内涵上会显得更直观一些。由于在 SPSS 中使用这两种变量中任意一种变量进行聚类的结果是一样的，因此本节使用的是原始数据中“出生日期”这一变量名进行聚类分析的结果）。

本例是对变量进行分类的 R 型聚类。运用 SPSS18.0 软件包，在已打开的数据文件窗口中，根据下列操作步骤（如图 15－3 所示）可以求解本例的系统聚类分析结果。

1.分析→分类→系统聚类
2.变量框：年龄，教育水平，当前薪金，起始薪金，雇佣时间 经验
3.分群：变量
4.统计量：（1）方案范围
　　　　（2）最小聚类数：2
　　　　（3）最大聚类数：5
　　　　（4）继续
5.绘制：（1）树状图
　　　　（2）继续
6.方法：（1）区间：下拉菜单中选择“Pearson 相关性”
　　　　（2）继续
7.确定

图 15-3 在 SPSS 中设置系统聚类分析程序各步骤示意图

六、聚类分析的结果解释

利用 SPSS 软件包分析之后，得到结果摘选分析如下：

表15-14 聚类的进度表（Agglomeration Schedule）

阶	群集组合		系数	首次出现阶群集		下一阶
	群集 1	群集 2		群集 1	群集 2	
1	3	4	.880	0	0	2
2	2	3	.647	0	1	3
3	1	2	.145	0	2	4
4	1	5	.016	3	0	5
5	1	6	-.221	4	0	0

表 15-14 是聚类的进度表，第一列表示聚类的步骤顺序，第二、三列表示每一步将哪两个类合并，新合成的类的序号用两个类中更小的类表示，第四列表示聚合系数，是被合并的两类之间的距离或者相关，最后一列表示合并之后的新类下一次与其它类合并时出现的步骤顺序。本例中，按顺序的六个变量中，第 3 个变量和第 4 个变量首先合并，新类仍称第 3 类；新的第 3 类在第 2 步时与第 2 个变量合并后仍称为第 2 类。接下来新的第 2 类与第 1 个变量合并后仍称为第 1 类，而后新的第 1 类依次第 5 个变量和第 6 个变量合并。这是一种表格形式的进度表。

表15-15 各个变量分类归属表 （Cluster Membership）

案例	*5 群集*	*4 群集*	*3 群集*	*2 群集*
出生日期	1	1	1	1
教育水平（年）	2	2	1	1
当前薪金	3	2	1	1
起始薪金	3	2	1	1
雇佣时间（以月计）	4	3	2	1
经验（以月计）	5	4	3	2

表15-15是各个变量分成2-5类时不同情况下的归属。分成5类时，起始薪金和当前薪金两个变量的数字相同，表示在同一类；进一步分成4类时，教育水平，当前薪金，起始薪金合成新的一类。直到分成2类时，出生日期，教育水平，当前薪金，起始薪金以及工作时间合成一类，以前的工作经验单独成为另外一类。

冰柱图是报告聚类结果的一种图示方法。图中每行表示类的数目，每隔一列是一个案例或者变量，本例中是变量名。每一个变量名下都有 X，通过空列下面的 X 将变量联系起来，相连的变量表示合并为同一类，断开的变量名表示未归为一类。由图 15-4 可见，当类数为 5 时，同一行的变量名所对应的 X 中只有起始薪金和当前薪金两个变量之间有 X 连接，其余变量之间还是空白，说明起始薪金和当前薪金首先合并为一类。当类数为 4 时，同一行的 X 中，除了起始薪金和当前薪金之间有 X 连接之外，在当前薪金和教育水平之间也有 X 连接，说明当前薪金与起始薪金和教育水平可以进一步合并为一类。到类数为 1 时，除了经验这一变量之外，其他所有五个变量名下面的 X 之间都有 X 连接在一起，说明这些变量可以归到一类。

群集数	经验		雇佣时间		起始薪金		当前薪金		教育水平		出生日期
	X		X	X	X	X	X	X	X	X	X
	X		X	X	X	X	X	X	X	X	X
	X		X		X	X	X	X	X	X	X
	X		X		X	X	X	X	X		X
	X		X		X	X	X		X		X

图15-4 冰柱图

图 15-5 是树状聚类图。图中纵向是变量名，横向是转换过后的距离。横线代表变量或者类，通过纵线连接的两根横线表示两个变量或者类合成新的一类，连接两根横线的纵线所在的位置代表变量之间的距离。从图中可以看出，变量 3 和 4 在距离约为 1 左右合并成一类; 变量 3 与 4 合并之后的新类在距离约为 6 处与变量 2 合并成新的一类。而后在距离约 16 处与“出生日期”合并成一个新类；在约 20 处与雇佣时间合并为一类；最后在约 25 处所有变量合并为一类。从树状图可以根据距离的大小确定分类。例如，将距离设定在 10 附近时，此处画一条垂直线，正好将树根砍下，左边剩下三棵树干。最上面的一颗树干连接这变量 3，变量 4 和变量 2，下边三棵树对应着变量 1，变量 5 和变量 6。说明这时 6 个变量可以聚为四类，变量 3，4，2 聚为一类，其他三个变量各成为一类。

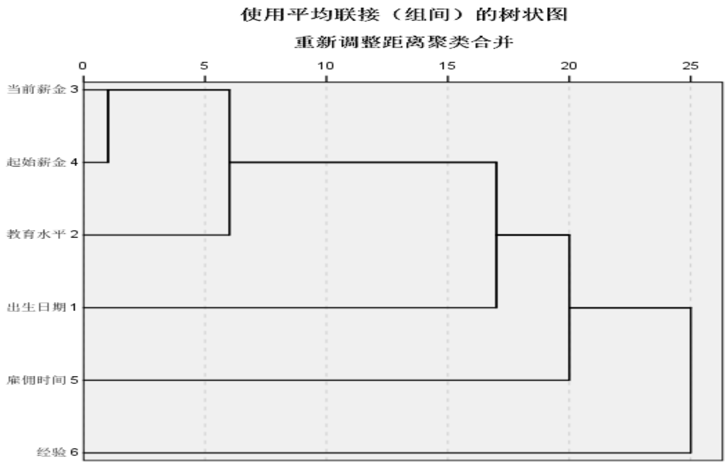


图15-5 树状聚类图

为了熟悉聚类分析的操作与解释，并了解不同的聚类方法对于聚类结果的影响，读者可以试用同一批数据分析，在去除“出生日期”变量之后，用组间链接法与用 Ward 方法进行聚类，看看聚类结果是否有不同。

第四节　因素分析

一、因素分析的思路

因素分析（factor analysis，又称为因子分析）是指从大量的相关变量中抽取出最基本的维度或因素并加以分析的一组多元统计分析方法的总称。这种统计方法主要是心理学家在心理学研究中建立和发展起来的。1904 年英国心理学家斯皮尔曼（C．Spearman）发表了 *General Intelli-gence, Objectively Determined and Measured* 一文，提出智力是由一般因素和众多特殊因素构成的，从而创立了因素分析的两因素（即一般因素与独特因素）方法。随着计算机技术的发展，因素分析方法在心理学，教育学，社会学，经济学等众多领域得到广泛的应用。在此基础上，还发展出了一种新的因素分析方法——验证性因素分析，而相应地，经典的因素分析方法被称为探索性因素分析。**所谓探索性因素分析**（exploratory factor analysis）是指在事先不知道可能的潜在因子的基础上,根据数

据进行内部结构探索的因素分析方法；所谓**验证性因素分析**（confirmatory factor analysis）则是指根据已有的理论或知识,在已知因子的情况下检验所搜集的数据资料是否支持某种理论假设的因子结构的因素分析方法。

因素分析的基本思想是通过对变量或个体之间相关系数矩阵的考察，寻找几个新的变量以再现这种变量之间或个体之间的内部关系，最终可以根据这些新的变量将原有变量或个体进行分类。这些新的变量是用来解释原有的观测到的变量的，我们称为潜变量，或者因素。考察变量之间相关关系的因素分析被称为 R 型因素分析，考察个体之间相关关系的因素分析被称为 Q 型因素分析。

本书只简单介绍探索性因素分析方法，对验证性因素分析感兴趣的同学请参阅相关文献或者结构方程模型的书籍。

二、因素分析的应用及要求

因素分析在教育与心理学中的应用主要可以在以下两个阶段的四个方面：

（1）在研究设计阶段或者问卷效果评估阶段，可以用来评估问卷的结构效度在心理学的研究中，许多特质都是潜在的，无法直接观测的，我们可以采用问卷的方法进行间接地测量。通过因素分析的方法，可以得到问卷中的试题是否确实反映所欲测量特质的维度，作为问卷的结构效度的一个证据。比如说，某人格理论认为人格可以从内外向性和情绪稳定性两个方面进行考察，若对据此编制的测验进行因素分析的结果确实有两个主要的因素，则这可以作为该测验具有结构效度的一个证据。

（2）在数据的统计分析阶段可以用来寻找变量间潜在结构，进行数据化简以及证实变量间的内在结构。因素分析的一个主要功能就是通过对观测的变量之间的相关的分析，寻找用以解释造成变量之间相关模式的潜在结构。将一系列的变量转化成少数几个共同因素之后，可以根据变量与因素之间的关系进一步将观测变量的取值转化成潜在因素得分，从而用少数几个潜在的因素得分代替原来较多的观测变量，达到简化数据的目的。根据某种理论或者已有知识对因子的个数和结构作出假设，可以用验证性因素分析的方法对数据是否支持假设进行验证。

采用因素分析方法可以得到隐藏在观测变量之下的潜在变量的内部结构，但是为了得到比较稳定的结果，进行分析的数据需要满足一定的条件：

各观测变量之间要有一定程度的相关。变量之间的相关系数全都不大时，变量之间不太可能有共同的公因素，此时没有必要进行因素分析。一般来说，变量之间相关系数矩阵中大多数值小于 0.3 时不适于进行因素分析。对于这一点，统计软件包 SPSS 中有专门的选项 KMO 统计量和 Bartlett 球形检验进行检查。KMO 统计量通过比较变量之间的简单相关系数和偏相关系数的大小而得到，它的取值范围在 0 到 1 之间。当变量之间有内在联系时，控制某个变量之后它与其它变量的偏相关系数会远远小于简单相关系数，此时 KMO 系数接近于 1。因此，KMO 统

计量越接近于 1 越好。一般来说，KMO 统计量在 0.90 以上最好，0.7 以上一般，0.5 以下时不宜做因素分析。Bartlett 球形检验是检验相关系数矩阵是否是单位矩阵，或者说各变量之间是否互相独立。若球形检验的结果不能拒绝虚无假设，则说明变量之间可能没有相关，不需要进行因素分析。

为了得到相对稳定的因素结构，因素分析方法希望样本量比较充足。有研究者提出，样本量与变量数之间的比率至少要在 5：1 以上，理想的样本量应是变量个数的 10-25 倍。而且不管变量个数如何，样本总量不应少于 100 人，原则上越大越好。

三、因素分析的模型及相关概念

因素分析的数学模型可以表示为：

$$\begin{cases} X_1 = a_{11}F_1 + a_{12}F_2 + ... + a_{1m}F_m + \varepsilon_1 \\ X_2 = a_{21}F_1 + a_{22}F_2 + ... + a_{2m}F_m + \varepsilon_2 \\ \qquad\qquad ...... \\ X_p = a_{p1}F_1 + a_{p2}F_2 + ... + a_{pm}F_m + \varepsilon_p \end{cases}$$

其中

X_i 表示第 i 个观测变量，F_j 表示第 j 个潜在因子，又称为公因子，是各个观测变量所共有的，它解释了观测变量之间的相关；ε_i 表示第 i 个独特因子，是每个观测变量所都有的，相当于多元回归中的残差项，表示该变量不能被公因子所解释的部分。a_{ij} 表示第 i 个观测变量在第 j 个潜在因子上的系数，又被称为因子负荷。因子负荷体现了潜在因子和各个观测变量之间的密切程度。当各潜在因子之间互不相关时，因子负荷就是潜在因子与观测变量之间的相关。其绝对值体现了因子对观测变量的影响程度，绝对值越大，说明该因子对观测变量的影响越大。

因素分析中还有一些常用的概念：

公共因子方差（common factor variance，又称共同度），是指观测变量方差中由公因子决定的比例，记为 h_i^2，其计算公式为：

$$h_i^2 = a_{i1}^2 + a_{i2}^2 + ... + a_{im}^2$$

公因子方差越大，说明变量对公因子的依赖程度大，变量的方差能够被各公因子解释的比率越高。

对应地，$1 - h_i^2$ 称为特异性方差，是变量中不能被公因子解释的那部分方差。

因子的方差贡献，是该因子能解释的各变量方差的总和，记为 g_j^2，其计算公式为：

$$g_j^2 = a_{1j}^2 + a_{2j}^2 + ... + a_{pj}^2$$

它是衡量公因子重要性的一个指标。g_j^2 越大，说明第 j 个公因子对变量方差的解释越多，它对变量的影响也越大。g_j^2 与变量总个数 p 的商，称为该因子的方差贡献比率。

四、进行因素分析中要注意的几个问题

（一）进行因素分析的基本步骤

一般来说，采用因素分析方法对数据的结果进行化简往往需要有以下步骤：

（1）计算所有变量的相关系数矩阵，并采用 KMO 统计量和球形检验来判断数据是否符合要求；

（2）根据一定的标准确定因子的个数，提取因子，计算因子负荷；

（3）考察所得因子的可解释性，必要时进行因子旋转，根据所得到的因子负荷进因子命名；

（4）如有必要，计算出因子得分以备进一步的分析使用。

（二）因子个数的确定

因素分析方法希望用尽可能少的潜在因素解释尽可能多的信息，但是多少个因子才算足够呢，一般有以下几种方法可以帮助我们做决策：

1. 因子方差所解释的比率，即因子的累积贡献率。

一般来说，提取因子的累积贡献率达到 80%-85%以上就很满意了。在有些情况下，为了尽可能简化数据，用尽可能少的模型解释原有变量的方差，因子累积贡献率在 50%也不是不可接受。

2. 特征根大于 1。

因为每个变量的方差为 1，而所求得的特征根实际上就对应于因子的方差贡献，因此留下的因子至少解释的方差不比一个变量所能解释的方差少，从而达到了简化数据的目的。其实这是一个最基本的要求，但是由于在 SPSS 中是缺省设置，因此成为应用最广泛的确定因子个数的方法。

3. 碎石检验（Scree test）。

这也是 SPSS 中提供的一个选项。这种检验以因子被提取的顺序为横轴，以相应特征根的取值作为纵轴高度，连线描图。因为抽取出的因子顺序是以特征根的大小进行排序的，所以连成的直线一定是逐步下降的。该曲线开始下降迅速，到后来慢慢变得平缓。碎石检验是将特征根变化平缓之前的那个点作为最后一个提取出的因子。

4. 平行检验。

还有统计学家提出一种模拟数据的方法来确定因子的个数。模拟产生一批样

本数和变量数都和实际数据相同的随机数据，针对这批模拟产生的随机数可以进行碎石检验。其结果与实际数据的碎石检验结果放在一起，两者交点处的特征根说明实际数据的特征根和模拟数据的特征根相等。大小超过随机数的特征根的因子被保留，这意味着变量内贡献超过随机数因子个数。这种方法无法在 SPSS 中直接实现。

5. 因素的可解释性。

在因素分析中，保留下的因子是否有意义，能否被解释，这也是确定因子时应该考虑的一个重要因子。

事实上，很少只使用一种方法来对因子的个数进行判断，有研究显示，特征根大于 1 的标准可能保留太多的因子，而碎石检验则有时留下的因子个数偏少。因此，可能的做法是选择因子个数的范围（例如 5-7 个），分别进行因子提取和旋转，结合因子的可解释性选择一个最合适的。另外有一个原则是一般一个因子至少有 3 个负荷较高，可解释的变量，否则可能意味着因子个数太多。

（三）因子的旋转

在探索性因素分析中，因子的提取是根据特征根的大小按顺序进行的，所以第一个因子也就是特征根最大（或者说对变量的方差贡献最大）的因子，绝大多数变量在第一个因子上的负荷都较大。这一点会影响因子的解释和命名。因子旋转是可以使我们避免这个尴尬的有效工具。所谓因子旋转（factor rotation）是指通过改变坐标轴的位置来重新分配各个因子所解释的方差比例，进而使变量在因子上的负荷更加清晰，因子结构更为简单、更易于解释的一种方法。

因子旋转不改变模型对数据的拟合程度，不改变每个变量的公因子方差，只是对于变量相关的另外一种解释。因子旋转有两大类：正交旋转与斜交旋转。所谓正交旋转，是指选择过后因子之间仍然不相关（正交），而斜交旋转之后，因子之间并不是正交的，而是允许有一定程度的相关。SPSS 中提供了三种正交旋转方法（Quartimax，Varimax，Equimax）和两种斜交旋转方法（Direct Oblimin 和 Promax）。

应根据研究的假设和研究的需要选择合适的旋转方法。当假设潜在的因子之间不相关，或者研究的主要目的是简化数据结构而不关注因子的实际含义时，可以选用正交选择的方法。当理论模型中的潜在因子是相互关联的，或者研究希望得到几个有意义的潜在因子时，则可以选用斜交旋转的方法。

五、利用 SPSS 软件包求解因素分析的实例

【例 15-4】以 SPSS 自带数据 employee data.sav 为例，试分析年龄（age），受教育年限（educ），现在薪金（salary），起薪（salbegin），工作时间（jobtime），以前的工作经验（prevexp）等变量的内部结构。

运用 SPSS18.0 软件包，在已打开的数据文件窗口中，根据图 15-6 所示的操作步骤可以求解本例的因素分析结果。

1.分析→降维→因子分析
2.变量框：出生日期，教育水平，当前薪金，起始薪金，雇佣时间，经验
3.描述：（1）相关矩阵框中选择“KMO and Bartlett 的球形度检验”
（2）继续
4.抽取：（1）输出框中选择“碎石图”
（2）继续
5.旋转：（1）方法框中选择“直接”
（2）继续
6.确定

图 15-6 在 SPSS 中设置因素分析程序各步骤示意图

六、因素分析的部分结果解释

部分结果分析如下：

表15-16 KMO检验以及Bartlett球形检验的结果（KMO and Bartlett's Test）

取样足够度的 Kaiser-Meyer-Olkin 度量。		0.617
Bartlett 的球形度检验	近似卡方	1593.791
	df	15
	Sig.	0.000

表 15-16 是 KMO 检验以及 Bartlett 球形检验的结果。KMO 越接近于 1 越好，本例中统计量的值 0.617，勉强可行。Bartlett 球形检验的伴随概率（sig.）小于 0.001，说明可以拒绝相关矩阵是单位矩阵的假设，变量之间是有一定相关的。由此可见，对这几个变量进行因素分析是满足假设，可行的。

表15-17 被公因素解释的方差表（Communalities）

	初始	提取
出生日期	1.000	.894
教育水平（年）	1.000	.733
当前薪金	1.000	.895
起始薪金	1.000	.911
雇佣时间（以月计）	1.000	.998
经验（以月计）	1.000	.899

提取方法：主成份分析。

每个变量原始的方差都为1，被公因素解释的方差列在“提取（Extraction）”列中，该值越接近1说明被公因素解释的部分越大，特异性方差越小。由表15-17可看到，“被雇佣后的工作时间”的共同度为99.8%，基本可以由公因子解释。

表15-18 被方差所解释的特征根表（Total Variance Explained）

成份	初始特征值			提取平方和载入			旋转平方和载入[a]
	合计	方差的 %	累积 %	合计	方差的 %	累积 %	合计
1	2.572	42.870	42.870	2.572	42.870	42.870	2.489
2	1.754	29.225	72.095	1.754	29.225	72.095	1.909
3	1.004	16.739	88.834	1.004	16.739	88.834	1.013
4	.374	6.232	95.066				
5	.194	3.240	98.306				
6	.102	1.694	100.000				

提取方法：主成份分析。

a．使成份相关联后，便无法通过添加平方和载入来获得总方差。

表15-18中左半部分第一列是6个特征根的值，即每个因素所能解释的总方差，它们与变量总个数之比即因素所能解释的方差占总方差的比率，放在第三列，第四列是累计的解释方差比。右边部分是根据SPSS缺省的特征根大于1的方法确定的因素数所能解释的方差。例子中有3个公因素的特征根大于1，且其余的三个特征根都比1小很多，加上三个因素能够解释总方差的88.83%，因此选择抽取3个因素是合理的。

碎石图显示这些特征根的变化都是陡峭的（图 15-9），变化平缓之处特征根

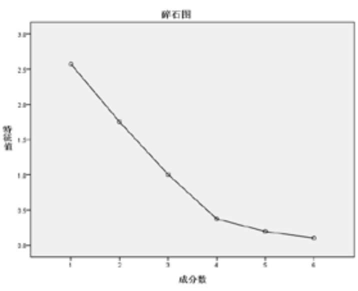


图 15-9 碎石图

已小于 1，因此抽取 3 个因素是合理的。

在进行碎石检验之后报告的是使用主成分法抽取的未经旋转的初始因素负荷矩阵（component matrix）。因为本例中使用斜交旋转方法，SPSS 结果随后报告三个矩阵：因素模式矩阵（pattern matrix），因素结构矩阵（structure matrix）以及因素相关矩阵（component correlation matrix）。

表15-19 初始因素负荷矩阵表（Component Matrix(a)）

	成份		
	1	2	3
出生日期	.437	-.838	-.010
教育水平（年）	.854	.055	.015
当前薪金	.894	.311	.013
起始薪金	.835	.449	-.111
雇佣时间（以月计）	.050	.088	.994
经验（以月计）	-.391	.861	-.059

提取方法 :主成份。

a. 已提取了 3 个成份。

初始因子负荷矩阵（表 15-19）是采用主成分方法所得到的因子负荷矩阵，主成分分析的结果也可以利用这个表得到。在这个矩阵中，所得到的因子之间是互相独立的，而且所有变量在同一个特征根的负荷的平方和等于相应的特征根。同一个变量在各个因子上的负荷的平方和等于该变量的公共因子方差。由于初始负荷矩阵中第一个因子的特征根最大，所以往往各个变量在第一个因子上的负荷都不低，这会造成因子命名的困难。为了得到更简明的结构去描述变量之间的相关，往往还要进行因子旋转。

表15-20 因素模式矩阵表Pattern Matrix(a)

	成份		
	1	2	3
出生日期	.030	-.940	-.045
教育水平（年）	.797	-.220	.035
当前薪金	.945	.012	.047
起始薪金	.962	.169	-.071
雇佣时间（以月计）	.003	.022	.999
经验（以月计）	.027	.951	-.022

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的斜交旋转法。

a. 旋转在 4 次迭代后收敛。

在斜交旋转中因子负荷矩阵（表15-20）又称因子模式矩阵，通过旋转使得变量在因子上的负荷或者尽可能接近于1，或者尽可能接近于0。由表15-20可以看出，受教育水平，当前薪金以及起始薪金三个变量只在第一个因子上负荷高，以前的工作经验和年龄两个变量只在第二个因子上负荷高，工作时间只在第三个因子上负荷高。

表15-21 因素结构矩阵表Structure Matrix

	成份		
	1	2	3
出生日期	.152	-.944	-.042
教育水平（年）	.827	-.325	.067
当前薪金	.945	-.112	.085
起始薪金	.937	.043	-.033
雇佣时间（以月计）	.040	.020	.999
经验（以月计）	-.098	.947	-.023

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的斜交旋转法。

因子结构矩阵（表 15-21）是描述变量和因子之间的相关关系。当因子之间正交时，因子负荷实际上就是因子与变量之间的相关。因此，在正交旋转时，因子模式矩阵和因子结构矩阵相同，SPSS 中记为 rotated component matrix。由上表可以看出，因子 1 主要与受教育水平，当前薪金和起始薪金相关很高，与其它变量相关很低，因子 1 可能是反映收入水平的潜变量；因子 2 主要与以前的工作经历以及年龄相关很高，与其它变量的相关都很低，因子 2 可能主要是反映年龄的潜变量；而因子 3 只与工作时间相关很高，与其它变量相关都低，这是一个反映在本公司工作时间的潜变量。而且我们注意到共同度的表格中工作时间的共同度几乎为1，其特征根也接近于1，可知因子3实际上就相当于工作时间变量。

由此可见，因子分析的结果可以将这六个变量分成三大类，同一类的变量在同一个因子上的负荷高，在其它因子的负荷低。可以将因子分析的结果与例15-3 中对变量聚类的分析结果联系起来相互印证。

因子相关矩阵给出了三个潜在因子之间的相关系数。表15-22表示三个因子之间相关接近于0，说明在允许因子相关时，抽取的三个因子也是接近互相正交的。可以看出，工作时间变量于其它因子相关几乎为0。回头检查工作时间与其它变量之间的相关可以看出，相关都接近于0，而且不显著。因此，在本例中将工作时间变量取出，对因子结构影响不大。读者可以自行练习。

表15-22 潜在因素相关系数表Component Correlation Matrix

成份	1	2	3
1	1.000	-.131	.040
2	-.131	1.000	-.002
3	.040	-.002	1.000

提取方法 :主成份。

旋转法 :具有 Kaiser 标准化的斜交旋转法。

但是需要指出的是，在实际的因子分析实践中，有时增加或者删除一个变量，因子的结构可能发生比较大的变化，因此变量的增加或删除需要重新做因素分析的全过程。

因素分析得到因子与变量之间的关系之后，可以通过观测变量来描述因子，因此每个个体都可以由它们在观测变量上的取值计算得到相应的因子得分，这一点也可以在SPSS中直接做到并保存为新变量。对于因子得分可以作进一步的统计分析。

小 结

本章主要介绍了多元方差分析，典型相关分析，聚类分析以及因素分析等几种比较常用的多元统计方法的思路以及SPSS操作及其结果的分析。

多元方差分析主要用于考察自变量在多个因变量上是否有差异。一般要求因变量之间有一定程度的相关，在这种情况下多元方差分析的结果与多个一元方差分析的结果会有不同。典型相关分析主要用于考察两组变量之间的相关，其基本思想是寻找各组变量的线性组合（称为典型变量），使得两组变量的线性组合之间的相关（称为典型相关系数）尽可能大。聚类分析主要是用于将样品或者变量按照近似程度进行分类。可以采用距离或这相似性程度来描述样品或变量之间的接近程度。常用的聚类方法有层次聚类和迭代聚类两种，其中层次聚类可以对变量进行，也可以对个体观察值进行，但是个体观察值或变量一旦归入某类，其结果就不可能改变；迭代聚类只能对个体观察值进行，个体观察值或变量的分类结果可以调整，但是需要预先知道类别数。因素分析是寻找新变量以重新再现变量之间内部关系的一种降维方法，一般要求变量之间有一定程度的相关，它可以用来探索变量之间的内部结构。在获得初始因素负荷矩阵之后，为了更好地给因子命名和简化结构，往往需要进行因素旋转。根据变量在因素上的负荷高低，可以对观测变量进行分类，也可以将观测变量转化为因素得分作进一步的分析。

关键术语：

多元统计分析方法：研究多个自变量与因变量相互关系的统计理论和分析方法。

多元方差分析：同时考察一个或多个分类变量在多个相关的因变量上是否有差异的统计方法。

典型相关分析：先将较多变量转化为少数几个典型变量，再通过其间的典型相关系数来综合描述两组多元随机变量之间关系的统计分析方法。

聚类分析：是一种采用数值的方法对个体观察值或者变量进行分类的多元统计方法。

Q型聚类：对个体观察值点进行的聚类。

R型聚类：对变量进行的聚类。

系统聚类：又称作谱系聚类或层次聚类，是将每个样品各自看成独立的一类，然后根据相似性高或距离近的原则逐一合并，最终归为一类的聚类方法。

动态聚类：先将所有个体观察值进行粗略的初始分类，然后根据某种原则对初始分类进行调整，直到分类结果稳定时为止的聚类方法。

因素分析：是指从大量的相关变量中抽取出最基本的维度或因素并加以分析的一组多元统计分析方法的总称。

探索性因素分析：在事先不知道可能的潜在因子的基础上,根据数据进行内部结构的探索，这种因素分析被称为探索性因素分析。

验证性因素分析：根据已有的理论或知识,在已知因子的情况下检验所搜集的数据资料是否支持某种理论假设的因子结构，这种因素分析被称为验证性因素分析。

公共因子方差（共同度）：是指观测变量方差中由公因子决定的比例。

因子旋转：通过改变坐标轴的位置来重新分配各个因子所解释的方差比例，进而使变量在因子上的负荷更加清晰，因子结构更为简单、更易于解释的一种方法

思考与练习：

1. 多元方差分析的适用条件是什么？它与多因素方差分析有什么不同？
2. 典型相关是用来处理什么情况的？它的基本思路是什么？
3. 聚类分析中系统聚类和动态聚类各有什么优缺点？
4. 因素分析的基本步骤如何？它与R型聚类有何区别与联系？