

教育与心理研究中

数据分析方法

潘玉进 编著



科学出版社

www.sciencep.com

教育与心理研究中 数据分析方法

科学出版社

北京

内 容 简 介

本书从教育与心理研究人员必备的数据分析方法入手,阐述了数据采集、录入、整理、显示、推断等方面的内容,具体内容包括:教育与心理研究概述、抽样设计、数据的采集录入与整理、描述统计分析、推断统计分析、*T*检验、方差分析、非参数检验、相关分析、回归分析、因素分析、信度分析与多选题分析。

本书关注实用性和操作性,使读者能在具体的研究情境中体会数据分析方法的应用。书中并不涉及公式的推导和定理的证明,重点介绍数据分析方法的原理、使用条件、SPSS 操作步骤和 SPSS 输出结果的解读,以提高读者数据的整理能力和分析能力。

本书的读者对象是从事教育与心理问题研究的教师、研究人员或本科生、研究生。

本书实例的 SPSS 数据文件可从 <http://jpkc.wzu.edu.cn/xltjx> 下载。

图书在版编目(CIP)数据

教育与心理研究中数据分析方法/潘玉进编著. —北京:科学出版社, 2010

ISBN 978-7-03-028784-7

I. ①教… II. ①潘… III. ①统计分析—统计程序—应用—教育统计—多元分析 ②统计分析—统计程序—应用—心理统计—多元分析 IV. ①G40-051②B841.2

中国版本图书馆 CIP 数据核字(2010)第 166156 号

责任编辑:沈力匀 / 责任校对:耿耘
责任印制:吕春珉 / 封面设计:东方人华平面设计部

科学出版社出版

北京东黄城根北街 16 号
邮政编码:100717

<http://www.sciencep.com>

骏杰印刷厂印刷

科学出版社发行 各地新华书店经销

*

2010 年 9 月第 一 版 开本:787×1092 1/16
2010 年 9 月第一次印刷 印张:25
印数:1—3 000 字数:600 000

定价:40.00 元

(如有印装质量问题,我社负责调换〈环伟〉)

销售部电话 010-62140850 编辑部电话 010-62135235 (HP04)

版权所有,侵权必究

举报电话:010-64030229; 010-64034315; 13501151303

前 言

受 T. Kunn (1962) 的范式概念和理论的影响, 在教育与心理研究领域, 人们也归纳出了许多研究范式, 大致可以分为实证主义的量化研究范式和建构主义的质化研究范式两个极端。虽然两者“殊途同归”, 都是要揭示蕴含在教育与心理现象背后的本质与规律, 但质化研究范式侧重于通过归纳、生成假设和理论的探究模式获得知识, 它更需要研究者具有敏锐的观察力和独特的创新思维, 对研究者理论素养的要求也相对较高; 而且研究者往往从个人的经验出发, 用思辨的方法对教育与心理现象进行研究, 其研究成果可能具有较大的主观随意性。相对而言, 在量化研究范式中, 由于数据挖掘技术的原理和方法已在数理统计学得到推导和证明, 因而只要所采集的数据可靠, 所选择的处理方法得当, 研究者无须非常严密的思维, 凭借数据挖掘技术, 就可以较为容易地从大量的数据资料中概括出蕴含在其背后的本质规律。这一点对于尚处在在职前教育阶段的大学生和长年工作在第一线的中小学教师来说尤为重要。

掌握数据分析方法的意义有二。首先, 在教育与心理研究过程中, 当选择数据分析方法时, 由于对自变量、因变量、控制变量的操作与控制方法不同, 被试的选择方法与分组方法不同, 其所采用的数据分析方法也大不相同。因此, 数据分析方法的学习过程实际上也是掌握研究方法的过程。其次, 掌握数据分析方法可以顺利阅读运用统计方法进行定量分析的研究文献。在现代教育与心理研究文献中, 大多采用各种统计方法来表述、说明与解释其研究成果。如果不掌握统计的符号、概念及思维方法, 就不能顺利阅读有关文献, 无法学习先进的研究成果。

本书的撰写有两个特点。

一是侧重于帮助读者根据教育与心理研究的目的和内容, 选择合适的数据分析方法, 并运用 SPSS for Windows 统计软件包得到研究结论。至于统计推断过程中所涉及的定理、公式以及复杂的计算过程, 只要了解其基本原理即可, 因为这些定理和公式都已在数理统计学中得到证明和推导, 复杂的计算过程也已由 SPSS 代劳。研究者更应关注的是如何选择恰当的数据分析方法来分析研究所采集的各种数据, 如何准确理解 SPSS 输出的数据处理结果。我经常给我的学生讲的一个例子, 我们这些从事教育与心理研究的人就好像厨师。对于一个厨师而言, 除了要了解各种刀具的不同性能外, 更重要的是要练好刀法, 至于刀具是如何设计、如何制造的, 那是刀具设计师和工程师的事情, 是制造工厂的事情。同样, 对于从事教育与心理研究的我们来说, 除了要了解数据分析的基本原理外, 最重要的就是要如何用好数据分析方法这个工具, 为研究实践服务, 至于公式的推导和定理的证明, 那是专门从事数理统计学研究的人的事情。

二是帮助读者将教育与心理的研究实践与 SPSS for Windows 统计软件包有机地结合起来。虽然我们将统计推断中复杂的计算过程交由 SPSS 来运算, 但本书并不是 SPSS 统计软件包的使用说明书。我们将根据数据分析方法的不同类型, 逐个说明每种数据分析方法的内涵和使用条件, 并在此基础上, 说明如何建立 SPSS 数据文件, 调用哪些 SPSS

过程和命令，并进行哪些必要的设置，以及如何解读 SPSS 输出的分析结果。

衷心地希望本书能对在读学生和广大中小学教师有所帮助！

言 语

潘玉进

2010年8月8日

自20世纪80年代以来，随着计算机的广泛应用，教育研究中的定量研究方法得到了迅速发展。在定量研究方法中，统计方法是应用最广泛、最普遍的方法。然而，由于统计方法的复杂性，许多教育研究者往往感到困惑和不知所措。本书正是在这样的背景下编写而成的。本书旨在帮助教育研究者了解统计方法的基本原理和应用，掌握 SPSS 软件的操作方法，并能对 SPSS 输出的分析结果进行正确的解读。本书共分 10 章，主要内容包括：统计推断的基本原理、参数估计、假设检验、方差分析、回归分析、多元统计等。本书力求做到概念清晰、重点突出、循序渐进、由浅入深。本书可作为教育专业及相关专业的教材，也可供从事教育研究工作的教师参考。

本书在编写过程中，参考了国内外许多优秀的统计教材和文献，并得到了许多同行专家的指导和帮助。在此，我表示衷心的感谢。同时，我也希望广大读者能对本书提出宝贵的意见和建议，以便今后进一步修订和完善。本书的出版得到了许多领导和同事的支持，在此，我也表示衷心的感谢。本书的出版，是我多年从事统计教学和研究的结晶，也是我从事教育研究的成果。本书的出版，不仅是我个人的幸事，也是广大教育研究者的幸事。本书的出版，将为广大教育研究者提供一本实用的统计工具书，也将为教育研究事业的发展做出应有的贡献。

目 录

第一章 教育与心理研究概述	1
第一节 教育与心理实验研究	1
第二节 数据的误差	3
第三节 数据的类型	5
一、点计数据与度量数据	5
二、离散数据与连续数据	5
三、称名数据、顺序数据、等距数据与比率数据	5
第四节 概率与概率分布	7
一、概率的一般概念	7
二、二项分布	8
三、正态分布	11
第二章 抽样设计	16
第一节 抽样设计概述	16
一、概述	16
二、抽样设计的原则	19
第二节 概率抽样方法	20
一、概述	20
二、常用的概率抽样方法	20
第三节 非概率抽样方法	22
一、概述	22
二、常用的非概率抽样方法	23
第三章 数据的采集、录入与整理	26
第一节 数据的采集与录入	26
一、数据的采集	26
二、数据的录入	29
第二节 数据的整理	36
一、数据的搜索	36
二、变量的插入与删除	37
三、个案的插入和删除	37
四、根据已有变量生成新变量	38
五、变量排序	40

六、数据的排序	43
七、数据的行列转置	44
八、数据文件的分组	45
九、数据文件的筛选	47
十、数据文件的合并	48
第四章 描述统计分析	51
第一节 数字特征量	51
一、集中量	51
二、差异量	52
三、相关量	57
第二节 统计图表	60
一、条形图	60
二、直方图	65
三、面积图	67
四、饼图	69
五、线图	71
六、箱图	73
七、频数分布表	75
第三节 交叉表分析	79
一、概述	79
二、SPSS 操作过程	79
三、实例	83
第五章 推断统计分析	85
第一节 推断统计概述	85
一、统计推断的意义	85
二、统计推断的内容	85
第二节 抽样分布	86
一、抽样分布的概念	86
二、样本平均数的抽样分布	88
三、样本平均数与总体平均数离差统计量的形态	88
第三节 参数估计	89
一、点估计和区间估计的概念	90
二、区间估计的原理和方法	90
第四节 假设检验	91
一、假设检验的基本原理	91
二、假设检验的几个基本概念	94

第六章 T 检验	98
第一节 单一样本 T 检验	98
一、概述	98
二、SPSS 操作过程	99
三、实例	99
第二节 总体平均数差异的显著性检验	102
一、概述	102
二、统计检验量	102
第三节 相关样本 T 检验	103
一、概述	103
二、SPSS 操作过程	104
三、实例	104
第四节 独立样本 T 检验	107
一、概述	107
二、SPSS 操作过程	108
三、实例	109
第七章 方差分析	115
第一节 方差分析的基本原理	115
一、概述	115
二、方差分析的基本原理	115
三、方差分析的假设检验	117
四、方差分析中的几个术语	117
五、方差分析过程	119
第二节 单因素 ANOVA 过程	121
一、概述	121
二、SPSS 操作过程	121
三、实例	124
第三节 单变量过程	131
一、概述	131
二、SPSS 操作过程	132
三、实例	137
第四节 多变量过程	155
一、概述	155
二、SPSS 操作过程	156
三、实例	157
第五节 重复测量过程	168
一、概述	168

二、SPSS 操作过程·····	169
三、实例·····	171
第六节 协方差分析·····	204
一、概述·····	204
二、SPSS 操作过程·····	205
三、实例·····	205
第八章 非参数检验·····	226
第一节 概述·····	226
一、参数检验与非参数检验·····	226
二、非参数检验的种类·····	226
第二节 χ^2 检验·····	227
一、概述·····	227
二、拟合优度 χ^2 检验的 SPSS 操作过程·····	228
三、独立性或同质性 χ^2 检验的 SPSS 操作过程·····	230
四、实例·····	230
第三节 单一样本 K-S 检验·····	236
一、概述·····	236
二、SPSS 操作过程·····	236
三、实例·····	237
第四节 两个独立样本的非参数检验·····	239
一、概述·····	239
二、SPSS 操作过程·····	241
三、实例·····	242
第五节 多个独立样本的非参数检验·····	247
一、概述·····	247
二、SPSS 操作过程·····	249
三、实例·····	249
第六节 两个相关样本的非参数检验·····	253
一、概述·····	253
二、SPSS 操作过程·····	254
三、实例·····	255
第七节 多个相关样本的非参数检验·····	259
一、概述·····	259
二、SPSS 操作过程·····	261
三、实例·····	261

第九章 相关分析	266
第一节 相关分析的概念与过程.....	266
一、概述.....	266
二、相关分析的 SPSS 过程.....	266
第二节 积距相关与等级相关.....	267
一、概述.....	267
二、SPSS 操作过程.....	269
三、实例.....	271
第三节 质与量的相关和品质相关.....	275
一、概述.....	275
二、SPSS 操作过程.....	276
三、实例.....	277
第四节 偏相关分析.....	284
一、概述.....	284
二、SPSS 操作过程.....	285
三、实例.....	285
第十章 回归分析	289
第一节 线性回归.....	290
一、概述.....	290
二、SPSS 操作过程.....	295
三、实例.....	300
四、含虚拟自变量的回归分析.....	309
第二节 曲线估计.....	315
一、概述.....	315
二、SPSS 操作过程.....	315
三、实例.....	318
第三节 逻辑回归.....	322
一、二分变量逻辑回归分析.....	322
二、多项逻辑回归分析.....	338
三、顺序变量逻辑回归分析.....	346
第十一章 因子分析	351
第一节 因子分析的基本原理.....	351
一、概述.....	351
二、因子分析的基本原理.....	351

第二节 因子分析过程	353
一、概述	353
二、SPSS 操作过程	353
三、实例	357
第十二章 信度分析与多选题分析	366
第一节 同质性信度	366
一、概述	366
二、SPSS 操作过程	366
三、实例	368
第二节 分半信度	371
一、概述	371
二、SPSS 操作过程	372
三、实例	372
第三节 再测信度	372
一、概述	372
二、SPSS 操作过程	373
三、实例	373
第四节 多选题分析	374
一、概述	374
二、SPSS 操作过程	375
三、实例	378
附录 正态分布表	383
主要参考文献	387

第一章 教育与心理研究概述

第一节 教育与心理实验研究

科学研究的根本目的是为了探索事物之间的因果关系，而实验法是探索事物之间因果关系最有效的研究方法之一。教育与心理实验研究也是如此，它是研究者按照研究目的充分地控制实验的环境，创设一定的实验条件，科学地选择研究对象，以确立自变量与因变量之间因果关系的一种研究方法。为确保教育与心理实验能顺利进行，必须科学、合理地进行实验设计，主要包括变量的操作与控制（即自变量如何操作、因变量如何检测、控制变量如何控制等）、被试的取样与分组（即被试在哪个范围内取样、如何取样、取多少被试、被试如何分组等）、数据的采集与处理（即数据如何采集、用何种统计方法进行分析等），其中“数据的采集与处理”主要涉及运用数理统计学的原理与方法，研究如何收集、整理和分析教育与心理研究中的数字资料，并以此为依据进行统计推断，揭示教育与心理现象所蕴含的客观规律。

例如要研究小学生的孤独感与父母教养方式的关系，其研究对象应该是所有的小学生，从地理维度上看，不仅包括研究者所在的学校、县、市，还应该包括所在的省，乃至全国所有的小学生；从时间维度上看，不仅包括现在的小学生，更应该包括未来若干年的小学生，否则将由于研究结果不能推广应用而使研究变得毫无意义。要全面观测如此庞大的研究对象，需要投入大量的人力、物力、财力与时间，这对一个小小的课题组来说是很困难的，而且对于未来若干年的小学生而言也是不可能的。在这种情况下，教育与心理统计就大有用武之地了。研究者可以根据研究目的，在研究对象的范围内适量地随机抽取一小部分小学生，比如随机抽取 200 个小学生进行观测，测量这 200 个小学生的孤独感及其父母教养方式，然后根据数理统计学中业已证明的有关原理与方法在一定的可靠性程度上（比如在 95% 的可靠性程度上）推测所有小学生的孤独感与父母教养方式的关系。

在这里，所有的小学生在统计学上称为“总体”，是研究的对象；而从中随机抽取的 200 个小学生则称为来自于这个总体的“样本”，是在研究中实际观测的对象。也就是说，可以根据从总体中随机抽取的一定量的样本数据来推测总体的特征，这就是统计推断的基本思想，如图 1-1 所示。因此，统计推断的重要意义就在于对不可能获得的总体，能对其各种特征及分布性质做出一定可靠程度的估计和推测。

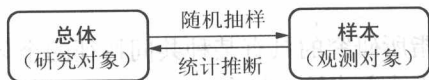


图 1-1 统计推断的基本思想

在数据的分析与处理过程中，将涉及随机现象与随机事件、总体与样本、参数与统计量等基本概念。

1. 随机现象与随机事件

在自然界和人类社会中，有时一旦某种条件实现，就必然会发生某种现象，这称为确定性现象。如化学中强酸与强碱的中和反应，或是一个盒子中有 10 个完全相同的白球，搅拌后从中任意摸取一球，必是白球等。但也有些现象是非确定性的，也就是说，在一定的条件实现后，可能发生也可能不发生的现象，也即在相同的条件下其出现结果具有不确定性、无法确切预知的现象，这就是随机现象（random phenomenon），它具有三个特征：

- (1) 一次试验有多种已知的可能结果。
- (2) 在试验之前其结果无法预料。
- (3) 在相同的条件下可重复试验。

例如，从某班级随机抽取 2 名学生，观测其性别，有三种已知的可能结果：二男、二女、一男一女；究竟是哪种结果，试验之前无法预料；在相同的条件下（每次抽样并观测后归队）可以重复抽取（试验）。又如，往上抛掷一枚硬币，有两种已知的可能结果：正面朝上和背面朝上；究竟哪一面朝上，试验之前无法预料；相同的条件下（钢性平面、硬币质地均匀、举到一定的高度）可以重复抛多次，这些都是随机现象。

随机现象的某种可能结果，称为随机事件（random event）。随机事件在一次试验中，可能发生，也可能不发生，表现出随机性；但是，在大量的重复试验中，它们又表现出统计规律，即随着试验次数的增加，某随机事件 A 发生的次数 f （也称为频数）在试验总次数 N 中所占的比率 f/N （这个比率称为该随机事件的频率，用 $W_{(A)}$ 表示，即 $W_{(A)} = f/N$ ）会逐渐稳定地逼近某一恒定值。

例如，设硬币的正面朝上为随机事件 A，背面朝上为随机事件 B。在向上抛掷一次硬币时，随机事件 A 可能发生，也可能不发生，但如果重复抛掷许多次，随机事件 A 的发生就会呈现出某种规律性，即它出现的频率接近 0.5，如表 1-1 所示。

表 1-1 抛掷硬币试验的可能结果

随机事件		试验次数					频率所逼近的恒定值
		10000	20000	30000	40000	50000	
正面朝上 (A)	频数	4875	9922	14941	19934	24946	—
	频率	0.4875	0.4961	0.4980	0.4984	0.4989	0.5
背面朝上 (B)	频数	5125	10078	15059	20066	25054	—
	频率	0.5125	0.5039	0.5020	0.5016	0.5011	0.5

2. 总体与样本

总体（population）是指所研究的具有某种共同特征的个体总和。总体可以是人，或是人所具有的人格特征、活动方式等，如全市高一学生、全省初二女生、沿海地区初中生的智力发展水平等；也可以是物，如所有的常用字、某次考试的全部试卷，某次测验

的所有分数等。

组成总体的基本单元称为个体 (individual), 如一个学生、一份试卷、一个测验分数等。总体分为有限总体和无限总体。包含有限个体的总体称为有限总体 (finite population), 如 2009 年某校中学生、2010 年某市的小学中高年级学生、2010 年参加全国高考的所有考生等; 包含无限个体的总体称为无限总体 (infinite population), 如所有的高一学生, 理论上包括了现在以及未来若干年的所有高一学生, 也包括全校、全县、全省, 乃至全国所有的高一学生。

如前文所述, 要对总体中的所有个体都进行观测, 不仅是不可能的, 也是不必要的, 因为这要花费大量的人力、物力、财力和时间, 一般难以实现。因此, 必须从总体中随机抽取一小部分个体进行观测, 再根据这部分个体所提供的信息在一定的可靠性程度上对总体的特征与分布性质进行推断, 如图 1-1 所示。

从总体中抽取的作为观测对象的一部分个体称为样本 (sample)。样本所包含的个体数目称为样本容量 (sample size), 一般用 n 表示。如某样本包含 65 个个体, 记为 $n=65$ 。

总体与样本是相对的。如研究某校学生的智力水平, 那么全校学生的智力水平就是总体, 从中随机抽取的 50 个学生的智力水平就是样本; 但如要研究全市学生的智力水平, 那么该校的学生智力水平就可以作为样本。

3. 参数与统计量

描述总体分布特征的量数, 称为参数 (parameter)。例如, 描述总体集中趋势的有总体平均数 (μ), 描述总体内个体间离散程度的有总体标准差 (σ), 描述两个总体之间相关关系的有总体相关系数 (ρ) 等。

描述样本分布特征的量数, 称为统计量 (statistic)。例如, 描述样本集中趋势的有算术平均数 (mean)、中位数 (median)、众数 (mode) 等; 描述样本离散程度的有四分位数 (quartile)、差异系数 (coefficient of variation)、标准差 (standard deviation) 等; 描述两个样本之间相关关系的有相关系数 (correlation coefficient) 等。

根据样本统计量来估计与推断总体参数, 正是统计推断的内容之一。

第二节 数据的误差

教育与心理统计中所说的误差, 泛指观测值与真值之差, 主要有以下三种。

1. 系统误差

在数据采集过程中, 由于仪器不准确, 指导语有暗示性, 或对某些标准掌握过宽、过严等原因, 导致数据成倾向性地偏大或偏小而引起的误差叫系统误差 (systematic error)。如每年的高考, 虽然各科试卷都有统一的评分标准, 但由于各地在掌握评分标准上有宽、严的差异, 以致评分标准掌握得宽的地区, 考生的得分会呈倾向性地偏高, 而评分标准掌握得严的地区考生的得分会普遍地偏低, 像这样的误差就属于系统误差。

系统误差有一定的倾向性, 如恒定性: 在多次测试中, 如果观测值比真值偏高就

普通偏高,比真值偏低则都普遍偏低;累增性:用同一方式多次观测时,累积误差会越来越大;周期性:如果观测系统的某部分发生偏差时,每到观测这部分时都会发生系统误差。

系统误差影响着原始数据的准确性,必须加以控制。但由于系统误差难以从原始数据中分析出来,因此其发现和控制较其他误差更困难。控制系统误差的措施主要是仔细分析观测的方法、量表及仪器,查明产生系统误差的原因,然后调整观测系统,重新进行观测;此外还可以改用另一种方法进行观测来确定是否存在系统误差。

2. 随机误差

在数据采集过程中,由于存在一些不易发现或无法控制的偶然因素,致使同一对象经同一方式观测多次,其结果都不相同,这样产生的误差叫随机误差(random error)。

随机误差一般都由众多微小的偶然因素造成。努力达到观测系统的精确性,可以缩小随机误差,但无法完全避免,无论观测的方法如何正确,仪器如何精密,量表如何规范、科学,随机误差总是存在的,但应控制在允许的范围之内。

随机误差虽不能完全消除,但它却有一定的统计规律性,这表现在以下几方面:

- (1) 随机误差不恒定,有时为正,有时为负。
- (2) 正负误差出现的可能性大致相等。
- (3) 小误差出现的可能性大于大误差出现的可能性,特别大的误差一般不会出现。
- (4) 同一对象被观测多次时,随着观测次数的增加,误差的算术平均数逐渐接近于零。

3. 抽样误差

随机样本的统计量与总体参数之差,称为抽样误差(sampling error)。例如,样本平均数与总体平均数的差异、样本标准差与总体标准差的差异等。抽样误差越小,则样本对总体的代表性越强;抽样误差越大,则样本对总体的代表性越弱。所以,抽样误差是测定样本代表性的指标之一。

在教育与心理研究中,即使严格控制了系统误差,并把随机误差控制在允许的范围之内,但随机样本的结构不可能与总体结构完全一致,所以,抽样误差不可避免。

抽样误差受样本容量、总体标准差和抽样方法的制约。在其他条件相同的情况下,样本容量越大,抽样误差越小,样本容量越小,抽样误差越大,样本容量与抽样误差成反比;总体标准差越大,抽样误差也越大,总体标准差越小,抽样误差也越小,总体标准差与抽样误差成正比;在样本容量相同时,分层随机抽样比简单随机抽样的抽样误差小。

样本统计量与总体参数既有差别,又有明显的一致趋势,是抽样误差的显著特点。这一特点表明样本与总体的辩证关系。总体所具有的必然性寓于随机获得的、带偶然性的样本之中,并通过样本数据反映出来。从个别数据看,误差有大有小,好像捉摸不定;但从大量数据看,从整体看,它是有规律性的。抽样误差虽然不能避免,但因它是随机产生的,故可根据概率统计原理估计其大小和范围,并通过抽样程序加以控制,将其缩

小到最低限度和控制允许的范围之内。

第三节 数据的类型

随机现象有多种可能结果，这种能表示随机现象各种可能结果的变量称为随机变量(random variable)，统计处理的变量都是些随机变量。例如，学生的性别、智商、学科考试成绩，教师的年龄、教龄、教学水平等。可用 X 、 Y 等符号分别表示不同的随机变量。例如，用 X 表示某班语文测验分数这一随机变量；用 Y 表示数学分数这一随机变量。而该班每个学生的语文分数可以分别用 X_1, X_2, \dots, X_n 表示，数学分数可以分别用 Y_1, Y_2, \dots, Y_n 表示。

数据(data)是随机变量的观测值，它是用来描述对客观事物观察测量结果的数值。当对某个随机变量进行观测时，事先不能预料会取到什么值；一旦某个值被取定，就称这个值为随机变量的一个观测值，即数据(也是一个随机事件)。

数据的种类不同，统计处理的方法也不同。根据不同的分类标准，数据可分为不同的类型。

一、点计数据与度量数据

根据数据的来源，可将数据分为点计数据和度量数据。点计数据(count data)是指通过计算个数所得到的数据。如学校数、班级数、学生数、教师数、教室数、图书册数等。点计数据一般取整数形式。度量数据(measurement data)是指用一定的测量工具对事物进行测量，或者按照一定的测量标准给对象赋值而得到的数据。如身高、体重、学科成绩、智商、反应时间、记忆保持量等。

二、离散数据与连续数据

根据数据的取值情况，可将数据分为离散数据和连续数据。取值个数有限的数，称为离散数据(discrete data)。这种数据的单位是独立的，两个单位之间不能再划分成更小的单位，一般用整数表示。例如，学生人数、某门课程不及格的人数、等级制的学科成绩等。取值个数无限的数据，称为连续数据(continuous data)。它们可能的取值范围能连续充满某一个区间，数据的单位之间可以再划分成无限多个细小的单位，一般可以用小数表示。例如，学生的身高、体重、智商、百分制的学科成绩等。

三、称名数据、顺序数据、等距数据与比率数据

根据数据的测量水平，可将数据分为称名数据、顺序数据、等距数据与比率数据。

1. 称名数据

称名数据(nominal data)只表示某一事物与其他事物在属性上的不同。它既不能排序，也没有相等单位，更没有绝对零点。例如，运动员的号码、学生的学号、性别、衣

服的颜色、不同的教学方法等。在这里，只给不同属性、不同类别的事物以不同的名称，这些名称的排列顺序无关紧要，在处理这类数据时，既不能排序，也不能进行加减乘除运算。

2. 顺序数据

顺序数据 (ordinal data) 是指可以就事物的某一属性的多少或大小依次排序，但没有相等单位和绝对零点的数据资料。例如，比赛中的第一、二、三名，学科成绩的优、良、中、合格、待合格等。对于这类数据，不仅可以从所赋予的名称来区别各事物之间的不同，而且可以排出它们在某一特质方面的多少或大小。顺序数据与称名数据的主要区别在于它有“方向次序”。在处理这类数据时，仅能进行排序，但不能进行加减乘除运算。

应注意的是，这里所指的“排序”与计算机软件（如 Excel）中所涉及的“排序”完全是两回事。这里的“排序”是按照事物本身属性的大小或多少来进行，而计算机软件中的排序依据的是人为制定的规则，如英文字符是根据其 ASCII 码的大小进行排序，中文字符则是根据其内码的大小进行排序。

3. 等距数据

等距数据 (interval data) 是指具有相等单位，但无绝对零点的数据资料。例如气温，假设周一的气温为 0°C ，周二的气温为 5°C ，周三的气温为 10°C ，首先，周一气温的 0°C ，并不表示气温真的为零，没有温度，它只是一个相对的零点；其次，周一、周二、周三的气温各不相同；第三，周三的气温高于周二，周二的气温高于周一；第四，周三与周二的气温之差，等于周二与周一的气温之差。再如百分制的数学成绩，因学业考试考核的是学生已掌握的数学知识与技能，如果某个学生考了零分，这个零分并不表示该生已掌握的数学知识技能真的为零，它只是一个相对的零点。因此，等距数据没有绝对零点，它只是一个相对的数值，而不能作比率性描述。在对这类数据进行统计处理时，不仅能排序，也能进行加减运算，但不能进行乘除运算。

4. 比率数据

比率数据 (ratio data) 是指具有相等单位和绝对零点的数据资料。这类数据可以作比率性描述，在统计处理时，既可进行加减运算，也可进行乘除运算。例如，身高、体重等。假如一个成年人的身高是 176cm ，一个儿童的身高是 88cm ，那么，该成年人的身高比该儿童高 88cm ，同时也是该儿童身高的 2 倍。在处理这类数据时，不仅能排序，也能进行加减乘除运算。

不同测量水平的数据之间可以进行转换，但只有比率数据或等距数据可以转换为顺序数据，而且转换后数据所包含的信息量将会损耗。例如，百分制的学科成绩（等距数据）可以按照一定的规则转换为等级制的学科成绩（顺序数据），例如，60 以下为“待合格”，60~69 为“合格”，70~79 为“中”，80~89 为“良”，90 及以上为“优”，但应该看到，转换后的数据只保留了彼此间大小的顺序关系，而相差的量已被抛弃。