魏魏交大　百年书香
www.jiaodapress.com.cn
bookinfo@sjtu.edu.cn

策划编辑　张宇晗
责任编辑　胡思佳
封面设计　刘文东

统计学
TONGJIXUE

扫描二维码
关注上海交通大学出版社
官方微信

ISBN 978-7-313-33719-1
9 787313 337191 >
定价：58.00元

统计学

主编　睢华蕾

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

高等院校经济管理类系列教材
“互联网＋”创新型教材

统计学
TONGJIXUE

主编　睢华蕾

上海交通大學出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

高等院校经济管理类系列教材
“互联网 + ”创新型教材

统计学

TONGJIXUE

主　编　睢华蕾
副主编　徐　梅　李小巍　席晓娟

上海交通大学出版社
SHANGHAI JIAO TONG UNIVERSITY PRESS

内容提要

本书以统计学中的描述统计与推断统计的方法体系为基础构建内容框架，包括统计与统计数据、数据的搜集、数据的可视化基础、数据的描述统计、推断统计导论、参数估计、假设检验、列联分析、方差分析、相关关系与回归分析、主成分分析与因子分析等十一章内容，系统讲解了统计设计、统计调查、统计整理和统计分析等环节的重要理论和实践技能。

本书既可作为高等院校经济学、管理学等专业的教材，也可作为相关领域实务工作者的参考用书。

图书在版编目（CIP）数据

统计学 / 睢华蕾主编. -- 上海 : 上海交通大学出版社, 2025.10. -- ISBN 978-7-313-33719-1

Ⅰ. C8

中国国家版本馆 CIP 数据核字第 2025DP4307 号

统计学

TONGJIXUE

主　　编：睢华蕾

出版发行：上海交通大学出版社　　地　　址：上海市番禺路 951 号

邮政编码：200030　　电　　话：021-64071208

印　　制：三河市龙大印装有限公司　　经　　销：全国新华书店

开　　本：787 mm×1 092 mm　1/16　　印　　张：17.5

字　　数：505 千字

版　　次：2025 年 10 月第 1 版　　印　　次：2025 年 10 月第 1 次印刷

书　　号：ISBN 978-7-313-33719-1

定　　价：58.00 元

前言

数据已成为驱动全球经济发展、社会治理和人际交往方式变革的重要力量。党的二十大报告多次提及数字经济、现代产业体系、数字贸易、教育和文化的数字化转型，以及数据安全等议题，全面展现了以数据为关键要素推动社会各领域创新发展的战略部署。

在数字经济时代，数据搜集、存储、分析处理、呈现成为一个部门或单位的必要基础工作。因此，培养兼具专业理论、数据思维与统计分析能力，能够跨领域整合专业技术与数据分析技术的人才是高等教育阶段数字化人才培养的目标。

统计学为高校构建数字化人才培养新路径提供支撑，重点培养学生的数据分析能力，包括复杂数据处理、规律探究及问题解决能力。“培养什么人、怎样培养人、为谁培养人”是教育的根本问题，统计学应培养学生的数据思维与创新能力，为中国式现代化建设输送人才。基于此，本书既针对初步掌握数理统计与概率论等基础知识的学生，也兼顾法学等文科专业的学生，以实际数据分析为纽带，结合软件实操，帮助学生在习得统计分析技术的同时，初步建立数据思维，逐步形成严谨规范的科学态度。

本书在内容安排上主要具有以下特色。

(1)融理论教学于实践案例之中。本书将教学侧重点放在统计方法的应用层面，通过“背景启题”“知识拓展”“实战演练”等栏目提供丰富的统计学理论方法和实战技巧，帮助学生破解知识难点和重点，培养学生灵活运用统计方法解决实际问题的能力。

(2)兼顾理论性与可读性。本书设置了“案例分析”“课堂讨论”“问题思考”等栏目，力求将抽象的理论还原为生活工作场景，培养学生树立正确的统计理念与统计价值观，时刻关注中国经济发展动态，充分发挥育人作用。

(3)关注对统计结果的解读与软件的应用。本书详细讲解了各类统计方法在 Excel 和 SPSS 中的操作过程和步骤，结合案例给出了详细的统计结果解读，具有工具性和借鉴性。

本书各章的学时分配建议如下。

内　　容	理论学时	实践学时
第一章　统计与统计数据	2	
第二章　数据的搜集	2	2
第三章　数据的可视化基础	4	2
第四章　数据的描述统计	4	2
第五章　推断统计导论	2	

续表

内　　容	理论学时	实践学时
第六章　参数估计	4	2
第七章　假设检验	4	2
第八章　列联分析	4	2
第九章　方差分析	4	2
第十章　相关关系与回归分析	4	2
第十一章　主成分分析与因子分析	2	2
总学时(54)	36	18

本书由西北政法大学雎华蕾任主编，徐梅、李小巍、席晓娟任副主编，杨贾、李圣楠、周合欢、梁新翊参与编写。具体编写分工如下：雎华蕾编写第一章至第六章，徐梅编写第七章至第十一章，李小巍、席晓娟搜集并整理部分案例资料，杨贾、李圣楠、周合欢完成案例分析、数据分析与软件操作工作，梁新翊承担校阅工作。雎华蕾负责全书的统稿。

本书为陕西省高等教育学会、西北政法大学教学改革项目支持下的成果，历经多轮教学实践检验，授课效果良好，最终成稿是学科融合与教学相长的结晶。本书的出版得到了西北政法大学的资助，特此鸣谢。石红溶副教授提供了大量参考资料和具有可行性的修改意见，在此一并表示感谢。

由于编者水平有限，书中难免存在错误之处，恳请广大读者批评指正。

编　者

目录

第一章

统计与统计数据

学习目标

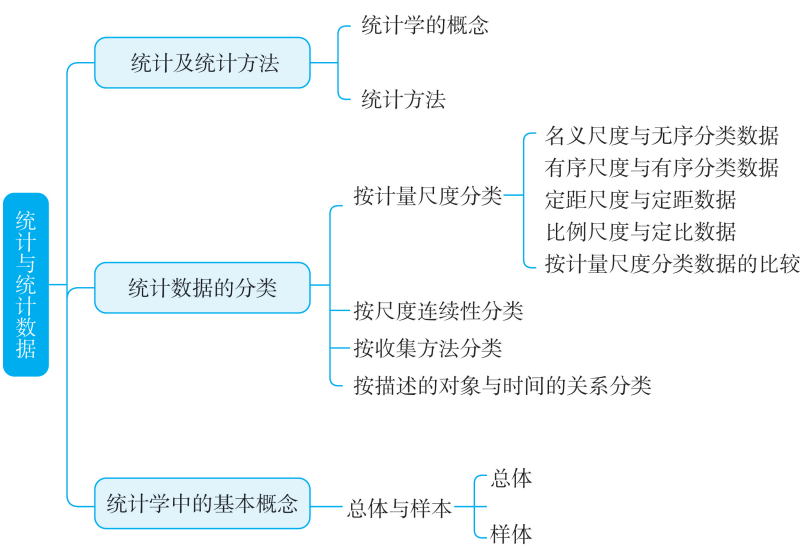
1. 理解统计与统计方法的基本概念；掌握总体、总体单位、样本、参数、统计量、标志与指标的概念。

2. 熟悉统计数据的分类标准（计量尺度、尺度连续性、收集方法、被描述的现象和时间的关系等），能够根据实际需求灵活区分数据类型。

素养目标

认识统计学在数据分析中的作用，理解统计方法在展现我国社会主义建设成就中的作用。

思维导图



案例导入

2024年4月16日，国务院新闻办公室举行新闻发布会，国家统计局副局长盛来运就2024年一季度国民经济运行情况作了介绍，并回答记者提问。以下为盛来运副局长答记者问的部分节选。

记者：

我们刚刚看到发布的数据，整体上来看，主要的经济指标在2024年一季度表现得比较不错。请问您如何评价2024年一季度的经济运行表现？

盛来运：

谢谢您提问。2024年一季度经济运行指标数据确实表现得不错。各地区、各部门认真贯彻落实党中央、国务院决策部署，加大宏观政策实施力度，政策靠前发力，狠抓落实，推动2024年一季度国民经济持续回升、开局良好。我想用四个关键词来评价一季度的经济运行情况。

第一个关键词是“持续回升”。2024年一季度，国民经济总体上延续了去年以来回升向好态势，生产需求主要指标稳中有升。从生产来看，第一产业增长总体稳定，增加值同比增长3.3%。第二产业增加值增长6%，其中规模以上工业增加值增速比2023年同期提升3.1个百分点，比2023年四季度提升0.1个百分点，呈现稳中有升。服务业增加值增长5%，住宿餐饮、交通旅游等接触型服务业继续保持较快增长。在2023年服务业高基数的基础上，2024年服务业继续稳定增长。

从三大需求来看，投资、消费、进出口几个指标增长总体稳定、稳中有升。2024年一季度固定资产投资同比增长4.5%，2023年全年增长3%，提升1.5个百分点。社会消费品零售总额增长4.7%，服务零售额增长10%。根据海关发布的数据，进出口增长5%，创6个季度以来新高。从三大需求来看，也是稳中有升。

第二个关键词是“起步平稳”。具体从增长、就业、物价、国际收支四大宏观指标来看2024一季度经济运行态势。国内生产总值(gross domestic product，GDP)同比增长5.3%，2023年四季度增长5.2%，也是稳中有升；全国城镇调查失业率平均值为5.2%，比2023年同期下降0.3个百分点，就业形势总体改善；居民消费价格指数(consumer price index，CPI)与2023年一季度持平，扣除食品和能源后的核心CPI同比上涨0.7%，与2023年各季度差不多；国际收支总体平衡。从以上四大宏观指标来看，2024年一季度经济运行总体稳定，起步平稳。

第三个关键词是“稳中有进”。国民经济在实现量的合理增长的同时，还实现了质的有效提升，高质量发展继续取得新进展。创新发展取得新成效。2024年一季度，规模以上高技术制造业增加值同比增长7.5%，比2023年四季度加快2.6个百分点。协调发展中产业结构和需求结构继续改善，内需对经济增长的贡献率达到85.5%，且产业和需求的内部结构也都有积极改善。绿色发展继续取得新成效，单位GDP能耗同比下降0.1%。高水平对外开放深入推进，我国对共建“一带一路”国家进出口总额增长5.5%。共享发展继续取得新进步。全国居民人均可支配收入增长6.2%。这些指标说明2024年一季度高质量发展继续取得新进展。

第四个关键词是“开局良好”。正是因为经济持续恢复、起步平稳、稳中有进，所以实现了经济良好开局。从经济运行主要指标来看，经济运行的稳定性、协调性增强，市场活力增强，市场经营主体信心提升。制造业采购经理指数(purchasing managers'index，PMI)3月份重新回升到景气区间，为50.8%。

当然，我们也清醒地认识到，目前外部形势仍然复杂严峻，国内正处在结构调整转型的关键阶段，市场经营主体的信心和经济回升的动力都需要进一步增强。下一步，要坚决贯彻落实党

中央、国务院关于促进经济稳定发展的决策部署，进一步加大政策落实力度，固本培元，持续增强市场信心和经济运行动力，巩固和增强经济持续回升向好的基础。谢谢！

资料来源：国家统计局. 国家统计局副局长就2024年一季度的国民经济运行情况答记者问[EB/OL]. (2024-04-16)[2025-07-01]. https://www.stats.gov.cn/sj/sjjd/202404/t20240416_1948587.html.（有改动）

请思考：

(1)盛来运在答记者问的过程中，从哪几个方面说明了2024年一季度经济运行情况？

(2)用来代表经济运行情况的关键指标有哪些？

(3)在我们日常生活中，还有哪些与统计有关的活动呢？

第一节　统计及统计方法

一、统计的概念

“统计”一词通常有三层含义，即统计活动、统计数据与统计科学。

统计活动是根据研究目的对现象的数量方面进行调查研究的实践活动，包括数据收集、数据处理、数据分析、数据解释及结果阐释五个基本环节。数据收集是指取得统计数据，数据处理是指将数据用图表等形式呈现，数据分析是指选择适当的统计方法研究数据，数据解释是指说明数并据分析结果，而从数据中提取有用的信息并得出客观结论就是指结果阐释。

统计活动的直接成果是统计数据。统计数据是指通过统计方法对现象的特征进行描述、计量和记录的结果，通常以汇总的形式展现。

统计科学简称统计学，统计学通过总结、归纳、概括为统计活动提供应遵循的理论与方法体系，是收集、处理、分析、解释数据并从数据中得出结论的科学。

由此可见，关于统计的三层含义并不是割裂的，而是存在紧密的内在联系：统计活动和统计数据是过程与成果的关系，统计活动和统计学则是实践与理论的辩证关系。

知识拓展

统计学的历史发展

统计学的历史发展生动诠释了统计学的三层含义。

1. 萌芽阶段

统计学在17世纪中叶至18世纪中叶萌芽于欧洲。创立时期主要有国势学派和政治算术学派。国势学派产生于17世纪的德国，以文字记述国家的显著事项，主要用对比分析的方法研究了解各国实力的强弱，为德国的君主政体服务。国势学派偏重对事物性质的解释，并不注重数量对比和数量计算，但其研究方法为统计学奠定了理论基础。特别是在外文中，“statistik”一词原本意为“国势”，这一概念后来被用来命名统计学(statistics)这门新兴学科，并沿用至今，成为全球公认的术语。随着资本主义市场经济的发展，对事物量的计算和分析越来越重要，国势学派后分化为图表学派和比较学派。

政治算术学派产生于17世纪中叶的英国，代表人物是威廉·配第(William Petty)和约翰·格朗特(John Graunt)。政治算术学派在分析社会经济问题时开始将统计方法与数据计算和推理方法结合起来。威廉·配第在1676年完成的《政治算术》中利用实际资

料对英国、法国、荷兰的国力进行数量对比分析，为统计学的形成和发展奠定了方法论基础。约翰·格朗特以伦敦教会发表的“死亡公报”为研究资料，在1662年发表的《关于死亡公报的自然和政治观察》中分析了60年来伦敦居民死亡的原因及人口变动的关系，首次提出通过大量观察可以发现新生儿性别比例具有稳定性和不同死因的比例等人口规律，并且第一次编制“生命表”，对死亡率与人口寿命进行分析，从而引起普遍的关注。

2. 发展时期

从18世纪末至19世纪末，统计学进入了发展时期，形成了两个主要学派：数理统计学派和社会统计学派。19世纪中叶，阿道夫·凯特勒把概率论引进统计学而形成数理统计学派，主张从总体中随机抽取部分或少数样本加以观察实验，用取得的数据进行推断，构建数学模型。这为分析社会经济现象提供了通用性的方法工具，并为其他学科研究服务。

社会统计学派产生于19世纪后半叶，创始人是德国经济学家、统计学家克尼斯(Knies)。早期社会统计学派认为，统计学是一门以社会经济活动为研究对象的实质性科学，而非数理统计学派所强调的通用方法。后来，社会统计学派融合了国势学派与政治算术学派的观点，沿着阿道夫·凯特勒的“基本统计理论”向前发展，将数量统计方法大量运用于解释社会经济现象，推动统计学发展成为既是实质性科学又是方法论的科学。

3. 快速发展时期

20世纪以来，统计学进入快速发展时期。现代所使用的“统计”一词被认为出自费舍尔(Fisher)所著《研究工作者的统计方法》一书中，这本著作指出，“统计学是指从观察到的样本中得来的值，目的是说明它来自群体的特征”。此后，统计学家们补充了许多有效分析数据的统计方法，并将这些新方法以自己的名字命名，这些统计方法一直被沿用到今天。

随着社会经济的发展、统计理论的完善及应用范围的推广，统计学仍在不断地发展。统计学由描述统计向推断统计发展，由社会、经济统计向多分支学科发展，发展成为有着许多分支学科的科学，成为通用的方法论；信息论、系统论、控制论与统计学相互渗透和结合，使统计学进一步发展并日趋完善；计算技术等一系列新技术、新方法在统计领域不断得到开发和应用，使统计学在现代化管理和社会生活中的地位变得日益重要。

4. 总结与展望

统计学是一门既古老又崭新的科学。说它古老，是因为它已有300多年的历史；说它崭新，是因为统计学走进了知识经济时代，仍在快速发展。梳理统计学的发展历史可以看出，国势学派和政治算术学派虽然最终走向了分化，但为统计学的发展提供了深刻的理论基础。无论是社会统计学派，还是数理统计学派，都或多或少地保留了它们的理论痕迹。社会统计学派虽然在发展初期为显示其独特性，对数理统计方法加以排斥，但最终还是大量使用数理统计方法来对社会经济现象加以阐释和分析，实际上统计学的发展是社会统计学和数量统计学相辅相成、共同作用的结果。今天，统计学拥有了更多更新的统计方法和手段，有了更多的研究对象和更广泛的应用领域，显示出更加重要的作用和更广阔的发展前景。

资料来源：邹庆华. 统计学的三个发展阶段及其性质分析[J]. 统计与决策，2003，(11)：15—16.

二、统计方法

统计学将理论与实践结合，既是一门实质性学科，也为统计分析提供方法论。统计方法可分为描述统计(descriptive statistics)与推断统计(inferential statistics)。描述统计研究的是针

对收集的数据，通过数据处理、汇总、图表描述、概括与分析等统计方法，得出简明扼要且具有一定意义的结论。

推断统计研究的是如何利用样本数据来推断总体特征的统计方法。例如，想要了解一个地区的宏观经济走势情况，没有必要对每一个经济主体进行测量，可采取一定的随机抽样的方法确定样本，通过测量样本中市场主体的相关指标来推断该地区的宏观经济走势。

开展推断统计主要基于以下几方面原因。

(1)调查并测量总体需要大量的时间和成本，而样本结果足以推断总体且更加经济。

(2)即便在大数据时代，先进的技术已解决了数据计算资源不足、数据采集限制及时效性等问题，但实际检测总体中的所有个体几乎不可能，甚至认为实践中针对总体的测量可能只是一个范围更大的抽样而已。

(3)某些测量本身具有破坏性，如测试灯泡的寿命、汽车碰撞试验等。因此，需要在总体中抽取部分个体测量样本数据，进而推断总体特征——这正是推断统计要解决的问题。

案例分析

2023 年 3 月 PMI 延续扩张走势

1. 制造业 PMI 持续位于景气区间

2023 年 3 月，制造业 PMI 为 51.9%，受上月高基数等因素影响，比上月下降 0.7 个百分点，但仍高于临界点(50%)，制造业保持扩张态势。

(1)生产指数和新订单指数分别为 54.6%和 53.6%，显示制造业产需两端继续扩张。

(2)企业采购量指数为 53.5%，显示企业采购意愿增强。

(3)各规模企业 PMI 保持扩张。大型企业 PMI 为 53.6%，中、小型企业 PMI 分别为 50.3%和 50.4%，持续两个月位于景气区间。

(4)重点行业 PMI 运行稳定。装备制造业、高技术制造业、高耗能行业和消费品行业 PMI 分别为 53.0%、51.2%、51.1%和 51.9%，继续保持扩张态势。

(5)生产经营活动预期指数为 55.5%，市场预期稳定向好。

2. 非制造业商务活动指数继续较快回升

2023 年 3 月，非制造业商务活动指数 58.2%，高于上月 1.9 个百分点，非制造业恢复发展步伐加快。

(1)服务业商务活动指数为 56.9%，高于上月 1.3 个百分点。

①从行业来看，零售、铁路运输、道路运输、航空运输、租赁及商务服务等行业商务活动指数高于 60.0%，表明近期居民消费和商旅出行意愿增强，相关行业市场活跃度较快回升。电信广播电视及卫星传输服务、互联网软件及信息技术服务等行业商务活动指数升至高位景气区间，行业运行稳中向好。

②从市场需求和预期来看，新订单指数为 58.5%，高于上月 3.8 个百分点，服务业市场需求释放继续加快。业务活动预期指数为 63.2%，连续 3 个月高于 63.0%，企业对市场预期持续向好。

(2)建筑业加速扩张。建筑业商务活动指数为 65.6%，高于上月 5.4 个百分点。

3. 综合 PMI 产出指数保持回升态势

2023 年 3 月，综合 PMI 产出指数分别为 57.0%，高于上月 0.6 个百分点，在较高景气区间继续上行，表明我国企业生产经营总体继续好转。

制造业 PMI、非制造业商务活动指数和综合 PMI 产出指数均保持在扩张区间，表明我国经济发展仍在企稳回升之中。但也要看到，企业发展过程中仍面临市场需求不足、资金紧张和运

营成本高等突出问题，我国经济回升基础有待进一步巩固。

资料来源：

[1] 国家统计局. 2023 年 3 月中国采购经理指数运行情况[EB/OL]. (2023-03-31)[2025-07-01]. https://www.stats.gov.cn/sj/zxfb/202303/t20230331_1938271.html.

[2] 国家统计局. 国家统计局服务业调查中心高级统计师赵庆河解读 2023 年 3 月中国采购经理指数[EB/OL]. (2023-03-31)[2025-07-01]. https://www.stats.gov.cn/sj/sjjd/202303/t20230331_1938272.html.(有改动)

简要分析：

(1)PMI 是指通过对企业采购经理的月度调查结果统计汇总、编制而成的指数，是国际上通用的监测宏观经济走势的先行性指数之一，具有较强的预测、预警作用。

(2)采购经理调查由国家统计局直属调查队具体组织实施。在调查时无须以全国所有的采购经理为调查对象，而是采用与规模成比例的概率抽样方法随机确定样本企业。

(3)采购经理调查问卷涉及企业采购、生产、流通等环节，将收集的数据经处理、汇总后测算 PMI，并与基期数据进行对比，分析现状与趋势。此过程属于描述统计分析。

(4)基于 PPS 结果开展采购经理问卷调查，利用样本数据推断我国经济生产经营运行情况及宏观走势，这属于推断统计范畴。

第二节　统计数据的分类

在统计学研究中，变量用于描述现象中某种随时间或空间变化的特征。例如，“案例分析”中的 PMI 就是一个变量。使用统计方法对现象的某种特征进行描述、计量和记录的结果，所得到的结果称为统计数据。也就是说，统计数据是变量在特定时间或空间条件下的具体取值。例如，国家统计局直属调查队每月发布的各维度 PMI 数据就属于典型的统计数据。

统计数据通常以统计表、统计图、统计报告、统计年鉴等形式呈现。随着统计学的不断发展，越来越多的数据成为统计数据，如平时看到的文字、图片、音频、视频等非结构化数据(包括部分具备结构特征的半结构化数据)，通过一定的方法进行量化(结构化)后，也能够运用统计方法进行处理和分析，这属于统计数据的范畴。

在大数据时代背景下，统计数据的来源不断拓展，体量不断扩大，传播速度急速提高，呈现出多样化的表现形式。如何将非结构化数据量化(结构化)，提取信息并与结构化数据对接，既是传统统计分析的挑战，也为统计学发展提供了新的机遇。

一、按计量尺度分类

计量尺度是指对计量对象量化时采用的具体标准，统计数据按计量尺度可分为分类数据与数值型数据，分类数据是指只能归于某一类别的非数字型数据，数值型数据是指数字测量的观察值，其结果表现为具体的数值。

进一步来说，计量尺度还可进一步被分为有名义尺度、有序尺度、定距尺度、比例尺度。其中，名义尺度测量产生的观测值为无序分类数据，有序尺度测量产生的观测值为有序分类数据，这两种数据被统称为分类数据。分类数据可以用数字来表示，但不应该被理解为普通数字；定距尺度与比例尺度原则上属于同一层次的计量尺度，区别较小，都是可以对数据进行精确计量的尺度，有时被统称为定量尺度。利用这种尺度测量时所得的变量被称为定量变量或数值型变量，定量变量的观测值被称为数值型数据。

(一)名义尺度与无序分类数据

名义尺度具有无序标识类别或按约定顺序排列的尺度,只能对现象的特征进行平行的分类或分组,这些类或组之间是并列或平等的、互斥的,且不存在着等级和顺序。

名义尺度可以通过类别来计数,按照名义尺度测量时所得的变量称为无序分类变量,无序分类变量的观测值形成无序分类数据,这类数据不能排序或测量,只能作为各类或各组的代码,因此只可以进行“=”或“≠”的运算,无法完成更高级别的运算。

例如,在问卷调查中,为了收集受访者人口统计学信息,在询问受访者生理性别时会设置“男”“女”两个选项,在数据化时通常将“男”设置为“1”,“女”设置为“0”,这,“1”和“0”既不能代表选项“男”比“女”大,更不能说明“1”大,只能作为区分选项“男”和“女”不同的编码。

有时,名义尺度可以按照约定顺序排序,即虽然约定了顺序,但不能反映各类或各组之间的优劣、确定特定等级的大小。例如,血型总是按照 A、B、O 的顺序进行表述,按照约定可以按照字母顺序进行排序,但血型本身没有优劣大小的区分。

(二)有序尺度与有序分类数据

有序尺度是具有有序标识类别的尺度。有序尺度比名义尺度高一个层次,是对现象的特征分类或分组之后对等级差或者顺序差的一种测量,这些类或组之间是相互排斥,没有重叠的,但存在着等级和顺序。值得注意的是,有序尺度所分成的类或组之间虽然可以排序或者排等级,但是不同类别或组别之间的差异不是很明确。也就是说,不同类别或组别之间可以比较等级的高低,但是高出多少无法确定,即等级差或者顺序差无法确定。

相应地,按照有序尺度测量所得的变量称为有序分类变量。有序分类变量的观测值形成有序分类数据。有序分类数据不仅可以作为代码区分不同的类或组,也可以反映各类或各组之间的顺序优劣,明确特定的等级,因此这类数据除了能进行“=”或“≠”的运算,还能进行“>”或“<”的运算,但不能明确各类或组之间数量的大小,因此无法完成更高级别的运算。

李克特量表被用于测量受访者对特定调查主题的综合态度和看法。例如,如果要测量满意程度,通常设置选项“非常满意”“比较满意”“不清楚”“比较不满意”与“非常不满意”,赋值为“5”“4”“3”“2”“1”,数量化后的统计数据能够说明受访者对特定调查主题的满意程度,可以排序。对“非常满意”与“比较满意”所赋的值“5”和“4”、对“比较满意”与“不清楚”所赋的值“4”和“3”都能计算出差为“1”,但“非常满意”与“比较满意”之间、“比较满意”与“不清楚”之间的差距并不相同。

(三)定距尺度与定距数据

定距尺度是指某一规定零点及相等间隔尺度值的连续尺度。定距尺度计量形成的数据被称为定距数据,这类数据不仅可以区分不同的类别和组别,还可以比较各类或各组的大小并排序,而且可以精确计算各类或各组之间差异的大小。

定距数据没有绝对零点,如摄氏度的零点是按照水的冰点规定的,是一个规定零点,并不是基于热力学的最低温度;“考试成绩为 0”不代表应试者完全没有相关的知识;“智商为 0”不代表此人完全没有智商等。因此定距数据可以计算“+”或“-”,但无法进行“×”或“÷”的运算。

(四)比例尺度与定比数据

比例尺度是具有一个绝对零点或自然零点及相等间隔尺度值的连续尺度。比例尺度计量形成的数据被称为定比数据。与定距数据相比,定比数据具有绝对零点或自然零点,即“0”代表

“没有”或“无”。例如，某项产品的销售量为“0”，则代表没有销售量，常见的长度、高度、利润、薪酬、产值等变量的观测值，都被认为是定比数据，可以完成“×”与“÷”的运算。

(五)按计量尺度分类数据的比较

在统计实务中，计量尺度通常决定了应该采用何种统计方法进行深入分析，特别是高计量尺度的数据可以向低计量尺度的数据转化。例如，在测量人的年龄时，既可以选择收集为数值型数据(如具体的岁数)，也可以将收集的数值型数据处理为“18岁以下”“18～35岁”“35～65岁”“65岁以上”形成有序分类数据，但反过来讲有序分类数据转化为比例数据是行不通的。在统计分析中，计量尺度的层级越高，越有利于提高测量的精确性，同时运用更多的统计方法。

此外，需要强调的是，数据归于哪一类计量尺度并没有完全达成一致，有时也会为了满足研究目的来改变它的类别。例如，使用李克特量表测量的综合态度和看法，聚类之后属于一个构面的题项可以计算平均值并完成回归分析，IQ数据也常常被研究人员作为定比数据来处理。

表1-1总结了常见的计量尺度的特征，用于理解按照计量尺度分类的统计数据。

表1-1　计量尺度的主要特征及举例

类　别	特　征	运　算	举　例
名义尺度	只做区分	“=”或“≠”	邮编、性别、肤色、国籍、民族
定序尺度	可以排序和排等级，但组间或类别间差异无法比较	“=”“≠”“>”“<”	喜好、态度、职称、成绩(5分制)
定距尺度	没有绝对零点，数据之间的精确差异可测量	“=”“≠”“>”“<”“+”“−”	试卷的分数、智商、经纬度、温度、年份、衣服的尺码
比例尺度	有绝对零点，不同的取值之间的比是有实际含义的	“=”“≠”“>”“<”“+”“−”“×”“÷”	身高、体重、薪酬、年龄、误差

二、按尺度连续性分类

具有连续可能值的尺度被称为连续尺度，只有一组或一系列不同值的尺度被称为离散尺度，根据尺度连续性的不同，统计数据可分为离散数据与连续数据。

离散数据是指在任意两个数据点之间的个数是有限的，只能用自然数或整数单位计算，而不能用小数计算。企业个数，职工人数，家庭人口数等指标的观测值都属于离散数据。

连续数据是指数据在任意两个数值之间还可以任意取值、数据连续不断、相邻两个数值可作无限分割的数据。如询问一个人的年龄时往往会得到一个整数单位的数字(如19岁)，在此基础上还能够继续细分下去，比如用年、月或者年、月、日等来计算更精确的年龄，这种能进一步细分用于精准测量与记录的就是连续数据。

三、按收集方法分类

统计数据按收集方法可分为观测数据和实验数据。观测数据是指通过调查或观测而收集到的数据。其特点是该数据是在没有对事物人为控制的条件下得到的，有关社会经济现象的统计数据几乎都是观测数据。如利用我国上市公司向社会公布的年度财务报告可以收集、汇总我

国上市公司财务状况、经营成果和现金流量相关财务数据。

实验数据是指在实验中通过控制实验对象而收集到的数据。在自然科学领域的大多数数据都是实验数据。

知识拓展

屠呦呦：一辈子潜心研究青蒿素

屠呦呦因发现青蒿素而获得诺贝尔生理学或医学奖。青蒿素的发现，为世界带来了一种全新的抗疟药，以青蒿素为基础的联合疗法已成为世界卫生组织推荐的疟疾治疗最佳方案，在全球范围内挽救了数百万人的生命。

青蒿素的发现历程并非一帆风顺。屠呦呦于1969年临危受命，接受国家疟疾防治项目“523”的抗疟研究任务，担任重要抗疟组组长。面对简陋的设备、匮乏的资源等挑战，屠呦呦带领团队翻阅大量古代医学典籍，收集了2 000多个方药，精选640个进行试验。最初对包括青蒿在内的多种中药进行试验，结果并不理想。

在经历多次失败后，受东晋葛洪《肘后备急方》中“青蒿一握，以水二升渍，绞取汁，尽服之”一句的启发，屠呦呦意识到温度可能是关键，于是开始尝试用低沸点溶剂提取，最终成功获得具有显著抗疟效果的青蒿素。为验证青蒿素的安全性和药效，屠呦呦和团队成员亲自试药，确认无严重毒副作用后，才进行临床验证青蒿素的安全性和药效。

在青蒿素研发的过程中，通过一系列科学实验和临床试验产生了大量的药品实验数据，对评价青蒿素的安全性与有效性提供了科学依据。

青蒿素及其衍生物制成的联合疗法有效降低了疟疾患者的死亡率，这一联合疗法也成为中医药现代化的里程碑，证明了传统中医药与现代科技结合的巨大潜力，促进了中医药的国际传播和认可。屠呦呦及其团队在艰苦条件下坚持不懈、勇于创新的精神，激励着无数科研工作者，展现了科学探索的价值和意义。

（作者整理）

随着社会科学研究方法的发展，利用观测数据并借助双重差分法、断点回归法、倾向得分匹配法、合成控制法等准自然实验方法来检验因果关系的做法日益普及。准自然实验主要通过在自然场景中运用特定程序（如对对照组无前测设计、非对等控制组设计等）来灵活控制实验对象。与真正的自然实验相比，因无法随机分配实验对象至实验组和控制组，准自然实验仅为近似随机实验，故其所得因果结论的效度略低于严格实验研究，但它在政策效应检验等研究主题上仍具有较高的效度。

四、按描述的对象与时间的关系分类

横截面数据也称静态数据，是对现象在某一时刻的变化情况进行的描述，是在相同或近似相同的时间点上收集的数据。这类数据通常是在不同的空间获得的，用于描述现象在某一时刻的变化。例如，2022年我国各地区生产总值于横截面数据。

时间序列数据也称动态数据，是对现象随着时间变化的结果进行的描述，是在不同时间收集到的数据。这类数据是按时间顺序收集到的，用于描述现象随时间变化的情况。例如，2010—2022年我国的GDP属于时间序列数据。

第三节　统计学中的基本概念

一、总体与样本

(一)总体

1. 总体的定义与范围

总体指统计研究对象的全体，通常由研究目的指向的具有某些或某种共同特征的事物组成的集合体。总体范围有时比较容易确定，如在关于中国工业企业的统计调查中，中国工业企业的全体属于总体。总体范围有时难以区分，这就需要以研究目的为导向来确定总体。例如，如果要调查潜在消费者对某款新产品的购买意向与态度，由于难以明确界定哪些人可能成为实际购买者，此时就需要根据研究目的来合理划定总体范围。

2. 总体单位与个体

根据研究目的，人为设定或客观存在的构成总体的相对独立的同质单位称为总体单位。总体是一个集合，总体单位就是构成这个集合的元素。如果总体单位是独立可计数的客观实体且不可进一步细分，则称之为个体。

在调查中国工业企业生产经营情况时，中国工业企业的全体是总体，每个工业企业就是一个总体单位，也是一个个体；在评估池塘的水质时，池塘的水体是总体，但构成这个总体的客观实体(水体)是不可计数。此时可以根据研究目的，人为定义总体单位，如“每 10 毫升水样”作为一个总体单位。

3. 有限总体与无限总体

根据包含的总体单位数目是否有限，总体可分为有限总体与无限总体。有限总体是指由数目有限的总体单位构成的总体。有限总体需满足两个条件：总体范围能够明确界定，总体单位是有限可数的。总体单位的数目被称为总体容量，通常记为 N。

无限总体是指构成总体的总体单位数是无限的。例如，在持续生产条件下，产品的总体规模会随着生产线的运行不断新增总体单位。同样，在科学实验中，如果实验可以无限重复进行，那么由实验数据构成的总体就属于无限总体。由此可见，无限总体通常指那些不断产生新总体单位的动态情形，即随时间推移不断扩展的总体。这类总体往往涉及时间因素或理论假设(如无限生产或实验)。在实际应用中，大多数总体通常是有限的，其总体容量(N)未必已知或未必可知。

(二)样本

样本是指从总体中抽取的一部分总体单位构成的集合。抽取样本的主要目的是利用样本信息来推断总体特征。例如，“案例分析”中提到的 PMI 调查中，采用 PPS 抽样方法随机选取样本，然后基于样本数据来推算全国 PMI 数值。样本中包含的总体单位数量称为样本量或样本规模，通常记为 n。当 $n=N$ 时，样本即为总体本身；当 $n=1$ 时，为最小样本量；通常情况下，n 介于 1 和 N 之间。

课堂讨论

讨论一下，在针对一个池塘水质的调查中，以该池塘的水体为总体，这个总体是有限总体还是无限总体？

二、参数与统计量

参数是用来描述总体特征的概括性数字度量，表示研究者希望了解的总体的某种特征。对于一个特定总体而言，总体数据通常是未知的但研究者非常关心的常数；而存在多个可类比但不相同的总体时，虽然参数名称相同，其取值可能不同，也就是说，不同总体对应着不同的参数。研究中常关心的参数包括总体平均数、总体标准差、总体比例等。在统计学中，常用希腊字母表示总体参数，如总体平均数用 μ(mu)表示，总体标准差用 σ(sigma)表示，总体比例用 π(pi)表示等。由于参数通常是未知常数，因而需要通过抽样，从样本中计算统计量来估计参数值。

统计量是用来描述样本特征的概括性数字度量，由样本数据计算得出。由于抽样具有随机性，只要样本量小于总体规模，在相同的抽样方案下便可能产生多个不同样本，从而导致统计量的取值随着样本变化而变化，因此统计量是样本的函数。研究者常关注的统计量包括样本平均数、样本标准差、样本比例等。样本统计量通常用英文字母表示，如样本平均数用 $\bar{x}$(x-bar)表示，样本标准差用 s 来表示，样本比例用 p 表示等。样本数据是已知的，因此统计量是可以计算的，借此可用于估计总体参数。例如，用样本平均数($\bar{x}$)估计总体平均数(μ)，用样本标准差(s)估计总体标准差(σ)，用样本比例(p)估计总体比例(π)。然而，样本统计量与总体参数之间往往存在差距，设法让二者之间的差距控制在可容忍范围内，正是统计通过样本已知统计量推断总体未知参数的核心过程。

三、标志与指标

(一)标志

1. 定义

标志是说明总体单位属性或特征的名称。每个总体单位从不同的角度考察，可以有许多属性或特征，例如，如果要了解中国工业企业生产经营情况，每一个中国工业企业都有所处的行业、所有制性质、规模、员工数、利润等特征，这些就是标志。总体中的一个中国工业企业就是总体单位，行业、所有制性质、规模、员工数、利润等标志说明该企业的属性与特征，该企业就是以上这些标志直接承担者。由此可见，标志依附于总体单位，总体单位是标志的载体。

2. 分类

(1)标志按变异情况可分为不变标志和可变标志。标志在总体单位之间有一定的具体表现，有的相同，有的则不尽相同，标志如果在总体各单位之间的具体表现完全相同就是不变标志，该标志就称为不变标志，否则就称为可变标志。

例如，在对中国工业企业生产经营情况进行统计研究的过程中，对所有的总体单位而言，工业企业这一特征(标志)是不变的，即不变标志。相应地，在不同的统计总体之间发生变化的标志被称为不变指标，如所处行业、所有制性质、规模、员工数、利润等表现工业企业不同维度特征的指标都是可变指标。

(2)标志按性质不同可分为品质标志和数量标志。凡是只能用语言文字表示、反映事物性质或属性的标志，称为品质标志；用数值表示的标志，称为数量标志。

(二)指标

统计指标简称为“指标”，通常被认为是表征现象特征的科学概念和具体数值(变量的名称与变量值)，随着统计理论和实践的不断发展，指标被认为不仅仅可以表征总体特征，而且可以表征样本、总体单位、个体的特征，即统计是表征集合特征的名称。

(三)标志与指标的关系

标志与指标在统计学中虽各自扮演不同角色,但在数据汇总与分析过程中,它们之间存在密切联系。

许多指标数值来源于总体单位的数量标志汇总。例如,在中国工业企业的统计中,一个地区工业企业的研发成果数量,是由该地区各企业的研发成果数量汇总而成的。前者作为表征地区研发活动水平的统计指标,后者则是企业这一总体单位的数量标志。

随着研究目的的变化,总体与总体单位之间的角色可能发生转化。例如,在对某地区工业企业的统计中,该地区的企业构成总体,每个企业是总体单位,企业研发成果为统计指标;而当研究对象扩大为全国工业企业时,该地区的工业企业则成为新的总体单位,其研发成果数量则转化为数量标志,反之亦然。

正因如此,在一定条件下,总体、总体单位、样本与个体之间可以相互转化,指标与标志的区别也随之变得不那么明显。多数情况下,它们仅是变量在不同语境下的不同称呼。实践中,统计工作通常以集合为研究对象,因此更常使用“指标”这一术语,“标志”一词的使用相对较少。

问题思考

1.什么是统计学?什么是描述统计和推断统计?

2.按照计量尺度来分,数据分为哪几种类型?尝试举例说明。

3.解释参数与统计量。

4.一家研究机构针对大学生消费习惯,随机从某市在校大学生中抽取500人作为样本开展调查,在这次调查中有关于大学生支付方式的题项,其中选项包括现金支付、电子支付与卡支付。

(1)请问此次调查的总体是什么?总体单位是什么?样本是什么?

(2)针对支付方式的题项所收集的数据,按照计量尺度来划分,属于什么类型的数据?

(3)按照描述的时间与对象的关系来划分,属于什么类型的数据?

实战演练

2024年一季度全国规模以上工业产能利用率为73.6%

2024年一季度,全国规模以上工业产能利用率为73.6%,比上年同期下降0.7个百分点,比上季度下降2.3个百分点。

分三大门类来看,2024年一季度,采矿业产能利用率为75.0%,比上年同期下降0.2个百分点;制造业产能利用率为73.8%,下降0.7个百分点;电力、热力、燃气及水生产和供应业产能利用率为71.2%,下降0.7个百分点。

从主要行业来看,2024年一季度,煤炭开采和洗选业产能利用率为71.6%,食品制造业为69.1%,纺织业为78.0%,化学原料和化学制品制造业为76.4%,非金属矿物制品业为62.0%,黑色金属冶炼和压延加工业为77.3%,有色金属冶炼和压延加工业为78.2%,通用设备制造业为78.2%,专用设备制造业为77.0%,汽车制造业为64.9%,电气机械和器材制造业为72.7%,计算机、通信和其他电子设备制造业为74.7%。

附注

(1)指标解释。

①产能利用率。产能利用率是指实际产出与生产能力(均以价值量计量)的比率。

②企业的实际产出。企业的实际产出是指企业报告期内的工业总产值。

③企业的生产能力。企业的生产能力是指报告期内，在劳动力、原材料、燃料、运输等保证供给的情况下，生产设备(机械)保持正常运行，企业可实现并能长期维持的产品产出。

(2)调查方法及范围。

①大中型企业全面调查，小微企业抽样调查，调查共涉及工业企业 11 万家左右。

②小微企业按抽样方法推算总体，与大中型企业调查数据合成，计算出全国工业产能利用率。

③本调查按季进行，数据未经季节调整。

(3)行业分类标准。执行国民经济行业分类标准(GB/T4754-2017)。

资料来源：国家统计局. 2024 年一季度全国规模以上工业产能利用率为 73.6%[EB/OL]. (2024-04-16)[2025-07-02]. https://www.stats.gov.cn/sj/zxfb/202404/t20240416_1948583.html.(有改动)

请思考并回答以下问题。

(1)按照计量尺度来划分，调查取得工业产能利用率数据是什么类型的数据？

(2)针对小微企业的抽样调查中，总体和总体单位是什么？样本量是多少？

(3)调查共涉及工业企业 11 万家左右，并没有覆盖全国所有规模以上工业企业，为什么能说明全国规模以上工业产能情况？