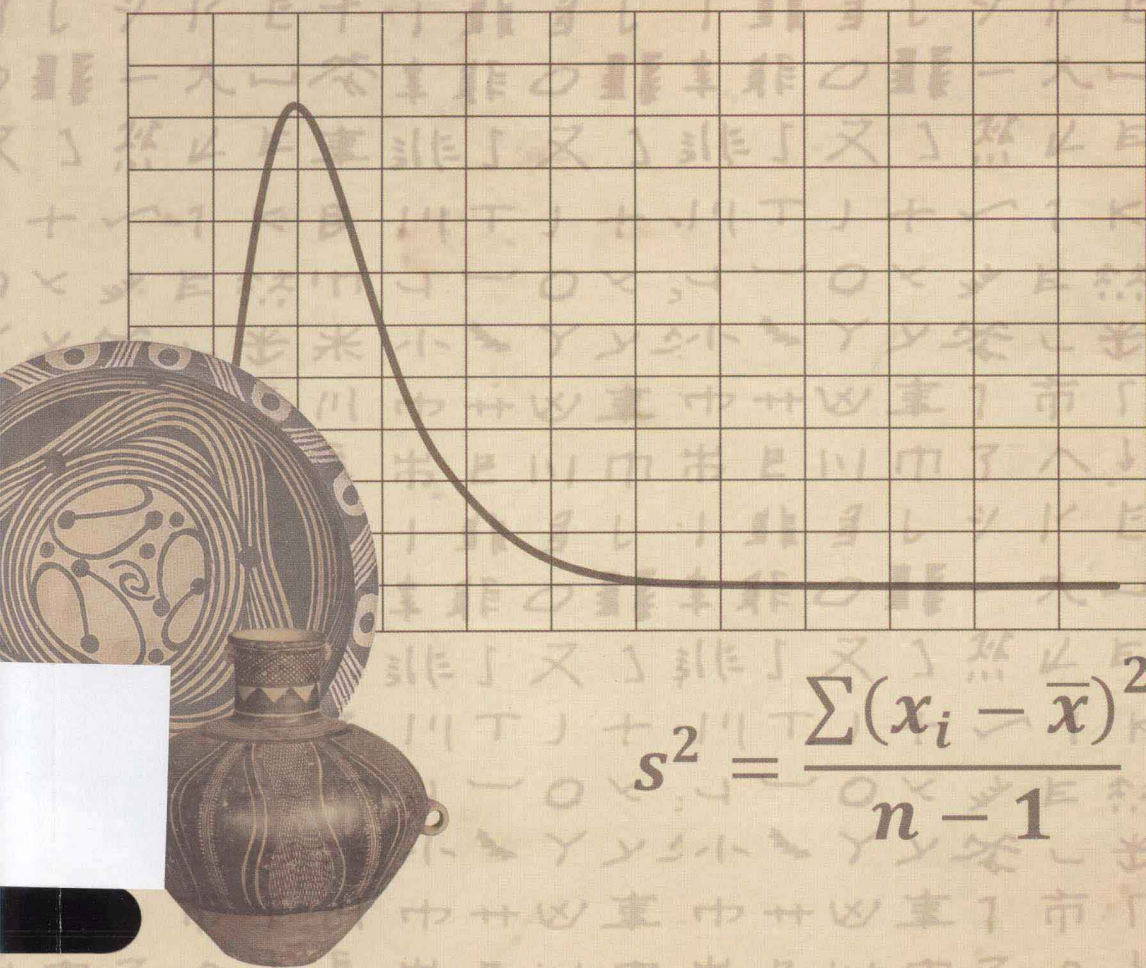


简明考古统计学

陈铁梅 陈建立 著



$$s^2 = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$



科学出版社

简明考古统计学

陈铁梅 陈建立 著

科学出版社

内 容 简 介

本书是陈建立以陈铁梅原著《定量考古学》(北京大学出版社 2005 年出版)为教材,多年讲授“定量考古学”课程的基础上,两人合作新编的简明修订版。全书共分十章,主要通过中国考古学的最新研究案例,遵循深入浅出、循序渐进和学以致用原则,系统介绍统计学的基本原理、方法和技术,以及相关的计算机应用程序,以便使读者在此基础上,认同统计学对考古研究的作用、激发对考古资料定量思考的兴趣,并将统计学方法应用到考古实际工作中。

本书较全面地介绍了二十多年来我国考古学定量研究的进展,适用于具有高中以上数学水平的考古、科技考古和文物保护等专业的研究人员和高校师生参考。

图书在版编目(CIP)数据

简明考古统计学 / 陈铁梅, 陈建立著. —北京: 科学出版社, 2013. 6
ISBN 978-7-03-037890-3

I. ①简… II. ①陈… ②陈… III. ①统计学 - 应用 - 考古 - 高等学校 - 教材 IV. ①K85

中国版本图书馆 CIP 数据核字 (2013) 第 131689 号

责任编辑: 闫向东 樊 鑫 / 责任校对: 郭瑞芝

责任印制: 赵德静 / 封面设计: 谭 硕 Charlotte Chen

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

北京通州皇家印刷厂 印刷

科学出版社发行 各地新华书店经销

*

2013 年 6 月第 一 版 开本: B5 (720 × 1000)

2013 年 6 月第一次印刷 印张: 15 3/4

字数: 292 000

定价: 55.00 元

(如有印装质量问题, 我社负责调换)

序

考古学研究的对象是不可再生的古代人类活动留下的遗存，所以考古工作者需要尽可能地从有限的考古资料中攫取更多的信息，若想达到这一目标，除了需要研究者观察材料时所具有的不同高度与角度、对资料更深刻的理解和认识以外，还需要有一些新的技术或手段。在考古学研究中应用数学方法，就像是为考古学增加了一个新的工具，通过对考古学资料的定量分析，可以观察到很多此前不能够发现或并不明确的现象和规律，为从考古资料中攫取更多的信息提供了可能，从而为考古学的解释提供更多的空间，得以使研究更加深入。不过能够保留到今天并被科学的考古发掘、整理出来的考古资料具有一定的偶然性，也就是具有一定的随机性，考古学研究实际上是从发现的带有随机性的考古遗存中去研究遗留下这些遗存的当时的人类活动。所以，在考古学研究中要使用数学方法，就必须考虑到考古资料的这种随机性。而在数学方法中就有通过随机性的资料来研究这些资料所反映的总体的情况，这就是统计学。这也是目前在国内外考古学领域应用数学方法进行定量分析与研究时运用的主要方法。

由于中国现行的教育体制，在高中时期学生就已文理分科，分别接受不同的训练，而高校中的考古专业往往以招收文科生为主，导致大部分考古工作者所受到的数学训练非常有限，这给考古工作者在自己的研究中应用统计学方法带来了一定的困难与障碍。不过由于计算机技术的发展和大量数学运算程序的开发，在运用统计学方法时，可以完全免去对大批量的数据进行运算的工作，因此对于中国的考古工作者来说，要想成功地运用统计学方法进行考古学研究，最重要的是需要了解有哪些统计学方法可以为考古研究所用。其基本的原理和适用的对象是什么？如何针对不同的研究资料、不同的研究目的，去选择最适当的统计学方法？如何解读统计学分析的结果？《简明考古统计学》就将为我国考古工作者解答这些问题。

《简明考古统计学》的作者之一陈铁梅老师，是我国最早从事定量考古

学研究的老一辈学者，从 20 世纪 80 年代开始，陈铁梅老师就利用统计学方法进行考古学研究，涉及考古类型学分析、考古遗存的分期、古代陶瓷产地等研究领域，同时为了推进统计学等数学方法在中国考古学中的应用，他最早在北京大学开设“定量考古学”课程，并于 2005 年出版了中国第一部针对考古学者编写的、详细介绍统计学等数学方法如何应用于考古学研究的专著——《定量考古学》。可以说陈铁梅老师是中国应用统计学等数学方法进行考古学研究的开创者，他这一研究领域的发展做出了卓越的贡献。

我第一次见到陈铁梅老师是在 20 世纪 90 年代，记得是在由北京大学举办的“二十一世纪中国考古学发展趋势高级研讨班”上，聆听了陈铁梅老师题为“自然科学与考古学”的讲座，讲座的具体内容今日已不能清楚的想起，只是记得陈老师用十分生动的漫画图像、深入浅出的语言，向大家介绍了一些自然科学技术的原理和其在考古学中的应用，给我留下了极其深刻的印象。后来我在尝试用数学方法做一些考古学研究时，曾向陈老师请教过很多问题。当陈老师的《定量考古学》出版后，我请当时已经退休的陈老师在出版社批量购书，以作为我开设的“计算机技术在考古学中的应用”这一课程的教材。此后由于各种原因，与陈铁梅老师的联系虽然没有中断，但也渐渐少了起来。2010 年，“全国第十一届考古与文物保护化学学术研讨会”在吉林大学召开，我惊喜地看到当时已年逾古稀的陈铁梅老师也来参会，就是在那次会议上，陈铁梅老师告诉我，他与陈建立老师正在针对《定量考古学》出版后听到的读者意见，以及陈建立长期应用该书作为定量考古学课程教材中注意到的某些问题，对其进行修订，以使此书更易于被考古工作者理解和接受。

现在，这本由陈铁梅老师和陈建立老师共同完成的《简明考古统计学》就要出版了，与《定量考古学》相比，此书有三个改变：第一，原书分为上下两编，分别为“考古研究中的基础统计学”和“多元统计方法在考古研究中的应用”，考虑到中国考古工作者的数学基础相对薄弱，而多元统计分析又比较复杂，因此，在此书中删去了原有的多元统计分析部分；第二，原书在介绍统计用软件的使用时，兼有 EXCEL 和 SPSS（Statistical Package for the Social Sciences，即社会科学统计软件包），考虑到 SPSS 使用起来也相对复杂，而 EXCEL 可以解决基础统计学的大部分问题，因此，删去了有关 SPSS 的使用部分；第三，增加了大量的考古学研究中应用统计学方法的研究案例，其中很多都是在 2005 年《定量考古学》出版以后中国考古学所取得的

新的研究成果。

《简明考古统计学》所做出的三个改变，体现出作者对在中国考古工作者中普及考古统计学方面所做出的努力。书中对统计学方法的介绍通俗易懂，免去了严谨的数学推导，重点介绍了方法的基本原理和适用对象；对如何正确运用统计学方法，运用这些方法对考古学资料有哪些要求，如何正确解读统计分析的结果等也给予了详尽的说明；书后的习题则为读者提供了绝好的实践机会；尤其是书中所列举的大量的中国考古学研究中应用统计学方法的研究案例，从旧石器考古到历史时期考古，从类型学分析到遗存分期，从对区域性田野考古调查资料的分析到对自然科学技术手段获取信息的再次解读等，涉及中国考古学的许多领域，不仅使我们了解到在考古学研究中运用统计学方法所具有的普遍意义，而且读起来也颇感亲切。相信在读完此书后，一定会有读者产生尝试进行这种研究实践的迫切愿望。

作为一本全面论述统计学在考古学中应用的力作，《简明考古统计学》是中国考古工作者在进行定量研究时的必读书。而随着中国考古学的发展，随着对考古资料进行定量分析与研究理念的日渐普及并被广大考古工作者所接受，《简明考古统计学》也会越发显示出其所具有的应用价值，以及其对中国考古学的重要贡献。

《简明考古统计学》共有十章内容，从对考古观测数据的最基本处理，到考古学的抽样问题，各部分既独立成章，相互间又存在着一定的逻辑关系。因此，此书可以有不同的阅读方式。读者既可以通篇浏览，在全面掌握各种统计学方法的基础上了解其在考古学中的应用，也可以根据自己的研究实践，选择相关章节进行阅读，从而得到有针对性的具体指导。

吉林大学边疆考古研究中心

滕铭予

2013年4月30日

前 言

定量化研究是 20 世纪下半叶以来社会科学和人文科学诸学科发展的总体趋势，而计算机科学的发展为这个发展趋势提供了技术基础。定量化研究对经济学、社会学的革命化影响是非常显著的，形成了新的学科分支，改变了学科的语境。甚至文字学、语言学和文学等似乎与数字毫无关系的学科也通过字频统计、字词和文法的关联比较，来探索文字、语言的演化和不同语言间的亲缘关系。

考古学是利用实物遗存资料去复原古代社会的科学。考古学研究墓葬、房址、文化层等不同层次的遗存。对每类遗存的描述是离不开定量属性的，离不开测量的，大至建筑和墓葬的位置、数量和规模，小至器物的数量、形状和材料的化学组成。至于描述人和动物骨骼的特征是大量测量数据和比例数的集合。更为重要的是，考古学并非以单件器物、单座墓葬或单个遗址作为研究对象，而是以器物的整个一个类型、墓葬群、遗址群作为自己的研究对象，并试图根据这些实物遗存的数据（称为样本），来推断关于整个地区一定时期人类文化（称为总体）一般性的结论。但是考古调查和发掘得到的资料相对于古代社会总是零星的、不完整的，即考古调查和发掘属抽样过程，往往是随机性的。根据这些资料（样本）来推断古代社会的情况（总体）属统计推断，需要应用概率统计学的原理和方法，由此得到的结论和知识也是带有统计性质的。这就是为什么考古研究需要统计学的知识。

我国考古学研究中应用定量方法虽相对较晚，但 20 世纪 80 年代后，已陆续可见利用定量统计方法对考古资料进行研究的尝试，如华北动物群的排序、雨台山墓葬的排序、史家墓地的分期、河南早商前后陶豆的分期、侯马乔村墓地陶器分期、两周随葬青铜容器的组合研究以及我国新石器时代墓葬的性比分析等。90 年代后一些大学和地方的考古教学和研究部门纷纷与国外单位合作广泛开展区域考古调查，包括日照、赤峰地区以及河南洹河流域、伊洛河流域、灵宝和洛阳盆地，颍河上游，山西垣曲盆地，陕西周原和周公

庙遗址以及四川成都平原等。设计并实施了多种抽样调查方案，并获得了大量关于所调查地区中遗址规模、持续时间和分布等多方面的数据和信息，建立了相应地区的考古地理信息系统。我国考古界应用了多种定量统计方法卓有成效地揭示了上述地区有关时段中人与自然环境的关系、各类考古学文化的演化以及不同文化间的交流和碰撞。最值得指出的是，考古工作者对定量研究和统计学观念和方法的认同日益增强，考古期刊中定量研究的论文已不再是凤毛麟角。特别是多位考古学者和古文字学者的专著中，贯穿了统计学的观念和定量分析的方法，例如张再兴的《西周金文文字系统论》、刘莉《中国新石器时代——迈向早期国家之路》、滕铭予《GIS 在半支箭河中游环境考古中的应用》和戴向明《陶器生产、聚落形态与社会变迁——新石器至早期青铜时代的垣曲盆地》等。定量和统计学方法在我国考古研究中的应用日益普及，作用越来越重要。

自 20 世纪 80 年代始，我国的北京大学和吉林大学等学校开设了定量考古学课程，但长期苦于缺乏相应的教材。2005 年陈铁梅编写出版了《定量考古学》（北京大学出版社），随后陈建立在北大讲授定量考古课程时始终以该书作为教材。它在帮助学生理解统计学等定量方法在考古学研究中的意义，帮助他们掌握统计学的内容与方法方面起到了有益的作用。但多年的教学实践和考古学界同仁对该书的意见反馈，也使我们认识到，该书作为定量考古领域我国第一本，也是至今唯一的专著和教材尚存在一些问题和不足。主要是部分内容偏繁偏深，使数学基础不强的考古工作者阅读起来不易理解。其次，该书主要介绍使用的 SPSS 统计软件也为多数考古工作者所不熟悉。因此，我们两人决定在对《定量考古学》作较大的删节、增补和修改的基础上，新编写本书。编写的原则是：①删繁就简，多元统计内容全部删除，并对基础统计内容的讲述更注意“浅出”，帮助读者更易学习和理解。②作为实例大量引用近年来我国考古研究中成功应用统计学方法的成果，以帮助读者学以致用。③全书介绍并使用我国读者的计算机中都安装的微软办公软件中的 EXCEL 程序来进行统计计算。

希望本书的出版能够让读者更直观、更容易地掌握基础统计学的基本思想和方法，并应用于考古研究实际，推进我国考古学研究的定量化进程。

目 录

序.....	(滕铭予)
前言.....	(v)
第一章 考古观测数据的最基本处理——随机现象的描述性统计.....	(1)
1.1 考古研究的对象与随机现象.....	(1)
1.1.1 随机变量的特点.....	(2)
1.1.2 随机变量的类型.....	(3)
1.2 连续型数值变量的描述性统计.....	(6)
1.3 数据组或样本的代表值：平均值和中位数.....	(7)
1.3.1 样本的平均值和中位数.....	(7)
1.3.2 用 EXCEL 软件计算样本的平均值和中位数.....	(8)
1.3.3 平均值和中位数的比较.....	(10)
1.4 数据组或样本的离散程度：方差、标准差和四分位差.....	(11)
1.4.1 样本方差和标准差的定义和计算， CV 系数	(11)
1.4.2 样本的四分位数和四分位差及其计算.....	(14)
1.4.3 标准差和四分位差的比较.....	(15)
1.5 样本中实体的频次分布和频率分布.....	(16)
1.5.1 样本中实体的频次分布表.....	(16)
1.5.2 使用 EXCEL 软件建立频次分布表.....	(17)
1.5.3 频次分布直方图和频率分布直方图.....	(19)
1.5.4 使用 EXCEL 软件绘制直方图和安装“数据分析”软件包	(20)
1.5.5 频率密度分布直方图和频率密度分布曲线.....	(23)
*1.5.6 定积分方法计算曲线下面积的基本概念简介	(25)
1.5.7 茎叶图.....	(26)
1.5.8 箱点图.....	(27)

1.5.9 分布图的比较和应用实例·····	(29)
第二章 样本平均值的性质和正态分布函数·····	(33)
2.1 样本平均值的三个重要性质·····	(34)
2.2 正态分布函数·····	(38)
2.2.1 正态分布函数的性质·····	(38)
2.2.2 标准型正态函数·····	(41)
2.2.3 标准型正态函数表·····	(42)
2.2.4 正态分布函数的应用实例·····	(44)
第三章 根据观测数据对总体平均值的估计·····	(52)
3.1 大样本情况下总体平均值的区间估计·····	(53)
3.1.1 大样本情况下总体平均值的区间估计·····	(53)
3.1.2 总体平均值区间估计中置信度、置信区间宽度和样品容量 三者间的关系·····	(55)
3.2 小样本条件下对总体平均值的区间估计与 t 分布函数介绍·····	(56)
3.2.1 t 分布函数介绍·····	(58)
3.2.2 小样本总体平均值的区间估计·····	(60)
3.3 怎样考察观测数据组是否接近正态分布·····	(61)
3.4 对有限总体平均值的区间估计和有限总体校正·····	(63)
第四章 关于总体平均值的假设检验·····	(66)
4.1 大样本情况下总体平均值的假设检验·····	(67)
4.1.1 《考工记》大刀之齐的检验——假设检验的基本原理和过程 ·····	(67)
4.1.2 墓主人死亡年代的 ^{14}C 测年验证——假设检验中的弃真错误 和纳伪错误·····	(70)
4.1.3 双侧检验和单侧检验·····	(75)
4.2 小样本情况下总体平均值的假设检验·····	(76)
4.3 两个独立样本总体平均值一致性的假设检验·····	(78)
4.3.1 大样本两总体平均值的一致性检验(以山东大仲家贝丘遗址 贝壳宽度的比较为例)·····	(79)
4.3.2 小样本两总体平均值的一致性检验·····	(80)
4.3.3 用 EXCEL 软件进行两总体平均值一致性检验·····	(84)

4.4 配对样本间总体平均值的一致性检验——以比较两个实验室的 测量数据为例·····	(84)
4.5 两总体方差一致性检验和 χ^2 分布函数·····	(88)
4.5.1 样本方差的分布和 χ^2 分布函数·····	(88)
*4.5.2 总体方差 σ^2 的区间估计·····	(89)
*4.5.3 两总体方差一致性的检验·····	(90)
4.6 小样本两总体平均值一致性的非参数假设检验·····	(92)
4.6.1 河南仰韶和龙山时期中心聚落面积一致性的秩和检验·····	(92)
4.6.2 河南仰韶和龙山时期中心聚落面积一致性的中位数检验·····	(96)
4.6.3 配对样本总体平均值一致性的符号检验·····	(98)
第五章 多总体平均值一致性的检验——一元方差分析·····	(100)
5.1 一元方差分析的原理、步骤以及 F 分布函数介绍·····	(100)
5.2 实例一：不同土壤肥瘠程度的地域中聚落平均面积的一致性 检验和 EXCEL 软件中的 ANOVA 程序说明·····	(103)
5.3 一元方差分析的前提和对分析结果的解释·····	(107)
5.4 实例二：东周青铜剑、戈和削平均含锡量的一致性检验·····	(108)
5.5 实例三：垣曲盆地陶器的生产从个体家庭向专业化作坊演化的 定量证据·····	(111)
5.6 用箱点图和子弹形图比较多个总体的中位数和分布的差异·····	(113)
5.6.1 东周青铜戈戟、剑和削含锡量的中值和分布比较·····	(114)
5.6.2 北京故宫收藏的宋代官窑传世品的产地溯源·····	(114)
5.6.3 子弹形图·····	(117)
第六章 比例关系和比例数的统计学研究·····	(120)
6.1 贝努里试验和二项式分布·····	(121)
6.1.1 什么是二项式分布·····	(121)
6.1.2 二项式分布的性质·····	(124)
6.2 墓地人骨男女性比正常或异常的统计判断·····	(125)
6.3 随葬工具是否与性别相关联以及对氏族内部性别分工的推断·····	(128)
6.3.1 判断墓地出土的随葬工具与墓主人的性别是否存在关联的 统计学方法·····	(130)
6.3.2 根据墓地资料对墓地所属文化的性别关联型工具比例数的	

区间估计·····	(137)
6.4 两总体比例值一致性的显著性检验——以两个文化的性别关联型工具的比例值为例·····	(140)
6.5 单总体的比例数估计中置信度、精密度和样本容量间的制约关系·····	(142)
6.6 考古调查中某类实体的缺失能否说明该类实体确实不存在·····	(144)
第七章 两个数值变量之间的关系——相关与回归·····	(147)
7.1 实体按两个数值变量分布的图形表述——散点图·····	(147)
7.2 线性回归的基本原理和皮尔逊相关系数·····	(152)
7.2.1 怎样建立线性回归方程, 系数 a 和 b 的确定·····	(152)
7.2.2 皮尔逊相关系数与线性回归方程的检验·····	(155)
7.2.3 用 EXCEL 软件计算皮尔逊相关系数和进行线性回归分析·····	(159)
7.3 相关分析的应用实例·····	(161)
7.3.1 仰韶文化陶器上刻划符号出现频率的相关性研究·····	(161)
7.3.2 两个实验室测量同一批瓷片的钾含量的相关性分析·····	(164)
7.3.3 相关分析考古应用的其他实例简介·····	(165)
7.4 相关分析中需要注意的一些问题·····	(166)
第八章 有序变量间的等级相关·····	(169)
8.1 斯皮尔曼等级相关·····	(169)
8.1.1 斯皮尔曼等级相关系数的计算和显著性检验·····	(169)
8.1.2 关于斯皮尔曼等级相关的几点说明·····	(171)
8.2 Gamma 等级相关系数——陕西史家墓地墓葬分期方案的比较·····	(173)
8.3 Kendall's τ_b 和 τ_c 等级相关系数——对 Gamma 等级相关系数的修正·····	(178)
8.4 有序变量的两条百分累加曲线之间的一致性检验·····	(180)
第九章 实体按名称变量的分布和名称变量间的关联·····	(184)
9.1 实体按单个名称变量分布的 χ^2 检验·····	(184)
9.1.1 实体按名称变量进行分布检验的过程和实例·····	(184)
9.1.2 检验结论的准确表述·····	(186)

9.1.3	样本容量对分布检验的影响·····	(187)
9.1.4	定量表征两个分布间相异程度的 ϕ 系数·····	(188)
9.1.5	实体按名称变量的频次分布检验在区域考古调查中的应用 实例·····	(188)
9.2	两个二元名称变量间关联的假设检验·····	(191)
9.2.1	二元变量间 2×2 四格交叉列联表假设检验的原理和过程 ·····	(191)
9.2.2	样本的容量对检验二元变量间关联的影响和关联强度系数 ·····	(193)
9.2.3	二元变量间关联检验的前提条件·····	(195)
9.2.4	二元变量间关联检验的连续性修正·····	(196)
9.2.5	二元名称变量间的关联与因果关系的探讨·····	(196)
9.3	多状态名称变量间的关联和关联强度系数 V ·····	(199)
第十章	考古研究中的抽样问题·····	(203)
10.1	抽样问题的一般性讨论·····	(203)
10.1.1	系统抽样方法·····	(204)
10.1.2	随机抽样方法·····	(205)
10.1.3	集团抽样方法·····	(207)
10.1.4	分层抽样方法·····	(207)
10.2	考古调查和勘探中的采样问题·····	(208)
10.3	考古调查实例·····	(211)
10.3.1	探孔和探方的密度和面积·····	(211)
10.3.2	赤峰考古调查·····	(213)
参考文献	·····	(216)
简明考古统计学习题	·····	(218)
附录一	标准型正态分布临界值表·····	(230)
附录二	t 分布临界值表 (双侧) ·····	(231)
附录三	χ^2 分布临界值表·····	(232)
附录四	不同的 df_1 和 df_2 取值条件下 α 与 F_α 间的对应关系表·····	(233)

第一章 考古观测数据的最基本处理

——随机现象的描述性统计

考古学是通过寻找、发掘、观测和研究古代人类活动的实物遗存来复原古代社会、探讨人类活动规律的一门学问。考古学家感兴趣的是古代社会、古代人类活动的整体，但他们所能掌握的实物遗存资料因千万年的沧海桑田总是零星的乃至是变形的，他们面临的任务是根据局部的知识去推断总体的规律，属于统计性的推断。此外，考古学研究的实物遗存不是单件器物、单座墓葬或单个遗址，而是器物群、墓葬群和遗址群。同一类遗存各实物的观测数据之间，或同一个实物多次观测结果之间一般是存在涨落和变化的，即属于统计学所研究的随机现象。因此考古学研究需要借助研究随机涨落现象的概率统计学。本章将介绍对观测数据组的最基础性的处理、研究方法，即描述性的统计。同时本章也将介绍怎样应用微软办公软件中 EXCEL 程序进行最基本的统计学计算。

1.1 考古研究的对象与随机现象

现实世界的客观规律，从预测的角度可以分为两类，即必然性的规律和随机性的规律。属于前者的小者如物体的自由落体运动：放开手中的石子，它一定垂直落地，而且可计算落地所需的时间和落地时石子的速度；大者则如太阳系中地球和月球的运动，我们不仅能预测日夜交替、月盈月亏，而且可准确预测在地球各地观察日月食发生的时间和过程。但是在自然界和特别是在人类社会中存在更多的现象，当对它们进行单次观测时，其进程和结果是不能准确预测的。这类现象的例子是很多的，如：①投掷一枚硬币，无法预测它落地时是正面向上还是反面向上，两种结果出现的可能性各为 50%。②随机找一个中国成年男子，我们不能预测他的身高为多少厘米，可能是高个，

也可能是矮个。③随机挖掘一座单人墓葬，墓主人的性别可能是男，也可能是女；其年龄段可能是少年、青年、壮年或老年。④某个地区某文化类型聚落的面积，在全面考古发掘前不能准确预测它的面积。⑤系统性区域考古调查中从每个采样单位（例如中美联合赤峰区域调查取 100×100 平方米为一个采样单位）地表所采集到某个时期陶片的数量也或多或少。上述这些现象称为随机现象，对随机现象进行观测的结果称为随机数据，或随机变量。考古研究的对象绝大多数属随机现象，其观测数据也属随机变量。概率统计学是专门研究随机现象的一门学科，属数学的一个分支学科，它对随机现象的不确定性给出明确和清晰的表述。

1.1.1 随机变量的特点

随机现象不服从必然性的规律，单次观测随机现象的结果无法准确预测，但是这并不意味着随机现象的观测结果是杂乱无章和无规律可循的。对于前面所举硬币的例子，假如我们将一枚硬币重复投掷十次（或一次同时投掷十枚同样的硬币），那么可以预测，在十次投掷中，正面或反面出现的次数大致是相近的，都在五次左右，即正面或反面出现次数的百分比都接近 50%。十次投掷结果均为正面或均为反面的可能性是极小的（可以计算其可能性约为千分之一）。随着投掷次数的增加，出现正面或反面次数的百分比则越接近 50%，而出现全部为正面或反面的可能性越来越低并很迅速地趋于零。在关于我国成年男子身高的例子中，如果我们用 167 ~ 175cm 的身高区间来预测一个随机抽取的男子的身高，那么预测正确的可能性应高于 50%，如果用一个更宽的区间（155 ~ 190cm）预测，那么有 99% 以上的把握是对的。另一方面如果随机找十名成年男子，预测他们的平均身高，还是用上述的两个区间来预测，预测的准确率将显著提高，应该有高于 80% 的把握可以确定十名男子的平均身高在 167 ~ 175cm 范围内。

从上面两个例子可以简要归纳随机现象的主要特点：

（1）单次观测的结果一般是难以准确预测的，而且单次观测结果也不能准确反映该类随机现象的普遍规律。

（2）对随机现象的研究必须要进行多次的重复观测，包括对同一个个体的某种属性在相同的条件下多次观测，或者对同一类型的多个个体的同一种

属性（例如多个男子的身高）进行群体观测。我们把所有被观测的个体的集合或所有观测数据的集合称为样本，而每个个体或每个观察数据是组成样本的实体，也可称为个案、样品等。在统计学中“样本”和“实体”都是基础性的概念。在英语中样本是 sample，而实体、个案和样品是 entity, case 和 specimen。样本与样品是两个完全不同的概念，样本是由多个样品组成的。对实体的某个属性进行多次观测的结果将得到一组数据，每一个实体有一个观测数据，因此，样本是由一组数据组成的，样本也称数据组。通过对数据组的分析来推测相关随机现象或随机变量的性质。

（3）多次重复观测的综合结果一定程度上能反映随机变量的客观规律性，称为统计性规律。统计性规律与必然性的规律不同，它不能预测一定会发生某种结果，而是预测某种结果会以多高的可能性（统计学中称之为概率）出现。例如，气象预报是对随机现象的预测，不少国家气象台预报的是第二天或未来几小时内发生某种天气现象，如晴和雨的概率多大。

（4）即使是根据多次重复观测的结果所推论的规律或所进行的预测，也仅是一种可能性预测，依然还可能预测错误，包括犯弃真或纳伪两类错误，正如预报晴天或降水均有可能与实际情况不符。但是对预测中犯错误的可能性，统计学中称为犯错误的概率，也是可以估测的，而且一定程度上也是可以控制的。

（5）观测的次数越多，观测的平均结果越稳定。根据观测结果所作推论越可靠，误差越小，犯预测错误的可能性也越小。反之，对基于少量几次观测所作的推论，需要持谨慎态度。“孤证不足信”也是这个意思。

以上是随机变量和随机规律的基本性质和特点，阅读了本书下面的章节后读者将对上述的内容有更深刻的理解。下面介绍随机变量的类型。

1.1.2 随机变量的类型

根据对随机变量观测的结果能否用数值来表示和观测结果能否按一定次序排列，随机变量可以分为三种类型，即数值变量、有序变量和名称变量。

（1）数值变量 观测结果能用数值表示的称为数值变量。如前面提到的人体的身高、聚落的面积、采样单位上采集到的陶片的数目等都是数值变量。此外，如从居址和墓地中出土的某种器物的数量及其百分比、某类器物的几

何尺寸和重量、陶瓷和青铜器的化学组成等都是数值变量。数值变量是最常见的随机变量，它可以直接参与通常的加、减、乘、除等数学运算，也可以比较两个观测结果之间的大小。

数值变量又分成离散型和连续型两类，最典型的例子是母鸡每天的下蛋量和母牛每天的出奶量，前者属离散型变量，因为它只能取 0、1、2 等正整数；而后者则可以取 0 ~ 20 升之间的任何整数和小数，因此属连续型变量。考古研究中墓葬中某种器物出现的数量只能是正整数，属离散型数值变量；而墓主人的身高、墓坑的深度、聚落的面积和青铜的含锡量等则可以用小数来表示，就是连续型数值变量了。处理两种类型的数值变量用不同的方法，连续型数值变量经常用后面要介绍的正态分布函数来处理，而离散型变量则常用二项式分布函数来处理。

(2) 有序变量 如果观测的结果不能用数值表示，而只是反映实体处于某种特定的状态，而这些状态又可以按某种规律排序的话，那么观测的结果就属于有序变量。例如，在分析墓葬出土的人骨的年龄组成时，把人骨按年龄段分成婴儿、儿童、少年、青年、壮年、中年和老年 7 个状态，这 7 个年龄属性是有顺序关系的，因此是墓主人的有序属性。墓主人的年龄就属于有序变量。此外，器物和墓葬的分期、地层的次序（第几自然层或第几文化层）和沉积物的粒度（黏土、粉沙、细沙、粗沙、砾石）等都是有序变量。

对于人骨年龄段，从婴儿到老年 7 个状态也可以用 1、2、3、4、5、6、7 七个数字来相应地描述，这时数字的大小反映顺序的位置，有序变量两个取值之间的减法运算是有意义的，差值表示两个状态相距几个序位。数学运算符大于“>”和小于“<”也是有意义的，它们表示状态的先后。但是加法和乘除法运算对于有序变量却是没有意义的。

还需要指出，有序变量两个相邻状态之间的“差距”不一定等距，如果等距则上升为“数量变化”，序位的差异并不表示在时间或空间上的数值差距。举例来说，地层有一定的次序，但各层的堆积厚度可以不一样，其堆积延续的时间，即时间跨度也完全可以不一样。两个相邻层位之间的时间间隔可能很接近也可能相距甚远。

(3) 名称变量 如果随机变量的观测结果不是数值，而是反映实体处于若干可能状态中的某个状态，而且各状态间不存在次序关系，那么这个随机变量属于名称变量。举例来说：墓主人的性别，或男或女，属两种状态的名