

军事医学课程系列教材

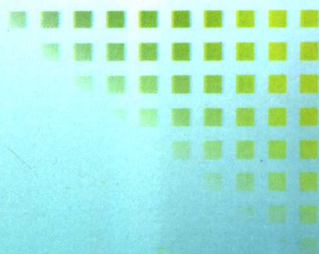
军事医学

统计学

J

JUNSHI YIXUE
TONGJIXUE

●主编 易东



第四军医大学出版社

军事医学课程系列教材

军事医学统计学

主 编 易 东

编 委 (以姓氏笔画为序)

王文昌 尹全焕 孙 鹏

刘 岭 伍亚舟 张彦琦

赵增炜 郭波涛 袁 伟

黄继东

学术秘书 郑永丽 敬培胜 张翼飞

第四军医大学出版社

图书在版编目(CIP)数据

军事医学统计学/易东主编. —西安:第四军医大学出版社, 2006. 8
ISBN 7 - 81086 - 293 - 6

I. 军… II. 易… III. 军事医学 - 医学统计 IV. R821.3

中国版本图书馆 CIP 数据核字(2006)第 087873 号

军事医学统计学

主 编 易 东
责任编辑 土丽艳 潘 峰
出版发行 第四军医大学出版社
地 址 西安市长乐西路 17 号(邮编:710032)
电 话 029 - 84776765
传 真 029 - 84776764
网 址 <http://press.fmmu.sn.cn>
印 刷 西安永惠印务有限公司
版 次 2006 年 8 月第 1 版 2006 年 8 月第 1 次印刷
开 本 787 × 1092 1/16
印 张 21.5
字 数 510 千字
书 号 ISBN 7 - 81086 - 293 - 6/R·221
定 价 40.00 元

(版权所有 盗版必究)

前 言

面向 21 世纪军事医学课程改革,是适应新形势下军事卫勤保障的需要,是部队培养合格的新型军事医学人才所采取的重要举措。军事医学课程系列教材的改革是整个课程改革的重要组成部分。本教材就是根据这一系列改革成果,结合军事医学和统计学的特点,经过三年的教学改革实践,在广泛征求意见和资料收集的基础上,经过编者大量辛勤劳动精心编写而成。

本教材有两个重要特色。一是强调了与军事医学的结合,主要体现在大量例子的引用多来自部队的实际数据,同时还增加了军队卫生统计工作的基本内容、方法、特点,详细介绍了军队平、战时卫生统计工作,以及部队医院统计等内容。二是内容丰富,深入浅出。全书共分上、下两篇,上篇强调一些基本的统计学方法,下篇介绍了现代统计分析方法,包括统计预测、评价,以及临床随访资料分析等现代统计分析方法等。另外,书中还以专门的章节介绍了军队平、战时卫生统计工作,以及部队医院统计工作等内容。

在教材编写的过程中,编者力求从实际应用的角度出发,侧重统计基本思想的介绍,尽量减少理论公式的推导,理论思路逻辑性强,内容丰富新颖,实用性强,覆盖面广,可作为大专生、本科生以及研究生的教材,也可作为医疗、科研、预防保健人员的实用参考书。

本教材的编写成员均具有丰富的教学经验,在编写过程中付出了辛勤劳动和热情。但囿于编者水平,书中的不足之处在所难免,欢迎广大读者批评指正。

编 者

2006 年 8 月

目 录

上篇 基本统计分析方法

第一章 军事医学统计学概述	(2)
第一节 军事医学统计学	(2)
第二节 常用名词概念	(5)
第三节 统计资料类型	(8)
第二章 资料描述性统计	(10)
第一节 相对数指标	(10)
第二节 平均数指标	(14)
第三节 变异指标	(22)
第三章 医学统计推断基础	(25)
第一节 随机变量的分布	(25)
第二节 正态分布	(27)
第三节 二项分布与 Poisson 分布	(33)
第四章 参数估计与参考值范围估计	(36)
第一节 标准误与可信区间估计	(36)
第二节 参考值范围估计	(41)
第五章 t 检验与单因素方差分析	(45)
第一节 假设检验概述	(45)
第二节 t 检验	(46)
第三节 单因素方差分析	(49)
第六章 列联表分析	(56)
第一节 四格表资料的 χ^2 检验	(56)
第二节 行 $\times$ 列表资料的 χ^2 检验	(63)
第三节 χ^2 分割	(66)
第七章 非参数统计分析方法	(69)
第一节 配对样本比较的 Wilcoxon 符号秩检验	(69)
第二节 两独立样本比较的 Wilcoxon 秩和检验	(72)
第三节 完全随机设计的 Kruskal - Wallis H 检验	(75)

第四节	随机单位组设计的 Friedman M 检验	(78)
第五节	Ridit 分析	(80)
第八章	回归与相关	(83)
第一节	直线回归	(83)
第二节	直线相关	(91)
第三节	等级相关	(94)
第九章	实验设计与调查设计	(96)
第一节	概述	(96)
第二节	实验设计	(99)
第三节	调查设计	(111)
第十章	统计表与统计图	(123)
第一节	统计表	(123)
第二节	统计图	(124)
第十一章	部队平时卫生统计	(133)
第一节	部队卫生统计工作的任务、作用与特点	(133)
第二节	部队人员健康统计	(134)
第三节	部队疾病统计	(136)
第四节	卫生工作统计	(139)
第十二章	部队医院卫生统计	(142)
第一节	军队医院统计工作的作用与特点	(142)
第二节	卫生资源统计	(143)
第三节	医疗工作数质量统计	(145)
第四节	综合效益统计	(149)
第十三章	战时卫生统计	(151)
第一节	战时卫生统计报表与主要登记格式	(151)
第二节	战时卫生减员统计与分析	(152)
第三节	战伤救治工作统计	(156)

下篇 高级统计分析方法

第十四章	多元线性回归与相关分析	(162)
第一节	多元线性回归	(162)
第二节	自变量筛选方法	(168)
第三节	多元线性回归的应用及注意事项	(173)
第四节	典型相关分析	(175)

第五节 典型相关系数的显著性检验	(179)
第六节 典型相关应用实例	(180)
第十五章 曲线拟合	(183)
第一节 常用的曲线函数	(183)
第二节 曲线拟合的线性化技术	(185)
第十六章 多因素方差分析	(189)
第一节 析因设计资料的方差分析	(189)
第二节 正交设计与方差分析	(195)
第三节 嵌套设计资料的方差分析	(201)
第四节 裂区设计资料的方差分析	(203)
第五节 两阶段交叉设计与方差分析	(205)
第六节 重复测量设计的方差分析	(207)
第十七章 协方差分析	(218)
第一节 协方差分析的基本原理	(218)
第二节 完全随机设计资料的协方差分析	(221)
第三节 协方差分析的应用条件及注意事项	(223)
第十八章 生存分析	(225)
第一节 生存分析中的基本概念	(225)
第二节 生存率的估计与生存曲线	(228)
第三节 生存曲线的对数秩检验	(233)
第四节 寿命表	(236)
第十九章 广义线性模型	(240)
第一节 Cox 比例风险回归模型	(240)
第二节 logistic 回归分析	(248)
第二十章 其他多元统计分析方法	(262)
第一节 判别分析	(262)
第二节 聚类分析	(268)
第三节 主成分分析	(276)
第四节 因子分析	(286)
第二十一章 统计预测与综合评价概述	(296)
第一节 统计预测概述	(296)
第二节 综合评价概述	(299)
附表 1 标准正态曲线下的面积表	(305)
附表 2 t 值表	(306)

附表 3 方差齐性检验 F 值表	(308)
附表 4(1) 方差分析 F 值表	(310)
附表 4(2) 方差分析 F 值表	(312)
附表 4(3) 方差分析 F 值表	(314)
附表 4(4) 方差分析 F 值表	(316)
附表 5 SNK 法两两比较 q 界值表	(318)
附表 6 χ^2 值表	(319)
附表 7 配对比较的符号秩和检验用 T 界值表	(321)
附表 8 两样本比较的秩和检验用 T 界值表	(322)
附表 9 H 界值表(三样本比较的秩和检验用)	(323)
附表 10 M 界值表(随机区组比较的秩和检验用)($P = 0.05$)	(324)
附表 11 r 值表	(325)
附表 12 r_s 界值表	(328)
附表 13 随机数字表	(329)
附表 14 样本均数与总体均数比较(或配对比较)所需样本含量	(331)
附表 15 两样本均数比较所需样本含量	(332)
附表 16 多个样本均数比较估计样本含量估计用 Ψ 值	(333)
附表 17(1) 两样本率比较所需样本含量(单侧)	(334)
附表 17(2) 两样本率比较所需样本含量(双侧)	(335)
附表 18 多个样本率比较估计样本含量用 λ 值	(336)

上

篇

第一章 统计学的意义

统计学的意义

统计学的意义

统计学的意义

基本统计分析方法

统计学的意义

统计学的意义

统计学的意义

第一章 军事医学统计学概述

第一节 军事医学统计学

一、军事医学统计学的重要性和任务

(一)军事医学统计学的意义和作用

医学统计学(medical statistics)是运用概率论与数理统计的原理及方法,结合医学实际,研究数字资料的收集、整理、分析与推断的一门学科。医学研究的主要对象是人和有关的影响因素。人本身是一种生物,具有生物的一般特性,即生物现象有较大的变异;但人又不同于一般生物,影响人体健康的因素除了自然因素外,还有心理、社会等因素。在人体研究中,生物学研究常用的损害性、破坏性的方法不能使用。由于上述原因,使医学研究变得更复杂,医学数字资料的收集、整理与分析也更困难。随着医学研究的发展,20世纪20年代后逐渐形成了医学统计学。60年代以来,医学研究的迅速发展又促进了医学统计学的迅速发展。电子计算机的发展、应用与普及为医学统计学提供了强有力的工具,使医学统计学进一步向深度和广度发展,成为医学研究中必不可少的手段。

军事医学统计学(military medical statistics)是医学统计学在军队卫生工作中的应用,其原理、方法和基本内容与医学统计学相同,但突出军队是武装集团的特点。军事医学统计的主要作用是及时准确地反映部队卫生工作情况的各种统计资料和统计分析结果,为各级军政首长和卫勤领导制订卫生工作规划、考核指导工作及科学管理和决策提供统计学依据。

(二)我军医学统计工作的基本内容

军事卫生统计工作是卫生统计学(health statistics)在部队卫生工作中的应用,其原理、方法和基本内容与卫生统计学相同,但突出了军队是武装集团的特点。它除介绍统计学的基本知识(如统计工作的步骤、常用统计指标的计算、统计表与统计图的绘制等)外,主要包括以下内容。

1. 部队平时卫生统计 是军队卫生统计中最基本的内容,无论部队、机关、医院、院校,都必须按统一规定的内容和时间完成登记、统计并逐级上报。其中主要有:①反映军队人员健康水平的健康统计;②反映军队人员疾病发生及其危害情况的疾病统计;③反映军队开展预防工作和健康教育情况的卫生防疫统计;④反映军队卫生资源利用和卫生力量状况的卫生业务统计等。军队卫生统计工作已形成一套完整的制度,有严格的规定,并已纳入军队卫生统计学的内容。

2. 医院卫生统计 其研究对象是医院管理工作中的各种随机现象,透过其偶然性探讨

其规律性。它的主要内容是:①医院卫生资源统计,目的是了解卫生资源(含人员、设备、经费与信息等)的投入、分配及利用是否合理与有效;②医疗工作数、质量统计,目的是了解工作量以及工作量与人员配备是否相适应,能否及时做出诊断与诊断是否正确,治疗是否有效、及时和彻底,抢救危重病人是否成功,护理技术操作是否正规、熟练,服务是否优质等;③医院综合效益统计,分社会效益与经济效益,前者是了解医疗、科研、人才培养以及为部队、为地方群众服务工作的情况,后者是了解医疗费用的收支和对上级拨款的使用情况等。近几年来,医院卫生统计发展很快,在医院信息管理系统和医院统计指标体系的研究与实践中取得了积极有效的成果,统计资料的搜集、整理、分析与信息反馈正逐步实现自动化,从而促进了医院管理的科学化。因此,医院卫生统计的内容已越来越充实。

3.战时卫生统计 由于战时情况紧急,部队行动快,因而要求统计工作必须内容简炼,填写方便,报告及时。其主要内容有:①部队减员统计,以反映指战员因伤、因病减员情况;②战伤救治工作统计,以反映各级救治机构的床位展开、救治效果及伤病员转运等情况;③伤员的伤情统计,以反映负伤的类别、部位及其原因等情况。海、空军及特种兵部队可根据自身特点适当修订或补充部分具体内容,但须报总后卫生部审查备案。此外,平时发生的各种灾害,如地震、洪水、火灾、空难等,其抢救中的卫生统计工作,也可参照战时卫生统计执行。

(三)军医掌握统计知识的重要性

医务人员既是卫生统计工作的执行者,又是受益者。军医掌握统计知识的重要性主要体现在以下几个方面:

1.医生的正确判断,如患者的病情诊断、治疗方法的选择与预后分析等,在很大程度上取决于统计信息的积累。

2.临床和卫生防疫中的许多观察数据具有不确定性,如患者的各项检查、化验数据,某一单位不同时期的发病人数等。没有一定的统计知识就不能正确地解释和评价这些数据。

3.大量的外界信息,如各种传播媒介、科技文献和情报资料,都伴随有直接或间接的统计信息,没有一定的统计常识,就无法正确理解和迅速吸收新知识。

4.军队卫生统计的许多第一手资料产生于军医之手,如门诊登记、体格检查记录、病历等。军医应该清楚如何从这些资料中提取出统计信息,并使之发挥更大的作用。

5.具备一定的统计知识,能对科研数据进行基本的统计分析,是每个医学科学研究者应具备的基本素质。

二、军队医学统计工作的要求

军队医学统计工作的基本要求是准确、完整、及时和保密。

(一)统计数据必须准确

统计数据的准确首先要求原始登记的准确,因此要求军医须正确理解医疗文书的各项内容并认真填写;其次是在数据的整理汇总过程中防止出现过失误差。负责统计工作的军医决不能在统计数据上弄虚作假,并且要勇于同篡改统计数字的不良行为作坚决斗争。

(二)统计资料力求完整

统计资料应能反映一个单位不同时期卫生工作全面情况,因此要求登记和统计的项目不能残缺不全。对于历年积累的统计资料,要定期进行整理、汇总和写出分析报告,并列入

移交。

(三)提供统计资料要及时

统计资料具有很强的时效性,在战时更是如此。各基层单位应按规定时间向上级单位报送统计数字,否则将影响全军统计资料的汇总。此外许多专题统计调查规定有统一的起始时间,如果延误了时间将影响整个统计调查的工作进度,甚至会危害整个调查结果的代表性和科学性。

(四)提高警惕、保守军事秘密

军队医学统计资料直接或间接地反映了部队的编制、人数、伤亡和疫情等军事秘密,如果麻痹大意或保管不当容易造成失密或泄密。因此,军医要提高警惕,对所掌握的统计资料要切实按保密制度严加保管。因学术交流须公开发表统计资料时,要防止泄密。在对外报告时,单位名称要用代号,尽量避免使用绝对数。有些疫情资料的公布,还需经国家有关部门的批准。

三、军队医学统计工作的步骤

军队医学统计工作的步骤与其他行业的统计工作的步骤基本相同,即先有一个周密、完整的统计设计,然后再按照设计的要求收集资料、整理资料和分析资料。

(一)统计设计

在医学研究设计中,除了医学专业设计外,还需要对数字资料的收集、整理、分析的方法和过程进行必要的设计,称为统计设计。医学研究一般分为调查与实验两类,统计设计也相应地分为调查设计与实验设计。由于临床试验研究的对象主要是病人,既不能像实验研究那样进行随机抽样与分组,也很难严格控制非实验因素,因而临床试验设计与一般实验设计有所不同。

(二)收集资料

收集资料是根据统计设计的要求采集原始数据的过程。由于原始数据关系到最终统计分析的正确性,而且原始数据中有错误很难事后补救,所以对收集资料的每一个环节必须高度重视。

1. 资料的来源

(1)经常性资料:经常性资料(regular data)来自于常规的数据采集系统。在我国,主要是国家、军队规定的统计报告(报表)制度和日常医疗卫生工作登记制度,如出生、死亡登记、疫情报告和统计报表制度;我军平时军队卫生统计报表制度、军医门诊(巡诊)登记制度、医院病案书写和管理制度;战时伤票登记制度等。

国家的法律、军队的纪律、卫生统计常设机构和专(兼)职统计人员为经常性统计资料的准确性提供了有力的保障。我军关于军队卫生统计报表的纪律是:①对报表项目和格式不得擅自修改;②按规定日期和程序上报;③不得伪造数据;④保守机密。

我军团以上卫勤机关设专人分管统计工作,驻军以上医院设有专职统计员或病案统计室,每个军医要自觉接受他们的指导和监督,严格执行各项登记、统计规章制度,为保证各项原始记录的正确、完整作出自己的努力。

(2)专项调查资料:专项调查资料(adhoc data)是为特定目的而专门收集的资料。

2. 数据的采集方式 数据的采集方式主要有以下三种:①直接使用测量工具,如磅秤、

体温表、血压计、血气分析仪等;②主观判断,如常见病诊断、临床疗效评价等;③借助测量仪器进行判断,如阅看 X 线片等。

从数据质量看,直接用仪器测量的数据受主观因素的影响较小,比较可靠。因此,在设计调查项目时应尽量选择测量指标。如果在调查项目中有主观判断的项目,数据采集人要统一判别或评价标准,必要时进行集体培训。

(三)整理资料

整理资料是将原始记录按统计的要求进行归纳分组的过程。原始记录经整理后,可以将原始记录转变为有分析价值的数字信息,并且容易显现出隐藏在原始记录中的统计规律。此外,在资料的整理过程中,通过逻辑检查和简单计算,如经过简单的归纳和整理,计算均数、率等统计指标,这样能够及时发现、更正或删除错误的原始记录,以保证下一步统计分析的正确性。

(四)分析资料

医学研究的最终目的是通过样本的信息去推断总体的情况,包括两个重要领域:一是参数估计,即用样本统计量来估计总体中相应的参数;二是假设检验,即依据资料性质和所需解决的问题,建立统计假设,然后采用适当的检验方法,根据样本是否支持所作的假设,来决定该假设应当拒绝或不拒绝。

分析资料是在分组的基础上,计算有关统计指标,然后结合各专业的研究背景和业务知识对统计指标的大小和差别做出合理的分析、解释。

分析资料主要有以下三个步骤:

1. 选择正确的统计指标。这里最主要的判别标准就是要根据资料的具体分布情况来正确选择,使计算的统计指标能反映出该资料的特征。

2. 为便于分析和比较,可将分组计算后的统计指标列在统计表内或作统计图(详见第十章)。

3. 对经常性资料和普查资料,可直接比较统计指标的大小和组间差别,并做出合理的分析和解释。如果是根据抽样调查资料推论全体观察对象的特征,则需要估计统计指标的抽样误差,并对组间差别作假设检验,有关统计方法将在以后各章介绍。

第二节 常用名词概念

这里先介绍一些常用的医学统计学的基本概念,从中也可了解到本课程的基本思路和总体框架。

(一)同质与变异

同质(homogeneity),系指事物的性质、影响条件或背景相同或非常相近;变异(variation),系指相同条件下的同类个体之间某一方面发展的不平衡性。例如研究入伍新兵的身体状况,同性别、同年龄的新兵,在统计上称为“同质”观察单位;但是,一批同性别、同年龄的新兵,他们的身高、体重、胸围等却各不相同,互有差异,这就是变异。又如在同样条件下接受同一种疗法的某病患者,可认为是“同质”观察单位;但他们的疗效并不一样,有的被治愈,有的略有好转,还有的反而恶化甚至死亡,这也称作“变异”。个体变异可表现为定量,如身高的高低、体重的轻重等;亦有表现为定性的,如用某药治疗某种病人后的痊愈、好转、

恶化等。

同质观察单位之间的个体变异,是生物的重要特征,是偶然性的表现,是由于生物体内外环境中多种因素的综合影响造成的,其中许多因素是未知的,也是难以控制的。统计研究的是有变异的事物,其任务就是在同质分组的基础上,通过对个体变异的研究,透过偶然现象,反映同质事物的本质特征和规律。

(二)总体与样本

总体(population)是根据研究目的确定的同质的研究对象的全体。更确切地说,是性质相同的所有观察单位某种变量值的集合。例如研究某地某年正常成人的血压值,则研究对象是该年的正常成人,观察单位是每个人,变量是血压,变量值是测得的血压值,该地该年全部正常成人的血压值就构成一个总体。它的同质基础是同一地区、同一年份、同为正常成人。这里的总体只包括有限个观察单位,称为有限总体(finite population)。有时总体是设想的,如研究高血压患者用某药治疗后的血压,这里总体的同质基础是同为高血压患者,同用某药治疗,包括设想该药治疗的所有高血压患者,其观察单位数显然是不确定的无限的,称为无限总体(infinite population)。

医学研究中,很多是无限总体,要直接研究总体的情况是不可能的。即使对有限总体来说,若包括的观察单位数过多,直接研究总体也是耗费人力、财力很大的,有时也是不可能和不必要的,如成批罐头食品的生产,不可能将所有罐头逐个作食品卫生检查。所以在实际工作中,常常是从其样本的信息来推断总体的性质。

样本(sample)是从总体中随机抽取的部分观察单位,其实测值构成样本。样本性质的研究是对总体性质研究的基础,如在对某地区成年人血压的研究中,可从该地某年的正常成人中,随机抽取 200 人,分别测定其血压值,组成样本,计算样本均数,用来估计该地该年正常成人血压的总体均数。这种方法称为抽样研究,这是一种很常用、极其重要的科学研究方法。

要做好抽样研究,必须注意①样本的可靠性:首先要明确划清总体的同质范围,也就是确定被研究对象的同质基础,包括时间、空间、条件等。如观察某种疫苗预防接种的效果,首先要确定易感人群,接种组和未接种组的界限清楚,未完成全程接种的人如何划分等也要作出规定。②样本的代表性:必须遵循随机抽样的原则,必须保证样本中有足够的观察单位数,即样本含量要足够大。

(三)参数与统计量

参数(parameter)是用来表示总体分布特征的统计数字。统计中常用的总体参数有描述总体中心位置或集中趋势的总体平均数、中位数和众数等;有描述总体离散度的总体标准差、极差和四分位数间距等。研究总体的特征往往是医学统计研究的目的,若能直接观察总体的每一部分无疑是最理想的,但事实上是做不到的,主要因为:①总体中的个体数无穷多;②总体随时间和空间不断变化,如某病患者不断发生、其转归情况也不断变化;③研究方法具有破坏性;④用于某项研究的技术力量、设备条件或资金不足等,限制了人们直接了解总体全貌。故人们只能从总体中抽取一部分具有代表性的样本进行检测,从而对整个总体进行推断。这也是整个概率统计研究的最基本的思路和方法。

统计量(statistic)是我们对样本作了若干次观察后,依据观察值推算出的反映样本分布特性的一些量。例如,为了了解健康成年男子每升血液中白细胞数而进行检验,由所测得

的一系列数值算出一个平均数,这个平均数就是一个“样本统计量”。样本标准差和样本率也是统计量。

从研究总体中抽取部分观察单位构成具有代表性的样本,用样本统计量推断总体参数的方法称为抽样研究(sampling research)。抽样研究有两个目的:其一是估计总体参数;其二是判断某个随机变量是否取自某一预期或假设的总体。它们也是参数估计与统计量关系研究的主要内容。

(四)确定性现象与随机现象

在一定条件下,必定会发生或必定不会发生某一种结果的现象,称为确定性现象(decided phenomenon)。其表现结果为两种形式相反的事件:肯定发生某种结果的叫做必然事件,肯定不发生某种结果的叫做不可能事件。例如在适当温度下经过一定时间孵化,正常受精鸡蛋必然会孵出小鸡来,而石头是不可能孵出小鸡来的。

在同样条件下,可能出现两种或多种结果,而事先不能确定发生哪一种结果的现象,称为随机现象(random phenomenon)。其表现结果叫做随机事件。例如胎儿的性别,可能是男,也可能是女,事先不能肯定,这是一个随机现象,而结果胎儿为男性,则是一个随机事件。随机现象有两个主要特征:①随机性,即它的发生是随机的;②规律性,即它每次发生的可能性的的大小是确定的。在医学研究中,大量的随机现象,通过对某随机现象的研究,找出其规律性。

(五)频率与概率

1. 频率(frequency) 一个随机试验有几种可能结果,在重复进行试验时,个别结果看来是偶然发生的,但当重复试验次数相当大时,总有某种规律性出现。例如,投掷一枚硬币,结果不外乎出现“正面”与“反面”两种,历史上有些人对此作过试验并得到如下结果:

实验者	投掷次数	出现“正面”次数	频率
Buffon	4040	2048	0.5069
k. Pearson	12000	6019	0.5016
k. Pearson	24000	12012	0.5005

可见,在相同条件下重复试验,试验结果为“正面”或“反面”虽不能事先断定,但我们知道试验的所有可能结果只有两种。在重复多次后,出现“正面”(或“反面”)这个结果的次数与试验次数的百分比称为频率,当试验次数逐渐增多时,其频率在一个常数(0.50)附近摆动,且摆动的幅度随试验次数的增加而变小,并逐渐出现稳定性(越来越接近0.50),这就是定义随机事件的概率的客观基础。

2. 概率(probability) 概率是描述随机事件发生可能性大小的一个度量。设在相同条件下,独立地重复 n 次试验,随机事件 A 出现 f 次,则称 f/n 为随机事件 A 出现的频率。当 n 逐渐增大时,频率 f/n 始终在一个常数左右作微小摆动,则称该常数为随机事件 A 的概率,可记为 $P(A)$,简记为 P 。在实际工作中,当概率不易求得时,只要观察单位数足够多,可将频率作为概率的估计值。但在观察单位数较少时,频率的波动性很大,用于估计概率是不可靠的。

随机事件概率的大小界于0与1之间,即 $0 \leq P \leq 1$,常用小数或百分数表示。 P 越接近1,表示事件发生的可能性越大, P 越接近0,表示事件发生的可能性越小。在一定条件

下必然发生的事件称为必然事件,必然事件发生的概率 $P=1$;在一定条件下不可能发生的事件称为不可能事件,不可能事件发生的概率 $P=0$,这两类事件具有确定性,不是随机事件,但可视为随机事件的特例。统计分析中将 $P \leq 0.05$ 称为小概率事件,表示在一次实验或观察中该事件发生的可能性很小,可视为很可能不发生。

(六)误差

观察值与真值之差称为误差(error)。误差主要有三类:

1. 过失误差(gross error) 又称粗差,由观察者的错误造成的误差。如记录误差、计算误差或错用分析方法等。这类误差是不允许的,必须消除。

2. 系统误差(systematic error) 由于受试对象、研究者、仪器设备、研究方法、非实验因素影响等原因造成的,有一定倾向性或规律性的误差。如抽样不匀、仪器未校正、操作技术不当、计算方法不完善等。这类误差应通过严格的设计和严密的技术措施尽力消除或减到最小。

3. 随机误差(random error) 或称偶然误差,是排除了上述误差后,由于众多无法控制的因素引起的,大小及方向不一的随机变化。随机误差是不可避免的,它服从一定的分布规律,可以用概率统计方法处理。

由于总体中各观察单位间存在个体变异,抽样研究中得到的样本不会完全反映总体的信息。样本统计量与总体参数间存在的差异称为抽样误差(sampling error),在随机误差中,最重要的是抽样误差。显然,抽样误差愈小,用样本推断总体的精确度就愈高,反之亦然。由于生物的个体变异是客观存在的,因而抽样误差是不可避免的,但抽样误差是可以控制的。抽样误差有一定的规律性,在没有人或外来因素干扰的条件下,如果样本不是太小,则抽样误差将会时大时小而小多大小,或正或负而正负近似各半,随着抽样次数的增加,各抽样误差的代数和逐渐趋向于“0”。因此,大多数来自样本的平均指标或相对指标都围绕在相应总体指标的上下。在根据样本估计总体时,这是必须正确领会的一个基本概念。

第三节 统计资料类型

确定总体之后,研究者应对每个观察单位的某项特征进行测量和观察,这种特征能表现观察单位的变异性,称为变量(variable)。对变量的测得值称为变量值(value of variable)或观察值(observed value),变量的取值若是连续的称为连续变量(continuous variable),若是离散的称为离散变量(discrete variable)。以人为观察单位调查某地某年7岁正常儿童的生长发育状况,性别、身高、体重等都可视为变量,性别有男有女,身高可高可矮,体重可轻可重,不同个体不尽相同,这种个体间的差异称为变异,这些变异来源于一些已知的或未知的,甚至是某些不可控制的因素所导致的随机误差。变量的观察结果可以是定性的,例如属男属女;也可以是定量的,例如身高的厘米数。按资料属性或定量,可将它们分为以下几种类型:

1. 计量资料(measurement data)又称定量资料(quantitative data),或数值变量(numerical variable)。为测定每个观察单位某项指标量的大小,经整理后而获得的资料。其变量值是定量的,表现为数值大小,一般有度量衡单位。如调查某地某年7岁正常男童的身体发育状况,以人为观察单位,每个人的身高(cm)、体重(kg)、血压(Kpa)、坐高指数(% ,坐高/身

高)等均属此类资料。

2. 计数资料(enumeration data), 又称定性资料(qualitative data)或无序分类的分类资料(unordered categorical data)。为将观察单位按某种属性或类别分组计数, 分组汇总各组观察单位数后得到的资料。其变量值是定性的, 表现为互不相容的属性, 而且一般是非数字的, 例如性别的男女, 试验结果的阳性阴性, 家族史的有无等等。分两种情形:

(1) 二项分类: 如检查某小学学生大便中的蛔虫, 以每个学生为观察单位, 结果可报告为蛔虫卵阴性或阳性; 如观察某药治疗某病患者的疗效, 以每个患者为观察单位, 结果可归纳为治愈与未愈两类。两类间相互对立, 互不相容。

(2) 多项分类: 如观察某人群的血型分布, 以人为观察单位, 结果可分为 A 型、B 型、AB 型与 O 型, 为互不相容的四个类别。

3. 等级资料(ranked data), 又称半定量资料(semi-quantitative data)或有序分类的分类资料(ordinal categorical data)。为将观察单位按某个指标量的大小分成等级或按某种属性的不同程度分成等级后分组计数, 分类汇总各组观察单位数后而得到的资料。其变量值具有半定量性质, 表现为等级大小或属性程度, 如观察某人群某血清反应, 以人为观察单位, 根据反应强度, 结果可分 -、±、+、++ 四级; 又如观察用某药治疗某病患者的疗效, 以每名患者为观察单位, 结果可分为治愈、显效、好转、无效四级等。

实际上, 资料类型的划分, 是根据研究目的而确定的, 而统计分析方法的选择, 又是与资料类型密切联系的。在资料分析过程中, 根据需要在有关专业理论指导下, 各类资料间可以互相转化, 以满足不同统计分析方法的要求。例如, 以人为观察单位观察某人群脉搏数(次/分钟), 属计量资料; 若根据医学专业理论, 定义脉搏数在 60~100 次/分钟为正常, <60 次/分钟或 >100 次/分钟为异常, 按“正常”与“异常”两种属性分别清点人数, 汇总后可转化为计数资料; 若进一步定义脉搏数 <60 次/分钟为缓脉, >100 次/分钟为速脉, 按“缓脉”, “正常”与“速脉”三个等级分别清点人数, 汇总后可转化为等级资料。以上的例子是先获取计量资料后向计数资料或等级资料的转化, 只要能在专业理论的支持下, 确定不同属性或不同等级的数量界限, 这种转化是不难实现的, 这提示我们在研究设计中, 对于能测量的指标, 尽可能设计为定量指标, 这将为分析中的资料转化带来方便; 另一方面, 对于那些原本为计数或等级的资料, 在资料分析过程中, 为满足某些统计分析方法的要求(如各类回归分析的要求), 有时要在有关理论和实践的指导下设法转化为计量资料, 称为资料或指标的量化。

(易 东)