

医学科研统计方法

杨开济 编著



北京部队后勤部卫生部

014996

95815

医学科研统计方法

杨开济 编著



014996 /R195YKJ

北京部队后勤部卫生部

1983

前 言

早在 1964 年曾为我院开办的主治军医以上技术骨干医学统计方法专题讲座,编写过约三十万字的讲义。以后曾多次举办过不同规格的学习班,由于统计方法不断有新进展,在讲课时,也不断增加一些新的内容。1978 年从新编写了约十万字讲义。1980 年根据我院的军医进修班讲课和参考需要,又在此基础上作了较大的补充修改,共得十一章四十余万字,自 1980 年至 1982 年曾陆续在我院《医学资料选编》院刊上分六次刊登,但由于是在不同年度,由不同印刷厂承担排印的,各章之间风格不尽一致,有的章节错误较多。这次在我区后勤卫生部首长支持下,有机会又作了全面修订,编印成书,供我区各医院和科研单位学习参考之用。这一小书如能对普及统计方法知识,提高医学科研水平,多少有所帮助,则本书写作的目的就算已经达到。

关于统计方法在医学科研中的重要性,已为愈来愈多的人所认识,但由于这些方法是数理统计学根据近代概率论理论,用数学方法推导而来的。这些都属于变量数学,正像恩格斯指出的,不外乎是辩证法在数学中的应用。在实际应用这些方法时,虽然并不太困难,一般有中学数学基础,都可以掌握。但深入理解其辩证逻辑实质,正确而合理的加以应用,则并不太容易。本书有鉴于此,对统计方法在应用过程中的几个主要环节,归纳为数量与质量,变量与常量,偶然与必然,局部与整体,错误与正确,否定与肯定六个辩证关系,并用以指导各种统计方法的实际应用。这一想法是否妥当,殷切希望有关专家批评帮助。

考虑到医学科研,特别是临床科研的实际需要,在方法选择上尽量选用不太复杂而效率较高的方法,并用简单例题说明其具体处理步骤。每章均有练习题,并附答案,便于自行验算。

本书的编写,始终得到我院领导的关怀支持,也得到我院图书资料室,训练队和传染科同志们在各方面的协助,均此致以谢意。

由于编著者水平有限,错误之处一定不多,望读者和专家们批评指正。

编 著 者

1983 年 2 月于北京部队总医院

目 录

第1章 导论	1
1.1 医学科研与统计方法	1
1.2 数据	2
1.3 数据处理	4
1.4 统计推断的基本思想	5
1.5 统计检验的验证逻辑	7
1.6 实验设计与科研	9
第1章练习题 1—4	10
第2章 常用统计特征值	11
2.1 数据整理及频数分布表	11
2.2 平均数及其他集中统计量	12
2.3 标准差及其他离散统计量	15
2.4 小样本中平均数与标准差的计算	17
2.5 大样本中平均数与标准差的计算	19
2.6 标准误	20
2.7 百分率及其标准差	21
第2章练习题 5—14	23
第3章 t 检验——平均数比较	26
3.1 t 检验概说	26
3.2 样本平均数与总体平均数比较	27
3.3 两组样本平均数比较	27
3.4 成对的样本平均数比较	29
3.5 大样本的平均数比较	30
3.6 大样本百分率比较	32
3.7 两组方差不同质的样本平均数比较	33
第3章练习题 15—26	37
第4章 F 检验——方差分析	39
4.1 F 检验概说	39
4.2 单因多组方差分析	40
4.3 方差分析中的各组平均数比较	43
4.4 二因多组方差分析	47
4.5 二因多组中的交互影响	52
4.6 二因多组方差分析中的缺项处理	54
4.7 三因多组方差分析	57
4.8 拉丁方区组设计中的方差分析	62
第4章练习题 27—36	65
第5章 χ^2 检验——计数数据比较	67

5.1 χ^2 检验概说	67
5.2 四格表 (2×2 列联表)	68
5.3. $2 \times k$ 列联表	71
5.4 $1 \times k$ 列联表	74
5.5 $R \times C$ 列联表	77
5.6 部分的 χ^2 检验	80
5.7 几组计数资料的归并	81
5.8 四格表中冒险率的直接计算	84
5.9 关联性检验——属性相关	86
第5章练习题 37—46	88
第6章 置信区间及舍弃检验	90
6.1 平均数的置信区间	90
6.2 观测数据的置信区间	91
6.3 舍弃检验	92
6.4 百分率的置信区间	95
第6章练习题 47—52	98
第7章 理论分布及数据的分布型变换	100
7.1 理论分布	100
7.2 二项分布	100
一、二项式展开	100
二、二项分布的平均数及标准差	102
7.3 二项分布的应用	103
7.4 样本分布是否符合二项分布	105
7.5 波松分布	107
7.6 正态分布	110
7.7 正态分布的配合	112
7.8 样本分布是否符合正态分布	114
7.9 非正态型数据的分布型变换	116
一、二项型数据的变换处理	117
二、百分数据的变换处理	119
三、波松型数据的变换处理	122
四、指数型数据的变换处理	123
第7章练习题 53—60	123
第8章 回归与相关	125
8.1 相关数据	125
8.2 回归方程式	125
8.3 大样本中回归方程式的计算	128
8.4 回归系数的显著性	130
8.5 两回归系数比较	132
8.6 调整值的计算	134
8.7 线性回归的置信区间	136
8.8 相关系数	138

8.9 大样本中相关系数的计算	141
8.10 相关系数显著性	143
8.11 两相关系数的比较	144
8.12 相关系数的置信区间	147
第8章练习题 61—65	147
第9章 非参数统计检验	149
9.1 非参数统计检验概说	149
9.2 符号检验	150
9.3 差数秩检验 (Wilcoxon 法)	154
9.4 秩和检验 (Wilcoxon 法)	158
9.5 H 检验——单因多组秩检验	163
9.6 χ^2 检验——二因多组秩检验	166
9.7 Spearman 秩相关(差方和法)	171
9.8 Kendall 秩相关(顺逆和差法)	175
9.9 符合度检验	179
第9章练习题 66—80	181
第10章 其他统计检验	184
10.1 Ridit 分析	184
10.2 判别分析	189
10.3 协方差分析	193
10.4 计数数据的序贯分析	198
10.5 测量数据的序贯分析	202
第10章练习题 81—94	205
第11章 实验设计概说	209
11.1 实验设计概说	209
11.2 实验设计模型	210
一、完全随机设计(单因多组设计)	211
二、有重复例的单因多组设计	213
三、随机区组设计(二因多组设计)	214
四、多因多组设计	215
五、拉丁方区组设计	218
六、希腊拉丁方区组设计	219
11.3 控制系统误差	220
11.4 随机化方法	221
11.5 抽样设计	225
一、单纯随机抽样	226
二、系统抽样	226
三、分层抽样	226
四、整群抽样	227
11.6 样本含量	229
一、有关样本含量的一般原则	229
二、样本含量估计	230
第11章练习题 95—100	233

附录一 统计用表	234
附表1 正态分布的上侧概率	234
2 t 分布临界值表	235
3 F 分布临界值表(用于方差同质性检验)	236
4(1) F 分布临界值表(用于方差分析) $\alpha = 0.05$	238
4(2) F 分布临界值表(用于方差分析) $\alpha = 0.01$	240
5 χ^2 分布临界值表	242
6 Grubbs 舍选法的 T_n 表	244
7(1) 百分率的 95% 置信上下限	245
7(2) 百分率的 99% 置信上下限	247
8 正态曲线之纵高 (z) 与面积 ($\alpha/2$)	249
9 正态符合度检验 (Kolmogorov-Smirnov 法)	249
10 反正弦角度变换表 ($\theta_i = \sin^{-1} \sqrt{p_i}$)	250
11 校正方根变换表	250
12 相关系数的各临界值	251
13 相关系数的 z 分布变换表	251
14 符号检验用表	252
15 差数秩检验 (Wilcoxon 法) 用表	253
16 秩和检验 (Wilcoxon 法) 用表	254
17 H 检验 (Kruskal-Wallis 法) 用表	258
18 χ^2 检验 (Friedman 法) 用表	259
19 同位秩 $t_2 - t$ 值换算表	259
20 Spearman 秩相关用表	260
21 Kendall 秩相关用表	260
22 随机数表	261
23(1) 两组平均数比较时的每组样本含量估计(双侧假设)	262
23(2) 两组平均数比较时的每组样本含量估计(单侧假设)	262
24(1) 两组百分率比较时样本含量估计(双侧假设)	263
24(2) 两组百分率比较时样本含量估计(单侧假设)	263
附录二 统计符号表	264
附录三 汉英术语对照	267
附录四 练习题答案	271
索引	281

第1章 导 论

内容提要: 本章主要从数量与质量,变量与常量,偶然与必然,局部与整体,错误与正确,否定与肯定等六个辩证关系,说明概率统计学方法在医学科研中的应用原理,并对医学科研中的数据处理和实验设计的基本内容,作了介绍。

1.1 医学科研与统计方法

1. 统计方法是什么? 扼要地说,统计方法是进行各类科学研究所共同必需的科学方法。这一方法,目前已广泛应用于各门科学和生产实验如农业科学,物理学、生物学、工程技术等。此外,在社会科学,企业管理,以至在日常生活中,都分别得到了有效利用。

2. 科研为什么要用统计方法? 这是适应科研和生产需要决定的。从科学发展来看,除数学外,一般都有以下三个特点:(1)由定性研究向定量研究和计量科学迅速发展,(2)由经验、半经验科学向实验科学发展。(3)自觉地应用逻辑推理,特别是归纳推理的方法,(如求同法,求异法、混合法、剩余法,因果法等)获取新的知识。这就需要有相应的数学工具,因而促进了推断统计学或概率统计学(数理统计学)的产生和发展。并提供了一系列应用统计方法。廿世纪以来,推断统计学的发展,出现了样本推断,特别是小样本推断理论,以及实验设计分支学科的创立,判别函数理论,以及近几十年非参数统计方法的出现,近年正交设计方法在国内工农业生产实验中的广泛应用,都为科研和生产作出了很大的贡献。当然,迄今的统计方法还不是十分完备的,有不少科研项目还缺乏有效的统计方法,这就给统计学家和各专业工作者提出了方法学研究的新任务。

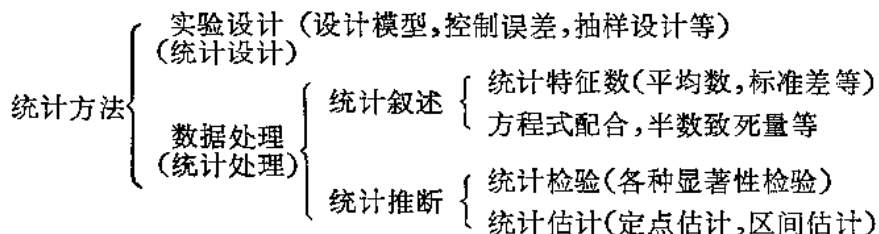
3. 医学科研为什么也要应用统计方法? 这也是根据医学发展需要决定的。医学科研包括有临床医学,基础医学,和预防医学三方面,三方面的情况有所不同,基础医学和生物科学是姊妹学科,情况类似,统计方法的应用,为时较早并已取得了不少实践经验。预防医学科研过去一向采用卫生统计学,即属于叙述统计学的一些方法,近年也逐步应用了推断统计学的科研方法。临床医学一向被认为属于经验科学,研究方法长年停留在经验积累,医案报告上,但随着其它学科的迅速发展,临床医学也逐步向“实验医学”,并进一步将向“计量医学”发展。为加速提高临床医学水平,应加强基础医学研究,创建更多的实验动物模型,但这并不能代替临床科研,临床医学本身的研究,仍然迫切要求进一步加强。尽管目前统计方法用于临床研究的经验还不够丰富,方法还不够完备,需要研究创新。但应用已有的统计原理和方法,对提高临床观察和总结的科学性,特别对帮助整理和提高中医中药的临床经验,还是大有可为的。

近年,优选法也有人试图用于医学科研,考虑到人和生物的变异性都较大,在应用这一方法时也应结合统计方法进行(统计优选法)。

4. 在医学科研中要用那些统计方法? 统计方法所包括的范围很广,根据医学科研特



点,应用的主要有实验设计和数据处理两大部分。在数据处理中,包括叙述统计学和推断统计学两方面的一些方法,都要应用,但主要是后者,特别是其中的各种小样本统计检验方法。此外,有时也要应用区间估计、相关与回归,实验方程式,半数致死量等方法。概括如下:



5. 怎样学习和正确应用统计方法? 由于统计方法来源于概率统计学,概率统计学属于高等数学,公式和定理推导需要一定的数学基础,有些人因此也对统计方法望而生畏。其实,学习应用方法,并不困难,如果我们对概率统计学中的高等数学推导部分暂时不作深究,那么,只要有中等数学水平,经过几次实际应用练习,应用方法是可以学会的。但能掌握它的精神实质,正确应用,却不一定很容易。这是因为概率统计学是一门变量数学,“本质上不外是辩证法在数学方面的应用,”(恩格斯)。在学习统计方法时,对其数学推导部分,可以暂时不问,但对于在数学推导中蕴藏着的辩证法,必须有所了解。不然的话,在使用中就可能会产生误解,以至误用。我们体会需要弄清以下六个关系,即数量与质量,变量与常量,偶然与必然,局部与整体,错误与正确,否定与肯定。以下将结合统计方法中的一些基本概念,加以讨论。我们相信,弄清这些关系,一定会有助于这些统计方法的正确理解,更有效地发挥它在科研和生产中的工具作用。

1.2 数 据

1. 什么是数据? 观测数据简称数据,数据中蕴藏着一定的统计信息,是进行科研和实验的重要物质基础。数据是某一事物数量的数量表现。“任何质量都表现为一定的数量,没有数量也就没有质量。”(毛泽东)。在应用统计方法时,一定要弄清数量与质量的对立统一和互相转化关系。例如科研或实验,在这个意义来说,也就是将质量转化为一定数量,即数据的一个过程。在实验设计时就要选定最能反映事物本质的数量指标,并用合理方法得出适当的数据。在数据处理时,要用适当的统计方法进行处理,这又是由数量再转化为质量的过程。统计方法也是处理数量与质量互相转化的科学方法。

2. 数据有那些种类? 根据事物特性和科研需要,须用各种不同的数据表达,但作为基础的数据有以下三种:

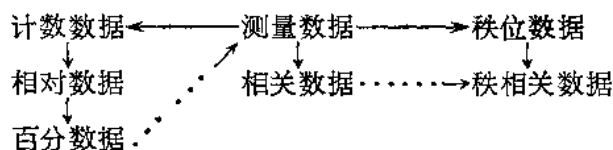
(1) **计数数据:** 指按一定的事物属性标准取得的事物个数或次数,如患者例数,治愈例数,动物头数,实验重复次数等。

(2) **测量数据:** 指按一定的度量衡计量的数据,如身长(厘米),体重(公斤),生化定量(毫克%)等等。

(3) **秩位数据(分级数据):** 指按一定的数量大小或属性优劣标准排列名次或顺位的数据。如一组测量数据由小到大,或由大到小的名次排列等。

由上列基础数据又可以派生出一些其它数据,如相关数据(如身长与体重并列的数据),相对数据(如比率数据),百分数据(如各种百分率,治愈率、阳性率等)。

各种数据在一定条件下,可以互相变换,如下:



3. 数据有什么数学特性? 数据作为一种数量,有以下两个数学特性:

1) 变异性: 数据是一种变量,而不是常量,这是一个重要的统计学基本概念。但必须从**变量与常量**的辩证关系上,正确地加以理解。每一观测数据从孤立的个体看,是固定的数量,并用常量数学进行运算。但从数据与实际事物本质相联系的群体来看,又不是固定不变的常量,而是一种可变的变量,举例如下:

(1) 调查某地的一组若干例 20 岁青年男子身长或体重,可以发现这一组测量数据,不是等齐一致,而是参差不齐的。

(2) 同一实验在若干次重复观测中所得数据,通常是有差异的。

(3) 同一疗法在不同患者身上的疗效,不会相同。同一疗法在不同单位,甚至在同一单位不同时间的治愈率,通常也不会是完全一致的。

(4) 调查若干两子女家庭,可以发现子女的男女分配和先后顺序,也都是不一致的。

由上可以窥知各种观测数据虽然变化形式不同,但都是可变的,这些在数或量上的变异,统称**变差**。数量观测出现变差,是一种**偶然现象**,而这种现象却是宇宙间客观存在着的一种普遍现象,称为**随机现象**。也可以理解为运动着的物质在数量上的一种客观反映。有人把观测当中出现的变差,一律看成是观测上的差错或不精确,是一种误解。观测变差中,固然有些属于工作粗糙,有些则属于仪器或方法有偏倚所致,但更多的属于正常应有的现象。这在学习和应用统计方法中,是首先应予弄清的一个基本概念。

2) 规律性: 客观存在着的变差,它的变动不是杂乱无章,而是有确定的规律可以遵循的。一组观测数据,如果这些数据不是凭主观意愿,挑拣取舍的,而是按一种叫作**随机**的原则和方法,尽量按客观本来面貌取得的,经过整理,我们可以发现其中确有某种规律性,观测例数愈多,这种规律性就愈趋明显。例如,一组 100 例某地 20 岁男子身长测量数据,有高有低,参差不齐,但如果将身长的高度分为若干组级,记录每一高度组级的例数(在统计学上称为**频数**),这些可以构成一种**频数分布**,各频数除以总频数,则构成相对频数分布,或**频率分布**。如果以身长高低为横轴(X 轴)、频数多少为纵轴(Y 轴),则可绘制成频数分布或频率分布的直方图,联结各直方顶点,可构成一种折线图。总频数愈多,分组级愈细,折线愈趋平滑,总频数为无穷大时,频率分布就形成一条光滑曲线,形成一种特定的**理论分布**。这种理论分布从直观看,为中间高两边低的一条山形曲线,这条曲线有很多特点,例如平均值位于曲线顶点(即频数最多或频率最大的一点)的 X 轴上,标准差与曲线内所包含的面积成函数关系等,这种理论分布称为**正态分布**,曲线称为**正态曲线**。正态分布是最多见的一种“连续型理论分布”,适用于测量数据,是统计学最主要的一种理论基础。十七世纪数学家最初发现这一所谓“**大数定律**”时,曾惊叹为神的意志。理论分布的发现,说明在变量中有不变的规律。变差是一种偶然现象,但是在偶然性中又有确定的

必然性,或者说必然的规律透过偶然的变差中表现出来。这也是一个重要的统计学概念,我们在应用统计方法时,必须弄清这一偶然与必然的辩证关系,才能得到正确的应用。例如我们要求在实验设计,即在抽取样本、获得数据时,之所以必须遵循随机的原则和随机化的方法,主要也是为了适应偶然现象客观存在的实际情况,不然就不符合实际,科研结果也就失去了客观性。

理论分布除上述的正态分布属于连续型分布以外,还有各种离散型理论分布,其中最常见有二项分布,如男女分布,成功与失败,阳性与阴性,有效与无效等等计数数据,如经整理,排成频数分布,也可以看出接近理论的二项分布规律,观察例数愈多,就愈逼近。波松分布是例数很大,而某一现象出现例次数很小的特殊的二项分布,适用于工业生产的废品数,各类工作事故差错次数,以及放射线 α, β 粒子,血细胞中嗜碱细胞,或嗜酸细胞计数等。包括以上正态分布,二项分布,波松分布在内的各种理论分布,都是统计学,以及统计方法的重要理论基础。

1.3 数据处理

1. 怎样进行数据处理? 数据既然是反映事物质量的一种数量表现,是一种统计信息,如何把这一信息正确地、充分地表达出来,最合理的方法是进行统计学处理。从统计学的发展看,有以下两种方法:

(1) 统计叙述:这是原来叙述统计学(如卫生统计学)主要采用的处理方法。通过大数例观察,用各种统计图、或统计表、或统计特征数等,直观地反映数据分布的某些规律性。近代科研和实验的用例都较少,只用这一方法已不能满足需要,但仍可作为初步的数据整理,或辅助手段加以利用,例如平均数、标准差和标准误以及相关系数等的计算,是进行下一步统计推断(如 t 检验, F 检验等)的必要步骤等。

(2) 统计推断:这是推断统计学(概率统计学)主要采用的方法。统计推断又有统计检验和统计估计两种,两者方法不尽相同,但又有密切联系。统计估计又有定点估计和区间估计两种、在医学科研中以区间估计的用途较多,如置信区间,或舍弃极限的划定等。统计检验有多种方法,如 χ^2 检验(适用于计数数据的处理,或百分率的比较等), t 检验(适用于两组平均数比较等), F 检验(适用于多组,或更复杂的多因多组比较或分析等),以及各种非参数统计检验(如各种秩位检验,适用于对秩位数据或测量数据转变为秩位数据的分析,也适用于部分的计数数据的分析,方法一般都简便易行)。各种统计检验都有各自的性能和适应症,要掌握这些特点,正确地加以利用。

2. 为什么要进行统计推断? 简单说,这是因为数据的变差的来源,不是一种,而至少是有以下内外两个方面的来源决定的。这两种来源的变差,或称误差的是:

(1) **条件误差:**这是一种外源性的变差,是因外部条件不同所引起的变差,以两组疗法比较为例,因治疗条件不同,在治愈率和平均治愈天数上出现的差别,属于条件误差,外部条件可以是单一的,但往往是多种因子的综合。

(2) **随机误差:**又称偶然误差,是一种内源性的变差,属于事物本身物质运动反映出的变差,但也包括某些难以避免的观测误差。如两组疗法比较中,同一治疗组的内部各例,在治愈与否,或在治愈天数上表现出来的变差。

在科研或实验中,这两种变差纠缠在一起,往往难以用简单的直观方法,立即加以区分。例如,在两组疗效比较中,两组的每个病例的治愈与否,或愈疗天数长短,既有由于治疗条件带来的条件误差,又有不是由于治疗条件,而是患者本身来源的随机误差,简单地把两者都看成是单一的条件误差,显然是不正确的。正确区分这两种误差的办法是合理应用各种统计推断的方法,有些人在疗效评价时,只简单比较两组治愈率的高低,或只比较两组平均治愈天数的长短,就下结论。这是一种“直观判断”的方法,这样判断的结果,除了在观察例数很多,或随机误差很小(如对无生物物的理化学实验)的特殊情况下,也可能与统计推断的结论一致。但在医学科研,特别是临床医学科研,由于一般用例较少,而生物学变差又都较大,两者结论往往不一致,而直观判断的结论,是不可信的。这也是我们一再强调对数据要作统计学处理,特别是要作统计推断的理由所在。

1.4 统计推断的基本思想

1. 统计推断的基本思想是什么? 统计推断的基本思想是**样本推断**。前面已经讨论过,近代科研的一个共同特点是典型调查,或模型实验,通过有充分代表性的典型或模型的建立,并对这一典型或模型的分析 and 推理,对某一事物,或某一问题的一般规律,作出合理的判断。这种推理,是由已知的一个局部对未知整体的一种推理判断或估计,在逻辑上属于一种不完全归纳推理,或称局部推理。这里存在着一种**局部与整体**的关系问题,也必须用辩证的观点处理好这一关系。不完全归纳推理,或局部推理的结论,在逻辑上必定都是相对的,不是绝对的,不可避免地存在或多或少的“或然性”(不肯定性),也就是说,其中可能包含着或多或少程度不同的推断错误成分。

推断统计学把上述的局部与整体,以及局部推理过程,通过数学推导,构成一整套样本与总体的概念,以及样本推断的理论和方法。

推断统计学把科研中的一个典型或模型,称为一个随机样本,简称**样本**。而把科研中要解决的问题,称为**总体**。近年,为了更适合科研的实际情况,又将总体再分为**目的总体**和**抽样总体**两种。所谓目的总体是指研究或想了解某一事物的全部,这种总体可以是理论性的,也可以是实在的。所谓抽样总体则是我们可以从中抽取样本的一种总体。例如,我

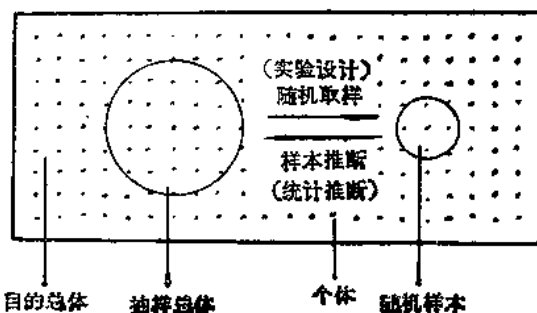


图 1.4-1 样本与总体的关系示意图

们想调查某一地区的乙型肝炎抗原无症状携带者的阳性率,这是一种目的总体,而我们例如计划在这一地区十个不同单位,分别按一定比例各随机抽取若干人进行调查。这十个单位的全部人员,则是一个抽样总体。而从十个单位抽取的总人数,在调查研究中则看成是一个样本。其间的关系如上列的示意图:

图中总体和样本内的每一个小圈点,代表一个统计个体,在样本中则表现为每一个观测数据。

2. 什么是小样本和大样本? 样本的大小是按一个样本中所包含的个体数(即频数,或

例数)的多少区分的(样本含量)。一般将个体数在 30 例以下的,称为**小样本**; 100 例以上者称为**大样本**; 30—100 例之间的样本,一般也按小样本处理。区分大小样本,主要是因为所选用的统计方法,有所不同。在医学科研中,特别强调搞好实验设计,严密控制系统误差,抽取少而精的小样本,并认为这样更能获得满意的结果。这种小样本称为**精确样本**。对小样本的处理必须选用小样本方法,而不宜采用大样本方法。但对大样本,则小样本方法同样可以适用,只是会增加一些计算手续而已。

3. **样本推断可能发生那些推断错误?** 如上所述,样本推断是一种局部推理,可能出现程度不同的推断错误。推断统计学将可能出现的推断错误,按其性质划分为以下两类:

(1) **第一类错误**: 从科研的予期结果来说,又是一种**假阳性错误**(或称误判),例如我们评价 a 、 b 两组疗法的疗效有优劣差别($a \neq b$)时,实际上却是没有差别($a = b$)的。发生第一类错误的可能性大小,用概率表达,称为**第一类错误概率**(α)。

(2) **第二类错误**: 是否定科研的予期结果时可能犯的推断错误。所以又是一种**假阴性错误**(或称漏判),例如我们评价两组疗法的疗效无优劣差别($a = b$)时,实际上却是有差别($a \neq b$)的。发生第二类错误的可能性大小,用概率表达,称为**第二类错误概率**(β)。

设阳性结论为 $a \neq b$, 阴性结论为 $a = b$; 以上关系,概括如下表:

表 1.4-1 样本推断与推断错误

样 本 推 断	总 体	判 断	概 平
阳性结论 ($a \neq b$)	阳 性 ($a \neq b$) 阴 性 ($a = b$)	正确 第一类错误	$1 - \alpha$ α
阴性结论 ($a = b$)	阳 性 ($a \neq b$) 阴 性 ($a = b$)	第二类错误 正确	β $1 - \beta$

在样本推断中,发生第二类错误的原因,往往是因为观察例数不足,或因所用的统计检验方法不够敏感所致。克服的办法是增加例数,继续观察,或改用更敏感的统计检验方法等。

产生第一类错误的原因,往往是没有对数据作必要统计检验,仅仅凭平均数和百分率的表面差别,就冒然下结论所带来的。偶尔也因统计检验方法选用不当引起的,例如,对小样本(一般指 30 例以下的样本)误用了大样本统计检验方法等。克服的办法是进行适当的统计检验,通过这些统计检验,都可以具体算出可能犯第一类错误的概率(α)。“**概率**”(P)为表达某一事件出现可能性大小的一种一般性数量尺度,并约定为 0—1 之间的正小数值 [$1 \geq P \geq 0$], 也可以用百分率(0—100%)表达,称为“**百分概率**”。可能犯第一类错误的概率,我们采用了**冒险率**这一专用名词加以表达,以区别于一般泛泛使用的概率,用 P 或 α 表示。

克服第一类错误最有效的办法,是根据冒险率的大小来下结论,上文已讨论到,样本推断是局部推理,结论都是相对的,或多或少都可能出现推断错误,也就是说,不同程度地会有一定的冒险率。但当冒险率小到一定程度时,在实际应用中,又可以允许有条件地下阳性结论,并承认这样的推理判断是正确的。在这一推断过程中,又必须正确理解**错误与正确的辩证关系**,才不致产生误解。我们在下科研结论时,承认存在有推断错误的可能性,并列举其冒险率,正是为了保证结论的正确性和可信性。否则,反而是不正确、不可信的。

4. 什么叫显著性? 统计检验也称显著性检验, 所谓显著性, 是推断统计学的一个重要概念, 是指推断结论是否可信而言的。可信程度的大小称为显著性水平。显著性水平取决于冒险率的大小。冒险率愈小, 显著性水平愈高。现在一般公认的显著性水平的分级标准如下:

表 1.4-2 冒险率与显著水平

冒 险 率	显 著 性 水 平	统 计 学 意 义	符 号
$P > 0.05$ (或 5%)	无显著性	无意义	— (或 NS)
$P \leq 0.05$ (或 5%)	有显著性	有意义	*
$P \leq 0.01$ (或 1%)	有非常显著性	非常有意义	**
[#] $P \leq 0.001$ (或 0.1%)	有高度显著性	高度有意义	***

[#] 这一显著性水平一般不常用。

[注] 显著性水平原用“显著”, “不显著”, 或“非常显著”等表达术语, 但看来容易产生误解, 故统一修改如上表。例如两组疗法的治愈率有显著性差别, 并无两组治愈率差距大小的涵义。另外, 无显著性差别, 也不等于是两组治愈率无差别, 而是其中包括两组无差别, 和虽有差别但因例数少等原因未能将差别显示出来的两种可能性。

上述的显著性水平分级, 只是一般约定的标准, 也允许适当灵活掌握, 但在结论后面要附记冒险率。

5. 如何确定冒险率? 在进行统计检验时, 由求得的统计量(如 χ^2 、 t 、 F 等)和自由度, 即可代入各自的概率密度函数求得冒险率, 但这一计算过于繁重, 在实用上可查有关量表, 确定冒险率所在范围, 即可达到确定显著性水平, 作出统计结论的目的。

6. 什么是统计估计? 如前所述, 除统计检验外, 统计估计是另一种统计推断方法。从一组调查数据用统计方法, 可得出平均数, 标准差, 百分率, 相关系数等等统计特征值。这些特征值都是由一个样本得出的, 是一种样本特征值, 如样本平均数, 样本标准差等等, 又是用统计方法得出的, 统称为统计量。根据样本和总体理论, 可以认为这些样本特征值都是从相应的总体特征值抽出的无数个样本特征值中的一个, 我们进行科研所要探索的目的是总体特征值, 而我们所能得到的却是样本特征值, 由样本特征值估计总体特征值, 同样也是一种局部推理, 推断中也出现一定的冒险率。这些总体特征值, 一般称为总体量, 又称参数, 所以统计估计又称参数估计, 详见后文第6章。

1.5 统计检验的验证逻辑

1. 统计检验是怎样验证结论的? 统计检验验证结论的方法, 实际上包括以下两步的一个逻辑证明过程。第一步, 先预设一个完全否定预期结果的统计假设, 称为无效假设。同时, 另设一个与此互相对立, 互相排斥, 也就是完全肯定预期结果的统计假设, 称为对立假设。举例说明如下:

(1) 无效假设 $[H_0]$: 以两组疗法(a 、 b)比较为例, 预期结果为两组疗效有优劣差别 ($a \neq b$), 这时的无效假设为两组疗效无优劣差别 ($H_0: a = b$)。

(2) 对立假设 $[H_A]$: 为两组疗效有优劣差别, ($H_A: a \neq b$)。

第二步, 采用在数学证明中常用的一种“反证法”, 进行逻辑证明。即用否定无效假设的方法, 证明对立假设的成立。但当无效假设不能否定时, 就承认无效假设, 否定对立假设。统计检验验证的过程, 也是一个对一对统计假设决定取舍的过程, 所以统计检验也称

为**假设检验**。这一验证过程,实质上是一种**否定与肯定**的相互关系问题,也必须用辩证的观点,正确地加以理解和应用。

两类不同的推断错误,在假设检验中的关系如下表(以两组疗法比较为例):

表 1.5 1 假设检验与推断错误

假 设 检 验	总 体	判 断	概 率
否 定 $H_0: a=b$ (承 认 $H_A: a \neq b$)	$a \neq b$ $a=b$	正 确 第一类错误	$1-\alpha$ α
承 认 $H_0: a=b$ (否 定 $H_A: a \neq b$)	$a \neq b$ $a=b$	第 二 类 错 误 正 确	β $1-\beta$

由上表可以看出,统计检验所得的冒险率,也就是否定无效假设时的错误概率。如上例两组疗法的疗效差别“有显著性”,就是意味着可以以 5% 或 5% 以下的冒险率,否定无效假设,承认对立假设。“有非常显著性”意味着可以以 1% 或 1% 以下的冒险率,否定无效假设。“无显著性”意味着否定无效假设的冒险率大于 5%,因而不宜否定无效假设。

2. 什么叫单侧假设和双侧假设? 以往的统计假设一般都采用双侧假设,但随着科研和生产的需要,近年采用单侧假设的,也逐渐增多。两种统计假设的区别如下:

(1) 双侧假设: 以 a, b 两组疗效比较为例:

无效假设 $H_0: a = b$,

对立假设 $H_A: a \neq b$ (有 $a > b$, 和 $a < b$ 两种可能)

(2) 单侧假设: 可分以下两种情况两种可能:

1° 当预知 a 疗法至少不会劣于 b 疗法, 即 $a \leq b$ 时, 这时只有 $a=b$ 和 $a > b$ 两种可能

无效假设 $H_0: a = b$ (但 $a \leq b$)

对立假设 $H_A: a \neq b$ (即 $a > b$)

2° 当预知 b 疗法至少不会劣于 a 疗法, 即 $a \geq b$ 时, 这时只有 $a=b$, 和 $a < b$ 两种可能:

无效假设 $H_0: a=b$ (但 $a \geq b$)

对立假设 $H_A: a \neq b$ (即 $a < b$)

两种统计假设在确定显著水平时, 所要求的冒险率有所不同, 以 t 检验为例, 双侧假设的冒险率为正负两部分之和, 而单侧假设则仅取正或负的一部分。常用的 t 检验专用表是按双侧假设要求编制的, 采用单侧假设时, 取冒险率的一半 ($\alpha/2$) 确定显著性水平。两种统计假设的显著性水平与冒险率的关系, 对照如下表和下图:

表 1.5-2 两种统计假设检验的临界值比较

显 著 水 平	双 侧 假 设	单 侧 假 设
无显著性 ($P > 0.05$)	$t < t_{0.05}$	$t < t_{0.10}$
有显著性 ($P \leq 0.05$)	$t \geq t_{0.05}$	$t \geq t_{0.10}$
有非常显著性 ($P \leq 0.01$)	$t \geq t_{0.01}$	$t \geq t_{0.02}$

在科研和实验中, 如果没有确实把握预知只有一侧可能时(例如在两组比较中, 没有确实把握预知 $a > b$ 或 $a < b$ 时), 则宁可采用双侧假设, 不用单侧假设。

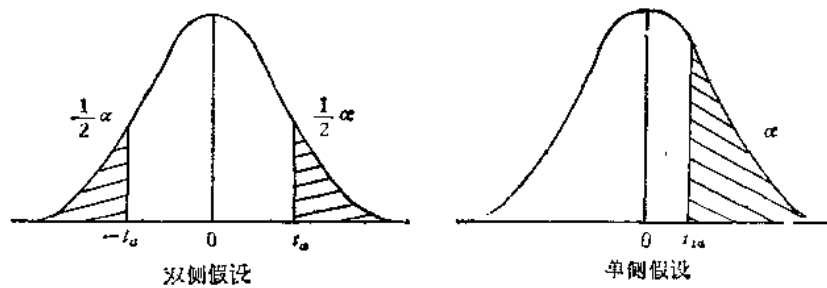


图 1.5-1 两种统计假设检验冒险率示意图

1.6 实验设计与科研

1. 实验设计在科研中有何重要意义? 对科研所得数据进行合理的统计学处理,固然十分必要,但科研目的是否能够达到,结论是否正确,并不完全取决于数据处理,更重要是数据的来源是否合理可靠。从这个意义来说,实验设计在整个科研中,是一个关键,往往会决定科研的成败。

2. 实验设计包括那些主要内容? 实验设计是推断统计学的一个分支,主要解决由随机总体抽取随机样本的一套科学方法,并且还注意到使以后的数据处理能够更顺利地进行,获得必要的结果,有利于提高科研的精度和效率。它的主要内容有以下三方面:

(1) 提供各种**设计模型**,供各种科研和实验不同的目的和要求,适当选用。例如,多因多组的设计模型可供各种复杂的实验设计之用、拉丁方设计,正交设计,混杂设计等都是些效率高的设计模型,可以用较少的例次数,获得较大的成果。

(2) 控制各种系统误差,避免可能出现的**偏倚**,以保证样本的随机性,从而保证科研结论的正确性。其中,主要对实验中的各个环节采用各种行之有效的**随机化方法**,例如,利用**随机数表**等工具进行随机抽样,如随机配置,随机排列,以及对照实验,盲法的应用等。

(3) 估计适当的实验例次数,实验的例次数过少,容易出现“抽样误差”,过多又容易出现“不均匀误差”,也会增加人力,物力时间的浪费。例次数的估计是一个复杂的问题,实验设计在这一方面也提供了一些可供选用的估计方法。

3. 统计方法与专业的关系如何? 统计方法包括实验设计和数据处理两大部分,对科研或实验都是必要的,也是十分有效的。但在科研和实验的整个过程中,统计方法都必须密切结合专业进行,特别是在根据科研的目的提出**实验设计**的阶段,也就是在统计学上由目的总体选定抽样总体的阶段,以及在数据处理后,如何根据统计结论回答科研的目的,也就是由抽样总体判断目的总体的阶段,都不能靠统计方法,而只能依靠丰富的专业知识,进行选定和判断。科研或实验与统计方法的关系是本体与方法的关系,两者必须密切结合,但又必须分清主次,科研工作中有时需要专业工作者和统计学家合作,但也必须以专业工作为主,不能主客颠倒。由此看来,各个专业工作者了解一定的统计学知识,掌握常用的统计方法,对促进专业科研工作的进展是大有好处的。科研实验过程与统计方法应用的关系如下:

科研目的→实验设计→实验数据→数据处理→科研结论

随机抽样

样本推断

目的总体→抽样总体→随机样本→随样总体→目的总体

表 1-1 各种统计检验方法应用

数据种类	变 量	因 子	分 组	方 法
测量数据	单 变 量	单 因	单 组 配对比较 两组比较 多组比较	定点估计, 区间估计 t 检验 F 检验, 符合度检验 方差分析
		多 因	多组比较	方差分析
	二 变 量	单 因	多 组	相关, 回归
		多 因	多 组	协方差分析
	多 变 量	单 因	多 组	复相关, 复回归
		多 因	多 组	复协方差分析
秩位数据 (测量数据)	单 变 量	单 因	配对比较 两组比较 多组比较	差数秩检验 秩和检验 H 检验
		二 因	多组比较	χ^2 检验
	二 变 量	单 因	多 组	秩相关
计数数据	单 变 量	单 因	1×2 列联表 1×k 列联表	χ^2 检验, 符号检验 χ^2 检验,
		二 因	2×2 2×k R×C	χ^2 检验, Ridit 分析 χ^2 检验, Ridit 分析 χ^2 检验, Ridit 分析

第1章 练 习 题

题 1 下列事项应该用那一种数据表达?

(1)治愈天数 (2)治愈人数 (3)五例中三例治愈 (4)儿童胸围坐高测定 (5)十大名酒评比

题 2 下列为 10 例急性肝炎的血清总胆红素 (mg/dl) 测量数据,问如何转变为计数数据?

3.2 1.6 0.8 4.5 0.7

0.7 0.8 1.5 0.9 2.3

题 3 下列论断是否正确?为什么?

(1)某药治疗某病三例全部治愈,因而肯定某药的治愈率为 100%。

(2)旧法治疗某病的治愈率为 60%,新法的治愈率为 80%,因而肯定新法疗效优于旧法。

题 4 下列统计检验的显著性水平如何确定?

(1)a, b 两疗法治愈天数,经 t 检验,查表,冒险率(P)在 0.05—0.10 之间 ($P \approx 0.08$)。

(2)已知 a 疗法至少不会比 b 疗法差,经 t 检验,同样查表,冒险率同上。