

医用数理統計方法

郭祖超等編

人民卫生出版社

医用数理統計方法

郭 祖 超 等 編

許 世 瑾 審 校
李 光 蔭

人民卫生出版社

一九六三年·北京

内 容 提 要

本书通过实际資料介绍了数理統計方法在医学科学研究中的具体应用, 适当闡明了有关原理, 并系統介绍了实验設計。

主要讀者对象是医学科学研究人員、高等医药院校教师、高級医务人員等。

医用数理統計方法

开本: $850 \times 1168/32$ 印张: $16\frac{1}{4}$ 插頁: 1 字数: 448千字

郭 祖 超 等 編

人 民 卫 生 出 版 社 出 版

(北京書刊出版业营业許可証出字第〇四六号)

· 北京崇文区线子胡同三十六号 ·

北 京 市 印 刷 一 厂 印 刷

新华书店北京发行所发行 · 各地新华书店經售

統一書号: 14048·2729 1963年6月第1版—第1次印刷

定 价: 2.20 元 印 数: 1—2,800

目 次

第一章 資料之搜集与整理	1
1.1 科学研究与数理統計方法	1
1.2 实验設計中之一般注意事項	2
1.3 分組法	6
1.4 統計表	15
1.5 統計图	24
小 結	35
練 习 題	36
参考文献	38
第二章 計数資料	40
2.1 运用相对数时的注意点	40
2.2 由样本相对数估計全体相对数	42
2.3 χ^2 測驗之意义	47
2.4 四格表	56
2.5 $2 \times K$ 表	60
2.6 $R \times C$ 表中独立性(或联系性)之測驗	64
2.7 連續性之校正	65
2.8 $2 \times 2 \times 2$ 表	71
2.9 四格表中机率之直接計算法	74
小 結	78
練 习 題	80
参考文献	82
第三章 平均数与变异指标	84
3.1 平均数与均数的概念	84
3.2 馬克思、列宁、斯大林关于均数的理論	84
3.3 小样本中均数之計算法	86
3.4 大样本中均数之求法	88

3.5	中位数	90
3.6	众数	92
3.7	变异指标的意义	93
3.8	全距(或称变异幅度)	93
3.9	四分位数间距	93
3.10	均差(或称平均直线差)	95
3.11	标准差(离差均方根)的意义	96
3.12	自由度	97
3.13	小样本中标准差之计算法	100
3.14	大样本中标准差之计算法	104
3.15	变异系数	106
	小 结	107
	练习題	109
	参考文献	112
第四章	二项分配与普哇松分配	114
4.1	二项分配	114
4.2	二项分配的应用	119
4.3	样本分配与二项分配之比較	124
4.4	二项分配一致性之測驗	125
4.5	最可能数(液体中細菌密度之估計)	127
4.6	普哇松(Poisson)分配	132
4.7	样本分配与普哇松分配之比較	136
	小 结	138
	练习題	139
	参考文献	140
第五章	常态分配与半数致死量	141
5.1	常态分配	141
5.2	常态曲线之面积	144
5.3	样本均数的分配	152
5.4	应用常态分配进行显著性測驗的例子	155
5.5	常态曲线之配合法	159
5.6	对称性的測驗	163
5.7	常态性的測驗	165
5.8	常态分配中全距与标准差的关系	167

5.9	半数致死量(LD_{50})与机率单位(Probit)的意义	169
5.10	由目测法求半数致死量	179
5.11	由累計法求半数致死量	181
	小 結	183
	練 习 題	185
	参考文献	187
第六章 均数之显著性		189
6.1	样本均数、变异数及标准差的分配	189
6.2	t 分配	189
6.3	均数显著性的測驗	194
6.4	两均数相差的显著性	197
6.5	个别比較与团体的比較	199
6.6	变异数一致性的測驗	200
6.7	全体标准差不相等时两均数相差的显著性	202
6.8	样本的大小	205
	小 結	207
	練 习 題	208
	参考文献	210
第七章 直綫迴归		211
7.1	迴归方程式的計算法	211
7.2	迴归系数的显著性	215
7.3	营养实验中“每克飼料所增体重”計算法的商榷	218
7.4	修正值	220
7.5	两种全体中的抽样	221
7.6	大样本中迴归方程式的計算法	226
7.7	直綫迴归的可信限	232
7.8	倚变量平方和分析	237
7.9	由加权直綫迴归法求半数致死量	239
	小 結	244
	練 习 題	246
	参考文献	249
第八章 相关		250
8.1	相关系数	250
8.2	小样本中相关系数的計算法	253

8.3	常态双变量全体的特性	254
8.4	相关系数的抽样变异及显著性测验	258
8.5	两相关系数相差之显著性与相关系数之合并及可信限	261
8.6	相关与迴归	266
8.7	相关与离均差之平方和	268
	小 結	269
	練習題	270
	参考文献	272
第九章	变异数分析(一)	274
9.1	变异数分析的意义及计算方法	274
9.2	随机单位組中的变异数分析	284
9.3	各組均数間之相互比較	286
9.4	拉丁方中之变异数分析	289
9.5	随机单位組与拉丁方中缺項之估計	291
9.6	参数之估計	297
9.7	平衡不完全单位組中之变异数分析	301
9.8	变异数一致性的測驗	308
9.9	变异数为間杂性时的近似 F 測驗法	311
9.10	变数代換后之变异数分析	313
	小 結	320
	練習題	321
	参考文献	323
第十章	变异数分析(二)	325
10.1	单一自由度的比較	325
10.2	2×2 复因实验	331
10.3	X 为等距时迴归中正交的比較	333
10.4	3×4 复因实验中的迴归	338
10.5	生物檢定中的平行綫檢定法	342
	小 結	348
	練習題	349
	参考文献	352
第十一章	共变数	353
11.1	在完全随机的实验中有两种处理时的共变数	353
11.2	共变数分析的計算法	361

11.3	在完全随机的实验中有几种处理时的共变数	362
11.4	随机单位组的共变数	368
11.5	复因实验中的共变数	371
	小 结	374
	练习題	375
	参考文献	377
✓ 第十二章	多元迴归	378
12.1	含有三个变数的資料	378
12.2	三元迴归方程式的計算法	380
12.3	多元相关	382
12.4	估計之誤差	383
12.5	显著性測驗及可信限	385
12.6	部分相关	389
12.7	綫性方程組的高斯解法	391
12.8	含有四个以上变数的資料	392
12.9	四元迴归的計算法、显著性測驗及可信限	394
12.10	R^2 的显著性測驗、四元和三元迴归的比較	400
12.11	多元共变数	402
12.12	計算图(諾模图)的繪制	405
	小 结	410
	练习題	411
	参考文献	412
✓ 第十三章	曲綫迴归	414
13.1	引言	414
13.2	简单拋物綫的配合法	415
13.3	三次以上拋物綫的配合法	419
13.4	最小二乘法配合指数曲綫	423
13.5	目測法配合指数曲綫	428
13.6	双曲綫的配合法	433
13.7	逐步近似法对曲綫的校正	441
13.8	有限差数法	446
13.9	Logistic 曲綫的配合法	452
13.10	曲綫迴归的显著性測驗	458
	小 结	464

練習題	465
參考文獻	466
第十四章 實驗設計	467
14.1 随机数目表的使用	467
14.2 完全随机的設計	482
14.3 配偶組的設計	483
14.4 随机单位組的設計	483
14.5 拉丁方的設計	485
14.6 不完全单位組的設計	491
14.7 复因實驗的設計	500
14.8 样本的大小	504
小 結	515
參考文獻	516

第一章 資料之搜集与整理

1.1 科学研究与数理統計方法 “科学的使命是洞察过程的深处，深入現象的本质，透过次要的、不固定的联系，揭露出主要的、固定的联系；从而发现現象的規律^{(1)*}”。所謂次要的、不固定的联系是指偶然性，而主要的、固定的联系則指必然性。医师們都知道，在同样的医疗条件下，用同一种药物来治疗某病，病人甲痊癒，而病人乙不幸死亡。这是因为病人本身具有某些不同的特点。当药物的作用与病人体内有利于痊癒的某些特点相結合时，便产生效果。当病人体内这些特点缺乏或占劣势时，医疗便不产生效果。如果我們仅观察很少数的病人，便不易正确断定該药物对某病是否有疗效；但当观察病人的数量足够多时，就可以得出該药物在一定程度上有疗效的結論。这說明：“必然性是不可避免地通过大量的偶然性为自己开辟道路的”。“科学之所以能成为精确的科学，就是因为它們透过偶然性揭示了严格的必然性，揭示了現象的运动和发展的規律”⁽¹⁾。

医学上也經常用动物实验来观察一种或比較几种药物的疗效。在实验中，每組所用的动物数量往往是不多的，少則十头左右，多的也不过数十头。实验者是想通过这几十头动物的实验結果来推論到該药物对这种动物在大量应用时的情况的。实验中所用的几十头动物在数理統計上叫做样本（亦称抽样总体），所推論的全部叫做全体（亦称全及总体）。数理統計的基本任务在于从部分推論全部，即从样本推論全体。由于現象的具体进程常受許多偶然因素（即次要的、不固定的联系）的影响，而表现出与必然性常有或大或小的离异，因此凭实验或观察的样本，不易很肯定地推断

* 括号内数字系参考文献序碼，見每章之末。

出所研究現象本身的必然性或必然联系。但是偶然性是可以用数学上的机率来衡度的。而数理統計方法就是在掌握或然(偶然)現象的規律性的基础上,利用机率来探討样本与全体間的关系的,从而使科研人員可能作出比較正确的結論。

在科学研究中,恰当地运用数理統計的方法,可以帮助我們透过偶然性来揭示必然性。一提到数理統計,也許有人就联想到高等数学上的許多符号和一連串的公式演化。不錯,这些是属于数理統計的范畴,但本书的內容是把数理統計的方法,具体应用到医学科学研究中去,因此每个公式都是通过实例来介紹的。至于公式的来源和詳細演化过程,讀者于必要时可参考数理統計专书。

把数理統計的方法应用于医学科学研究中去,首先要考虑是否用得上,是否用得恰当。用不上而硬用或用得不恰当,都会变成数字游戏。例如狗的神經型不同,把許多狗的条件反射記錄放在一起求平均数是不妥当的。但若研究在某种营养条件下,狗(或其他动物)的体重增长情况,那么同年龄同性別同种狗的体重是可以平均的,这时可以不考虑神經型的問題,因为神經型对于体重增长并不是一个决定性的因素。

1.2 实验設計中之一般注意事項 每个研究題目应有詳細的研究計劃,其中也包括实验設計部分,所謂实验設計包含下列內容:

1. 拟定相互比較的处理:如营养实验中各种飼料,药理实验中的不同药物或某药物的各种剂量等;
2. 确定实验单位:指实验所用动物或血液标本等;
3. 确定适当規則,将各实验单位分配到各种处理中去;
4. 列举对每个实验单位应作的各种測量或其他記錄。

当然,上述內容首先須視研究的目的任务而定,而数理統計的知識可以帮助研究人員在考虑上列問題时更为确当。要知道,残缺不全甚至錯誤的資料决不能靠統計方法来补救。要获得完整、正确的資料,从而得出可靠的結論,必須注意实验設計。在本书最后,对于实验設計,将作較有系統的敘述,这里只把实验設計中一般注意事項作一初步介紹。

1. 当被研究的处理数較多时,須有計劃地安排每次实验用几种处理和用哪些处理,以便相互比較;某单位曾研究 20 多种金属化合物对于抑制流行性感胃病毒的效果。由于金属化合物种类較多,不便同时进行,所以分作好多次实验,每次实验比較 2—3 种金属化合物和一个对照組(即每次实验有 3—4 組),每組用动物 3—4 头。这样作为过筛性研究是可以的。但对照組过多,各对照組間結果很不一致,不在同一次实验里的各金属化合物也不易比較。要获得可靠結論,尙須作进一步研究。因此我們对于这个实验可提出下列建議:(甲)就該研究已进行的情况来看,根据初步分析結果,有 5—6 种金属化合物可能有抑制流行性感胃病毒的作用。在进一步研究时可把这 5—6 种金属化合物加上一个对照組重做实验,并增加每組的动物数;(乙)假如該研究尙未进行,由于被研究的处理数(金属化合物种数加对照一种)有 20 几种之多,而每次实验只能做几种,这样就必須有計劃地安排每次实验所用的处理数和用哪些处理,这样各处理之間才能相互比較。具体方法可參閱第十四章第 14.6 节。

2. 相互比較的各組,除对实验因素作有計劃的变化外,其他因素应尽量使之相似;在自然界中,一切現象都是处在相互联系、相互制約中的。但自然科学家为了明确因果关系,“必須把两个現象和其余的現象隔离开来,孤立起来(虽然这两个現象和其余的現象也有联系)并把它們孤立地加以研究⁽²⁾”。因此在对照組与实验組之間必須使其他条件(如动物的年龄、性别、原始体重、健康状况以及环境条件、观察時間等等)尽量相似,而让实验因素单独作有計劃的变化,以測定該因素对实验結果的影响。有某单位在某市郊区定期噴洒 666,城区則不噴洒,两地均按时捕捉孑孓与成蚊,計其密度,以观察噴洒 666 的效果。这个实验的缺点是两地环境条件不一样,郊区水田很多,城区則街道房屋較多,水田很少。如果說,水田很多的郊区噴洒后蚊虫显著减少,足見 666 杀虫的效果。但若进一步問,究竟噴洒后蚊虫或孑孓密度减少了百分之几? 这个实验結果,便无法回答了。又某实验因动物須分三批供应,每批动物数約 50 头。研究者把第一批动物作对照組,第二批动物进行

甲种实验,第三批动物进行乙种实验,先后历时约三个月,这样设计也是不合适的。正确的设计应该是每批都做对照与甲乙两种实验,每批每组动物各为16头。

3. 各组所用实验单位必须有相当的数量,不能只用一个: 某实验将血吸虫尾蚴50条分别接种于若干动物体内,然后注射锑剂治疗,经过一定时期后将动物杀死,检视其体内血吸虫成虫的数目。在同时期内另用动物一头,也接种50条血吸虫尾蚴,不加任何治疗,观察相同时期后杀死,检视其体内成虫数。这一头动物的结果,是不能用作对照的正确依据的。因为生物现象的个别差异很大,用几头相似的健康动物,各接种血吸虫尾蚴50条,经过一定时期后,各个动物体内所生长的成虫数往往是各不相同的。所以一头动物的实验结果,常不能代表好多动物的情况(若作为典型的一例,那是另一问题,它不属于实验设计的范围以内。若实验时仅用一头动物,也无从确定它是否典型)。在设计类似的实验时,最好使实验组与对照组所用动物数相等,不要因为想节省几头动物,而在对照组中仅用一头。

4. 各实验单位应随机分配到各组(处理)中去,不能有先入为主的成见: 实验室所用若干组动物,从统计方面说,只是几个样本。从样本的结果,我们希望推论全面的情况。所推论的全部称为全体。要使样本能够代表全体,那么在抽样时应依据随机原则,不能有先入为主的成见。例如有白鼠40头,雌雄各半,年龄大致相同,拟分成四组,各组给以一种饲料。分组时可将同性别的白鼠,按体重的轻重,顺序排列。假定雌鼠为1—20号,雄鼠为21—40号。然后用红绿黄蓝四种颜色的球分别代表甲、乙、丙、丁四种饲料。各种色球以相等数量(假定各为5个,或各为10个)盛入同一口袋内。将球彻底混和后,摸出一球若为绿色,则第1号白鼠给以乙种饲料。将球放回袋内,再摸第二球,若为黄色,则第2号白鼠给以丙种饲料。如此放回再摸,若仍得绿色或黄色,因与前两次相同,故即将球放回袋内,以免重复。若继续摸得者为红球,则第3号白鼠给以甲种饲料。于是在第1—4号鼠中,剩下第4号鼠给以丁种饲料。再用同样方法指定第5—8号鼠的饲料。余依此类推。

(實驗設計中常用“随机数目表”以代替摸球法，該表及其用法，本书末章将予介紹)。这样分配的結果，哪一头鼠吃什么飼料是随机指定的，而且每种飼料的鼠数相等(上例各为 10 头)，其平均原始体重也大致相等。由于各組白鼠除飼料不同外，其他主要条件大体相似，所以实验結果是可以相互比較的。与随机原則相违反的是有成見的指定方法。例如曾有某細菌学实验將大量細菌接种在体重輕的动物体内，以促其及早死亡，而將小量細菌接种在較重的动物体内，以減輕其症狀或减少死亡。这种設計是錯誤的，因为故意加强了两組間的差別，夸大了事实。不論是有意地或无意地夸大或縮小了两組(或几組)間的差別，結果都带来偏性。在实验結果里有了偏性，就可能作出錯誤的結論。要避免偏性，必須进行随机分配。

5. 必要的記錄决不能遺漏：某实验研究白鼠注射消毒牛乳后生理上的变化，其中有体重一項。研究者將一批健康白鼠求得其平均体重，將另一批白鼠注射消毒牛乳，分別在 24、48、72 小时后測量其体重。結果“白鼠在注射后 24 小时内体重迅速降低，迄 72 小时后才漸趋恢复”。这里存在一个問題，即注射牛乳的白鼠，在注射前的体重是否与健康白鼠相似。万一注射前的平均体重較健康白鼠者为輕，那么上述結論便无所依据了。如果研究者能在注射牛乳前將各鼠的体重加以測量并記錄下来，这个問題便很容易解决了，但在事后是无法补救的。所以在动手实验以前，应仔細考虑哪些是必不可少的記錄；这些記錄必須詳細記載，否則事后很难弥补。

如果一个动物需先后受到甲、乙、丙、丁、戊五种不同的处理(处理次序应随机排列)，而在实验过程中不致引起动物死亡的話，那么每个动物都应该接受这五种处理，不应有多有少。可是我們也看到有些实验，其第一号动物仅受到甲、乙、丙三种处理，第二号动物受到乙、丁两种处理，余类推。这样无計劃地或多或少、残缺不全的記錄，就減損了实验的价值。

此外，实验記錄应有統一的格式，格式的拟訂要便于填写和整理。同一項目的度量衡单位必須統一。所用符号，如“+、++、+++”

等应有明确的定义。綜合性研究由几个部門协作进行时,对于实验单位(动物、粪便标本等)必須統一編号。某校曾进行腸道传染病調查,由微生物学、寄生虫学、流行病学三个教研組分工合作。最初对粪便标本和被調查者姓名,各教研組分別編号。进行了一个短时期,才发现三方面的結果很难碰头,才改用統一編号。这些經驗教訓也是值得我們吸取的。

1.3 分組法 搜集資料后首先要經過严格的审查。假如在实验过程中有一动物发生意外死亡,通过尸体解剖及其他各方面证据,查明与实验因素(处理)无关,那么該动物的所有記錄应予删除。又如因仪器发生故障,致記錄錯誤者亦应删除。但研究者不能先有成見,把合于他意图的資料留下,不合意的資料刪去,这样做是不对的。

資料經過审查以后就可着手整理。整理的第一步是分組。所謂分組是根据被研究現象的特性,将所有实验单位划归各組。要分組正确,必須查明被研究現象的本质。換句話說,被研究現象的本质是分組法的基础。因此我們必須选择与被研究現象的本质有关的特性作为分組的依据。分組时所依据的特性称为分組标志。

茲以各种蚊子幼虫孳生地的資料(此系某研究机关之調查結果)为例,說明統計分組的意义和作用。表1.1把34种蚊子幼虫孳生地罗列起来,未加任何分組,因此不容易看出問題。表1.2系按照各种蚊子的生活习性,特別是产卵地点,将34种孳生地区分为四大类——小容器、大容器、清水池塘和污水沟坑;每类下面添一“小計”。

經過分組以后,由表1.2可以看出尖音庫蚊淡色变种,在任何有水場所都能孳生,而尤以污水沟、坑为多。吉浦伊蚊幼虫仅在小容器与大容器內发现;而背点伊蚊与刺扰伊蚊几乎都在清水池塘与污水沟坑中孳生;中华按蚊亦然(按中华按蚊通常在清水池塘中孳生,此处結果略有出入,尚待研究)。上述几种蚊子的情况可摘要列如表1.3。

上文已提到,被研究現象的本质是分組法的基础。分組时如果誤用了非本质的标志,往往会导致錯誤的結論。例如日本血吸

表 1.2 在各种蚊子孳生地捕获幼虫的次数

蚊 子 孳 生 地			尖音庫蚊	淡色变种	三節庫蚊	二節庫蚊	翅斑庫蚊	莫地斯脫蚊	庫食庫蚊	瓦干庫蚊	吉浦伊蚊	高麗伊蚊	背点伊蚊	刺扰伊蚊	中华按蚊	潘氏按蚊
小 容 器	小砂鉄石花	紅鍋鍋槽盆	1 1 6		1	1				1	3 1 1 2	2	1			
	小 計		8		1	1				1	7	2	1			
大 容 器	鉄銅醬臭肥荷太平破花卉	箱紅缸桶盆(桶)船盆	3 5 3 9 4 39 1 8		1						3 3				1	3
	小 計		72		6				3	3	6				2	3
清 水 池 塘	葦薺花水	塘地田塘池(清)池井沟	14 8 5 5 29 3 14	11 1 16 2 6 13		2	1	3 4	2	7 5 1 3		16 1 5	10 1 3	2 1 17 1 3 8		
	小 計		78	56	2	1	7	4	16			30	18	34		
污 水 沟 坑	水临时蓄粪貯污河工地	坑(污)坑坑池沟河沟水	51 30 8 3 1 4 34 5 3 4 1	45 8 5 4				3	2	8 6 1 1 1 6 1 1 1		10 3 2 4	2 1 1 1	21 2 2	4	
	小 計		144	82			3	2	26			26	8	31	7	