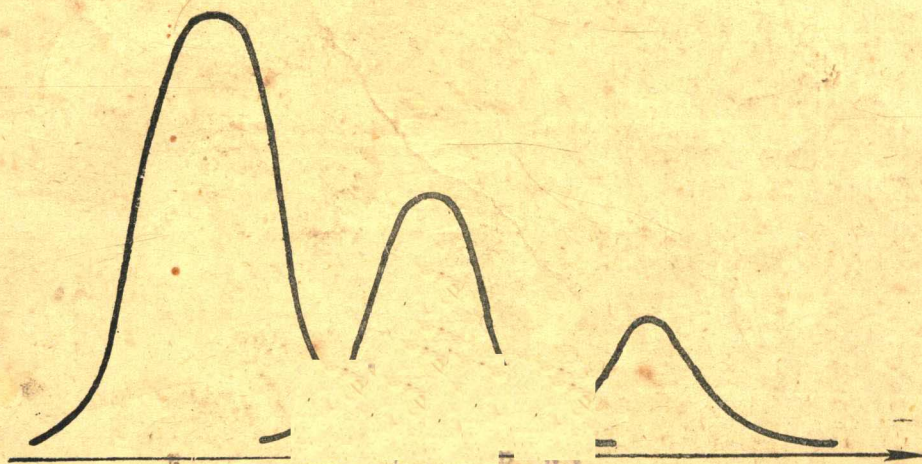


临床检验统计方法

LIN CHUANG JIAN YAN TONG JI FANG FA



吉林省临床检验中心

一九八五年六月

临床检验统计方法

编者 吕凤昌 苏凯 伍炯如

审者 贾广炎

吉林省临床检验中心

内 容 简 介

本书是以临床检验人员为主要读者的入门教材，比较全面地讲述了统计学方法在临床检验工作中的应用。本书侧重于统计学中最基本的概念和一些常用的方法，其特点是，所有的例题都与临检专业工作密切相关，并结合DS—5型袖珍计算机，具体说明了如何运用计算机进行统计学处理。对书中所涉及到的计算公式，一般都直接给出，再结合实例加以说明，尽量避免了繁冗的数学推导，通俗易懂，便于自学。

前 言

随着医学科学的迅速发展，对于临床检验的质量要求越来越高。特别是概率论与数理统计学在临床检验质量控制中的应用，使质量控制真正成为一门独立科学。此外，统计学作为一种手段，对于临床检验工作中的数据处理，方法的选择，以及科研工作起着重要的作用。为了搞好质控工作，帮助检验工作人员提高统计学理论水平，以适应迅速发展的检验事业，我们以多次举办的生化质量控制学习班讲授内容为基础，对已出版的《临床检验统计》进行了较大幅度的增补，并吸收了其它单位的经验，经过反复修改充实，重新编写成册，以供检验工作人员参考。

本书的编写工作是在吉林省卫生厅，中华医学会吉林省分会及吉林省人民医院的领导下完成的，同时得到了白求恩医科大学卫生系统统计教研室王广仪副教授，张志军及白淑清老师的支持和帮助。书中所有插图均由贾笠同志绘制，高俐、孙庆霞同志参加了校对工作。在此谨向支持我们工作的同志表示衷心感谢！

由于我们并非从事统计学工作的专业人员，因此在编写过程中难免出现错误，望读者批评指正。

编 者

目 录

绪 论	(1)
第一章 概率的基本知识	(3)
§ 1 随机事件	(3)
§ 2 随机事件的规律性	(4)
§ 3 概率	(5)
§ 4 概率的计算	(6)
第二章 随机变量	(6)
§ 1 随机变量的概念	(6)
§ 2 总体和样本	(7)
§ 3 抽样的随机化及其方法	(8)
§ 4 常见的误差	(10)
第三章 袖珍计算机的使用	(11)
§ 1 DS—5 型袖珍计算机	(11)
§ 2 操作系统	(11)
§ 3 显示系统	(14)
§ 4 功能系统	(16)
§ 5 计算示例	(43)
第四章 资料的分析	(55)
§ 1 资料收集	(55)
§ 2 资料来源	(55)
§ 3 资料整理	(58)
§ 4 资料特点	(61)
第五章 统计量与参数	(61)
§ 1 统计量与参数的意义	(61)
§ 2 平均数	(62)
§ 3 中位数与百分位数	(71)
§ 4 全距与方差	(73)
§ 5 标准差	(76)
§ 6 变异系数	(82)
§ 7 正态分布	(83)
§ 8 标准误	(90)
§ 9 数据的舍弃	(96)

第六章 t检验	(97)
§ 1 概述	(97)
§ 2 显著性检验的步骤	(97)
§ 3 t分布	(99)
§ 4 u检验	(107)
§ 5 t' 检验	(108)
第七章 方差分析	(111)
§ 1 概述	(111)
§ 2 单因素分组的多个样本均数差异的显著性检验	(112)
§ 3 双因素分组的多个样本均数差异的显著性检验	(113)
§ 4 W检验	(115)
第八章 相对数(率与比)的分析	(121)
§ 1 率	(122)
§ 2 比	(123)
§ 3 注意事项	(125)
第九章 χ^2检验(卡方检验)	(126)
§ 1 四格表资料的 χ^2 检验	(127)
§ 2 四格表专用公式计算 χ^2 值	(129)
§ 3 四格表校正 χ^2 值的计算	(130)
§ 4 行×列表的 χ^2 检验	(132)
§ 5 配对资料的 χ^2 检验	(136)
第十章 计数资料的显著性检验	(137)
§ 1 2×2 表资料的差异显著性检验	(137)
§ 2 一般行×列表资料的差异显著性检验	(138)
§ 3 4×2 表资料的差异显著性检验	(139)
第十一章 回归分析	(140)
§ 1 回归分析的意义	(140)
§ 2 直线回归	(141)
§ 3 曲线回归	(159)
第十二章 标准曲线	(170)
§ 1 标准曲线的制备	(170)
§ 2 标准曲线的绘制	(171)
第十三章 确定正常值范围	(188)
§ 1 概述	(188)
§ 2 抽取有代表性的正常人样本	(190)
§ 3 确定正常值范围	(190)
§ 4 计算方法	(193)
第十四章 统计表与统计图的制作	(197)

§ 1	统计表的制作	(198)
§ 2	统计图的制作	(200)
附表 1:	对平均数作抽样调查, S/δ 取不同的数值时所需样本大小 n 表 I ($\alpha = 0.05$)	(203)
附表 2:	对平均数作抽样调查, S/δ 取不同的数值时所需样本大小 n 表 II ($\alpha = 0.01$)	(203)
附表 3:	对率作抽样调查时所需样本大小 n 表 I ($\alpha = 0.05$)	(204)
附表 4:	对率作抽样调查时所需样本大小 n 表 II ($\alpha = 0.01$)	(204)
附表 5:	对样本均数与总体均数的差别作显著性作检验时所需样本大小 n 表	(205)
附表 6:	对两个样本均数的差别作显著性检验时所需样本大小 n 表	(206)
附表 7:	对两个率的差别作显著性检验时每组所需样本大小 n 表 I ($\alpha = 0.05$, $\beta = 0.10$)	(207)
附表 8:	对两个率的差别作显著性检验时每组所需样本大小 n 表 II ($\alpha = 0.01$, $\beta = 0.05$)	(208)
附表 9:	容许限系数 K 值表	(209)
附表 10:	异常值的舍弃 α 检验用表	(209)
附表 11:	t 值表	(210)
附表 12:	F 值表 I (方差齐性检验用)	(211)
附表 13:	F 值表 II (方差齐性检验用)	(212)
附表 14:	F 值表 I	(213)
附表 15:	F 值表 II	(214)
附表 16:	相当于概率 0.05 与 0.01 的卡方值 (χ^2) 表	(215)
附表 17:	相关系数 r 的界限值表	(215)
附表 18:	正态性 D 检验界值表	(216)

绪 论

临床检验统计方法是用数量说明事物本质和发展规律的科学，是认识社会和自然现象数量关系的重要工具，是用数字反映事物的本来面貌的科学，亦是衡量临床检验质量水平的有逻辑性和规律性的工具。

随着医学检验科学的发展，特别是生物学，生物化学的发展，实验室大量的工作（包括日常工作），需要用临床检验统计方法进行处理。在临床检验工作中，由定性的实验方法已逐步发展为定量的检验方法。目前，由于机械化、自动化等现代化仪器的生产与应用，临床检验工作已经发展到较高的水平。因而临床检验统计方法在临床检验中的作用，就更为重要，临床检验的日常工作与临床检验统计方法亦有密切关系。例如，评价实验方法的准确性、灵敏性；确定允许误差、正常值范围；说明抽样误差；衡量实验的干扰因素、影响因素；绘制各类标准曲线；实验室的质量控制等，都必须用临床检验统计方法进行处理。

临床检验统计方法的指标，有平均数、标准差、标准误、变异系数、推测统计方法、t检验、u检验、F检验、相关回归、卡方检验等，在临床检验和科研工作中，这些指标都被广泛的应用。

在科学工作中，各类科学都有自己的统计方法，医学各科也有自己的统计方法。当前临床检验质量控制工作，也需要有自己的统计方法，临床检验统计方法就是临床检验专业的统计方法。

应用统计方法于临床检验研究时，必须从唯物的观点，以临床检验学为基础，而不能只凭常识或数字的表面值，判断实验研究的差异。应用统计处理，需注意考虑抽样误差。有正确的统计方法的观点，才能发挥作用，否则判断常是错误的。

当做一项科学研究时，常需有足够的实验数据，才能反映出事物的本质及其规律。这时尤需注意，决不能应付、凑合实验数据，如实验数据少或实验数据的概率分布不符，也会得出错误结论。

正确地运用统计方法，能够帮助我们认识客观事物，阐明事物内部固有的逻辑规律，用于临床检验实践，绝不应将统计方法视为万能。统计方法不能改变事物的本来面貌，更不能将原来不存在的问题、不存在的规律创造出来。在进行一项科研、一项实验、两种检验事物的比较等，如事先不做充分的考虑和设计，收集一些杂乱的甚至是不正确的资料数据，而希望用统计方法来解决或弥补，对于这种情况，统计方法是无能为力的。就是能对付出统计指标，也常是“事与愿违”。只凭主观愿望出发，滥用统计方法去凑合预先想象的结论，是违背唯物论认识论的，对于原始数据，一定要真实可靠，不能随意编造。

检验统计方法与其他科学一样，只能阐明而不能创造事物的规律，只有有计划地收集资料，正确、合理地整理资料和分析资料，才能得出符合事物客观规律的结论，才能

用于指导临床检验工作。

临床检验工作中的许多实际问题，都是随机现象，可用随机变量描述。然而要清楚一个随机变量，就必须知其概率分布或至少要知其数字特征，但这件事很不易做到，如某日的检验结果经过严格的质量控制（统计方法），作到不失控才能报告结果。但是在不进行质量控制的情况下，要想知道当日检验结果有多少是准确的，有多少是错误的，概率是多少，显然用一般的检查方法是不行的，因为不能逐个的都做重复试验，观察其数字特征，求其概率分布，同时这样做耗费人力物力，时间也不允许。

检验工作通常是检查人体的体液（血液、尿液、脑脊液等）中某种化学物质的质和量的变化。卫生检验学中，是检查某地区水质情况，或某冰棍厂的冰棍质量，但不能把要检查的对象全部取来检验。同样，在临床检验中，也不能把一个人的血液全部取出来检验。把一口井的水、一个冰棍厂的全部冰棍都取来检验，这是不可能的，总是要从中抽取一部分进行检验，而根据这一部分来估计全体。例如，要确定吉林省血流变学各项检验正常值，只能在吉林省各地区抽取一部分正常人测定其血流变值，不能把吉林省所有的正常人都进行血流变学检查。这样，如何根据从研究的对象的全体中，抽取一部分个体进行检验或观察到的数据，对研究对象的全体的客观规律性，作出种种合理的估计和推断，也是临床检验统计要研究的问题。

临床检验统计方法是用数字说明检验结果的准确性或不准确性，是认识检验事件及数量关系的工具。临床检验大量的工作，是对人体内各种物质的检验和测定，其间为疾病的诊断、治疗、愈后，提供客观的数字资料；是用数字反映机体的生理及病理变化，如计数某病的白细胞数是15,000/微升，说明该病人有炎症；测定血尿素氮为100毫克/毫升，反映该病人有酸中毒，此即用数字反映出疾病的面貌。一种疾病、一批体检、一项科学试验、两种或两种以上实验方法的比较、正常值确定、标准曲线制备等的检验事件，都要测定出大量的数据。而用临床检验统计的方法来处理这些数据，就能够反映出资料内在的固有的规律。

第一章 概率的基本知识

§1 随机事件

在自然界和实际生活中所出现的事件（现象），一般可以概括为两种类型：一类称为必然事件，一类称为随机事件。

必然事件是指在一定条件下必然发生或者必然不发生的事件。例如用手将一枚硬币向上空抛去，必然要落回地面；在现实生活中，一个人的生命必然要死亡；在一个大气压下，纯水加热到 100°C 时，必然要沸腾；在十进制运算中，必然有 $1 + 1 = 2$ 的等式等等。这里应当特别注意，只有在以一定的条件作为前提，才能谈论事件是必然发生还是必然不发生。比方说，“在一个大气压下纯水加热到 100°C ”是条件，“水的沸腾”是事件。每当条件实现时，事件必然发生，这种事件就是必然事件，用“S”表示。有的事件在一定条件下必然不发生，比如“在一个大气压下纯水加热到 75°C ”的条件下，“水的沸腾”这个事件就必然不发生。这种每当条件实现时必然不发生的事件称为不可能事件，用“ \emptyset ”表示。必然事件和不可能事件的共同特点是，在一定条件下，可以肯定该事件是否发生，也就是说，它们发生与不发生都是必然的和可以预测的。

随机事件是在条件实现时可能发生也可能不发生的事件，也就是说，在一定条件下，无法预测该事件是否能发生。比方说，用手将一枚硬币向上空抛去，落回地面后，可能发生“正面向上”的事件，也可能发生“背面向上”的事件。究竟是哪一种，事先是无法预测的。这是随机事件的典型例子。

检验人体内的各种材料（如血液、尿液、脑脊液等），先要采样，这在临床上称为“收集标本”。在收集标本前，无法知道是总体中的哪一部分，将会得到什么样的结果，这也是“随机事件”。

例1—1：吉林省临床检验中心组织10个实验室，分别测定同一份标本的谷丙转氨酶（GPT），其测得值见表1—1。

表1—1：10个实验室谷丙转氨酶的测得值

实验室编号	1	2	3	4	5	6	7	8	9	10	真值
测定结果 (单位)	36	35	37	100	34	38	35	34	98	36	35

由表1—1可见，同一实验，同一标本，在10个实验室测定的10个测得值中，有8个数值接近真值，称为“准确值”，有2个数值偏差很大，称为“错误值”。如从这10个测得值中，任意取5个测得值，则可能发生5个测得值，都是“准确值”，也可能不发生；又可能发生5个测得值中有4个是“准确值”，1个是“错误值”，也可能不发生；再可能发生5个测得值中有3个是“准确值”，2个是“错误值”，也可能不发

生，这样的事件，都是随机事件，用“A、B、C、……”表示。

从10个测得值中，因为只有2个是“错误值”，所以当任意取5个测得值时，不可能发生5个测得值全是“错误值”，亦不可能发生5个测得值中有3个是“错误值”。这是不可能事件。

§ 2 随机事件的规律性

从表面上看，随机事件的发生与否似乎全为偶然性，而无规可循。但实际上并不如此。

在临床检验的各项实验中测得的数值，可能“准确”，也可能“错误”。就“准确”而言，可能发生，也可能不发生，操作者全然不知。出现错误好象完全是偶然的，这种测定的错误，通常称为“偶然误差”。但是如果经过多次或长时间的观察（例如开展质量控制工作），在对大量现象的归纳和整理中，会发现这种表面上的偶然性始终受其内部隐蔽着的规律所支配。如在操作严格的条件下，多次测得值的平均数就接近真值。许多实验的测量值，当测定次数多到一定限度时，其频数呈“正态分布”（正态分布是统计学中一种重要分布，待后详叙）。由此可见，对于随机事件发生的可能性大小，是能够用某种数量来描述和表达的。

例1—2：测定长春市、老年人的血脂含量，以观察高脂血症的发病情况，由于测定（调查）人数的不同，得出的发生率亦不同，见表1—2。

表1—2：长春市、老年人高脂血症发生率

测定人数 (n)	高脂血症人数 (n _A)	发生率(频率) $\left(\frac{n_A}{n}\right)$
5	0	0
10	2	0.200
50	6	0.120
170	19	0.110
800	59	0.074
900	78	0.086
1300	106	0.081
2000	176	0.088

表1—2中的发生率（频率），等于高脂血症的检出人数与调查人数之比。

由表1—2中的数字可见，高脂血症的发生人数的多少，具有偶然性（随机出现），但随测定（调查）人数的增多，则发现高脂血症的发生率（频率），在0.08附近波动，且逐渐趋于稳定。

临床检验在一般情况下，当要考查随机事件A发生的可能性大小时，只要在相同的条件下，操作严格，做大量的重复实验，会得出一种结果，就是随实验次数的增多，随机事件A发生的频率，总是会稳定在某一数值的附近。这种频率的稳定，是每次实验本

身所固有，不随人们的主观意愿而改变，是一种固有的属性。随机事件A发生的可能性的大小，正是此频率的稳定值。因此我们可以通过此频率的稳定值，来认识随机事件A。

§3 概 率

在相同的条件下，严格、重复操作 n 次实验（测定），以 n_A 表示 n 次实验（测定）中随机事件A发生的次数，当实验（测定）次数很大时，则频率 $\frac{n_A}{n}$ 稳定在某一数值P的附近波动，随实验次数增多，这种波动的幅度越来越小，我们把数值P称为随机事件A的概率。

概率（Probability）用P表示。

随机事件的概率用 P_A, P_B, P_C, \dots 表示，必然事件的概率用 P_s 表示，不可能事件的概率用 P_ϕ 表示。

随机事件A有确定的概率，可用下式表示：

$$P_A = P$$

其中P是一个确定的数值。

概率可成为描述事件发生大小的一个数量。如例1—2，长春市中、老年人高脂血症发生的可能性为0.08（8%），特别是在实验次数 n 很大时，频率 n_A/n 就会近似于高脂血症人数（事件A）的概率 P_A 。

即
$$P_A \approx \frac{n_A}{n}$$

由于
$$0 \leq n_A \leq n$$

所以
$$0 \leq \frac{n_A}{n} \leq 1$$

这就是说，对于任何随机事件A，都有 $0 \leq P_A \leq 1$ ，即概率的数值范围在0到1之间。很显然，必然事件的概率 $P_s = 1$ ，不可能事件的概率 $P_\phi = 0$ 。概率P越接近1，表明事件发生的可能性越大；概率P越接近0，表明事件发生的可能性越小。

临床检验的各项实验，从标本的收集，到测定值的出现，整个过程都是随机事件，因而必然性和不可能性都不存在。

在临床检验统计方法中，常用的有 $P \leq 0.05$ 和 $P \leq 0.01$ 。 $P \leq 0.05$ ，表明事件发生的可能性等于或小于0.05（5%），作为事物的差别有显著意义的界限； $P \leq 0.01$ 表明事件发生的可能性等于或小于0.01（1%），作为事物的差别有非常显著意义的界限（待后详叙）。

我们必须准确地理解概率的定义，千万不要把概率简单地看作为频率、百分数等等。另外，在实验中，随机事件A的概率值，只能是一个，而不是几个。例如在例1—2的实验中，每一次测定都得一个频率，但不能随便取一个都当作概率，而只能把在相同条件下，当测定次数充分大时的稳定频率看作概率 P_A 的近似值。又如在某次白细胞分类计数时，淋巴细胞为80%。这就是说，在100个白细胞中，有80个淋巴细胞，因而得出淋巴细胞占80%，在数学上称为百分率，不是概率。只有在多次观察（测量）后，此淋巴细胞

的百分率总是在80%这样一个水平时，才能说此病人患淋巴细胞型白血病的概率为0.8 (80%)。

§ 4 概率的计算

实际上，概率的准确值，在多数情况下是难以得到的，只能在大量重复实验所得的频率进行估计。在随机事件中，可用计算的方法求其概率。例如，用手向上空抛去的一枚硬币，当落回地面时，“正面向上”的概率是多少？很显然，投掷硬币只有两种不同的结果，即“正面向上”或“背面向上”，这两种不同的结果，出现的机会相等，每种出现的结果，称为“基本事件”。

计算公式：

$$P_A = \frac{K}{n} \quad \text{公式 1—1}$$

式中： P_A 为事件A的概率；K为事件A包含的基本事件；n为事件的总数。

设硬币落回地面时“正面向上”的基本事件 $K = 1$ ，基本事件总数 $n = 2$ ，

$$\text{则 } P_{\text{正面向上}} = \frac{1}{2}。$$

例 1—3：在木箱中有20瓶化学试剂，其中有2瓶已经变质，自木箱中任意取出1瓶，问取出的变质化学试剂的概率是多少？

由木箱中任意取出1瓶，则会有20种结果，每瓶被取出的机会是相等。

变质的化学试剂基本事件 $K = 2$ ，变质和不变质的化学试剂的基本事件总和 $n = 20$ 。

将有关数值，代入公式 1—1。

$$P_{\text{变质试剂}} = \frac{2}{20} = \frac{1}{10}$$

临床检验工作的大量测得值，都是随机事件，其准确值和错误值，都有相同的机会出现，特别是在没有质量控制的手工操作情况下， $P_{\text{准确值}} = \frac{1}{2}$ 。因而在临床检验统计学的应用，以及开展质量控制工作，都是很重要的。

第二章 随机变量

§ 1 随机变量的概念

在第一章中，已叙述了随机事件和概率的性质，这使我们对随机事件的统计规律性能有一个初步的认识。今将随机事件的数量性特征——随机变量进行叙述。下面举几个例子来说明随机变量：

例 2—1：任意从一批尿标本中取出50个，其中酮体“++++”的标本数是一个随机变量，它可能取值为0、1、2、……50。

例 2—2：某化验室在一小时内接到化验单的次数是一个随机变量，可取值为0、1、2、……。

例 2—3：从一批待测血糖的标本中任取一个，其含糖量是一个随机变量，可能取值范围，大约是70毫克/100毫升（或以下）~150毫克/100毫升（或以上）。

由此可见，随机变量是一个变量，但所取值是随着偶然机会而变化的，象这样在一定范围内随机地取值的变量就称为随机变量，用“X”表示。临床检验中各项实验的测得值，都有这样的特点。随着实验次数的不同而致测得值不同；即使在相同的条件下，同一份标本经多次测定，每次的测得值也在变，这种变量都是随机变量。

随机变量按其取值范围的性质，可以分为两类：有的随机变量只能取有限或可数多个值，而且所能取的值可以一一地排列起来（如例 2—1 和例 2—2），这种随机变量称为“离散型”；而有一种随机变量所能取的值是充满某个区间，或者说，可以在某个区间内连续地取值（如例 2—3），这种随机变量称为“连续型”。

§ 2 总体和样本

临床检验统计主要是研究检验事件中的每一总体的数量特征，如人体内血糖、血脂、血钾等。每一种化学成分，每一总体，都有其数量特征，这些数量特征自然对应一些随机变量值X。随机变量值X的全体取值，称为总体。

总体是相同性质的个体所组成的集体。总体包含的个体数量是无穷大，这常是设想或是抽象的。

临床检验的日常工作，是测量病人体内物质（血液、尿液、脑脊液等）的质量或数量的变化。病人体内的血液，24小时的尿液量等，都可设想为总体。水质检验就可将一口井的水，一个湖内的水，视为总体。

在一个总体，如长春市中年知识分子的血糖浓度，要测定（考虑）血糖随机变量值X中，提出 X_1 、 X_2 、 X_3 …… X_n 个知识分子的血糖，此为总体X（知识分子的血糖）中的n个样本，亦即从总体中抽取的部分个体，称为“样本”。

从总体中抽取部分个体，称为“抽样”，所抽得的部分，称为“样本”，在一个样本中可以含有不同的个体数。样本所包含的个体数，称为“样本含量”。临床检验的每份标本（检验物）都是由总体中抽取的一部分，如血液标本、尿液标本、脑脊液标本等，都称为“样本”。

样本 X_1 、 X_2 、 X_3 …… X_n ，必须是从总体中随机抽取的可能结果，可视为是n个随机变量值，在每次抽取之后，都有具体的数值，此具体数值，称为“样本值”。

我们应该注意，样本 X_1 、 X_2 、 X_3 …… X_n ，是具有双重意义的。它有时是指具体抽取后的样本值，有时是指任何一次抽取后的结果个数，即是n个随机变量值，称为“样本含量”。

临床检验统计，就是根据样本 X_1 、 X_2 、 X_3 …… X_n 的数据性质，对总体X的某种（些）特性进行估计和推断。如吉林省为100名知识分子进行血糖测定，这100名知识分子血糖含量，有 X_1 、 X_2 、 X_3 …… X_{100} 的随机样本值，同时又可视为“样本含量”为100，根据这100个样本值的性质（大、小、正常、异常），推断吉林省几万名知识分子血糖含量的特征（高、低），这是用样本的性质决定总体的特征，因此要求抽取的样

本，能很好的反映总体的特征，也就必须对样本的如何选取，提出严格要求。

在临床统计学上，通常要求样本 X_1 、 X_2 、 X_3 …… X_n 等是相互独立的，且又与总体有相同概率分布的随机样本。临床检验的样本，绝大多数随机样本，即随机重复抽样，对于总体 X ，取值是无限个或有限个，但取值多时，不放回抽样，可视为是放回抽样。例如，取某低血糖病人血液10次，即 X_1 、 X_2 、 X_3 …… X_{10} 样本，分别测其血糖含量（毫克/100毫升），推断总体 X （血液含量）。当再取 X_{11} 样本时，也应视为总体 X 中含有 X_1 、 X_2 、 X_3 …… X_{10} （已放回总体中）。对于这样的随机样本与总体应有相同的概率分布，如再取 X_{12} 样本时，低血糖病人已经治愈，或在点滴葡萄糖时， X_{12} 样本的血糖含量可增高，则 X_{12} 样本值与总体 X （低血糖）概率分布不相同，这对于推断低血糖水平，没有统计意义。

§ 3 抽样的随机化及其方法

随机化是指在抽取样本时，应使总体中的个体都有同等被抽取的机会。这是抽样研究和抽样分配的一个十分重要的原则。从研究对象的总体中抽取样本进行研究时，为了对总体有较好的代表性，并能使抽样误差的大小可用统计方法估计，故抽取样本时必须使样遵守这个原则。就是在抽取样本以前，要使每个单位（总体中的个体）都有同等的机会被抽取，也就是机会均等。

临床检验抽取样本时，随机化的意义十分重要，特别是有许多样本在抽取时有一定的要求，如空腹采血，服药前留取标本，清晨采血，留取24小时尿液等，都有一定的要求，但又必须注意随机要求，有时只注意实际的要求，不注意随机意义，也会得出错误结论。如某单位有职工5000人，要了解职工的肝炎发病率，测定血清转氨酶。组织者为了满足实际要求和随机要求，通知×月×日早晨5—6时，抽取100人空腹血液测定转氨酶，大家自愿来抽血，抽取至100人为止。经过计算，得出该单位职工患肝炎率为58%的结论。很显然，组织者在这里虽然满足了实际要求，但是却没有注意到怎样做才算是“随机化”，因为按照这种自愿抽血检验的通知，在来者中间往往是怀疑自己有肝炎的人居多，而那些自认为身体健康的人可能不愿意来。这种抽样不随机化的现象，很可能会导致不正确的结论。

因为一般的数理统计方法都是在随机抽样的基础上推演而来，在实验中，测量值的对象要分组时，也必须用随机方法，将各个“对象”有同等的机会分配到各组，这样就不至于人为地造成各组间的不齐同性。随机化的具体方法有抽签法和随机率字表法。

（一）抽 签 法

此法是将测量值写在签上，按资料分析的目的进行抽取，有简便易行的优点。

例如，要了解2000名50岁以上老年人糖尿病发病情况，拟用抽签法调查100人，可将欲调查的2000人，每一人名写在一个签上，则有2000个签，将这些名签放一容器内，充分混匀，随机抽取100个名签，与这100个名签相同姓名的老年人，就是所要调查的100人，亦即单纯随机抽样的一个样本。

（二）随机数字表法

此法是一种由许多随机数字排列起来的表格，这与抽签的道理是相同的，只不过不

用事先写名签，如表2—1。

表2—1： 随机数目表

28	64	190	37	95	193	160	99	168	125
167	18	165	97	62	166	9	181	75	102
109	73	103	69	179	66	42	195	30	72
36	163	118	20	34	155	198	107	96	38
78	117	107	40	188	111	159	134	39	189
186	145	90	94	3	139	148	50	124	127
53	154	82	110	115	197	77	52	100	105
183	101	33	106	91	87	56	63	41	65
132	161	176	169	67	122	191	184	180	135
114	43	29	32	199	147	192	27	131	47
17	130	16	4	133	158	68	57	123	21
49	142	44	2	71	22	81	140	104	46
137	174	92	182	144	153	138	113	23	151
60	88	6	156	54	76	12	5	178	59
143	26	15	170	14	150	119	164	194	152
149	177	84	61	19	58	129	120	31	86
187	45	108	70	171	185	51	173	83	141
98	146	126	175	11	7	93	128	35	89
80	79	200	74	55	8	24	85	13	196
10	116	25	48	162	1	121	136	172	112

随机数字表的用法，当人数较多时，常是将要了解的人群都编号，然后应用随机数字表，进行随机抽样。例如，欲了解吉林省某单位1000名50岁以上老年人患糖尿病的情况，拟抽取120名进行血糖检验，以观察患尿糖病的情况。首先将1000名老年人都编成号，在随机表中任意指定一个数，由这个数（一位数）开始向任意方向摘录数字，以四个数为一组，共摘录120组，这些四位数大于1000—2000者，均减去1000，大于2000—3000者减去2000，大于3000—4000者减去3000，余类推，使每个数字都不大于1000。例如，表2—1中的第六行第一列的1开始，向右摘录的数字是1861、4590、9431、3914、8501、2412、7531、5482、1101、1519、7775、2100、1051……等，分别减去1000、2000、3000、4000……等。所得之差，分别为861、590、431、914、501、412、531、482、101、519、775、100、51……等，可得出120个数字，与数字相对应的老年人，就是随机抽取的样本。

随机数字表的用法，当人数较少（200名以内）时，则可直接抽取。例如，欲抽50名作样本，可以任意选用2组，与2组编号相对应的人，就是随机抽取的样本。

§ 4 常见的误差

在临床检验工作中，对于误差的研究和控制是重要的内容。误差是指测得值与真值的差距，亦是指样本的指标与总体指标的差距，主要有系统误差，随机误差和抽样误差。

(一) 系统误差——系统误差是在收集资料过程中，由于仪器不准，标准试剂长期不标定等，使测定结果呈倾向性的偏大或偏小。系统误差可影响原始资料数据的准确性，应尽力避免，如已发生，要及时查明原因，予以改正。

(二) 随机误差——随机误差是在实验过程中，仪器和试剂已经校正，但由于各种偶然因素造成的测定结果的过高或过低。随机误差亦称为偶然误差。随机误差是不可避免的，但应努力作到仪器性能（线性、灵敏性、准确性）及操作方法稳定，使其控制在一定的允许范围内，可用临床检验统计方法进行控制。

随机误差包括过失误差，过失误差是一种不负责任、粗枝大叶、科学性不强所致的误差，如收集资料过程中取错标本，数据填写错误等。这种误差，在统计过程中应避免。

(三) 抽样误差——抽样误差是由于抽样所引起的差异。总体中的各个体存在差异，因而在同一总体中随机抽取若干个体（含量）相同的样本，而各样本的均数（或率）相互亦不同，这些样本间的差异，同时反映了样本与总体间的差异。一般样本越大则抽样误差越小，越与总体的情况相接近，正确性越高，也就是越能阐明事物的规律性。

临床检验工作总是由总体中抽取一部分做为样本，再根据样本观察（测量）出的结果推论总体的情况。例如，临床检验由病人体内采取血液（静脉或末梢）标本，用以说明或推论该病人总的血液内物质质量或数量的变化。食品检验随机抽取数支冰棍用以推测一个冰棍厂的冰棍质量。由于同在一个总体内个体之间存在差异，同一病人的血液，在不同时间、不同病情，血液中的个体差异是永远存在的，同一冰棍厂的冰棍，按严格要求质量亦均不同，存在个体差异。因而样本的均数和总体的均数之间，各个样本均数之间存在着一定的差异，这种差异是由抽样所引起。抽样误差在临床检验抽样中是不可避免的，但是抽样误差有一定的规律性。在理论上，当样本均数无限多时，样本均数的平均数应等于总体平均数。

在实际工作中，由于种种客观条件的限制，我们不可能观察无限多的对象，而只能从一个较小的样本资料进行分析研究，得出一系列的统计结果。一般来说，样本越小，则抽样误差相应地越大。如果仅仅凭数字的表面值进行判断，不考虑偶然性问题，往往会导致错误的结论。所以我们应该学会正确地应用统计学的原理和方法，才能辩明哪些数据（样本值）有实际意义，哪些数据（样本值）是由偶然性因素造成的。如实验的偶然误差产生的样本值，取样错误造成的样本值及与实际不符的样本值等等，都是“偶然性的样本值”。如吉林省临床检验中心调查19个医院生化室血糖标准液的情况，其中有一个医院送检的血糖标准液样本值为0毫克/100毫升，经了解其样本取错了。这当然是偶然性问题，不应考虑，算作偶然误差。