

计算药理学

相秉仁 编著
安登魁 审阅

中国医药科技出版社

911

内 容 提 要

本书系统介绍计算机技术、数学在药学（特别是药物分析）中的各种应用方法，主要内容有计算机和数理统计基本知识、计算分光光度法、聚类分析、最优化技术、因子分析和模式识别等。为方便读者实际应用，附有化学统计分析、分光光度分析、色谱优化、药物设计、药动学等方面的应用程序 33 个。可作为药学院校本科生、研究生的教学参考书，也可供药学、化学工作者参考。

计 算 药 学

相秉仁 编著
安登魁 审阅

*

中国医药科技出版社 出版发行

(北京西直门外北礼士路甲 38 号)

江苏省地矿局测绘大队印刷厂 印刷

*

开本 787×1092mm 1/16 印张 34 1/8

字数 460 千字 印数 1-2000

1990 年 12 月第 1 版 1990 年 12 月第 1 次印刷

ISBN 7-5067-0343-2/R·0293

定价: 9.80 元

前 言

科学的发展,技术的进步,先进仪器和设备的使用,给人们提供了越来越多的数据和信息,这就对数据和信息的处理加工能力提出了更高的要求。显然,由于自然力的限制,人类仅靠自身的力量越来越不能满足这一要求。而电子计算机特别是微型机的广泛应用,为这一问题的解决提供了可能。在药学各领域中,诸如实验条件优化、实验数据处理和结果评价、有机结构解析、波谱分辨、药物结构和生物活性间定量关系研究、药物动力学研究、生药鉴定以及智能模拟等方面,计算机技术均获得了成功的应用,使药学科学的研究方法发生了深刻的变化,从而孕育出一门新的药学分支——计算药学。所谓计算药学,就是计算机在药学领域中应用的科学。它是计算机技术、应用数学和药学相互渗透的结果,是研究如何将计算机技术应用于药学领域的一门方法学,它和最近崛起的计算物理、计算流体力学、计算几何和计算化学等新的学科分支一样,都是计算机广泛和深入应用而结出的硕果。计算药学的形成和发展,将为实验技术的更新、数据和信息的处理加工乃至整个药学科学开辟了新的前程。

计算药学是一门新的边缘学科,它涉及学科面很广。它的历史不长而且仍在迅速发展之中。因此,本书远不能包含它的全部内容。为便于读者掌握和开发计算机在药学中应用的方法,在编排上首先着重介绍最必需的计算机以及数学方面的基本知识,然后逐一介绍计算机技术在药学中应用的有关方法。为使初学者能更好地领会本书内容并将其应用于自己的工作,本书附列了大量的应用程序以供读者参考。

安登魁教授七十年代中期就致力于计算机在药学(主要是药物分析)中应用方面的研究,并指导我们这方面的科研和教学工作。在他的指导和帮助下,在为本科生和研究生授课讲稿的基础上,结合近几年的科研工作,作者执笔完成了本书。安登魁教授还审阅了本书的全文。特此向悉心指导并付出辛勤劳动的安登魁教授致以最衷心的感谢。

在本书编写中,中国药科大学的房杏春、郭寅龙、徐建平、马勇健、冯芳、张尊建和杨永健等同志先后为本书编制或改编了有关应用程序。张尊建和郭寅龙两同志修改、整理了全部程序,并在 IBM-PC 机上通过。此外,郭寅龙、杨永健和张尊建等同志参加了本书的微机排版工作;开封医学专科学校王明治副教授曾给予不少帮助。谨此一并表示谢忱。

最后,向中国药科大学药物分析教研室和研究室的全体老师、计算中心的全体老师以及南京大学忻新泉教授对作者工作的指导和帮助表示衷心的感谢。

作者

目 次

第一章 电子计算机基本知识	1
§ 1 电子计算机的基本概况	1
§ 2 微型计算机系统	4
第二章 实验数据的统计处理	7
§ 1 样本的数字特征	7
§ 2 t-检验在样品测试中的应用	8
一、平均值的置信界限	8
二、平均值与标准值的比较	9
三、两平均值的比较	10
四、配对比较	12
五、t-检验应用程序	14
§ 3 单因素方差分析及应用程序	18
§ 4 交叉分组的方差分析	23
一、交叉分组的双因素试验	23
二、有交互作用的双因素试验	26
第三章 矩阵与向量的简单知识	31
§ 1 矩阵及其基本运算	31
§ 2 线性方程组	36
一、线性方程组的基本知识	36
二、线性方程组求解法	40
三、全主元高斯消去法程序	44
四、误差分析	46
§ 3 特征值与特征向量	50
一、特征值与特征向量	50
二、用雅可比法求算实对称矩阵的特征值和特征向量	52
三、求算实对称矩阵特征值和特征向量的雅可比法程序	53
第四章 回归分析与曲线拟合	58
§ 1 一元线性回归分析	58
一、经验公式与最小二乘法	58
二、相关性检验	59
三、回归线的精度	61
四、回归方程的稳定性	62
五、两条回归直线的比较	63
六、化曲线为直线的回归问题	64
七、一元线性回归程序	64
§ 2 多元线性回归	66
一、基本原理	66

二、回归问题的方差分析和统计检验	69
三、多元线性回归程序	75
§ 3 逐步回归	78
一、最优回归方程的选择	78
二、逐步回归的计算方法	79
三、逐步回归的一些问题	81
四、逐步回归程序	86
§ 4 正交多项式回归	90
一、多项式回归	90
二、正交多项式回归	91
§ 5 曲线拟合	96
一、多项式拟合	98
二、高斯-牛顿法	99
三、高斯-牛顿法求算药动学模型参数程序	102
四、麦夸尔特法	104
§ 6 回归分析在药学中的应用	106
第五章 复方药物制剂的计算分光光度法	109
§ 1 双波长分光光度法	109
一、等吸收点法测定原理	109
二、系数倍率法用于二组分混合溶液测定时的原理	110
三、系数倍率法用于三组分混合溶液测定时的原理	110
四、系数倍率法用于紫外-可见分光光度法测定时的主要步骤	112
五、一元线性回归在系数倍率法中的应用	112
六、两组分系数倍率法波长对选择程序	113
七、三组分系数倍率法波长对选择程序	115
§ 2 三波长分光光度法	118
一、基本原理	118
二、三波长法中波长组选择程序	120
§ 3 多波长直线回归法	124
一、两组分测定原理	124
二、三组分测定原理	125
三、两组分多波长直线回归法程序	126
四、三组分多波长直线回归法程序	128
§ 4 导数分光光度法	131
一、导数分光光度法的基本原理与特点	131
二、导数法中间波长选择程序	133
§ 5 正交函数分光光度法	136
一、基本原理	136
二、正交函数法条件选择程序	139

§ 6 线性方程组、最小二乘和 P-矩阵分光光度法	141
一、线性方程组法	142
二、最小二乘法	143
三、P-矩阵法	144
四、P-矩阵法程序	145
§ 7 线性规划分光光度法	149
一、基本原理	149
二、线性规划法程序	151
§ 8 卡尔曼滤波法	154
一、卡尔曼滤波法在多组分分光光度法中应用的基本原理	154
二、卡尔曼滤波法程序	157
第六章 聚类分析	160
§ 1 聚类统计量	160
一、距离	160
二、相似系数	162
§ 2 系统聚类法	163
一、最短距离法	163
二、最长距离法	165
三、中间距离法	165
四、重心法	166
五、类平均法	167
六、离差平方和法	167
七、系统聚类法的性质	168
八、系统聚类法程序	170
§ 3 动态聚类法	183
§ 4 聚类分析在药学中的应用	187
第七章 最优化技术	189
§ 1 单纯形法	189
一、基本原理	189
二、计算步骤	192
三、色谱优化指标与单纯形法的应用	195
四、改良单纯形法(实验法)程序	197
五、改良单纯形法求算药动学模型参数的程序	201
§ 2 窗口图解技术	204
一、窗口图解技术及其在 GLC 分析上的应用	205
二、窗口图解技术在 HPLC 分析上的应用	208
三、窗口图解法程序	210
§ 3 混合物设计统计技术和叠加分辨率图示法	212
一、混合物设计统计技术	212

二、叠加分辨率图示法 ORM	219
三、混合物设计统计技术程序	220
§ 4 纸层析和薄层层析溶剂系统最佳组合的数值分类-信息量选取法	222
一、层析系统的信息量评价	222
二、数值分类-信息量选取法	222
三、数值分类-信息量选取法程序	223
§ 5 系统矩阵组合法	226
一、逐步组合法	227
二、择优组合法	227
三、系统矩阵逐步组合法程序	229
第八章 因子分析	233
§ 1 主成分分析	233
一、主成分分析的一般数学模型	234
二、主成分的导出	236
三、主成分分析的数值算例	239
四、NIPALS(非线性迭代偏最小二乘)算法	240
§ 2 因子分析	241
一、主因子方法	241
二、主成分分析与因子分析的关系	245
三、因子数的确定	245
四、方差极大正交旋转	246
五、因子得分	249
六、因子分析的计算步骤	251
七、因子分析程序	255
§ 3 目标因子分析	264
一、基本概念	265
二、目标变换技术	269
三、目标因子分析数值算例	272
四、目标因子分析程序	279
§ 4 偏最小二乘法	287
一、基本原理	287
二、偏最小二乘法程序	288
§ 5 因子分析在药学中的应用	294
第九章 模式识别	296
§ 1 模式识别的基本概念	296
§ 2 最小距离判决法	298
一、最小距离判决法	298
二、最近邻域判决法	299
§ 3 线性学习机法	300

一、线性学习机法	300
二、线性学习机法程序	303
§ 4 最小二乘的最小距离判决法	304
§ 5 判别分析	306
一、贝叶斯判别准则	306
二、Fisher 型线性判别分析	307
三、贝叶斯判别分析	309
四、逐步多类判别分析	314
§ 6 PRIMA 法	315
§ 7 SIMCA 法	318
一、基本原理	318
二、SIMCA 法程序	321
§ 8 映射技术	333
一、特征向量投影法	334
二、非线性映射技术	334
三、非线性映射程序	336
§ 9 模式识别在药学中的应用	346
附表	348
表 1 t 分布的双侧分位数 (t_α) 表	348
表 2 F 检验的临界值 (F_α) 表	349
主要参考文献	354

第一章 电子计算机基本知识

§ 1 电子计算机的基本概况

电子计算机从原理上可分为二种类型：模拟电子计算机和数字电子计算机。模拟电子计算机是一种以连续变化的电压来表示被运算量的电子计算机；数字电子计算机是一种以数字形式的量值在机器内部进行运算的电子计算机。在这二种类型的基础上把模拟技术和数字技术灵活地结合起来的电子计算机称为混合式电子计算机。

由于数字电子计算机运用得最为广泛，所以通常称呼的电子计算机或计算机即指数字电子计算机。

从用途上讲，计算机又分为通用计算机和专用计算机。通用机又分为巨型机、中型机、小型机和微型机等。由于微型机具有成本低、体积小、功能强的特点，已广泛应用于药学实验室中。因此，本章着重于微型计算机的讨论。

电子计算机(以下简称计算机)俗称电脑，是一种能够把信息自动高速存储和加工的电子设备。它不但能进行加、减、乘、除等数学运算，而且能对参加运算的数据和计算结果进行逻辑判断，并且具有记忆能力。因此，它能自动、快速地解决各种数学问题和逻辑问题。

为了说明计算机的组成，让我们来研究一下人们利用算盘算题的过程。

假定用 W 克重的 Na_2CO_3 标定 HCl 溶液时，消耗了 HCl 溶液 V 毫升，则 HCl 的当量浓度可通过下式计算：

$$N = \frac{W}{\frac{VE}{1000}}$$

其中 E 为物质的克当量。

这个题目的计算步骤为：

第一步 根据给定题目确定计算步骤和方法，并且把计算公式、解题步骤和原始数据写在纸上。

计算公式： $N = W / (VE / 1000)$

解题步骤：先做 $V \times E$ ，其次做 $VE / 1000$ ，最后做 $N = W / (VE / 1000)$

原始数据为： $W, V, E, 1000$

计算结果为： N

第二步 按先乘除后加减的原则在算盘上进行计算。

做 $V \times E$ ，将得到的中间结果写在纸上；

做 $VE / 1000$ ，将得到的中间结果写在纸上；

计算 $N = W / (VE / 1000)$ ，得到结果。

第三步 把结果在纸上记录下来，全部计算最后完成。写在纸上为 N 。

从上述过程可知，为完成计算必须具备：

记忆部分 用来存放原始数据、计算步骤、中间结果和最终结果。这里是纸和人的记忆中枢。

运算部分 用来完成运算。这里是算盘。

控制部分 用来控制整个计算过程，这里相当于人的意志中枢。

输入输出部分 将信息送给意志中枢，将数据写在纸上。这里相当于眼、手和笔。

计算机就是模拟上述解题过程的自动机。因此，它也应具有与上述功能相应的几个组成部分。在计算机中分别称为存贮器、运算器、控制器和输入输出设备。其组成框图如图1所示。

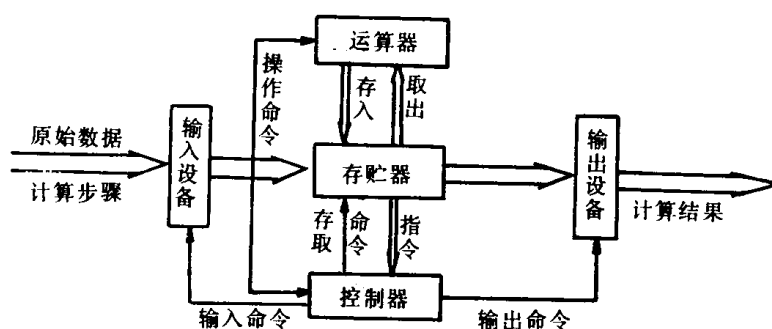


图1 电子计算机组成框图

存贮器：具有记忆功能，用来存贮计算步骤、原始数据、中间结果和最终结果。简言之，就是用来存贮程序和数据。

运算器：用来进行加、减、乘、除等算术运算及逻辑运算。

输入设备：解题程序及原始数据通过输入设备转换成计算机能够识别的代码送到存贮器中保存起来。

输出设备：计算结果或人们所需要的其它信息通过输出设备传送出来，如打印在纸上。

控制器：控制计算机各部分按人们预先规定的计算步骤(即事先编好的程序)自动地进行操作。如控制运算器进行运算，控制运算器和存贮器之间的交换，控制输入输出设备的工作等。

通常，我们把组成计算机的这五个部分，称为硬件。

实际上，计算机除了这五部分外，还必须有电源等设备。而且，有的计算机各部分紧密而有机地结合在一起，很难机械地加以划分。但是，从原理上讲，计算机总可以看作是由上述这五部分组成的。

和别的设备不同，计算机不只是有了适当的硬设备和有训练的操作人员就能使它得以运行。它还需要有存放在机器里的程序。告诉它要作什么，怎样进行运算。因此，为了让计算机能按人的要求去工作，就必须事先把计算方法和步骤翻译成机器能懂的语言，即机器语言，这种语言的翻译工作就是使用机器语言编制解题程序。接着，把编好的程序连同原始数据设法通过输入设备送到存贮器中存起来。然后，启动计算机，计算机按人的要求自动地进行全部计算工作。计算机计算完毕后，通过输出设备送出计算结果。

相对于硬件而言，我们把各种各样的程序称作软件。

只有硬件和软件同时具备，计算机才能自动地、快速地、连续地工作，完成各种各样的工作任务。

现在，我们仍从 $N = W / (VE / 1000)$ 为例，说明计算机是如何工作的。

第一步 将编好的解题步骤(以后称为程序), 原始数据通过输入设备送到存贮器存起来。

第二步 进行计算

1. 从存贮器取出 V , 送到运算器中;
2. 从存贮器取出 E , 送到运算器中, 与运算器中的 V 相乘, 中间结果 VE 放在运算器中;
3. 从存贮器取出 1000, 送到运算器中, 进行 $VE / 1000$ 运算; 中间结果放在运算器中;
4. 从存贮器中取出 W , 进行 $W / (VE / 1000)$ 运算, 这时, 运算器里暂存着最终结果 $N = W / (VE / 1000)$ 。
5. 将最终结果 N 送到存贮器。

第三步 由输出设备将最终结果 $N = W / (VE / 1000)$ 打印在纸上。

第四步 停机。

计算机的最基本功能是进行数的计算和处理加工。我们日常使用的是十进制的数。但在计算机中, 数是以器件的物理状态来表示的。例如电路的“开”、“关”这两个状态。因此, 为了使表示更为方便和可靠, 在计算机中主要采用了二进制数字系统。计算机只认识二进制数, 即 0 和 1 表示的数。换句话说机器处理所有的数, 都要用二进制系统来表示; 构成计算机信息的所有字母、符号都要变换为计算机能够识别和处理的二进制编码(即 0 和 1 的特定组合)。

在电子计算机中, 输入的信息, 如符号、字母、数字等都要化成电子计算机能够识别的二进制数码, 这种表示信息的二进制数码就称为代码。在计算机中, 不仅信息用代码来表示, 对计算机进行操作的指令, 存贮的地址也都用代码来表示。

计算机中, 表示数目的代码有原码、补码和反码三种。

由于计算机仅使用两个数字(0 和 1), 因此, 二进制的“位”是计算机的最小信息单位。但是, 当中央处理器向存贮器传送或从存贮器取出信息时, 不是用单个的“位”, 而是用“字节”(byte)和“字”(word)等较大的信息单位来工作。

在计算机中, 作为一个整体来处理或运算的一串数码, 称为一个计算机字, 简称字, 它是由一定数的位组成。在计算机的运算器、控制器中, 通常都是以此为单位进行传送的。字出现在不同的地方, 其含义是不相同的。例如, 送往控制器去的字是指令, 而送往运算器去的字就是一个数。

每个计算机字可以代表一个十进制数码; 三个十进制数; 两个字母或几个特殊的符号; 或一个英文单字。每个字所包含的位数称为字长。根据计算机的不同, 字长有固定的和可变的两种。固定字长, 即字长度不论什么情况都是固定不变的。可变字长, 则在一定范围内其长度是可变的。

计算机字长, 是指它一次可处理的二进制数字的数目。计算机处理数据的速率, 自然和它一次加工的位数以及进行运算的快慢有关。如果一台计算机的字长是另一台计算机的两倍, 即使两台计算机的速度相同, 在相同的时间内, 前者所做的工作是后者的两倍。

一个计算机字所具有的位数取决于所用计算机的设计和规模。一般地, 大型计算机的字长为 32-64 位, 小型计算机为 12-32 位, 而微型计算机为 4-16 位。字长是衡量计算机

性能的一个重要因素。

计算机字通常分为若干字节。每个字节一般是八位。字节象字一样也是由一定数的位组成，是构成信息的一个小单位，并作为一个整体来参加操作。当计算机字包含大量信息位时，使用字节作为处理信息的单位比较方便。

在微型计算机中，通常用字节的多少来表示存储器的存储量。

数在机器中的表示形式称为机器数。机器数用具有两个状态的寄存单元(如触发器、磁心等)来表示。规定这些寄存单元的两个状态分别对应二进制数码的“0”和“1”。

计算机进行运算时，必须按一定方法确定小数点的位置，针对小数点的处理，机器数可分为定点机器数和浮点机器数两大类。与之对应的计算机，前者称为定点计算机，后者称为浮点机。此外，机器数还可分为原码、反码、补码等各种形式。

§ 2 微型计算机系统

一、微处理器(MPU)

半导体器件已经发展到大规模集成电路，把整个中央处理器(CPU)集成到一块或几块芯片上，这就是通常所说的微处理器(简称 MPU)或中央处理器(CPU)。一般微处理器功能组织中包括算术逻辑部件、若干个专用和通用寄存器以及对程序流进行控制的逻辑部件等。有的微处理器内部带有时钟发生器。

二、微型计算机

以微处理器为中心加上存储器和输入输出接口电路通过系统总线接口所组成的计算机，叫作微型计算机，基本结构如图 2 所示。它是一个独立的子系统，一般由少量的大规模集成电路组成。这些大规模集成电路可装在一块插件板上，甚至一个微型计算机就可以是一块芯片。

存储器包括两种功能的存储器

1. 随机存储器(RAM)，存放各种现场的输入或输出数据，中间计算结果，并与外存储器交换信息，信息可以随机存入和取出。

2. 只读存储器(ROM)，ROM 里的信息在使用时是不能改变的，即不能写入，只能读出。故一般用来存放固定的程序，如微型机的管理程序、监控程序、汇编程序等。

早期的 ROM 是由厂家按某种固定线路制造的，即所谓固化了的，造好后就只能读不能改变。为了便于用户根据自己的需要来写 ROM，就发展了一种称为可程序的只读存储器(PROM)，可由用户对它进行编写。但这种 ROM，用户只能写一次。

为了适应科研工作的需要，希望 ROM 能根据需要写，也希望能把已写上去的内容擦去，然后再写，能改写多次。这就产生了可擦去的可程序只读存储器(EPROM)，和电可擦去的可程序只读存储器(EEPROM)。但是它们写的时候需要一些额外条件，故使用时仍作为只读存储器来用。

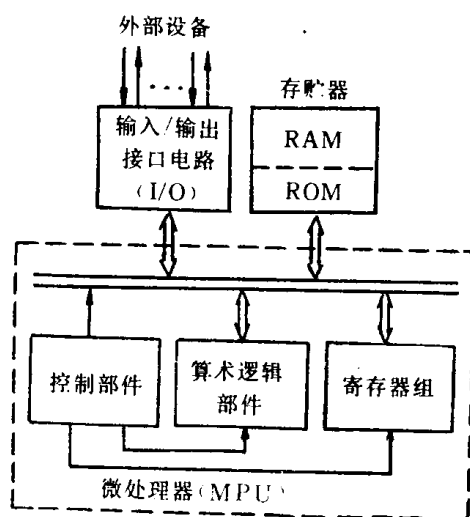


图 2 微型计算机基本结构

ROM 电路比 RAM 简单，故而集成度更高，成本更低。而且有一个重大优点就是当电源去掉以后，它的信息是不丢失的，而 RAM 则不然。

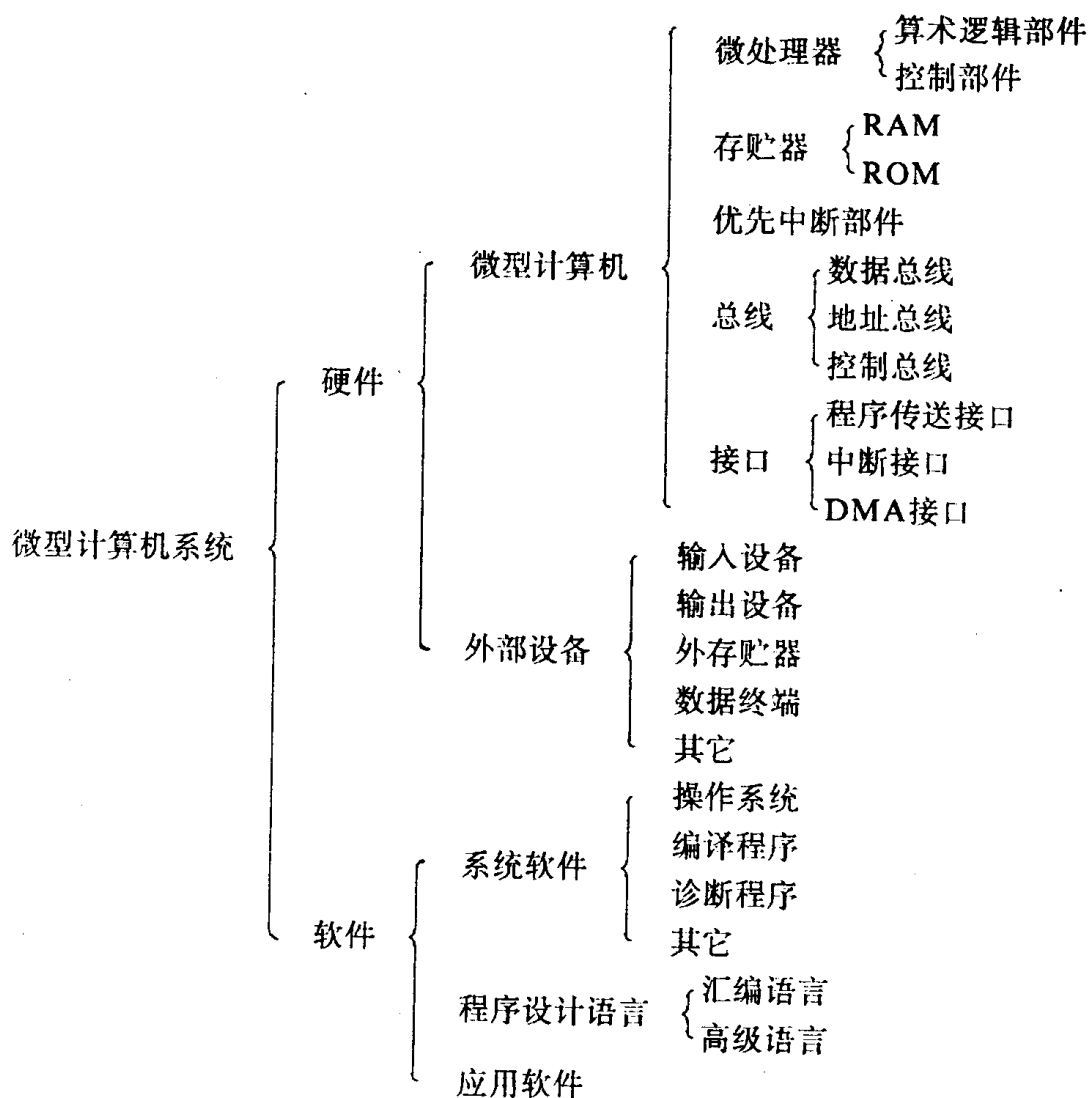
三、单板微型计算机

单板微型计算机就是把微型计算机的全部电路——CPU、ROM、RAM、I/O 接口电路以及其它一些辅助电路，集装在一块印刷电路板上，如 TP801 型单板机是以 8 位微处理器 Z80 为中心组装而成的 8 位单板机。

还有一类以计算机系统总线为基础的单板机，又称为单板(机)模块。主要为某些厂家提供插件产品，以便构成不同功能微型计算机系统。

四、微型计算机系统

一个完整的微型计算机系统包括硬件和软件两大部分。其具体组成如下：



五、程序设计语言

用计算机解题，必须先编好程序。既然要编程序，这就存在着以什么形式来表达程序的问题，说得更通俗些，就是用什么“语言”来表达程序的问题。计算机进行计算的过程就是执行一系列指令的过程，每个型号的计算机，都有自己规定的各种指令，指令的全体称

为指令系统。指令可以看作计算机的“语言”，这种语言称为机器语言。

在计算机问世的初期，是用机器指令来编写程序的。各种型号计算机的机器指令往往是互不相同的，而机器指令是很难记忆和掌握的，使用起来十分不方便。编写这种程序工作很费时，也很容易出错。将程序从一台计算机转到另一台计算机时，就要另编程序。编写这种程序的程序员要经过很长时间的训练。于是，很快就出现了以符号代替指令数码和存贮地址的“汇编语言”，比机器指令容易记忆。但汇编语言仍然不很直观，不易编写程序，也不易交流。

五十年代中期开始，人们开始发展高级程序设计语言。如 FORTRAN, ALGOL, COBOL, BASIC, PL / 1, FORTH, PASCAL, APL, SNOBOL, LISP, PL / M, PLZ, Ada, BLISS, PROLOG, C 语言等等相继出现。这些语言有一个共同特点，就是比较接近数学和人们的自然语言，对计算过程刻画比较清楚、直观。

BASIC 语言是微型计算机中最常用的语言。它的主要特点：①小巧灵活，简单易懂，使用方便。②具有会话性，形象地说，使用者可以与计算机进行交谈，因此称 BASIC 语言为人机对话语言。

BASIC 语言虽然简单但不贫乏。由于药学工作者大都使用微型机，因此，它是计算药学工作中目前最常用的语言。建议读者从有关专著中学习 BASIC 语言^[12]。

第二章 实验数据的统计处理

§1 样本的数字特征

在实际工作中，常常是从研究对象中抽出一部分样品进行试验，并从这些样品的试验结果来研究对象的整体的某个或某些性质。理论上，应该研究的全部对象称为“总体”，而从这些全部研究对象中抽出的一部分样品称为“样本”。事实上，只能从总体中抽出一部分样品作为直接试验的对象，以期推断和了解总体。统计学方法就是用于解决如何从样本来研究总体的问题。为此，先介绍几个常见的样本特征数。

一、平均数

平均数是表示一组观察值的平均水平和集中趋势的统计指标。平均数有多种，如算术均数、几何均数、中位数等。其中最常用的是算术均数。

算术均数在统计学中常称为“样本均数”，并记为 $\bar{x}$ 。对于一批数据 $x_1, x_2, \dots, x_n$ 有

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

有时候，为了减少计算，把数据 $x_1, x_2, \dots, x_n$ 按大小次序排列，用排在正中间的一个数表示总体的平均水平，称它为中位数。当 n 为奇数时，正中间的数只有一个；当 n 为偶数时，正中间的数有两个，在后一种情形，中位数等于这两个数的算术平均值。

二、极差

在实际工作中，只反映平均水平经常是不够的，例如，甲乙两人对某一样品进行分析，得到的测定值分别为：

甲：2.9、2.9、3.0、3.1、3.1

乙：2.7、2.8、3.0、3.2、3.3

它们测得的平均值均为3.0，但乙的测定结果波动性大，因此我们说，乙的分析测定水平比甲差，所以数据的波动大小也是一个重要的指标。如何度量波动的大小呢？一个简单的方法是用极差。极差是指一组测定中最大值和最小值之差，即

$$R = \max\{x_1, x_2, \dots, x_n\} - \min\{x_1, x_2, \dots, x_n\}$$

式中 $\max\{x_1, x_2, \dots, x_n\}$ 和 $\min\{x_1, x_2, \dots, x_n\}$ 分别表示测定值 $x_1, x_2, \dots, x_n$ 中最大值和最小值。

极差也叫范围，是表征观测值离散程度的各种统计特征数中最简单的一种。

三、标准差

由于极差没有充分利用数据提供的信息，因此反映实际情况的精密度较差，于是人们常用另一个统计量——样本标准偏差，简称标准差。

$$S = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}$$

其中 $\bar{x}$ 为均值， n 为测定次数。

标准差是一个表示测量精度的好方法，是反映一组观察值离散程度的统计指标，它不仅取决于测定中每一个数据，而且对一组测定中的极值反应比较灵敏。显然比极差反映问题精确；在均数及单位相同的条件下， S 越大，离散程度越大，反之亦然。以前面甲乙两

人的测定值为例, 前者的标准差为 0.10, 后者为 0.25, 说明甲的测定结果比乙好, 事实上也是如此。

标准差公式中的分子 $\sum (x - \bar{x})^2$ (注: 以后为了书写方便而又不致弄错的情况下, 将 $\sum_{i=1}^n (x_i - \bar{x})^2$ 简写为 $\sum (x - \bar{x})^2$) 是各测定值与均数的差值的平方的总和, 简记为 l_{xx}

$$l_{xx} = \sum (x - \bar{x})^2 = \sum (x - \bar{x})(x - \bar{x})$$

为了获得较好的数字效果, 根据数学推演, 由下式来计算 l_{xx}

$$l_{xx} = \sum x^2 - \frac{1}{n} (\sum x)^2$$

样本标准差的平方 S^2 称为样本方差, 也可用来表示离散程度的大小。

四、变异系数

当两组观察值单位不同, 或单位相同但均数相差较大时, 不能用标准差来比较两组间变异大小, 应将标准差化为不含单位的纯数, 称为变异系数或者叫做相对标准差, 通常用百分数表示:

$$CV = \frac{S}{\bar{x}} \times 100\%$$

式中 S 为标准差, $\bar{x}$ 为均数, CV 是变异系数。

例 1, 对一个气相色谱的实验人员进行技术考核。他连续注射样品共十次, 每次 0.5 μ l 量, 得色谱峰高分别为:

142.1 147.0 146.2 145.2 143.8 146.2 147.3 150.3 145.9 151.8(mm)

试通过计算峰高的变异系数, 对他的技术水平进行评价(有经验的色谱工作者很容易将变异系数控制在 1% 以内)。

解: $\bar{x} = (142.1 + 147.0 + \dots + 145.9 + 151.8) / 10 = 146.6(\text{mm})$

$$S = \sqrt{l_{xx} / (n - 1)}$$

$$= \sqrt{\frac{(142.1^2 + 147.0^2 + \dots + 151.8^2) - (142.1 + 147.0 + \dots + 151.8)^2}{10 - 1}}$$

$$= 2.83(\text{mm})$$

$$CV = \frac{S}{\bar{x}} \times 100\% = \frac{2.83}{146.6} \times 100\% = 1.93\%$$

经与 1% 比较, 可认为该实验人员进样技术还不够稳定, 不够熟练。

§ 2 t-检验在样品测试中的应用

一、平均值的置信界限

在药学工作中, 总是把测定数据的平均值作为结果来报告。但是平均值的可靠性是相对的, 仅仅报告一个平均值不能明确地说明测定的可靠性。因此, 人们期望使其值以指定的概率落在测定平均值附近的一个界限内, 这个界限称为置信界限(Confidence Limit), 简写为 CL。例如在元素分析中, 某样品中含碳量的测定, 报告为:

$$\%C(95\%CL) = 35.43 \pm 0.12$$

这个报告能比较确切地说明:①测定的平均值为 35.43%,②有 95%的把握认为样品含碳量落在 $35.43 \pm 0.12\%$ 这个范围内。这样说明问题是合理的,既不绝对化,而又明确。

上例中,要求 95%把握,这通常称为置信水平(Confidence Level),表示对可靠性要求的准则。在药学工作中常按 95%或 99%置信水平来要求。上面例子中的 $\pm 0.12\%$ 称为置信区间(Confidence Interval)。

小样本测定的平均值的置信界限的计算由下式给出:

95%置信界限:

$$\bar{x} \pm t_{0.05, f} S_{\bar{x}}$$

其中 $S_{\bar{x}}$ 为平均值的标准偏差

$$S_{\bar{x}} = \frac{S}{\sqrt{n}}$$

其中 S 为标准差, n 是测定次数。

99%置信界限:

$$\bar{x} \pm t_{0.01, f} S_{\bar{x}}$$

例 1 测定样品中某成分的含量,先测定两次,测得的百分含量为 1.12 和 1.15;再测定三次,测得的数据为 1.11, 1.16 和 1.12,试分别按两次测定和五次测定的数来表示平均值的置信界限。

解:两次测定时

$$\bar{x} = \frac{1.12 + 1.15}{2} = 1.135$$

$$S = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = 0.021$$

$$S_{\bar{x}} = \frac{S}{\sqrt{n}} = \frac{0.021}{\sqrt{2}} = 0.0148$$

$$f = n - 1 = 1, \quad t_{0.05, 1} = 12.706$$

$$\therefore \bar{x} \pm t_{0.05, 1} S_{\bar{x}} = 1.14 \pm 0.19$$

五次测定时

$$\bar{x} = \frac{1.12 + 1.15 + 1.11 + 1.16 + 1.12}{5} = 1.13$$

$$S = \sqrt{\frac{\sum(x - \bar{x})^2}{n - 1}} = 0.022$$

$$S_{\bar{x}} = 0.00984$$

$$f = n - 1 = 4, \quad t_{0.05, 4} = 2.776$$

$$\therefore \bar{x} \pm t_{0.05, 4} S_{\bar{x}} = 1.13 \pm 0.03$$

通过上面两种结果的比较,可以看出两次测定结果的置信区间要比五次测定结果的置信区间大。或者说,测定次数愈多,其置信界限愈小,平均值的可靠性愈高。

二、平均值与标准值的比较

因为标准样品的含量是已知的,可以将其名义值(或称标准值)视为真值,即视为总体均值,现在的任务是确定测定平均值和标准样品名义值是否一致。这时, t-检验计算公式