

统计质量控制实用指南

100 例

陈国铭 编著

中国石化出版社

图书在版编目(CIP)数据

统计质量控制实用指南 100 例/陈国铭编著.
—北京:中国石化出版社,2002
ISBN 7-80164-240-6

I. 统… II. 陈… III. ①数理统计-质量控制-指南
②电子表格系统, Excel-应用-数理统计-指南
IV. 0212-62

中国版本图书馆 CIP 数据核字(2002)第 031497 号

中国石化出版社出版发行

地址:北京市东城区安定门外大街 58 号

邮编:100011 电话:(010)84271850

<http://www.sinopec-press.com>

E-mail: press@sinopec.com.cn

北京精美实华图文制作中心排版

河北省徐水县印刷厂印刷

新华书店北京发行所经销

*

787×1092 毫米 16 开本 20 印张 506 千字 印 1—2000

2002 年 7 月第 1 版 2002 年 7 月第 1 次印刷

定价:45.00 元

前 言

数理统计技术是根据概率论原理和随机变量的分布规律收集、整理和分析来自社会和自然界各方面的受到随机性影响的数据,并进行统计推断,从而用数据为正确决策提供科学依据的一种应用数学技术。在质量管理的产品策划、市场研究、过程控制、质量保证和质量改进等诸方面都有实际应用价值。GB/T 19000 idt ISO 9000 - 2000 族质量管理标准中提出的八项质量管理原则之一就是基于事实的决策方法,要求有效的决策要建立在数据和信息分析的基础上,把统计技术的应用放在了很重要的位置。

综观世界质量管理历史,经过了检验质量管理、统计质量控制而发展到全面质量管理,而中国在推行全面质量管理过程中明显地逾越了统计质量控制阶段,使得数理统计技术没有引起足够的重视,再加上统计技术理论的深奥,使人望而却步。实际上,数理统计技术是一门应用数学,重在实际应用,而不是更多的理论推导。目前对其理论研究方面的书籍颇多,而对实际应用指导尚欠不足。为此,本书尝试应用实例解析的方法,补充实际应用指导之不足,主要从实际应用出发,只需懂得基本原理,省略复杂理论推导。并尽可能利用计算机现有程序进行解析,希望达到最佳效果。

本书的特点:一是应用实例解析,再用基本原理讨论,力图使读者能够独立进行实际操作,无需追究深奥理论;二是内容全面,包括了数据收集整理、抽样检验、检验和估计、控制图、方差分析、相关和回归分析、实验设计等主要方面的实际应用内容,不仅有作者自己编写的案例,还将收集的国内外许多应用实例重新进行了整理、编写,读者可以根据自己的需要有选择地研究;三是本书的实例计算几乎全部采用了计算机计算,主要是应用了 Microsoft Excel 工作表中有关统计的函数公式,并详细描述应用步骤,读者不仅可以了解数理统计的应用原理,还可以学会用 Microsoft Excel 工作表的解析方法,避免了许多复杂的手工计算。为了满足部分读者的要求,大部分实例保留了手工计算的步骤和方法,供需要时对比参考。

由于著者水平有限,书中难免有各种错误,其解题步骤也未必完全符合规定要求,仅供读者参考。书中有未尽之处,请提出批评指导。

编著者

2001 年 8 月于北京

目 录

一、数据的收集和整理

例 1	表示数据特征的基本参数	(1)
例 2	用计算机 Excel 工作表计算数据特征参数	(4)
例 3	求数据的数学期望值(1)	(7)
例 4	求数据的数学期望值(2)	(8)
例 5	方差的计算(1)	(8)
例 6	方差的计算(2)	(11)
例 7	排列图	(11)
例 8	直方图	(13)
例 9	相关图	(15)
例 10	分层法	(17)
例 11	饼图法	(19)
例 12	因果图(特性因素图)	(20)
例 13	已知总体标准偏差 σ 时的异常值判断	(21)
例 14	未知总体标准偏差 σ 时的异常值判断	(23)
例 15	不确定度	(24)

二、概率和数理统计基础

例 16	随机现象和随机事件	(31)
例 17	概率的确定方法之一：古典方法	(32)
例 18	概率的确定方法之一：统计方法	(34)
例 19	概率的乘法	(34)
例 20	连续型数值的概率分布：正态分布	(36)
例 21	离散型数值的概率分布：超几何分布	(38)
例 22	三种离散型数值的概率分布	(40)
例 23	统计量的分布： $\bar{x}$ 分布	(42)
例 24	统计量的分布： χ^2 分布	(44)
例 25	统计量的分布： t 分布	(46)
例 26	统计量的分布： F 分布	(47)
例 27	Microsoft Excel 工作表中的统计函数	(50)
例 28	用 Excel 工作表画离散型数值的分布图	(52)

三、估计和检验

例 29	一个正态总体均值的检验(u 检验)	(56)
例 30	两个正态总体均值差的检验	(59)
例 31	一个正态总体均值的检验(u 检验)及检出力的计算	(63)
例 32	标准偏差未知时正态总体均值的检验(t 检验)	(66)
例 33	两个正态总体方差的检验(F 检验)	(68)
例 34	正态总体方差检验(χ^2 检验 1)	(69)
例 35	正态总体方差的检验(χ^2 检验 2)	(70)
例 36	总体不合格品率的检验	(71)
例 37	总体不合格品率差的检验	(72)
例 38	秩和检验	(73)
例 39	总体平均值差的检验和估计(简易分析 1)	(75)
例 40	两个总体方差的差异的检验(简易分析 2)	(77)
例 41	两个总体均值的差异的检验(简易分析 3)	(78)
例 42	序位一致系数和 Friedman 检验(简易分析 4)	(79)

四、抽样检验

例 43	计数值抽样检验原理和抽样特性曲线(OC 曲线)	(81)
例 44	抽样分布在质量体系审核中的应用	(84)
例 45	保证连续批不合格率的逐批检查计数正常二次抽样检验方案	(88)
例 46	逐批检查计数抽样一次抽样方案	(91)
例 47	逐批检查计数抽样二次抽样方案	(93)
例 48	周期检查的计数抽样方案	(98)
例 49	孤立批的不合格品率的计数标准型一次抽样方案设计	(101)
例 50	计数型百分比抽样方案	(103)
例 51	规定下限的平均值的计量标准型一次抽样方案(1)	(108)
例 52	规定下限的平均值的计量标准型一次抽样检验方案(2)	(111)
例 53	平均值的计量标准型一次抽样方案	(112)
例 54	方差已知控制双侧的计量标准型一次抽样检验	(113)
例 55	σ 未知和已知的保证批不合格率的计量标准型一次抽样检验	(115)

五、方差分析

例 56	单因素方差分析一(重复试验次数相等)	(120)
------	--------------------	---------

例 57	单因素方差分析二(重复试验次数相等)	(124)
例 58	单因素方差分析(重复试验次数不等)	(127)
例 59	双因素方差分析(样本数相等无重复试验)	(130)
例 60	双因素方差分析(重复试验次数相等)	(135)
例 61	双因素方差分析(样本大小一样的计数值方差分析)	(139)
例 62	三因素方差分析(无重复试验)	(141)
例 63	三因素方差分析(重复试验次数相等)	(145)
例 64	误差分析	(153)

六、回归分析

例 65	一元回归分析(调整操作)	(160)
例 66	一元回归分析(控制)	(164)
例 67	一元回归分析(供方评价)	(167)
例 68	一元回归分析(过程质量控制)	(175)
例 69	一元回归分析(过程控制)	(178)
例 70	二元回归分析(产品质量分析)	(182)
例 71	二元回归分析(过程控制和误差来源分析)	(185)
例 72	二元回归分析(控制和预测)	(189)
例 73	回归系数检验	(193)

七、控制图

例 74	$\bar{x} - R_s$ 控制图	(199)
例 75	$\bar{\bar{x}} - R$ 控制图	(202)
例 76	$\bar{\bar{x}} - s$ 控制图	(210)
例 77	不合格品率控制图(p 控制图)	(214)
例 78	缺陷数控制图(C 控制图)	(217)
例 79	计数型累积和控制图(V 型模板)	(221)
例 80	计数型累积和控制图(固定判定界限)	(224)

八、试验计划法

例 81	拉丁方试验设计	(226)
例 82	三因素三水平正交试验(一)	(227)
例 83	三因素三水平正交试验(二)	(230)

例 84	三因素三水平正交试验(三)	(233)
例 85	四因素三水平正交试验(一)	(236)
例 86	四因素三水平正交试验(二)	(240)
例 87	四因素三水平正交试验(三)	(242)
例 88	五因素三水平正交试验	(245)
例 89	单指标四因素四位级正交试验	(254)
例 90	多指标四因素四位级正交试验	(257)

九、用 Excel 工作表制作分布数表

例 91	二项分布函数表及其制作	(262)
例 92	泊松分布函数表及其制作	(271)
例 93	标准正态分布密度函数表及其制作	(282)
例 94	标准正态分布函数表及其制作	(284)
例 95	标准正态分布单侧、双侧分位数表及其制作	(287)
例 96	t 分布函数表及其制作	(289)
例 97	t 分布单侧、双侧分位数表及其制作	(295)
例 98	χ^2 分布函数表及其制作	(298)
例 99	χ^2 分布上侧分位数表及其制作	(303)
例 100	F 分布分位数表及其制作	(306)

一、数据的收集和整理

在质量管理的策划、控制和改进活动中，需要大量的数据做依据，所谓“一切要用数据说话”，“基于事实的决策方法”，“有效决策是建立在数据和信息分析的基础上”就是这个道理。毛主席说，一切工作要做到“心中有数”，这个“数”也是这个道理。但是，人们从社会和自然界收集到的数据往往是杂乱无章的，甚至会有假数据，可见如何正确地收集数据、整理数据和分析数据是十分重要的。可以这样认为，不经科学地收集、整理和分析的数据是无用的数据，假数据比没有数据危害更大。掌握数理统计方法，首先要了解如何收集、整理和分析数据。

例 1. 表示数据特征的基本参数

某炼油厂测得 1997 年 11 月铂重整原料石脑油的干点数据如表 1-1。请计算最大值、最小值、众数、中值、平均值、样本标准偏差、总体标准偏差的估计值、样本方差、总体方差估计值、偏差平方和。

表 1-1 铂重整原料干点/℃

141	130	148	134	136	128	138	136	125
118	140	130	125	124	141	145	136	119
140	134	137	133	140	142	127	141	141
124	134	142	124	139	139	134	133	134
128	146	133	150	132	142	130	134	123
140	142	131	132	122	129	119	131	133
131	142	129	150	138	136	142	143	137
132	127	145	147	120	135	137	148	115
139	148	140	128	141	137	143	135	140
133	112	136	125	135	136	149	121	130

【解析】

可以分别用计算器和计算机方法进行计算，简述如下。

1. 计算器方法

凡带有统计功能的计算器(有 STAT 或 SD 功能)均可计算。首先按[STAT]或[SD]键，进入统计功能状态，将所有数据一个个输入，每输入一个，按一次[DATA]键，输完之后，计算完毕，可读取所有结果，就此例，结果如下：

按[n]键，读数为 90，即共输入 90 个数据，若不是，则应重新输入，直到正确为止；

按 $\bar{x}$ 键，读数为 134.57，即均值为 134.57；

按 $[\sigma_n]$ 或 $[s_n]$ 键,读数为 8.24,即计算样本的标准偏差为 8.24;

按 $[\sigma_{n-1}]$ 或 $[s_{n-1}]$ 键,读数为 8.29,即基于给定样本估算标准偏差为 8.29;

总体方差无偏估计 $s^2 = (s_{n-1})^2 = 8.29^2 = 68.72$

偏差平方和 $S = s^2 \times (n - 1) = 6116.1$

2. 用计算机方法

用 Excel 工作表很容易全部算出,步骤如下:

打开 Excel 工作表,将 90 个数据一一输入,在[插入]菜单中选用[函数],从[函数]下拉菜单函数分类中单击[统计],则出现一系列统计功能参数,从中选 MAX(最大值)、MIN(最小值)、MODE(众数)、MEDIAN(中值)、STDEVA(估算基于给定样本的标准偏差)、STDEVP(计算样本的总体标准偏差)、VAR(基于给定样本估算方差:无偏方差)、VARP(样本的总体方差)、DEVSQ(偏差平方和)、AVERAGE(平均值)即可直接算出。列入下表。

最大值	150	众数	136	偏差平方和	6116.1
最小值	112	中值	135	无偏方差	68.72
极 差	38	平均值	134.57	标准偏差	8.29

在用 Excel 工作表计算时,只要在要计算的单元格内写入上述公式即可求出,公式后边的括号内写入要计算的数据范围,详细计算方法见[例 2]。

【讨论】

数据在质量管理统计分析中有十分重要的作用。为了使用户得到满意的质量保证,生产者和经营者对产品质量特性和实现质量保证的活动必须做到心中有数。没有数据就无法进行定量分析,通过数据的整理、分析和判断,从中找出质量活动的规律,做出正确的质量决策,实现控制和提高产品质量的目的,因此可以说质量管理是基于数据的管理。

这是一组经实际测试某厂 1997 年 11 月份的铂重整原料的干点数据,显然,不加处理,只能是 90 个数,得不出什么结论。要处理,首先要研究数据的特征。在数理统计中通常要根据数据的波动性、规律性来解析这组数据。

1. 描述数据分布的中心位置的参数

描述数据分布的中心位置通常有几种表示方法,即众数(测定值中出现次数最多的一个数),用 M_0 表示;中位数(有序测定值中居中的一个数),用 $\tilde{x}$ 表示;算术平均值用 $\bar{x}$ 表示,即:

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} \quad (x_1, x_2, \cdots, x_n \text{ 为各次测定值, } n \text{ 为总次数 } 90)。$$

加权平均值用 $\bar{x}_f$ 表示,即:

$$\bar{x}_f = \frac{f_1 x_1 + f_2 x_2 + \cdots + f_k x_k}{\sum_{j=1}^k f_i} = \frac{\sum_{i=1}^k f_i x_i}{\sum_{j=1}^k f_i}$$

式中 $x_1, x_2, \cdots, x_k$ ——表示第 1, 2, $\cdots, k$ 个数据;

f_i ——表示 x_i 数据出现的频率。

本例中：

众数： $M_0 = 132, 136, 142$ ，（此数据若按众数的定义有三个都是同等多的数，众数 M_0 取三个数的中间值为 136）。

中位数： $\tilde{x} = 135$

算术平均值： $\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = 134.57$

加权平均值：

为简化加权平均值的计算，可分组列于表 1-2 中（分组原则见例 8）。

表 1-2 铂重整原料干点的频数分布表

组 界	组中值	频数 f_i	频率 f_i/n	累积频率 $F/\%$	组 界	组中值	频数 f_i	频率 f_i/n	累积频率 $F/\%$
111.5 ~ 116.5	114	2	0.02	2	131.5 ~ 136.5	134	23	0.26	58
116.5 ~ 121.5	119	5	0.06	8	136.5 ~ 141.5	139	20	0.22	80
121.5 ~ 126.5	124	8	0.09	17	141.5 ~ 146.5	144	11	0.12	92
126.5 ~ 131.5	129	14	0.15	32	146.5 ~ 151.5	149	7	0.08	100

$$\begin{aligned}\bar{x}_f &= \frac{114 \times 2 + 119 \times 5 + 124 \times 8 + 129 \times 14 + 134 \times 23 + 139 \times 20 + 144 \times 11 + 149 \times 7}{90} \\ &= 134.56\end{aligned}$$

由上述计算方式可知，在表示数据集中程度（或中心位置）的特性值中，算术平均值和加权平均值最为准确。而众数和中位数误差比较大，但计算容易，投入成本低，在某些情况下还是有适用价值的。

2. 数据的离散程度

数据的离散程度是反映数据波动的另一个特性值。产品质量数据离散程度的表示方法也有好几种。由简而繁排列有：极差、偏差和、偏差平方和、方差和标准偏差。铂重整原料的干点数据的离散程度用不同方法表示可计算如下：

极差 一组数据中最大值（ $x_{\max}$ ）与最小值（ $x_{\min}$ ）之差，记作 R 。

本例中 $R = x_{\max} - x_{\min} = 150 - 112 = 38$

虽然上述数据表面看起来杂乱无章，经分析，其均值为 134.57，且波动范围在 112 ~ 150 之间。极差 38。

极差是表示数据的离散程度最简单的方法，但是极差不能说明全部数据的分布情况。例如，1, 3, 5, 7, 9 和 1, 1, 8, 8, 9 两组数据最大值（ $x_{\max}$ ）与最小值（ $x_{\min}$ ）全一样，极差也一样，但各个数据的分布情况却不一样，于是引出表示离散程度的第二个概念：偏差和。

偏差 测定值与均值之间的差异，即 $D = x_i - \bar{x}$ 。

偏差和 所有测定值与均值之间的偏差的和，即：

$$(x_1 - \bar{x}) + (x_2 - \bar{x}) + \cdots + (x_n - \bar{x}) = (x_1 + x_2 + \cdots + x_n) - n\bar{x} = 0$$

显然，样本偏差之和不能反映数据的离散程度。若取其绝对值，可以算出平均偏差，即：绝对偏差的平均值。

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

由于引进了绝对值,计算不方便,很少应用。

表示离散程度的第三个概念是:偏差平方和。这样就可消除偏差有正负之分,使之成为非负性。以 S 表示。

偏差平方和

$$S = \sum_{i=1}^n (x_i - \bar{x})^2 \quad \text{在本例中 } S = 6116$$

显然,90 个数据,一个个计算,是非常复杂的,用计算器和计算机比较方便。由于偏差平方和引进了平方的计算,使得数据很大。表示离散程度的第四个概念,即平均偏差平方和。

平均偏差平方和 通常称为方差,用 s^2 表示。

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad \text{本例中 } s^2 = 68.72$$

方差的计算和实际数据还有一个平方的差异,在数值上难以对比衡量,因此又引进了表示离散程度的第五个概念,也就是经常使用的一个概念:标准偏差。

标准偏差 方差的平方根或称根方差,用 s 来表示。

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}} \quad \text{在本例中 } s = 8.29$$

这样,表示这组数据的一个基本概念比较清晰了,虽然看起来是一组杂乱无章数据,但它也体现了数据的波动性和规律性,波动范围在 112~150 之间,均值是 134.57。

例 2. 用计算机 Excel 工作表计算数据特征参数

请用计算机 Excel 工作表计算某油田原油含水抽样样本数据的最大值、最小值、平均值、众数、中值、偏差平方和、样本方差、无偏方差、样本标准偏差、估算的总体标准偏差。

共随机抽取了 20 个罐次,其含水量分别是 0.11, 0.10, 0.09, 0.15, 0.14, 0.12, 0.11, 0.13, 0.10, 0.10, 0.15, 0.13, 0.10, 0.14, 0.11, 0.16, 0.09, 0.08, 0.11, 0.13 (单位:%)。

【解析】

(1) 打开计算机,进入 Excel 工作表,将 20 个数据一一输入,并将所要计算的项目列入,见图 2-1。

(2) 先计算最大值。用鼠标选定准备写入最大值的 C8 处,点[插入]下拉菜单,并点[函数(F)],见图 2-1。

(3) 点击[函数(F)],将出现一菜单,见图 2-2。在[函数分类]栏中选[统计]项,然后

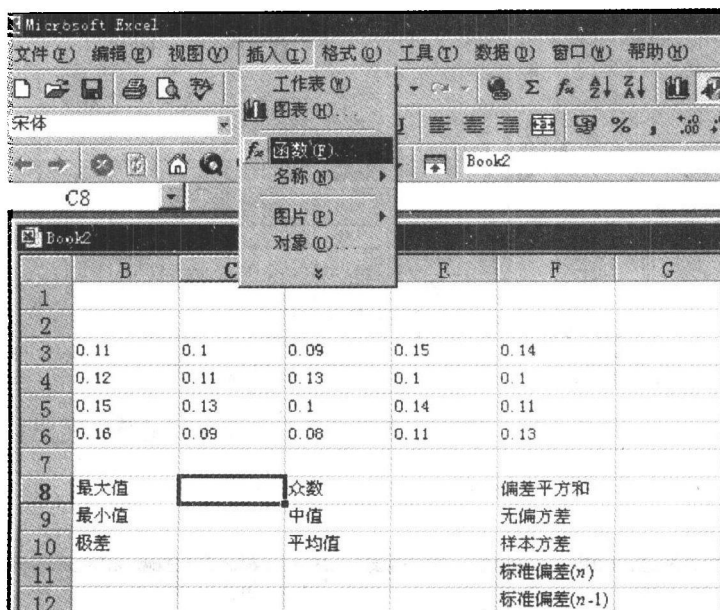


图 2-1 用 Excel 工作表计算数据特征的参数图

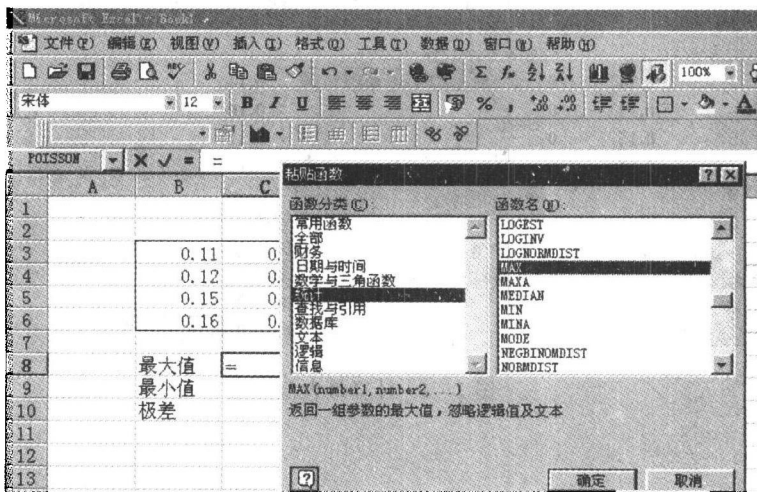


图 2-2 选最大值 MAX

在[函数名]栏中选[MAX]项。

(4) 点击[确定]键，在 Number1 中填入表示所列入统计数据范围 B3: F6，见图 2-3。

(5) 按确定键，则可计算出最大值为 0.16，见图 2-4。

以此方法，分别在 C9、E8 ~ E10、G8 ~ G12 位置写入 = MIN (B3: F6)、= MODE (B3: F6)、= MEDIAN(B3: F6)、= AVERAGE (B3: F6)、= DEVSQ (B3: F6)、= VARP (B3: F6)、= VAR (B3: F6)、= STDEVP (B3: F6)、= STDEVA (B3: F6)，则可分别计算出最小值、众数、中值、平均值、偏差平方和、无偏方差、方差、标准偏差(n)和标准偏差 ($n-1$) 的值。得出如下数据 见图 2-5(单位为: %)。

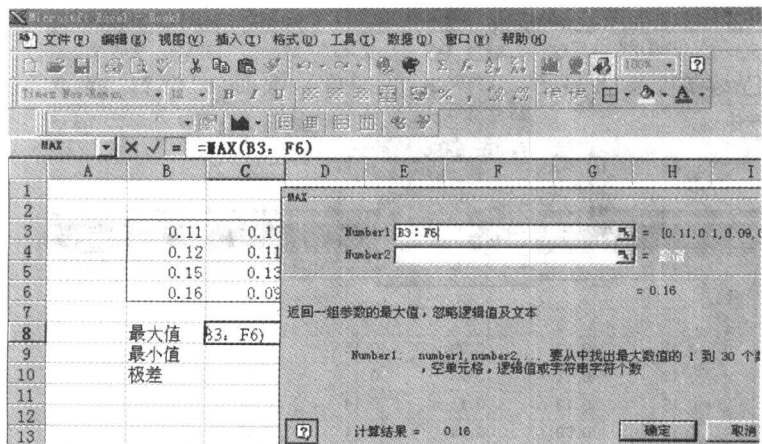


图 2-3 确定数据范围

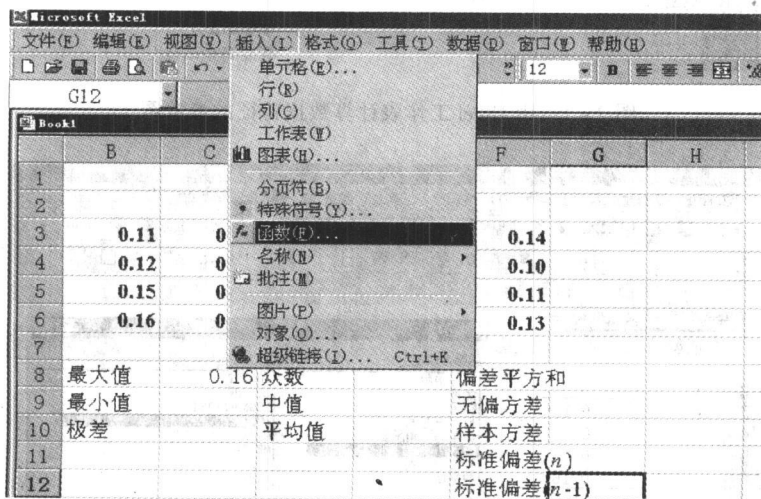


图 2-4 计算最大值

	A	B	C	D	E	F	G
1							
2							
3		0.11	0.1	0.09	0.15	0.14	
4		0.12	0.11	0.13	0.1	0.1	
5		0.15	0.13	0.1	0.14	0.11	
6		0.16	0.09	0.08	0.11	0.13	
7							
8		最大值	0.16	众数	0.11	偏差平方和	0.00978
9		最小值	0.08	中值	0.11	无偏方差	0.00051
10		极差	0.08	平均值	0.12	样本方差	0.00049
11						标准偏差(n)	0.022108
12						标准偏差(n-1)	0.022682

图 2-5 计算所有参数

不论用什么方法计算,将数据全部输入是必不可少的,这是计算准确性的基础,来不得半点马虎和偷懒。

【讨论】

计算机 Excel 工作表,几乎包括了所有数理统计简单的计算功能,既保证了计算的准确性,又提高了计算效率。本书所有例题主要依靠 Excel 工作表,主要图表也依靠 Excel 工作表,在以后的例题中还会遇到,望读者在学习数理统计方法的同时掌握计算机 Excel 工作表的简单应用,会提高对计算机应用的兴趣。

例 3 求数据的数学期望值(1)

日本东京都 1962 年 5 月 5 日发表的东京都彩票把 110000 号到 199999 号的 9 万张作为一组,共 3 组(27 万张)被售出,每张 100 日元,得奖规定如表 3-1 所示。

表 3-1 彩票获奖规定

等 级	奖金/(日元)	获奖个数	获奖概率
一等	3000000	1	1/270000
前后号奖	250000	2	2/270000
一等号但组别不同	50000	2	2/270000
和一等号差 1 个字	10000	132	132/270000
二等奖	100000	1	1/270000
三等奖	5000	54	54/270000
四等奖	1000	810	810/270000
五等奖	100	8100	8100/270000
六等奖	50	81000	81000/270000

请问每张彩票奖金 x 的期望值是多少。

【解析】

期望值就是总体均值,用 $E(x)$ 表示。

$$\begin{aligned}
 E(x) &= (3000000 \times 1 + 250000 \times 2 + 50000 \times 2 + 10000 \times 132 + 100000 \times 1 + \\
 &\quad 5000 \times 54 + 1000 \times 810 + 100 \times 8100 + 50 \times 81000) / 270000 \\
 &= (300 + 50 + 10 + 132 + 10 + 27 + 81 + 81 + 405) / 27 \\
 &= 1096 / 27 \\
 &= 40.59 \text{ (日元)}
 \end{aligned}$$

每张彩票奖金的期望值是 40.59 日元。

【讨论】

买彩票各国都有,每个国家买卖彩票的目的不同,但都是为了集资。改革开放以来,中国发行了福利彩票、体育彩票,也是为了集资。每个人买彩票的目的不同,大多数是为了捐献,也有的为了投资,希望有回报。每一个投资的人应该明白,投资是有风险的。设计彩票奖金额肯定不会赔钱,获得奖金的概率很小。如本题获 300 万日元奖金的概率只有 27 万分之一,平均每张彩票的回报是 40.59 日元,彩票发行公司每张彩票获利 59.31 日元。

例 4 求数据的数学期望值(2)

有一石油勘探区, 钻一口井完井费用是 10 万美元, 打一口干井耗费 8 万美元, 根据以往地质勘探统计资料预测显示, 在这个探区钻井的结果, 成功的可能性及利润情况汇总见表 4-1。

表 4-1 钻井利润

可能结果	概 率	纯利润/(万美元)	可能结果	概 率	纯利润/(万美元)
干井	0.30	-8	30 万桶储量	0.10	25.0
10 万桶储量	0.30	2.5	40 万桶储量	0.10	35.0
20 万桶储量	0.20	15.0			

求其钻一口井货币期望值。

【解析】

设纯利润金额为 x , 则:

$$E(x) = (-8) \times 0.30 + 2.5 \times 0.3 + 15.0 \times 0.20 + 25.0 \times 0.1 + 35.0 \times 0.1 = 7.35 \text{ 万美元}$$

由此可知, 如果不打井, 货币期望值为 0; 打井, 每打一口井货币期望值为 7.35 万美元。

【讨论】

在石油勘探中, 商业风险大小取决于对油田地质的评价和经济评价。首先要分析地质资料, 看是否有生油和储油条件, 再根据以往地质经验分析其勘探成功的概率。能发现有油不一定有商业价值, 还要看其储量大小, 预测今后原油价格趋势等, 才能分析是否有开采价值。因此在油田开发经济评价中, 不可忽视概率分析, 以提高石油勘探开发的抗风险能力。

例 5 方差的计算(1)

某服装厂承包一批工作服任务, 按合同规定, 所需布料由甲方提供。甲方提供的布料声称每一卷可制作 10 套工作服。服装厂为了调查甲方提供的布料长度是否可信, 对甲方提供的布料长度进行了测试调查。测试结果是, 一卷布料的长度 x 不是完全一样的, 有长有短, 但 x 服从正态分布; 由于每套工作服是量体裁衣, 制作 1 套工作服所需布料长度 y 也是不一样的, y 也服从正态分布, 其总体均值 μ 和总体标准偏差如表 5-1 所示。

表 5-1 总体均值与总体标准偏差

	μ/m	σ/m
1 卷布料的长度 x	30.5	0.50
一套工作服需布长 y	3.0	0.20

问题: 裁剪 9 套工作服时余下的布料长度是多少?

裁剪 9 套工作服时余下的布料可以制作第 10 套工作服的可能性是多少?

【解析】

这是一个理解正态分布及如何进行概率运算的问题。

1. 第一个问题的解题步骤

步骤 1 计算做 9 套工作服所需布料长度的期望值：

令 y_i 为第 i 个人的工作服需布料长度，那么，做 9 套工作服所需布料长度为：

$$y_1 + y_2 + \cdots + y_9 = \sum_{i=1}^9 y_i$$

期望值用 E 表示，其期望值为：

$$\begin{aligned} E\left(\sum_{i=1}^9 y_i\right) &= E(y_1 + y_2 + \cdots + y_9) \\ &= E(y_1) + E(y_2) + \cdots + E(y_9) \\ &= 9 \times 3.0 = 27.0(\text{m}) \end{aligned}$$

即做 9 套工作服所需布料长度的期望值为 27.0m。

步骤 2 计算做 9 套工作服所需布料长度的方差

$$\begin{aligned} s^2\left(\sum_{i=1}^9 y_i\right) &= s^2(y_1 + \cdots + y_9) \\ &= s^2(y_1) + s^2(y_2) + \cdots + s^2(y_9) \\ &= 9 \times 0.20^2 = 0.36 = 0.60^2(\text{m}^2) \end{aligned}$$

即做 9 套工作服所需布料长度的方差为 0.60m^2 。

步骤 3 计算做 9 套工作服后余下布料长度的期望值

根据已知，设 1 卷布料长度为 x ，期望值 $E(x) = 30.5\text{m}$ ；方差 $s^2(x) = 0.5^2 = 0.25\text{m}^2$ 。设制作了 9 套工作服余下的布料为 d ，则：

$$d = x - \sum_{i=1}^9 y_i$$

做 9 套工作服后余下布料长度的期望值为：

$$\begin{aligned} E(d) &= E\left(x - \sum_{i=1}^9 y_i\right) \\ &= E(x) - E(y_1 + y_2 + \cdots + y_9) \\ &= 30.5 - 27.0 = 3.5(\text{m}^2) \end{aligned}$$

步骤 4 计算做 9 套工作服后余下布料长度的方差

$$\begin{aligned} s^2\left(x - \sum_{i=1}^9 y_i\right) &= s^2(x) + s^2(y_1 + \cdots + y_9) \\ &= 0.25 + 0.36 = 0.61 = 0.781^2(\text{m}^2) \end{aligned}$$

由此可知，制作 9 套工作服余下布料长度的期望值为 3.5m，方差为 0.781^2m^2 。

2. 第二个问题的解题步骤

步骤 1 计算裁第 10 套工作服后所余布料的期望值和方差

裁剪 9 套工作服后余下布料长度为 d ，根据第一个问题可知：

$$E(d) = 3.5\text{m};$$

$$s^2(d) = 0.61 = 0.781^2\text{m}^2$$

因为第 10 套工作服所用布料的期望值是 3.0m，方差是 0.20m^2

所以裁第 10 套工作服后所余布料的期望值和方差分别是：

$$\begin{aligned} E(d - y_{10}) &= E(d) - E(y_{10}) \\ &= 3.5 - 3.0 = 0.5(\text{m}) \\ s^2(d - y_{10}) &= s^2(d) + s^2(y_{10}) \\ &= 0.61 + 0.04 = 0.65 = 0.806(\text{m}^2) \end{aligned}$$

步骤2 计算裁第 10 套工作服的可能性

根据题意，裁剪完 10 套工作服余下的布料长度应大于 0，即可裁第 10 套工作服。否则，

余下布料不可能再裁第 10 套了。余下布料长度 g 服从 $N(0.5, 0.806^2)$ 的正态分布(如图 5-1)，斜线部分为制作第 10 套工作服的概率。 g 服从 $N(0.5, 0.806^2)$ 正态分布，因

此统计量 $u_0 = \frac{g - 0.5}{\sqrt{0.806^2}}$ 服从 $N(0, 1^2)$ 标准正态分布。当 $g \geq 0$ 时， $u_0 \geq \frac{g - 0.5}{\sqrt{0.806^2}} = 0.620$ ，查正态分布表可知 $u_0 \geq$

-0.620 ；概率为： $1 - 0.268 = 0.732$ ，即可以制作第 10 套

工作服的概率是 73.2%。

【讨论】

(1) 本题计算很简单，但理论比较难以理解，题中所使用的参数均指总体的概念，是许多卷布测试统计的结果，不是指一、二卷布，因此要从整体考虑。在数理统计参数的表示方法上，通常总体均值用 μ 表示，总体方差用 σ^2 表示。而样本均值用 $\bar{x}$ 表示，样本方差用 s^2 表示。

(2) 要有分布的概念，如果简单从题意理解，既然一卷布是 30.5m，而做一套衣服要 3m，那么做 10 套肯定没问题。实际上 30.5m 和 3m 仅仅是一个均值，他们服从一定的分布，一卷布的长度偏差的范围如果按 $\pm 3\sigma$ 计，应为 $30.5 \pm 1.5\text{m}$ ，根据题意，一卷布最短可能是 $30.5 - 3\sigma = 29\text{m}$ ，最长可能是 $30.5 + 3\sigma = 32\text{m}$ 。每套工作服所需布料长度也不一样，如果做 10 套工作服的人都是篮球运动员，全部是高个子，需要的布料是每套 $3.0 + 3\sigma = 3.6\text{m}$ ，10 套即为 36m，显然一卷布不够，所以本题计算结果表示，做第 10 套的可能性不是 100%。

(3) 要注意方差的计算和期望值的计算不同，当求差的期望值和方差时，前者是用期望值的差来计算；而后者是用方差的和来计算。这是本题计算正确与错误的关键。

$$\begin{aligned} E(x - \sum_{i=1}^9 y_i) &= E(x) - E(y_1 + y_2 + \cdots + y_9) \\ s^2(x - \sum_{i=1}^9 y_i) &= s^2(x) + s^2(y_1 + \cdots + y_9) \\ E(d - y_{10}) &= E(d) - E(y_{10}) \\ s^2(d - y_{10}) &= s^2(d) + s^2(y_{10}) \end{aligned}$$

期望值(均值)的计算易于理解。对方差的计算，我们可以这样理解。一卷布的长度偏差的范围按 $\pm 3\sigma$ 计，应为 $30.5 \pm 1.5\text{m}$ ，做 9 套工作服所需布料长度的偏差的范围按 $\pm 3\sigma$ 计，