QC入门讲座(9)

统计检验和估计

〔日〕鐵　健司 编　谷津　進 著
王　瑞　坤　邸宏　译

中国经济出版社

QC入门讲座（9）

统计检验和估计

〔日〕 鐵 健司 编
谷津 進 著
王瑞坤 郧宏 译

中国经济出版社

内 容 简 介

本书是《QC入门讲座》之九，共分六章。阐述了数据分析的基本方法，对各种统计检验和估计分别进行了讲述，并概要叙述了统计估计理论，最后介绍了相关与回归的基础内容。

本书内容是质量管理方法中最难理解的部分，作者从实用性出发，力图通俗。每章后都附有研究课题，书后并有答案。可作为质量管理入门教材使用，也可供自学者参考。

责任编辑：肖玉平

统计检验和估计

〔日〕 铁健司 编
谷津进 著
王瑞坤 邸宏 译

*

中国经济出版社出版
（北京市翠微路22号）
北京京辉印刷厂印刷
新华书店北京发行所发行
各地新华书店经售

*

787×1092毫米 32开本 5 4/32印张 112千字
1987年9月第1版 1987年9月第1次印刷
印数：00,001—13,000
ISBN 7-5017-0002-8/F·115
统一书号：4395·115 定价：1.15元

前　言

现在，引进和推行全面质量管理的企业在不断增加，与之同时，数据分析事例与之成比例地增加。但是，从应用情况看，尚不能说有效地和熟练地掌握了统计分析方法。其中不正确地应用统计方法，甚至连进行分析的目的都不清楚的还不少。

本书讲的估计和检验是质量管理方法中最难理解的部分，对这部分内容许多人都持敬而远之的态度。但是，在分层的数据量不太多或想要掌握影响特性的主要原因的情况下，这种统计估计和检验的方法都是有效的，而且，作为实验计划法等其它统计方法的基础，也是一个重要的领域。

在应用检验和估计时，并不是通过机械计算得出结论就完事了，重要的是掌握计算过程所具有的具体意义和解释计算结果的物理意义。因此，本书不仅讲述计算程序，而且把重点放在计算的物理意义的解释和理论的直观性的解释上。

第一章虽讲的是数据分析的基本方法，但请您不要轻视这一章，要仔细地阅读。本章用各种事例说明运用数据说话的重要性和伴随数据而产生的离散是怎样一回事。

第二和第三章讲述的是关于计量值数据的估计和检验。这里，力图避免那些生硬的理论方面的内容，尽量采用以公式形式讲解的分析步骤，希望大家能理解每步前后内容。

第四章只概要叙述统计估计的理论，阅读有困难的读者可以跳过这一章，不阅读也可以。

第五章讲计数值的估计和检验。跳过第四章的读者可以只利用公式的分析步骤，读过第四章的读者会更容易理解第五章的理论根据。关于计数值分析常用的二项概率纸，因为只限于必要的最低限度说明，所以请参考本讲座之八——《工序分析的简易方法》。

本书主要是用分层数据的总体比较法来说明估计和检验的，因此，从全书系统性来看，第六章的相关和回归也许是有些脱节的题目。因为这一章也只限于讲一些基础内容，所以想要深入学习的读者，请阅读其它专著。

本书能够整理出版，多亏木暮正夫、真壁肇、三浦新、佐佐木修和赤尾洋二诸先生的帮助。借此机会，对从学生时代至今给予直接指导的各位先生深表感谢。

关于统计质量管理方法的实际应用，是在各公司和工厂的实际问题讨论会中学到的，这里对有关者致以谢意。

著　　者

1983年10月

目　录

1. 统计数据的分析方法

在质量管理中，一切都要根据事实来说话，这是非常重要的。要正确地掌握事实并作出客观的判断，以数据为依据的统计方法是不可缺少的。为了解决已暴露出来的问题或通过数据所发现的不合理情况，往往首先采取画特性要因图的方法。有人认为只要画出特性要因图，就可以了解所存在问题的主要原因，结果造成不去确认事实真相，就盲目地根据KKD（经验、第六感、气魄）来采取对策的现象是屡见不鲜的。这种态度是毫无意义的，我们所应取的态度是用数据来确认和判断出特性要因图上反映特性的原因是不是真正的原因。

本书是通过各要因条件的数据进行分层，说明取数据的重要性和分析数据的基本要领，然后再说明在统计上如何掌握所提出的要因对特性产生的影响。为慎重起见，本章反面事例中提到的人物、企业以及名称都是虚构的。

1.1 根据数据作判断

东京有一条叫做山手线的环形电车线。在一般人看来，可能认为正因为是环形线，所以电车必然环形运行，但实际情况果真如此吗？经仔细调查才发现，山手线是以品川为起点，经由新宿，至田端为终点的。并不是环形线，只是半环形。而另一半环形线是怎么运行的呢？通过图1.1可知，

从田端到东京是借用东北干线；从东京到品川是借用东海道干线，是这两个半环形结合起来才构成环形线的。实际上，表示线路起点的标志——○在上野一个也找不到，即使是东北干线，在注册上也是以东京为起点的。这真是“岁月和数据，不算不知道，一算吓一跳”啊！

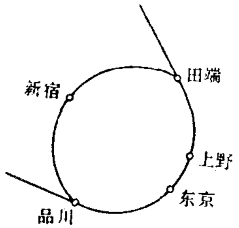


图 1.1　所谓的山手线

在质量管理中，在判断原材料、产品、批量、工序状态或作业方法等实际上处于怎样的状态时，不能凭想当然来下结论说“必然是这样的！”也不能以经验主义的态度来下结论说“从经验来看应该是这样的。”重要的是要用数据来说话。根据数据客观地掌握事实，并在此基础上作出恰当的判断。

【事例1.1】

某化学公司，产品收率波动很大，但无论如何也找不出其影响的原因。A科长大声向操作者问：“结果之所以有波动，是因为原因有离散。快把反应条件中的产生离散的地方给我查出来！例如反应温度是不是一定？”接着操作者回答A科长说：“反应温度是由计算机控制的，是绝对不会有问题的”。

不久，因为产品收率严重下降，甚至达到不得不停止生产的程度。因此，刚才说“没问题”的那位操作者取了各种数据，并且以抓根救命稻草的心情，安装了计量仪器，并测量了反应温度，结果意外地发现反应温度并不稳定。因此从对策出发安装了各种调节器，尽量使温度保持稳定，这样一来，产品收率稳定下来了。真是“数据胜于雄辩”。

1.2 数据的分布

认识到了用数据来说话的重要性，但如果不知道所取的数据的离散情况，那么这种数据即使取到了也是毫无意义的。

【事例1.2】

某建筑公司，围绕如何提高瓷砖的粘结强度这一课题进行攻关，他们不是单纯凭过去的经验，而是依据数据来进行判断解决。他们用各种方法对粘结着的瓷砖进行剥离，借以测出粘结强度，并将所测出的强度用数据记录下来，结果发现了比以前粘结强度更高的粘贴方法。但自从采用这种新方法进行粘贴瓷砖后，索赔问题倒比以前增加了。这究竟是什么原因呢？

这是因为没有考虑数据的离散而遭到失败的例子。该公司后来又取一些数据并整理得到如图 1.3 所示的结果。通过图可以看到，从平均来看，还是以前的粘贴方法粘结强度高，只是因为试验时是用偶而得到的一个新粘贴方法的数据

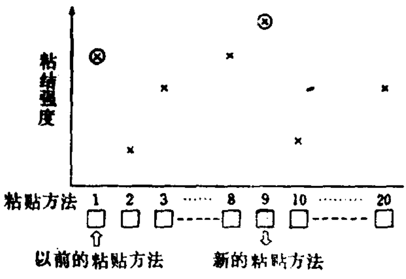


图 1.2 砖的粘结强度

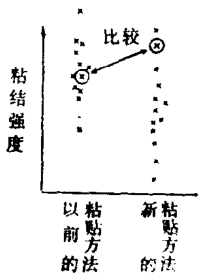


图 1.3 粘结强度的比较

作比较，所以得到了相反的结果。

在判断质量特性是否不同和判断究竟哪一特性好时，不能忽视数据中可能存在的离散现象，不能简单地相信表现在个别数据上的结果，要弄清比较结果的数据是否经常发生离散，或弄清比较结果是否就是如数据上所显示的结果，这是检验和估计的任务之一。

1.2.1 正态分布

当知道数据存在离散时，还要弄清该数据是以怎样的分布形态离散的。象产品尺寸、重量和强度这些数据，是以连续量得到的，这种情况下有一种典型分布即正态分布，其形态如图 1.4 所示，呈左右对称的“山形”。因同山形各处高度成正比，所以正态分布值很容易得出来。例如，设x_1处的高度为a，x_2处的高度为2a，则x_2附近值出现的比率（概率）比x_1附近值出现的概率要大 1 倍。因此，山顶处μ附近的值最容易出现，称μ附近的值为总体的平均值或总体平均。

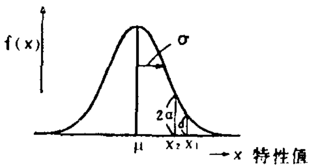


图 1.4 正态分布

这样作为连续量得到的数据，往往以同正态分布的山的高度成正比的概率来取总体平均值μ附近的各种值。这时，表示在平均值附近的离散程度的是总体标准离差或总标准偏差，以σ标记。把标准偏差平方即σ^2称为方差或总体方差。

往往有这种倾向，即总体平均越大，由此得到的特性数据也变大。若总体方差大则在平均值附近就会得到离散程度大的数据；相反，若总体方差小则在总体平均附近就会得到集中的值。

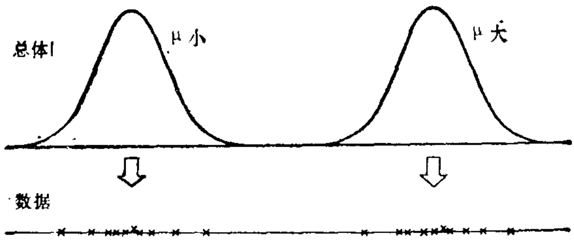


图 1.5　从不同平均值的总体中得到的数据

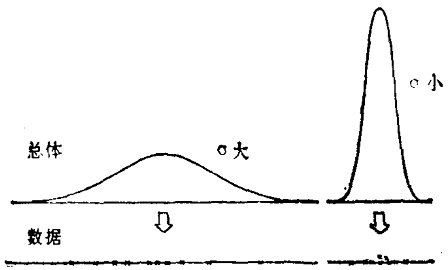


图 1.6　从不同标准偏差的总体中得到的数据

下面为供参考，将表示出现在 x 附近的值的概率的山的高度用下式来表示：

$$f(x)=\frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad (1.1)$$

式中，π 是圆周率（3.141…），e 是自然对数的底（2.718…）。

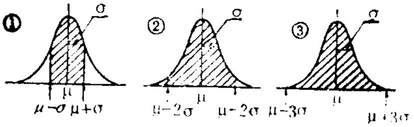
这样，在用正态分布图表示离散规律时，只要知道总体平均μ和标准偏差σ 的差，就能预计多大的数据将以怎样的概率出现。关于数据将在平均值附近的多大范围内出现的问题，有如下特点。

公式1

① $\mu-\sigma \leqslant x \leqslant \mu+\sigma$ 的概率，约68%

② $\mu-2\sigma \leqslant x \leqslant \mu+2\sigma$ 的概率，约95%

③ $\mu-3\sigma \leqslant x \leqslant \mu+3\sigma$ 的概率，约99.7%



这样，若不准确地掌握数据在多大范围内出现多大的离散，就会被其离散所迷惑。对这一点要格外小心，因一旦被这种离散所迷惑，就要做很多调整和调节作业，结果是事倍功半，甚至徒劳。

【事例1.3】

某面包厂在炉内连续地烤面包，班长在炉前看着烤面包的火候，当烤焦的面包出现时，班长就简单地认为是由于炉温过高所致，于是就将煤气阀门关小，借以使炉温降低。相反，当发现面包未烤透时，则开大煤气阀门，使炉温升高。尽管这样一天到晚不停地监视和调节，不良率仍达30%之多。一次，总厂质量管理骨干来指导工作并制定了某一措施，这样一来，不良率立刻下降到10%。这个所谓措施很简单，就是让那位班长离开炉子，质量管理骨干在炉上贴一张

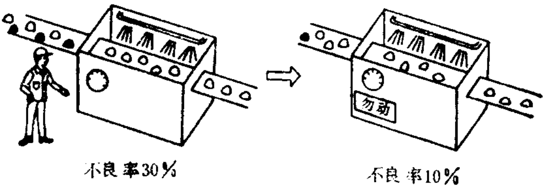


图 1.7　调节前后的比较

“勿动阀门！”的纸条。

【事例1.4】

某橡胶厂用自动切刀把挤压到园筒上的橡胶按一定重量切开。所切开的橡胶块重量的离散很大，不良率达5％。每当出现接近上限重量的橡胶块时，操作者就向减轻重量方向调整自动切刀的刻度；反之，当出现接近下限重量的橡胶块时，则向增加重量方向调整自动切刀的刻度。后来接受了质量管理骨干的建议，完全废弃了上述那种调整法，结果橡胶块重量的离散度降低了一半，从而减少了不良品。

因为不管在什么情况下，离散总是存在的，所以烤面包的火候适中程度也不例外。把火候适中程度作为横坐标，把面包达到了怎样一种程度的分布画成图（见图1.8）。因为这个工厂的工序能力低，至今偶而还会出现烤焦和欠火候的面包（约10％）。可是，如果总是一看到有烤焦的面包出现就去关小煤气阀门，那火候适中度就整个降低，就是说，火候适中度分布本身向左偏移了，所以欠火候的面包出现的比率倒会增大。当出现欠火候面包就去开大煤气阀门时，分布本身就向右偏移，这样则增加了面包烤焦的比率。

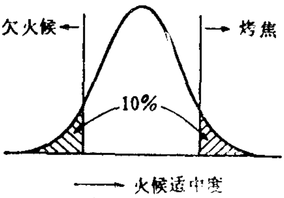


图 1.8 面包火候适中度的离散

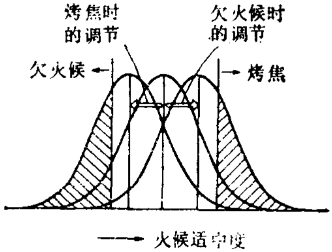


图 1.9 调节引起的分布移动

本来不采取任何措施就可以使不良率保持在10%左右的状态，但由于该班长所进行的调节作业，致使不良率上升到30%。后一个事例也是一样，本来不采取任何措施就会控制在某范围内的特性，也由于该操作者认为重量重就减少平均值，认为重量轻就升高平均值的调节作业而使分布自身失去平均，从而增大了离散度。

要消除类似上述事例的那种错误，必须掌握特性从哪到哪的离散范围，如果在某范围内并且是自然离散，就不必管它。相反，在这种情况下，如果急忙采取措施反而只能导致不良率提高、离散变大的结果。因此，即使为了恰当地进行判断，也要利用统计数据分析方法。

1.2.2 正态分布表的用法

正态分布时，如果k（≥0）值相等，那么不管μ和σ的值多大，x取$\mu+k\sigma$以上值的概率和取$\mu-k\sigma$以下值的概率等都是一定的。这个值根据书末附表1的正态分布表可以求得。例如，正态分布中离散特性为$\mu+1.92\sigma$以上的概率，可根据$k=1.92$栏得到0.0274（2.74%），相反，正态分布中离散

表 1.1 正态分布表（由k求P的表）*

k	0	1	2
⋮	(k=1.90) ↓	(k=1.91) ↓	(k=1.92) ↓
1.9	.0287	0281	.0274
2.0	.0228	.0222	.0217
⋮	↑ (k=2.00) ⋮	↑ (k=2.01) ⋮	↓ (k=2.02) ⋮

*注意：在有些表中，还有通过k求2P的情况。

特性值为$\mu+1.92\sigma$以下的概率可以用1减去此值求得：$1-0.0274=0.9726$(97.26%)。

取$\mu+2\sigma$以上值的概率与取$\mu-2\sigma$以下值的概率相同，可概据$k=2.00$栏得到0.0228。因而，$\mu\pm2\sigma$范围内的概率的正确计算为$1-0.0228\times2=0.9544$(95.44%)。同样，因$k=1.96$时，则$P=0.025$，所以公式1也会更准确地将95%的值计算到$\mu\pm1.96\sigma$范围内的。

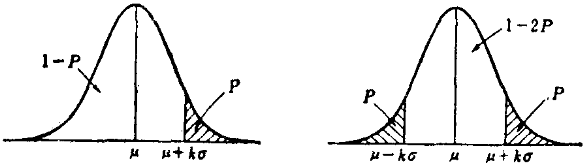


图 1.10　正态分布的各种概率

【例题1.1】

在平均值$\mu=20$，标准偏差$\sigma=7$的正态分布时，x值比34大的概率是多少？

【解】　因34比μ大，所以将此值表示为$\mu+k\sigma$的形式，k的值为

$$k=\frac{34-\mu}{\sigma}=\frac{34-20}{7}=2.00$$

因此，$P=0.0228$(2.28%)（见图1.11）

【例题1.2】

在平均值$\mu=10$，标准偏差$\sigma=0.5$的正态分布时，$x\leqslant9.04$的概率是多少？

【解】　因9.04比μ小，所以将此值表示为$\mu-k\sigma$的形式，k值为

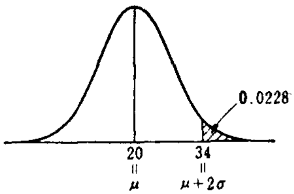


图 1.11　概率的计算

$$k=\frac{\mu-9.04}{\sigma}=\frac{10-9.04}{0.5}=1.92$$

所以，$P=0.0274$（2.74%）（见图1.12）

【例题1.3】

在$\mu=50$，标准偏差$\sigma=2$的正态分布时，x为46以上的概率是多少？

【解】 因46比μ小，故将此值表示为$\mu-k\sigma$的形式，k值为

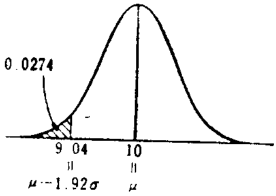


图 1.12 概率的计算

$$k=\frac{\mu-46}{\sigma}=\frac{50-46}{2}=2.00$$

因此可得x在46以下的概率为0.0228。相反，x在46以上的概率可用1减去0.0228得到：$1-0.0228=0.9772$（97.72%）（见图1.13）

【例题1.4】

在平均值$\mu=10$，标准偏差$\sigma=0.5$的正态分布时，$9.05\leqslant x\leqslant 11.01$的概率为多少？

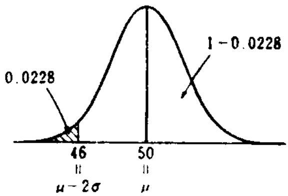


图 1.13 概率的计算

【解】首先因为11.01比μ大，故

$$k=\frac{11.01-\mu}{\sigma}=\frac{11.01-10}{0.5}=2.02$$

查表得，$x\geqslant 11.01$的概率为0.0217。其次，因9.05比μ小，故

$$k=\frac{\mu-9.05}{\sigma}=\frac{10-9.05}{0.5}=1.90$$

查表得$x\leqslant 9.05$的概率为0.0287。所以$9.05\leqslant x\leqslant 11.01$的概率为：