测量数据分析

吴世功　牛林山　编译

国防工业出版社

测量数据分析

吴世功
牛林山　编译

国防工业出版社

内容简介

本书以对话形式深入浅出地介绍了测量误差的估计方法和减小测量误差的试验设计法。重点介绍了信号与噪声之比（SN比）的求法，以及在测量技术、质量管理等领域的应用。全书共分十三章。其中，第二章到第九章介绍了测量误差的定量估计和经济性评价；第十章到第十三章从试验设计法的角度对测量误差的构成做了比较详尽的分析。另外，为使读者更好地理解本书的内容，我们特意还编写了第一章预备知识，叙述了最小二乘法、正交试验法和方差分析。

本书可供从事测量、计量、仪表设计、质量管理等工作的技术人员使用，也可供大专院校有关专业的师生参考，对具有中等文化程度的广大读者，可作为自学的入门书籍。

测量数据分析

吴世功
牛林山 编译

*

国防工业出版社出版

新华书店北京发行所发行　各地新华书店经售

国防工业出版社印刷厂印刷

*

787×1092 1/32　印张9 3/4　213千字

1988年8月第一版　1988年8月第一次印刷　印数：0，001—3，280册

ISBN 7-118-00139-2/P1　定价：3.25元

序

近代科研工作中包括了大量的测量技术。如何正确地评价测量方法的优劣及如何估算测量误差的大小等，一直都是从事理论研究和实际工作的科技人员极为关心的问题。本书将为读者提供解决上述问题的一个重要方法，这种方法就是信噪比（SN 比）方法。

在测量技术中，SN 比是测量设备准确度和测量方法精确度的一个重要指标。SN 比这一概念是美国普林斯顿大学盖特教授于一九六二年首先提出来的。近年来，在美国、西欧和日本等先进工业国已得到广泛应用，并取得了显著的经济效果。

日本 SN 比法的积极倡导、宣传和推行者是青山学院大学田口玄一教授，他曾发表了许多论文和专著，深受科技界和实业界的重视。

本书在中国电子学会可靠性分会何国伟副理事长的推荐和帮助下，根据田口玄一教授的“計測技術のためのデータ解析”技术讲座，参考国内有关著作编译而成。

编译过程中，对原文错误和欠妥之处作了更正和修改；对部分内容和章节作了调整和增删，文体采用对话形式，深入浅出，通俗易懂。但由于我们编译水平有限，错误之处在所难免，恳请读者批评指正。

编　译　者

目　　录

第一章　预备知识 ……………………………………… 1
§1-1　最小二乘法简介 ……………………………………… 1
§1-2　正交试验法 ……………………………………… 6
§1-3　方差分析 ……………………………………… 12
第二章　校准方法和测量误差 ……………………………………… 18
§2-1　测量误差综述 ……………………………………… 18
§2-2　各种校准方法 ……………………………………… 24
§2-3　误差方差的求法 ……………………………………… 29
第三章　比例校准和 SN 比 ……………………………………… 33
§3-1　测量方法的比较及其误差方差 ……………………………………… 33
§3-2　转动惯量的测定 ……………………………………… 35
§3-3　一般公式 ……………………………………… 42
§3-4　转动惯量的SN比 ……………………………………… 49
第四章　测量误差及其改进方法 ……………………………………… 53
§4-1　比例式校准 ……………………………………… 53
§4-2　一次（线性）式校准 ……………………………………… 57
§4-3　测量误差公式的有关注释 ……………………………………… 63
第五章　校准周期和误差波动 ……………………………………… 72
§5-1　零点校准和误差波动 ……………………………………… 72
§5-2　SN比的比较 ……………………………………… 75
§5-3　改变校准周期 ……………………………………… 84
§5-4　读数为差值（与真值之差）时，SN比的公式举例 ……………………………………… 87
第六章　信号因子水平的做法和 SN 比的比较 ……………………………………… 89
§6-1　信号因子及其水平 ……………………………………… 89

§6-2 信号因子水平的做法（以化学分析为例）……………………91
§6-3 信号因子水平的做法（以质量测量为例）……………………97
§6-4 其他问题……………………………………………………104
§6-5 转动惯量的信号因子问题…………………………………106
§6-6 轮箍的动平衡举例…………………………………………111
§6-7 信号因子水平的做法………………………………………115
§6-8 根据回归变量做信号因子的水平…………………………120
§6-9 有回归变量的SN比 ………………………………………122
§6-10 回归变量不能有效控制的情况 …………………………131
§6-11 信号因子的水平及其有关问题 …………………………136
§6-12 硬度测量的实例 …………………………………………138
§6-13 无重复测量的情况 ………………………………………144
§6-14 SN比置信区间表的使用方法举例………………………152
第七章 有关误差的一些问题 ……………………………………157
§7-1 内部干扰（内因）和外部干扰（外因） …………………157
§7-2 误差原因的分类……………………………………………158
§7-3 标准室的作用………………………………………………164
§7-4 试样误差问题………………………………………………168
第八章 误差因子的取法 …………………………………………172
§8-1 质量测量中的误差因子……………………………………172
§8-2 误差因子和SN比 …………………………………………177
§8-3 改变校准周期后的情况……………………………………185
第九章 改进SN比的试验设计 ……………………………………189
§9-1 信号因子和误差因子………………………………………189
§9-2 SN比的计算方法 …………………………………………192
§9-3 误差因子及其对策…………………………………………194
§9-4 SN比的置信区间（置信界限）……………………………199
§9-5 内侧正交表及控制因子的安排……………………………202
§9-6 外侧正交表及信号因子的安排……………………………206

§9-7 SN比的计算 ……………………………………………210
§9-8 方差分析……………………………………………………213
§9-9 显著因子的估算…………………………………………216
第十章 标准固有误差的试验设计 ……………………………218
§10-1 标准的误差 ………………………………………………218
§10-2 直读天平的试验 …………………………………………220
§10-3 方差分析 …………………………………………………223
§10-4 标示误差及其数据分析方法 ……………………………228
§10-5 求砝码标示值的误差方差的试验设计 …………………231
第十一章 标准的误差的求法 …………………………………235
§11-1 标准的标示值的误差 ……………………………………235
§11-2 环规举例 …………………………………………………241
§11-3 信号因子效果的多项式展开 ……………………………248
第十二章 标准的稳定性的试验设计及其数据分析 ……252
§12-1 标准的稳定性 ……………………………………………252
§12-2 稳定性的试验设计 ………………………………………253
§12-3 数据分析 …………………………………………………258
§12-4 变化量的预测 ……………………………………………261
§12-5 标准的校准周期 …………………………………………264
第十三章 二次校准式 …………………………………………267
§13-1 二次回归式 ………………………………………………267
§13-2 计算举例 …………………………………………………270
§13-3 校准方法 …………………………………………………276
§13-4 估计值的置信区间 ………………………………………280
附录 ………………………………………………………………283
附录Ⅰ 两个水平系列的正交表 ………………………………284
附录Ⅱ 三个水平系列的正交表 ………………………………288
附录Ⅲ χ^2分布表 ………………………………………………291
附录Ⅳ F分布表 ………………………………………………292
附录Ⅴ SN比的置信区间 (dB) ……………………………302

第一章 预备知识

§1-1 最小二乘法简介

甲 在工程上和测量中，经常会遇到某些变量之间具有一定的数量关系。例如，发动机燃烧室压力P与推力R就有一定的关系，即

$$R = a + bP$$

在进行涡轮流量校验时，频率f和流量Q就有一定的对应关系。有些关系通过理论分析，可以知道其所具有的某种形式，只是关系式中的一些参数要通过试验来求得。当然也有些关系是在试验后经统计数据处理才能发现。上述关系若是严格地一一相对应，那就比较好处理。但实际上，遇到的一些关系往往会受到仪器的精度和外界条件的影响，很难找到一一对应关系。在这种情况下，如何处理为好呢？

乙 下面介绍一种称为最小二乘法的方法。所谓最小二乘法，就是在试验中获得自变量与因变量的若干对应数据(x_1, y_1)，(x_2, y_2)，…，(x_n, y_n)时，要找出一个已知类型的函数$y=f(x)$，使得离差平方和$\sum\limits_{i=1}^{n}[y_i - f(x_i)]^2$为最小，这种方法称为最小二乘法。

甲 请你举一个例子讲一讲最小二乘法的应用好吗？

乙 可以。我想通过求涡轮流量计的校验曲线为例来说明最小二乘法的应用。在校验涡轮流量计时，对应不同的频

率 f 就有不同的流量 Q 值。设校验点为 n 时，则可得到下列的 n 对数值：

$$(f_1,\ Q_1),\ (f_2,\ Q_2),\ \cdots,\ (f_n,\ Q_n)$$

把这 n 对数值标画在分别以 f 和 Q 为横纵坐标的直角坐标系中，就可得到图1-1所示的对应关系。

从图 1-1 可以看出，频率和流量之间近似线性关系。如

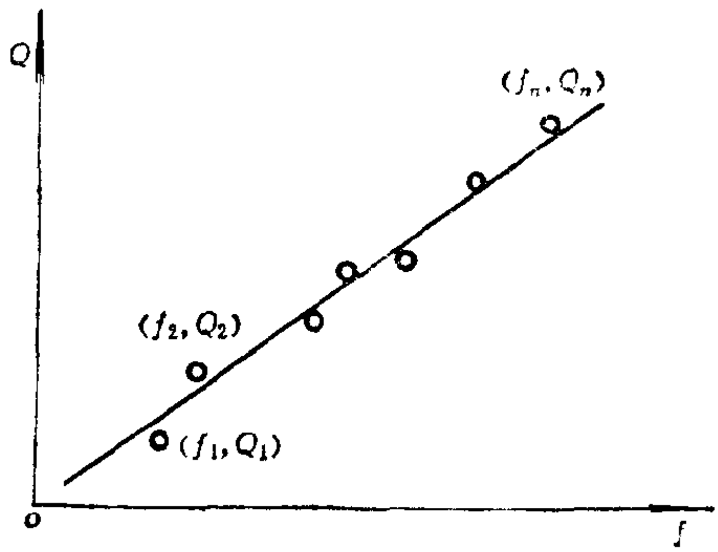

图1-1　频率 f 与流量 Q 的关系

果我们想找一条通过 n 个点的直线，一般是不可能的。因此，我们应寻找这样一条直线，即使之在 n 个点的间隙中通过时，距各个点的距离为最小。

甲　这里所说的“距各个点的距离为最小”，应该同上面所说的“使离差平方之和为最小”是一致的吧？

乙　是一致的。设这条直线为 $Q = a + bf$，其中 a 和 b 是待定的，可根据 n 对数值 $(f_i,\ Q_i)$（$i = 1,\ 2,\ \cdots,\ n$）估算出来。要求得 a 和 b 之值，并使函数 $L(a,\ b)$ 取最小值，即

$$L(a, b)=\sum_{i=1}^{n}[Q_i-(a+bf_i)]^2$$

为最小。

根据极小值的求法，将上式分别对 a 和 b 求偏导数，并令其等于零。

$$\begin{cases}\dfrac{\partial L(a, b)}{\partial a}=2\sum\limits_{i=1}^{n}[Q_i-(a+bf_i)](-1)=0\\[2ex]\dfrac{\partial L(a, b)}{\partial b}=2\sum\limits_{i=1}^{n}[Q_i-(a+bf_i)](-f_i)=0\end{cases}$$

将上面两式化简就可得到：

$$na+b\sum_{i=1}^{n}f_i=\sum_{i=1}^{n}Q_i \tag{1-1}$$

和

$$a\sum_{i=1}^{n}f_i+b\sum_{i=1}^{n}f_i=\sum_{i=1}^{n}Q_if_i \tag{1-2}$$

解上面两式，就可求得 a 和 b 的估计值：

$$\hat{b}=b=\frac{\sum\limits_{i=1}^{n}(f_i-\bar{f})(Q_i-\bar{Q})}{\sum\limits_{i=1}^{n}(f_i-\bar{f})}$$

$$=\frac{\sum\limits_{i=1}^{n}(f_i-\bar{f})Q_i}{\sum\limits_{i=1}^{n}(f_i-\bar{f})^2} \tag{1-3}$$

和

$$\hat{a}=a=\overline{Q}-\hat{b}\overline{f} \tag{1-4}$$

其中

$$\overline{f}=\frac{1}{n}\sum_{i=1}^{n}f_i$$

$$\overline{Q}=\frac{1}{n}\sum_{i=1}^{n}Q_i$$

由式（1-3）和（1-4）求得 $\hat{a}$ 和 $\hat{b}$，而且由它们构成的直线是最佳直线。

$$Q=\hat{a}+\hat{b}f$$

这时，可用

$$\sigma=\sqrt{\frac{\sum_{i=1}^{n}[Q_i-(\hat{a}+\hat{b}f_i)]^2}{n-2}} \tag{1-5}$$

来描述 n 个点偏离直线的情况。σ 越大，则表示点偏离直线就越远。

先计算出 $\overline{f}$ 和 $\overline{Q}$，然后照式（1-3）、（1-4）两式算出 $\hat{a}$ 和 $\hat{b}$，就可得出如下直线方程式：

$$Q=\hat{a}+\hat{b}f$$

甲　在数据处理中，什么情况下才能应用最小二乘法呢？

乙　严格地说，误差服从正态分布的情况下才能得以应用。但在与正态分布差异不太大的误差分布中，以及在一切误差都是相当小的任意分布中，实际上也常采用最小二乘法来进行数据处理。

甲 另一个问题，就是利用最小二乘法时，必须事先知道变量之间的函数关系。而函数关系中的各变量值的大小，正是最小二乘法要解决的。如果函数关系是近似的，或是经验方程式，那也可以利用最小二乘法吗？

乙 通常不行。之所以应该是准确的函数关系的原因，就在于假如采用近似或经验的方程式，可能在很大程度上抵消了最小二乘法所带来的好处。

甲 请你用一个具体例子来计算一下。

乙 比如在某试验中，得到如下一些数据。

	1	2	3	4	5	6	7	8	9	10	11
x	57.6	56.8	58.1	59.5	59.8	59.4	59.6	59.2	59.7	59.4	59.4
y	24.0	24.0	24.0	24.9	25.06	24.89	24.98	24.68	24.89	24.76	24.76

两个变量 x 和 y 之间是线性关系。首先求出 x 和 y 的平均值，即

$$\bar{x}=\frac{1}{11}\sum_{i=1}^{11}x_i$$

$$\bar{y}=\frac{1}{11}\sum_{i=1}^{11}y_i$$

然后再按式（1-3）、（1-4）两式，就可算得 a 和 b 的估计值：$\hat{a}=0.863$，$\hat{b}=0.403$。

从而求得：

$$y=0.863+0.403x$$

按照式（1-5）就可算出标准误差为

$$\sigma=0.145$$

§1-2 正交试验法

甲 据说在日本“正交试验法”已广为普及，并取得了十分显著的经济效果。请你简单介绍一下关于“正交试验法”的基本知识好吗？

乙 有关“正交试验法”的专门著作很多，因为要用到这方面的知识，所以为方便读者起见，我想在这里简单介绍一下。在测量技术中，被测对象往往要受到温度、湿度、压力和标准稳定性等许多因素的影响。其中，哪些是主要因素？哪些是次要因素？这都需要做各种各样的试验。至于如何来安排多因素试验，也是一个很值得研究的问题。试验安排得好，既可以减少试验次数，又能够得到满意的结果；反之试验安排得不好，试验次数既多，也得不到满意的结果。正交试验法是研究与处理多因素试验的一种科学方法。它利用一种现成的规格化的“正交表”，科学地挑选试验条件，合理地安排试验。这种方法的优点是，能在很多的试验条件中，选出影响指标的主要条件，仅仅通过少数几次试验，就能达到预期的目的。

甲 正交试验法的基本工具是规格化的“正交表”，而这种“正交表”具有什么特点呢？

乙 我先来介绍一种最简单的正交表$L_4(2^3)$，示于表1-1。

表1-1　正交表$L_4(2^3)$

试验号＼列号	1	2	3
1	1	1	1
2	1	2	2
3	2	1	2
4	2	2	1

首先说一下记号$L_4(2^3)$的含义：L代表正交表；L右下角的数字“4”表示有4个横行，即要做四次试验；括号内的指数“3”表示有3个纵列，即最多允许安排的因子个数是三个；括号内的数字“2”表示该表的主要部分只有两种数字，即因子有两个水平，称之为1水平和2水平。

表$L_4(2^3)$具有下述两种性质：

1. 每一列中，不同的数字出现的次数相等。这里不同的数字只有两个，即1和2，它们各出现两次。

2. 任何两列中，同一横行的两个数字看成有序数对（即左边的数放在前，右边的数放在后，按这种顺序排成的数对）时，每种数对出现的次数相等。这里有序数对共四种：(1，1)，(1，2)，(2，1)，(2，2)，它们各出现一次。

甲　具体来说，正交表怎么来使用？

乙　现举例来讲讲这个问题。例如，在某生产过程中，影响得率的因素有四种，每种因素都有两种状态，具体情况如下：

A反应温度	A_1　100°C	A_2　120°C
B反应时间	B_1　5小时	B_2　7小时
C混合比例	C_1　2:1	C_2　3:2
D反应压力	D_1　1个大气压	D_2　1.2个大气压

为了便于讨论，以后通称影响试验指标的因素为因子，用大写字母A，B，C，……来表示；每个因子可能处的状态称为水平，用表示该因子的字母加上脚码来表示，例如A_1，A_2，……表示A因子的第一水平，第二水平，……，等。

甲　按照这一规定，上面的问题就有四个因子，每个因子有两个水平，对吗？

乙　对。我们选用正交表$L_8(2^7)$●，其中包括8个试验。这8个试验是从$2\times2\times2\times2=16$种可能搭配中一次挑出的，只要条件许可，就能同时进行试验。

在1，2，4，7列上分别写上因子A、B、C、D，就得到了表1-2。

表1-2　正交表$L_8(2^7)$

列号 / 试验号	A 1	B 2	3	C 4	5	6	D 7
1	1	1	1	1	1	1	1
2	1	1	1	2	2	2	2
3	1	2	2	1	1	2	2
4	1	2	2	2	2	1	1
5	2	1	2	1	2	2	2
6	2	1	2	2	1	1	1
7	2	2	1	1	2	2	1
8	2	2	1	2	1	1	2

甲　为什么要这样安排四个因子呢？

乙　关于这个问题，以后要专门讲解。

甲　在表1-2的各因子列中，在数字“1”和“2”的位置分别填上该各因子的1水平和2水平，就可得到一张实

● 全书简记为L_8。——编译者

际使用的试验设计表了吧？

乙　是的。具体试验设计见表 1-3。当试验设计排定之后，就要按试验设计进行试验。不过，试验顺序可以随机选取。这样可以防止试验中因先后掌握不均所带来的干扰以及因外界环境变化所引起的系统误差。

表1-3　试验设计表

试验号 \ 因子	A (°C) 1	B (小时) 2	C (比例) 3	D (大气压) 4	试验结果 y_i(%)
1	100	5	2:1	1	96
2	100	5	3:2	1.2	105
3	100	7	2:1	1.2	101
4	100	7	3:2	1	104
5	120	5	2:1	1.2	101
6	120	5	3:2	1	106
7	120	7	2:1	1	93
8	120	7	3:2	1.2	98

甲　在这四个因子中，怎样判断哪些因子对得率影响大？哪些因子影响小呢？

乙　这里共进行了八次试验，因八次试验的条件各不相同，没有比较标准，故直接对 8 个数据进行比较是不行的。若将这 8 个数据进行适当组合，我们就可以发现还是能够进行比较的。

以因子 A 为例，A 的 1 水平出现在第1～4号试验中，这四个试验的平均得率为

$$\bar{A}_1=\frac{1}{4}(y_1+y_2+y_3+y_4)$$

$$=\frac{1}{4}(96+105+101+104)=101.5$$

A的2水平出现在第5～8号试验中，这四个试验的平均得率为

$$\overline{A}_2=\frac{1}{4}(y_5+y_6+y_7+y_8)$$

$$=\frac{1}{4}(101+106+93+98)=99.5$$

$\overline{A}_1$ 和 $\overline{A}_2$ 就可以进行比较，对于 A_1 条件下的四个试验和 A_2 条件下的四个试验来说，尽管其他条件在变动，但这种变动是平等的，因此，$\overline{A}_1$ 和 $\overline{A}_2$ 之间的差异则反映了A的两个水平的差异。即

$$\overline{A}_1-\overline{A}_2=101.5-99.5=2>0$$

所以说，因子A取 A_1 时，其平均得率较高。

同样，也可以对B、C、D的两个水平进行比较。比较时，可以 1 水平对应的四个试验为一组，以 2 水平对应的四个试验为另一组来进行，结果便可得到：

$$\overline{B}_1=\frac{1}{4}(y_1+y_2+y_5+y_6)=102.0$$

$$\overline{B}_2=\frac{1}{4}(y_3+y_4+y_7+y_8)=99.0$$

$$\overline{C}_1=\frac{1}{4}(y_1+y_3+y_5+y_7)=97.75$$

$$\overline{C}_2=\frac{1}{4}(y_2+y_4+y_6+y_8)=103.25$$

$$\overline{D}_1=\frac{1}{4}(y_1+y_4+y_6+y_7)=99.75$$

$$\overline{D}_2=\frac{1}{4}(y_2+y_3+y_5+y_8)=101.25$$

由上列数据可以得到：

$$\overline{B}_1-\overline{B}_2=3.0$$

$$\overline{C}_1-\overline{C}_2=-5.5$$

$$\overline{D}_1-\overline{D}_2=-1.5$$