

梁 昌 洪

西安电子科技大学出版社



人实验数据处理 谈 起

封面设计：傅化群

ISBN 7-5606-0474-9



9 787560 604749 >

ISBN 7-5606-0474-9/O · 0029

定 价：4.00 元



N₃/8

从实验数据处理谈起

梁 昌 洪

西安电子科技大学出版社

1996

(陕)新登 010 号

内 容 简 介

本书以简洁、实用的风格,介绍了实验数据处理领域的主要问题。它适合于理、工、医、农及广泛专业高校的大学生作为入门和应用的参考书,并且,对于研究生在此领域的深入和发展也将有所裨益。

全书配有计算机程序。

从实验数据处理谈起

梁昌洪

责任编辑 霍小齐

西安电子科技大学出版社出版发行

西安电子科技大学印刷厂印刷

新华书店经销

开本 787×1092 1/32 印张 2 10/32 字数 45 千字

1996 年 7 月第 1 版 1996 年 7 月第 1 次印刷 印数 1—1 000

ISBN 7-5606-0474-9/O·0029 定价: 4.00 元

序

数据处理是任何一门科学实验都必须涉及的课题。本身作为一门学科，百年来名著迭出。到现代，由于计算机技术的介入而活力倍增。不过，作为一个课题，它的主目标只是一个，即：如何从实验或考察所得的数据中，尽可能多、尽可能准、尽可能快地整理出所需的信息。这对于许多科技人员来说，是一种实用的要求。在面临一些任务的情况下，这种要求的迫切性往往更甚于理论上的透彻探讨。针对这种要求，手边备一本简明扼要、便于在实用中参考的小册子应当是合宜的。梁昌洪教授的《从实验数据处理谈起》正是从这个角度撰写的一本小册子。书中用尽可能简明易懂的方式把数据处理的要点介绍给读者，并备有“计算程序附录”。这对于那些一时

还不能花很多时间系统地学习这门学科而又需要在工作中使用的人们，对于一些数据处理的入门者，对于一些已经掌握了这门学科但在实用中希望随手翻阅参考的人们，可以说是提供了一本方便而实用的“案头本”。当然，就其作用而言，它是属于“简明备考”型，而不是“全面备检”型的。因此，对于比较复杂或专门的问题，则应当求助于相应的专著。在这种情况下，这本书应当可以看作是一个“引子”。

A stylized handwritten signature in black ink, consisting of several fluid, connected strokes. The signature appears to be '王炯' (Wang Jiong) with a surname '王' and a given name '炯'.

(王炯)

1996年6月于北京

目 录

序

1 引子	(1)
2 点约束	(4)
3 误差	(6)
4 平均法	(10)
5 最小二乘法	(13)
6 法向回归	(22)
7 权	(26)
8 坏点剔除	(29)
9 广义回归	(32)
10 全面最小二乘法	(36)
11 结语	(48)
程序附录	(51)
主题索引	(66)
参考文献	(67)

1

引 子

让我们做一个简单的电路实验。利用可变电源 E 测量流过 AB 的电流 I 和 AB 两端的电压 U 。并以此确定 U_0 和 R 两个参数。实验电路如图 1-1 所示。

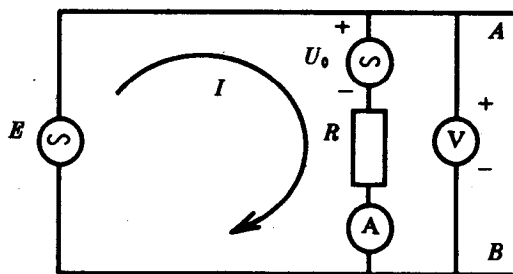


图 1-1 电路实验

这个实验的物理规律是简单且已知的。即利用 Ohm 定律有

2 从实验数据处理谈起

$$U = RI + U_0 \quad (1-1)$$

也就是说, U 和 R 成直线关系, 所求的 R 是直线的斜率, 而 U_0 则是在 U 上的截距, 如图 1-2 所示。

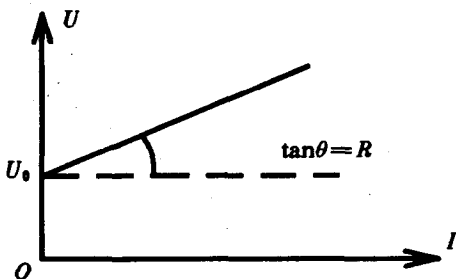


图 1-2 U 和 R 的函数关系

现在, 把测量而得的数据归纳于表 1-1 中, 并把问题具体提法总结成: 利用数据处理方法, 把表 1-1 中测量值的信息转化为要求的参量。

这个引子虽然简单, 但是却已概括了问题的对象和目标。我们把已知物理规律求参数的问题称之为回归 (regression)* (图 1-3)。

回归问题是数据处理的主要对象。

* 在本书中, 我们将把不知道物理规律求参数的问题称之为广义回归。

表 1-1 电路实验测量数据

i	$I_{(\text{mA})}$	$U_{(\text{V})}$	i	$I_{(\text{mA})}$	$U_{(\text{V})}$
1	2.00	5.85	6	12.00	10.88
2	4.00	7.08	7	14.00	12.04
3	6.00	8.16	8	16.00	12.95
4	8.00	8.95	9	18.00	14.05
5	10.00	10.15	10	20.00	14.85

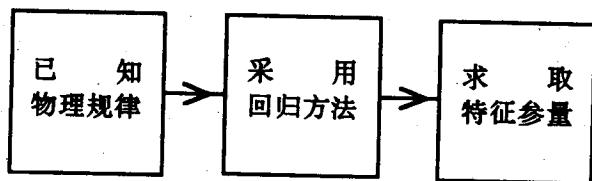


图 1-3 回归问题

2

点约束

点约束处理的思想十分简单：要求 m 个未知参量，则需要（也只需要）建立 m 个独立方程。如果方程数少于未知量数，即成为不定问题；反之，方程数多于未知量数，将称为超定问题。

引子中的电路实验，要求出 U_0 和 R ，所以只要有两次实验数据，例如第 i 次和第 j 次为

$$\begin{cases} U_i = RI_i + U_0 \\ U_j = RI_j + U_0 \end{cases} \quad (2-1)$$

很容易导得

$$\begin{cases} R = \frac{U_i - U_j}{I_i - I_j} \\ U_0 = \frac{U_j I_i - U_i I_j}{I_i - I_j} \end{cases} \quad (2-2)$$

但是，如果我们从表 1-1 的测量数据中随意提取 2 次，则问题就出现了

矛盾。例如， $i=1, j=2$ 有

$$\begin{cases} R = \frac{5.85 - 7.08}{2.00 - 4.00} \times 10^3 = 615.00 \, \Omega \\ U_0 = \frac{7.08 \times 2.00 - 5.85 \times 4.00}{2.00 - 4.00} = 4.62 \, \text{V} \end{cases} \quad (2-3)$$

而另取 $i=9, j=10$ ，又有：

$$\begin{cases} R = \frac{14.05 - 14.85}{18.00 - 20.00} \times 10^3 = 400.00 \, \Omega \\ U_0 = \frac{14.85 \times 18.00 - 14.05 \times 20.00}{18.00 - 20.00} = 6.85 \, \text{V} \end{cases} \quad (2-4)$$

选取不同的点得到的结果不同，而且离散度十分大，究其原因是因为所有实验所得的测量数据都存在误差，而点约束的思想则属于无误差处理。因此，点约束不是真正意义上的实验数据处理方法。

3

误差

现代数据处理是考虑了每次测量均存在各种误差时的**有误差处理**，因此它必须是多余量处理，即独立测量次数 n 必须大于待求未知量数 m 。

误差主要可分为三类：

1. 偶然误差——由于种种原因造成的随机误差。
2. 系统误差——由于仪表不准确、系统不理想（例如，电源内阻等）表现出固定或有规律的非随机误差。
3. 过失误差——出现机会极少，例如读表错误。

具体分类见图 3-1。

本书所述及的数据处理方法，主要是对**偶然误差**的处理，并同时尽可能地剔除过失误差。

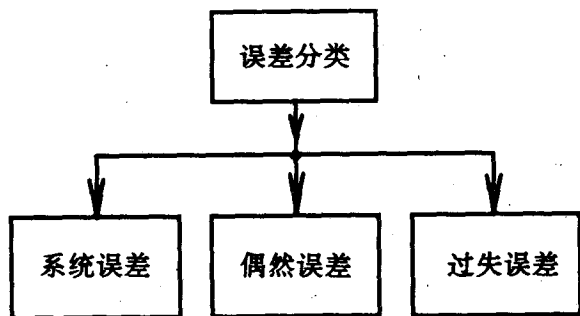


图 3-1 误差分类

已经知道，偶然误差分布满足正态（Gauss）规律，即误差 x 的概率密度 y 为

$$y = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{x^2}{2\sigma^2}\right\} \quad (3-1)$$

其中， σ 为方差，如图 3-2 所示。

偶然误差表现出的规律是

- (1) 正、负误差对称出现，机会相等；
- (2) 绝对值大的误差出现机会极为稀少。

以上两条亦称之为偶然误差公理。

由于测量仪表的系统偏差、精度限制和读数的偶然误差，使最后得到的结果数据只有有限位才有意义。因此，必须引进数据的有效数字概念。但是，在

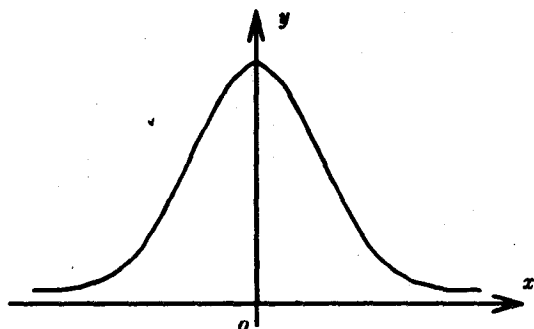


图 3-2 偶然误差的正态 (Gauss) 分布

定义有效数字时我们没有考虑系统偏差的影响, 举例来说, 对不准钟表的读数其有效数字不可能计及不准的因素。

我们定义, 仪表读数的最后 1 位 (如果是数字式仪表, 则规定为数字波动的 1 位) 是误差所在的 1 位, 也即有效数字的最后 1 位是偶然误差所在的 1 位。这样的做法兼顾了仪器的精度和读数的偶然误差。

例如 0~20 mA 量程的电流表, 其最小分度值是 0.1 mA。我们可以估读到分度的 1/10, 即如表 1-1, 有 2.00 mA 的测量数据。

有效数字以绝对误差为主要数据, 所以, 在作加减运算时将保持绝对误差位不变; 而当作乘除运算

时，一般按最少有效位（或再加上 1 位）较为合适。但具体情况还应具体分析。例如，引子中的 10 组数据，像 5.85 V 可以理解为 05.85 V，即所有电压（电流亦相同）均有 **4 位有效数字**。

4

平均法

平均法是测量数据偶然误差处理的最简单方法。它的基本思想是：利用全部 n 次测量数据信息建立 m 个方程，从而提取 m 个待求参量。对于引子中提出的实例，我们可以把全部 10 个数据分成两组：

I 组： $i = 1, 2, \dots, k = n/2 = 5$

II 组： $j = 1, 2, \dots, k = n/2 = 5$

每组 U_i 、 I_i 和 U_j 、 I_j 均采用平均方法获得，即

$$\begin{cases} U_i = \frac{1}{k} \sum_{i=1}^k U_i, \\ I_i = \frac{1}{k} \sum_{i=1}^k U_i, \end{cases} \quad \begin{cases} U_j = \frac{1}{k} \sum_{j=1}^k U_j, \\ I_j = \frac{1}{k} \sum_{j=1}^k U_j, \end{cases} \quad (4-1)$$

这样做能使构成 U_i 中的每个 U_i 由