回归分析中的计算问题

李　宁
（华东石油学院）

进行物理实验、取得实验数据的目的，是为了揭示自然界固有的客观规律，而实验数据处理的好坏，直接影响实验结果的可信度和对实验结果的解释．大量物理实验数据的处理，离不开回归分析方法．尽管目前论述如何进行回归分析的专著颇多，然而关于回归过程之每一步骤应怎样计算——这个看起来似乎不成问题的问题，却不应该地被疏忽了，即有意识或无意识地将实验数据(近似数)当成精确数，本能地使用初等代数的运算方法去求回归方程．结果导致有效数字随意保留或丢失；精确度随意提高或降低．自然，处理结果的可靠程度就可想而知了．

本文不想对所有回归方法中的计算步骤都加讨论，而只通过超声测量中求一元线性回归方程的两个实例，来说明数值分析在回归过程中的应用及其重要性．

一、实　例　1

在研究岩石声学性质时，往往要测量声波在岩石中的行走时间(下称走时)．走时折算到单位长上称时差，是声速的倒数． 时差与长度无关，而走时一般却是长度的线性函数． 用 T 表示走时，L 表示被测件长度，则 $T=f(L)$．针对一组不同的 $L(L_1, L_2, \cdots, L_N)$，测得一系列不同的 T $(T_1, T_2, \cdots, T_N)$，则用一元线性回归的方法，就可确定 $f(L)$ 的具体表达式．

在实验室用声速测定仪对一组长度不等的试件进行走时测量，记录数据如下：

L(厘米)	2.000	4.000	6.000	8.000	10.000
T(微秒)	5.4	8.9	12.1	15.4	19.0

根据文献[1]，将回归的格式写在下面．格式的由来及格式中各符号的意义均见[1]． 这里主要讨论格式中各步数字的运算．为方便起见，每行数据和算式前都编上了号码．

回归格式：

	L	T	L^2	T^2	$L\cdot T$
(1)	2.000	5.4	4.000	29	11
(2)	4.000	8.9	16.00	79	36
(3)	6.000	12.1	36.00	146	72.6
(4)	8.000	15.4	64.00	237	123
(5)	10.000	19.0	100.00	361	190
(6)	30.000	60.8	220.00	852	433

(7) $N=5$

(8) $\bar{L}=\dfrac{\sum L}{N}=\dfrac{30.000}{5}=6.000$

(9) $\bar{T}=\dfrac{\sum T}{N}=\dfrac{60.8}{5}=12.2$

(10) $\dfrac{(\sum L)^2}{N}=\dfrac{(30.000)^2}{5}=180.00$

(11) $\dfrac{(\sum T)^2}{N}=\dfrac{(60.8)^2}{5}=739$

(12) $\dfrac{(\sum L)(\sum T)}{N}=\dfrac{30.000\times 60.8}{5}$

$=\dfrac{30.00\times 60.8}{5}=365$

(13) $l_{LL}=\sum L^2-\dfrac{(\sum L)^2}{N}=220.00-180.00$

$=40.00$

(14) $l_{TT}=\sum T^2-\dfrac{(\sum T)^2}{N}=852-739$

$=113$

(15) $l_{LT} = \sum(L \cdot T) - \frac{(\sum L)(\sum T)}{N}$

$= 433 - 365 = 68$

(16) $b = l_{LT}/l_{LL} = 68/40.00 = 68/40.0 = 1.70$

(17) $a = \bar{T} - b\bar{L} = 12.2 - 1.70 \times 6.000$

$= 12.2 - 10.2 = 2.0$

(18) 回归方程：$T = f(L) = a + b \cdot L$

$= 2.0 + 1.70L$（微秒）

(19) 相关系数：$r = l_{LT}/\sqrt{l_{LL} \cdot l_{TT}}$

$= 68/\sqrt{40.00 \times 113}$

$= 68/\sqrt{4.52 \times 10^3}$

$= 68/67.2 = 1.0$

(20)声速：$c = 1/b = 1/1.70 = 5.88 \times 10^{-1}$

（厘米/微秒）$=5.88\times10^3$米/秒

现在具体分析上列格式中的各步计算.

先看(1)：$L = 2.000$ 这个数据具有四位有效数字,而 $T = 5.4$ 只有两位有效数字. 由近似数乘法法则，L^2 应该具有四位有效数字，故 $L^2 = (2.000)^2 = 4.000$；T^2 应该具有两位有效数字，但精确数 5.4 的平方是 29.16，故 $T^2 = (5.4)^2 = 29 \doteqdot 29.16$. 0.16 的舍去是按近似数的抹尾凑整法则进行的. $L \cdot T$ 的计算为：$L \cdot T = 2.000 \times 5.4 = 2.00 \times 5.4 = 11$.

(2)—(5)四行数据的运算方法同上.

(6) 的一行表示对上五行分别纵向求和. 需要指出的是 $\sum L^2$ 和 $\sum(L \cdot T)$ 的计算. 由近似数加法法则：$\sum L^2 = 16.00 + 36.00 + 64.00 + 100.00 + 4.000 = (16.00 + 36.00 + 64.00 + 100.00) + 4.000 = 216.00 + 4.000 = 220.00$；$\sum(L \cdot T) = 11 + 36 + 123 + 190 + 72.6 = (11 + 36 + 123 + 190) + 72.6 = 360 + 72.6 = 433$.

(7) 中 $N = 5$ 表示参加回归的点一共五个,因此 5 是一个精确数.

(8) 的一行表示对 L 求算术平均值. 由于 5 是精确数，故结果的有效位数应与 30.000 的有效位数(五位)相同. 考虑到 L 的测量精度在以厘米为单位时只达到小数点后第三位，故取结果 $\bar{L} = 6.000$，实际上保留了四位有效数字而不是五位.

(9)—(12) 各行计算原理同上，不赘述. (14)的一行进行减法运算，故参照近似数减法运算法则进行.

值得指出的是 (13) 和 (15). 不难看出，(13) 的减数和被减数都分别具有五位有效数字，但它们的差 40.00 却只有四位有效数字；(15) 的减数和被减数分别具有三位有效数字，而它们的差 68 却只剩下两位有效数字. 象这样的有效数字丢失，并不是测量结果本身所造成,而是由于减法运算所导致,故必须充分予以重视. 因为40.00和 68 这两个数下面还将参与除法运算[见 (16)]，这里有效数字位数的降低就意味着下面除法运算所得的商的有效数字位数的降低. 但下面我们会发现，40.00 和 68 同时参加一个除法运算步骤,由于40.00的有效数字位数较 68 多两位，故 68 是引起商的有效位数降低的决定因素.

正是充分注意了 68 这个数的由来,所以在 (16) 这行的计算结果里给出了 1.70 这个数而不是1.7. 现证明取1.70是合理的. 为此引用定理："若 A_1, A_2 分别为二近似数，E_{a1}，E_{a2} 为它们各自的绝对误差，ΔQ 为 A_1/A_2 的绝对误差,并令 $\omega = \max\{E_{a1}/A_1, E_{a2}/A_2\}$，则当条件：(1) $\omega \ll 1$；(2) $(E_{a1}/A_1) \ll (E_{a2}/A_2)$ 或 $(E_{a2}/A_2) \ll (E_{a1}/A_1)$ 被满足，有 $\Delta Q \leqslant (A_1/A_2)\omega$ 成立"[2]. 于是令：$A_1 = 68$，$A_2 = 40.0$；且有 $E_{a1} \leqslant 0.5$，$E_{a2} \leqslant 0.05$；那么$(E_{a1}/A_1) \leqslant 0.007$，$(E_{a2}/A_2) \leqslant 0.001$，故近似有 $(E_{a2}/A_2) \ll (E_{a1}/A_1)$ 和 $\omega = 0.007 \ll 1$，则式 $\Delta Q \leqslant (68/40.0) \times 0.007 = 0.01$ 成立. 这说明 $b = 1.70 \pm 0.01$，因而取 $b = 1.70$ 是合理的. 但需要意识到 1.70 在小数点后第二位上已有了误差.

(17)中同样出现了减法丢失有效数字的现象,但这里不必担心，因为 2.0 这个数在后面不再参与乘除运算.

(18)给出正确的回归方程.

要指出(19)的开方运算有效数字的保留问题. 对函数 $y = \sqrt{x}$，自变量 x 的直接误差

Δx 引起函数 y 的间接误差 Δy 是: $\Delta y \approx (1/2)x^{-\frac{1}{2}}\Delta x$. 具体到 $y=\sqrt{4.52\times10^3}$, 有 $x=4.52\times10^3$, 为三位有效数字, $\Delta x\leqslant5$. 所以

$$\Delta y\leqslant\frac{5}{2\sqrt{4.52\times10^3}}<0.04,$$

$$y=\sqrt{4.52\times10^3}=67.2$$

为三位有效数字. 另外, 68 这时又以分子形式参与除法运算, 且上述定理条件依然满足, 但商 1.0 在这里不取三位有效数字, 否则最低位上的误差将使相关系数 $r>1$, 不合理.

最后看一下(20)的结果. 上面分析了 1.70 的误差, 因而推知求出的声速 5.88×10^3 米/秒在第三位数字上是有误差的. 所以对声速测量的结果,只准确到前两位有效数字.

二、实 例 2

为了研究耦合界面对声速的影响, 可用同直径不同长度的圆柱形有机玻璃块按一定方式对接组合,在对接面上涂以硅油,人为地形成一个、两个等耦合界面. 现取 2、3、4、5、6、7、8、10 厘米的有机玻璃块各一块,先不做对接组合,测其走时. 然后将 2 厘米长的那块分别与 3、4、…8、10 厘米的各块做二组合对接, 形成一个耦合界面,再测其走时. 最后做三组合对接. 用本文的回归方法求出相应的声速为:

组合方式	不做组合	双组合	三组合
声速 (米/秒)	2.75×10^3	2.75×10^3	2.72×10^3

如果采用精确数的计算方法来求回归方程, 最后保留六位有效数字,那么,求出的声速是:

组合方式	不做组合	双组合	三组合
声速 (米/秒)	2761.92	2748.98	2753.20

显然前一个表是正确的, 因为根据一般的物理常识,界面增多,声速肯定要降低 (至少这种降低在仪器精度范围内测不出来),但决不会象后一个表给出的那样三组合(两个界面)的声速比双组合(一个界面)还高. 由此看来, 精确计算在处理近似数的过程中, 有时非但不能给出精确结果, 反而会掩盖问题的实质, 造成错误结论. 因此, 在回归过程中强调数值分析的重要性,就是十分必要的了.

参 考 文 献

[1] 中国科学院数学研究所数理统计组, 回归分析方法, 科学出版社, 1975 年.
[2] J. B. 斯卡勃罗, 数值分析, 科学出版社, 1958 年.

强磁场技术及其在科学研究中的应用

高 秉 钧
(中国科学院等离子体物理研究所)

强磁场的含义在技术上通常是指超过 3T ($1T=10^4G$) 的磁场.

产生强磁场的基本方法是让电流通过多匝线圈. 这项技术始于十九世纪, 正是这一发明导致了近代电力工业的出现.

本世纪三十年代末, 由于 F. Bitter 在美国麻省理工学院开创了水冷磁体设计技术[1], 从而使稳态强磁场的产生技术有了大的突破. 五十年代,许多国家开始建造强磁场实验装置;六十年代后, 大型的强磁场实验室在许多国家相继建立, 并广泛开展了强磁场下的科学研究工作. 目前, 水冷磁体产生的场强达到 23.5T. 与此同时,由于六十年代初,高临界参量 (H_{c_2}, T_c, J_c)II 类超导体的出现, 使相当数量的场强