# 第3章 模型拟合

## 引言

在数学建模过程中，需要根据不同目的分析数据．我们前面也已经看到，如何由假定引导出一个特殊形式的模型．例如，在第2章中，分析汽车刹车后能立即安全停稳的距离时，我们的假定引出了一个下列形式的子模型：

$$d_b = Cv^2$$

这里 $d_b$ 是使汽车停稳所要求的距离，$v$ 是刹车时刻汽车的速度，$C$ 是需要确定的比例常数．这样，就能够通过收集和分析足够的数据，确定假定是否是合理的．如果合理，则要确定常数 $C$，从一族刹车距离子模型 $y=Cv^2$ 中选择某一成员．

我们会遇到由不同的假定引导出不同的子模型的情况．例如，当研究质子在空气等介质中的运动时，能对阻力的特征做出不同的假定，如阻力正比于 $v$ 或 $v^2$ 等．甚至可完全忽略阻力．而另一个例子，当我们确定汽车消耗的燃料随着速度如何变化时，关于阻力的不同的假定，能够引导出预测的行驶里程按照 $C_1v^{-1}$ 或 $C_2v^{-2}$ 这样的规律变化的模型．这样产生的问题能按如下方式考虑：首先，用一些收集到的数据按某一方法选择 $C_1$ 和 $C_2$，该方法从每一族可能的曲线中选取一条能最好地拟合数据的曲线，然后选择一个最适合于研究的状态的模型．

当问题很复杂而难以建立能够解释该特殊情形的模型时，会有另一种情况．例如，当子模型涉及偏微分方程，而它们没有封闭形式的解时，那么构造一个主模型，在不使用计算机的情况下得到解答并进行分析的可能性很小．或者，问题中涉及的有显著影响的独立变量太多，人们甚至不想去构造一个明确的模型．在这种时候，为了在数据所在的范围内研究独立变量的行为，可能必须进行一些实验研究．

在分析一个数据集合时，前面的讨论实际上指明了三个可能需要解决的任务：

1. 按照一个或一些选出的模型类型对数据进行拟合．

2. 从一些已经拟合的类型中选取最合适的模型．例如，我们需要判断最佳拟合指数模型是否比一个最佳多项式模型更好．

3. 根据收集的数据做出预报．

在前两个任务中，可能存在一个或多个模型，似乎都能解释已观测到的行为．这一章将围绕着模型拟合来讨论这两种情形．在第三种情形中，不存在一个解释已观测到的行为的模型，而是存在一个数据点的集合，该集合能用来预测某个你所感兴趣的数量范围内的行为．从本质上来看，我们希望基于收集到的数据构造一个经验模型．第4章将围绕着内插来研究构造这样的经验模型．了解模型的拟合和内插在哲学上和数学上的差异是重要的．

### 模型的拟合和内插间的关系

我们来分析前面指出的三个任务，以便确定在每一情形必须做些什么．在任务1，必须明

确最佳模型的正确意义，以及由此而产生的需要解决的数学问题．在任务 2，为了比较不同类型的模型需要有一个判定准则．在任务 3 中，为了决定如何在观测的数据点间做出预测，也要明确一个判定准则．

要注意在上述每种情况中建模者的态度上的差异．在前两个模型拟合的任务中，要极力猜测出某种关系．建模者愿意接受模型和收集到的数据点间的某些偏差，以便有一个满意地解释所研究的问题的模型．实际上，建模者会预想到模型和数据两者都可能有误差．另一方面，在插值时，建模者会受到细心收集和分析过的数据的强力引导，曲线应追踪数据的趋向，在数据点间做出预测．这时，建模者一般很少会对插值曲线附加明确的意义．在各种情况，可能建模者最终都想用模型进行预测．然而，做模型拟合时，建模者更强调为数据提供模型，而做插值时，建模者对收集到的数据给予了更大的信任，而较少注意模型的形式意义．在此意义上，解释性的模型是理论推动的，而预测模型是数据推动的．

用一个例子说明前面的看法．假设为了研究两个变量 $y$ 和 $x$ 之间的关系，我们收集到图 3-1 所画出的数据，如果建模者仅根据图中的数据做出预测，他可以使用样条插值之类的技术(我们在第 4 章研究此技术，让光滑的多项式通过这些点．见图 3-2)．注意，图 3-2 内插曲线通过数据点，在观测点的整个范围内追踪变量间的趋势．

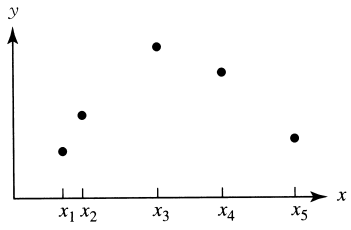


图 3-1　变量 $y$ 和 $x$ 的观测值

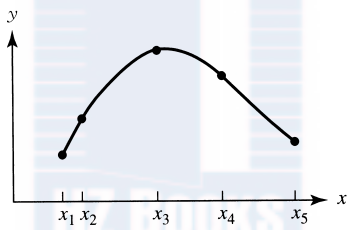


图 3-2　用光滑多项式对数据进行插值

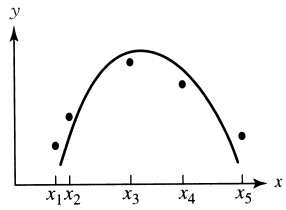


图 3-3　对数据点拟合抛物线 $y=C_1x^2+C_2x+C_3$

假设在研究图 3-1 所示的特定行为时，建模者做出假定，引导出一个 $y=C_1x^2+C_2x+C_3$ 形式的二次式模型或抛物线的预想．这时图 3-1 的数据将用于确定常数 $C_1$，$C_2$ 和 $C_3$，以选择最佳的抛物线(图 3-3)．抛物线可能与某些数据点甚至全部点有些偏离，这应是无关紧要的．要注意图 3-2 和图 3-3 中曲线在值 $x_1$ 和 $x_5$ 附近所做预测值间的差异．

建模者可能会发现在同一问题中需要拟合一个模型，同时还需要进行插值．一个给定类型的最佳拟合模型可能被证明是难于控制的甚至是不可能的，因为接下来的分析可能会涉及积分、微分之类的操作，这个时候模型可以用插值曲线(如多项式)代替．以便更容易对其微分或积分．例如为了简化后续的分析，一个用来对方波建模的阶梯函数可以替换为三角近似．在这些例子中，建模者希望用插值曲线近似并能贴近所代替的函数的基本特征．这种类型的插值通常称为**逼近**．这通常在导论性的数值分析课程中有介绍．

## 建模过程中的误差来源

在讨论基于曲线拟合还是基于插值的决策所依据的准则之前，我们需要考察建模过程中何处会引起误差．如果忽视考虑误差，过于信任中间结果，会在后续的阶段中招致失败的决策．我们

的目标是保证建模的整个过程在计算上是适宜的，并考虑到前面各步带来的累积误差的影响．

为了容易讨论，将误差分为下列几类：

1. 公式化的误差
2. 截断误差
3. 舍入误差
4. 测量误差

**公式化的误差** 可源于：一些变量可忽略的假设条件，或在各种子模型中描述变量间关系的过分简化．例如，为第 2 章的刹车距离确定一个子模型时，我们完全忽略了路的摩擦力，并假定由空气阻力引起的阻力有很简单的特征关系．即使在最佳模型中，也会有公式化的误差．

**截断误差** 可归因于解一个数学问题所用的数值方法．例如，可能要用幂级数表示的多项式近似 $\sin x$，

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \cdots$$

当级数截尾后用多项式近似时将引起一个误差．

**舍入误差** 是由计算时使用有限小数位的机器引起的．因为仅用有限位不能精确地表示全部的数，总是有舍入误差．例如，考虑一个 8 位的计算器或计算机，数 1/3 表示成 0.333 333 33，这样 3 倍的 1/3 是 0.999 999 99 而不是真实的值 1．误差 $10^{-8}$ 是舍入引起的．理想的实数 1/3 是小数 0.333 333 33 的无穷循环．但一个计算器或计算机能做的算术运算仅能用有限精度的数．当连续完成许多算术运算时，每一次有自己的舍入，累积的舍入能够显著地改变答案的数值．当我们使用计算装置时，舍入正是一个我们必须面对并谨慎处理的事情．

**测量误差** 是由数据收集过程中的不精确性引起的．不精确性可以包含：记录或报告一个数据时的人为错误，或实验室设备的测量精度限制等多种情况．例如，在刹车距离问题中，应预期到在反应距离和刹车距离的数据中会有应该考虑的测量误差．

## 3.1 用图形为数据拟合模型

假定建模者已做了某些假定，引出了某种模型．一般模型会包含一个或多个参数，要收集充足的数据来确定这些参数．现在来考虑数据收集的问题．

采集多少个数据点，要在观测它们的费用和模型所要求的精度间进行权衡．数据点至少需要与模型曲线中任意常数一样多．要用最佳的拟合方法，确定出每一个任意常数，要求有更多的点．将要使用的模型的范围决定了独立变量的区间端点．

在此区间中，数据点的跨度也是一个重要问题，因为区间中模型必须拟合得特别好的一部分可以用不等的跨度进行加权．在预期模型使用特别多的地方或独立变量会突然变化的地方应选取更多的数据点．

即使实验已精心设计，并极其细心地执行，建模者仍需在拟合模型前评估数据的精确性．数据是如何收集的？收集过程中测量设备的精度如何？有无有疑问的点？在评价或删除(替换)有疑问的数据时，应将一个数据点看做是一个置信区间而不是一个单独的点．图 3-4 表示了这一想法．每一区间的长度应与在数据收集过程中的误差的评估相一致．

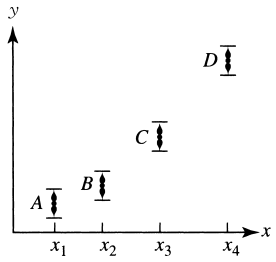


图 3-4 每一点看做一个可信的区间

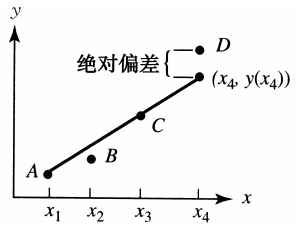


图 3-5 极小化拟合直线带来的绝对偏差和

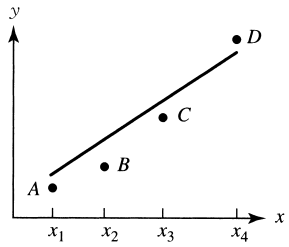


图 3-6 极小化拟合直线带来的最大偏差

## 对原始数据拟合视觉观测的模型

假设想要对图 3-4 所示数据拟合模型 $y=ax+b$. 应如何选择 $a$ 和 $b$，使直线最好地拟合数据？从图上看，当存在两个以上点时，不能期望它们均精确地处于一直线上．尽管一条直线精确地做出了变量 $x$ 和 $y$ 间关系的模型，一些数据点和直线间总存在一些纵向差异．我们称这些纵向差异为**绝对偏差**(图 3-5)．最佳拟合直线可极小化这些绝对偏差的和，这就引出图 3-5 中描画的模型．虽然在极小化绝对偏差和方面是成功的，但个别点的绝对偏差可能相当大，例如图 3-5 中的 $D$. 如果建模者对一个数据点的精确度有信心，则能由拟合的直线在该点的邻近处做出预测．再看另一种选择，按极小化任一点的最大偏差选择直线，对数据点运用这一准则，则会绘出图 3-6 的直线．

虽然这些对数据点拟合一条直线的视觉方法不是十分精确，但这些方法的不精确性往往与建模过程的精度相称．假定的粗糙和数据收集中涉及的不精确性，可能无法保证更先进的分析．在这种状态下，盲目应用 3.2 节给出的某一分析方法，可能会使模型远不如人们图形观测到的合适．对数据图形模型拟合的进一步视觉检查会立刻给出拟合是否好以及何处拟合得好的印象．然而，在由计算机解析地拟合大量数据的问题中，常会疏漏这些重要的考虑．因为建模过程的模型拟合部分比其他阶段似乎更为精细，更多分析，存在着不恰当地信任数值计算的倾向．

## 变换数据

多数人视觉上仅限于拟合直线，那如何用图示拟合曲线作为模型呢？例如，表 3-1 收集到的数据，猜想用 $y=Ce^x$ 形式的关系作为子模型．

**表 3-1 收集的数据**

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $y$ | 8.1 | 22.1 | 60.1 | 165 |

**表 3-2 由表 3-1 变换出的数据**

| $x$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| $\ln y$ | 2.1 | 3.1 | 4.1 | 5.1 |

模型说 $y$ 正比于 $e^x$，如果画一个 $y$ 对 $e^x$ 的图，几乎会近似于一条直线，图 3-7 描述了这一状况．由于画出的数据点近似地沿过原点的一条直线散布，可得到结论：假定正比是有道理的，从图上看，直线的斜率近似为

$$C=\frac{165-60.1}{54.6-20.1}\approx 3.0$$

现在先来考虑另一在多种问题中使用的技术，在方程 $y=Ce^x$ 的两边取对数，得

$$\ln y=\ln C+x$$

这一表达式在变量 $\ln y$ 和 $x$ 之间是一直线方程，数 $\ln C$ 是 $x=0$ 时的截距．变换后的数据列在表 3-2 中，图画在图 3-8 中．当有大量数据时，可使用半对数坐标纸或计算机画图．

从图 3-8，可近似确定截距 $\ln C$ 是 1.1，给出 $C=e^{1.1}\approx 3.0$，与前面一致．

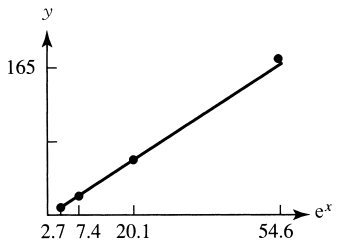


图 3-7 用表 3-1 的数据画的 $y$ 对 $e^x$ 的图

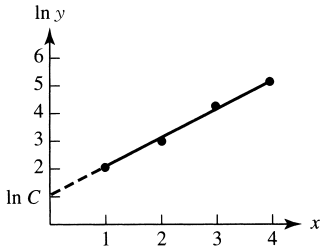


图 3-8 表 3-2 数据的 $\ln y$ 对 $x$ 的图

可执行各种其他曲线的类似的变换，使变换后产生的变量间形成线性关系．例如，如果 $y=x^a$，那么

$$\ln y = a\ln x$$

变换后的变量 $\ln y$ 和 $\ln x$ 间是线性关系．因此有大量数据画图时可用 log-log 坐标纸或计算机．

现在做一项重要的观察．假设我们像图 3-8 那样，做一个变换，画 $\ln y$ 对 $x$ 的图，找出直线，成功地极小化了变换后数据点的绝对偏差和，那么直线确定了 $\ln C$，逆转过程后为正比例常数 $C$. 虽然不是很明显，但在 $ke^x$ 形式的指数曲线族中，已得到的模型 $y=Ce^x$ 不是极小化原始数据点的极小化绝对偏差和的指数曲线(画 $y$ 对 $x$ 的图)．在后面的讨论中，在图示和分析中都将涉及这一重要的见解．当进行 $y=\ln x$ 形式的变换时，距离的概念受到破坏，虽然从图形分析观察到的固有限制来看，拟合是适宜的，但建模者必须认识到这个破坏，并且应该用图解核查模型，从图解中做出预测或结论，这里提及的是原始数据的 $y$ 对 $x$ 的图而不是变换变量的图．

现在用一个例子说明变换如何破坏了 $xy$ 平面的距离．考虑图 3-9 画出的数据，假定数据期望拟合一个 $y=Ce^{1/x}$ 形式的模型，使用前面的对数变换．有

$$\ln y = \frac{1}{x} + \ln C$$

按原始数据 $\ln y$ 对 $1/x$ 的点图如图 3-10 所示．从图可注意到变换如何破坏了原始数据点间的距离，它将全部点挤在一起．如果对图 3-10 中变换后的数据点拟合一条直线，绝对偏差相当小(即用图 3-10 的尺度计算是小的，而不是图 3-9 的尺度)．如果对图 3-9 的数据画出拟合模型 $y=Ce^{1/x}$，将看到曲线拟合数据效果相当差，如图 3-11 所示．

从上一例子可以看到，如果建模者使用变换时不是很小心，他可能会选中一个相当差的模型，在比较可选模型时，这一情况特别重要．必须与原始数据(我们例子中图解 3-9 所画的)一起进行全部比较，不然在选择最佳模型时会导致非常严重的错误．那样会导致可能会根据变换的特点选取最佳模型，而不是根据模型的价值以及拟合原始数据的程度，虽然从这幅图中可以很明显地看出进行变换的风险，但由于许多计算机程序在拟合数据时先进行了转换，建模者如果不是特别明白的话，还是会受到愚弄．如果建模者企图运用一些指示量，如绝对偏差和，来确定某一特别子模型是否适当或在备选的子模型中做选择，建模者首先必须明确这些指示量是

如何计算的.

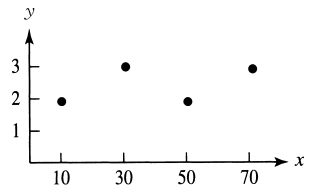


图 3-9 收集的数据点的图

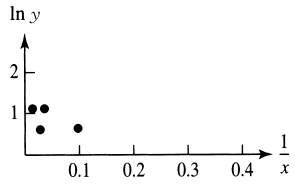


图 3-10 变换后数据点的图

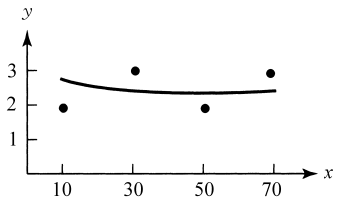


图 3-11 从图 3-10 取值 $\ln C=0.9$ 的 $y=Ce^{1/x}$ 曲线图

## 习题

1. 正常情况下，图 3-2 的模型会用来预测 $x_1$ 和 $x_5$ 之间的状况，用模型预测 $x$ 小于 $x_1$ 或大于 $x_5$ 时的 $y$ 会有什么危险？不妨设我们是在为投掷棒球的弹道建模.

2. 下表给出了在一根钢丝弹簧上施加拉力 $S$(单位：磅/平方英寸)后每英寸的伸长 $e$(单位：英寸/英寸). 画出数据，检验模型 $e=c_1S$，从图上估计 $c_1$.

| $S(\times10^{-3})$ | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $e(\times10^5)$ | 0 | 19 | 57 | 94 | 134 | 173 | 216 | 256 | 297 | 343 | 390 |

3. 下面的数据中，$x$ 是美国黄松在树身中部测得的直径(单位英寸)，$y$ 是体积的度量，即用 10 除后的板英尺数. 变换数据画图，检验模型 $y=ax^b$. 如果模型看似合适，从图上估计模型的参数 $a$ 和 $b$.

| $x$ | 17 | 19 | 20 | 22 | 23 | 25 | 28 | 31 | 32 | 33 | 36 | 37 | 38 | 39 | 41 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 19 | 25 | 32 | 51 | 57 | 71 | 113 | 141 | 123 | 187 | 192 | 205 | 252 | 259 | 294 |

4. 下面的数据，$V$ 代表一个平均的步行速率，$P$ 代表人群总体的人数. 我们希望知道是否能通过观测人们的步行速度来预测总体的人数. 画出数据图，猜测有何种关系？画出适当的变换数据，检验下列模型.

(a) $P=aV^b$

(b) $P=a\ln V$

| $V$ | 2.27 | 2.76 | 3.27 | 3.31 | 3.70 | 3.85 | 4.31 | 4.39 | 4.42 |
|---|---|---|---|---|---|---|---|---|---|
| $P$ | 2500 | 365 | 23 700 | 5491 | 14 000 | 78 200 | 70 700 | 138 000 | 304 500 |

| $V$ | 4.81 | 4.90 | 5.05 | 5.21 | 5.62 | 5.88 |
|---|---|---|---|---|---|---|
| $P$ | 341 948 | 49 375 | 260 200 | 867 023 | 1 340 000 | 1 092 759 |

5. 下面数据反映了在六个星期时间中果蝇群体的增长. 对一个适当的数据集画图. 检验下列模型，估计模型的参数.

(a) $P=c_1t$

(b) $P=ae^{bt}$

| $t$(天数) | 7 | 14 | 21 | 28 | 35 | 42 |
|---|---|---|---|---|---|---|
| $P$(观测到的果蝇的数目) | 8 | 41 | 133 | 250 | 280 | 297 |

6. 下面的数据表示以 1990 年为基准的(假设的)能源消费，画出数据图，画出变换数据. 检验模型 $Q=ae^{bt}$，用图形估出模型的参数.

| $x$ | 年 | 消费 $Q$ | $x$ | 年 | 消费 $Q$ |
|---|---|---|---|---|---|
| 0 | 1900 | 1.00 | 60 | 1960 | 66.69 |
| 10 | 1910 | 2.01 | 70 | 1970 | 134.29 |
| 20 | 1920 | 4.06 | 80 | 1980 | 270.43 |
| 30 | 1930 | 8.17 | 90 | 1990 | 544.57 |
| 40 | 1940 | 16.44 | 100 | 2000 | 1096.63 |
| 50 | 1950 | 33.12 | | | |

7. 1601 年，开普勒成为布拉格天文台的台长．开普勒曾经帮助 Tycho Brahe 收集了 13 年有关火星的相对运动的观察资料．到 1609 年开普勒已经形成了他的头两条定律：

ⅰ)每个行星都沿一条椭圆轨道运行，太阳在该椭圆的一个焦点处．

ⅱ)对每个行星来说，在相等的时间里该行星和太阳的连线扫过相等的面积．

开普勒花了许多年证实这些定律，并构造出第三条定律．该定律与运行轨道的周期和距离太阳的平均距离有关．

(a)使用当今的数据画出周期时间 $T$ 对平均距离 $r$ 的图．

| 行星 | 周期(天) | 到太阳的平均距离(百万公里) | 行星 | 周期(天) | 到太阳的平均距离(百万公里) |
|---|---|---|---|---|---|
| 水星 | 88 | 57.9 | 木星 | 4329 | 778.1 |
| 金星 | 225 | 108.2 | 土星 | 10 753 | 1428.2 |
| 地球 | 365 | 149.6 | 天王星 | 30 660 | 2837.9 |
| 火星 | 687 | 227.9 | 海王星 | 60 150 | 4488.9 |

(b)假定关系的形式为

$$T = Cr^a$$

画出 $\ln T$ 对 $\ln r$ 的图，用图确定参数 $C$ 和 $a$．模型合适吗？试构造开普勒第三定律．

## 3.2 模型拟合的解析方法

在这一节研究对收集的数据点拟合曲线的几个准则．每一准则提供了一个从给定的族中选出最佳的曲线的方法．依照这一准则，曲线最精确地代表了数据．另外还讨论了几个准则是如何相联系的．

### 切比雪夫近似准则

在前一节我们对一个数据点集用图形拟合一条直线，所用的最佳拟合准则之一是极小化直线到任一对应的数据点的最大距离．现在来分析这一几何结构．给定 $m$ 个数据点的集合$(x_i, y_i)$，$i=1, 2, \cdots, m$，用直线 $y=ax+b$ 拟合该集合，确定参数 $a$ 和 $b$，使任一数据点$(x_i, y_i)$和其对应的直线上的点$(x_i, ax_i+b)$间的距离最小，也就是对整个数据点集极小化最大绝对偏差 $|y_i-y(x_i)|$．现在将这一准则推广．

给定某种函数类型 $y=f(x)$和 $m$ 个数据点$(x_i, y_i)$的一个集合，对整个集合极小化最大绝对偏差 $|y_i-f(x_i)|$，即确定函数类型 $y=f(x)$的参数从而极小化数量

$$\text{Maximum}\ |y_i - f(x_i)| \quad i = 1, 2, \cdots, m \tag{3-1}$$

这一重要的准则常称为**切比雪夫**(Chebyshev)**近似准则**．切比雪夫准则的困难在于实际应用中通常很复杂，至少是仅用初等演算时很复杂．应用这一准则所产生的最优化问题可能需要

高级的数学方法，或者要用计算机的数值算法．

例如，设我们要度量图 3-12 表示的线段 $AB$，$BC$ 和 $AC$，假定你的测量产生估计 $AB=13$，$BC=7$，$AC=19$.

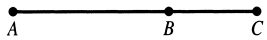

图 3-12 线段 $AC$ 分成 $AB$ 和 $BC$ 两段

可以预想到在一次实地的测量中会有矛盾的结果．这时，$AB$ 和 $BC$ 值加起来是 20 而不是估出的 $AC=19$. 现在用切比雪夫准则来解决这一个单位的差异，也就是用一个方法为三个线段指定数值，使得指定的和观测的任一对应数对间的最大偏差达到极小．假定对每一次测量有相同的信任度，这样每一测量值有相等的权值．这种情况下，差异应等可能地分配到每一线段，结果会预测 $AB=12\frac{2}{3}$，$BC=6\frac{2}{3}$，$AC=19\frac{1}{3}$. 每一绝对偏差为 1/3. 减少任一偏差会招致另一个偏差的增加(记住 $AB+BC$ 必须等于 $AC$). 现在把这一问题用公式表达出来．

令 $x_1$ 代表线段 $AB$ 长度的真值，$x_2$ 代表 $BC$ 的真值．为易于表示，令 $r_1$，$r_2$，$r_3$ 表示真值和测量值间的差异．即

$$x_1-13=r_1(\text{线段 } AB)$$
$$x_2-7=r_2(\text{线段 } BC)$$
$$x_1+x_2-19=r_3(\text{线段 } AC)$$

数值 $r_1$，$r_2$，$r_3$ 称为**残差**．注意，残差可以是正的也可以是负的，但绝对偏差总是正的．

如果用切比雪夫近似准则，应指定 $r_1$，$r_2$，$r_3$ 的值，使三个数值 $|r_1|$，$|r_2|$，$|r_3|$ 的最大者达到最小．如果记最大的数为 $r$，那么我们要求

最小化 $r$

约束有三个条件：

$$|r_1|\leqslant r \quad \text{或} \quad -r\leqslant r_1\leqslant r$$
$$|r_2|\leqslant r \quad \text{或} \quad -r\leqslant r_2\leqslant r$$
$$|r_3|\leqslant r \quad \text{或} \quad -r\leqslant r_3\leqslant r$$

这些条件的每一个可替换为两个不等式．例如 $|r_1|\leqslant r$ 能替换为 $r-r_1\geqslant 0$ 和 $r+r_1\geqslant 0$. 对每一条件均这样做，问题则叙述为经典的数学问题

最小化 $r$

满足约束条件

$$\begin{aligned}
r-x_1 \qquad\quad &+13\geqslant 0 \quad (r-r_1\geqslant 0)\\
r+x_1 \qquad\quad &-13\geqslant 0 \quad (r+r_1\geqslant 0)\\
r \qquad -x_2&+7\geqslant 0 \quad (r-r_2\geqslant 0)\\
r \qquad +x_2&-7\geqslant 0 \quad (r+r_2\geqslant 0)\\
r-x_1-x_2&+19\geqslant 0 \quad (r-r_3\geqslant 0)\\
r+x_1+x_2&-19\geqslant 0 \quad (r+r_3\geqslant 0)
\end{aligned}$$

这一问题称为**线性规划问题**．在第 7 章将讨论线性规划．即使很大的线性规划也能由计算机通过执行著名的**单纯形方法**的算法解出．在前面线段的例子中，单纯形方法产生 $r=1/3$ 的

极小值和 $x_1=12\frac{2}{3}$，$x_2=6\frac{2}{3}$.

推广这一过程，给定某一函数类型 $y=f(x)$，其参数待定，以及给定 $m$ 个数据点$(x_i, y_i)$的一个集合，并确定出残差为 $r_i=y_i-f(x_i)$. 如果 $r$ 代表这些残差的最大绝对值，那么问题表示成

$$\text{最小化 } r$$

满足约束条件

$$\left.\begin{aligned} r-r_i &\geqslant 0\\ r+r_i &\geqslant 0\end{aligned}\right\} \quad \text{对 } i=1,2,\cdots,m$$

虽然在第 7 章讨论线性规划，但这儿要强调这一过程产生的模型并不都是线性规则．例如，考虑拟合函数 $f(x)=\sin kx$. 还要注意许多计算机执行单纯形算法，仅允许变量为非负值．用简单的替换即可完成这一要求(参看习题 5).

我们将要看到，有另一准则能方便地解决最优化问题．正是由于这一原因，在对有限的数据点拟合一条曲线时不常使用切比雪夫准则．然而当极小化最大绝对偏差很重要的时候仍应考虑使用这一准则(在第 7 章考虑这一准则的几个应用)．进一步在用一函数代替一个区间上定义的另一个函数时，构成切比雪夫准则的原则是极其重要的，在该区间上两个函数间的最大差异必须达到最小．逼近论研究这一论题，而且导论性的数值分析中都有这部分内容．

## 极小化绝对偏差之和

在 3.1 节用图示为数据拟合直线时，准则之一是极小化数据点和拟合线上对应的点间绝对偏差的总和．这一准则可归纳为：给定某一函数类型 $y=f(x)$，以及 $m$ 个数据点$(x_i, y_i)$的集合，极小化绝对偏差 $|y_i-f(x_i)|$ 的和，也就是确定函数类型 $y=f(x)$的参数，极小化

$$\sum_{i=1}^{m}|y_i-f(x_i)| \tag{3-2}$$

如果令 $R_i=|y_i-f(x_i)|$，$i=1$，$2$，$\cdots$，$m$，代表每一绝对偏差，那么前面的准则(3-2)可解释成将由一条数量 $R_i$ 加在一起构成的一直线的长度极小化．图 3-13 说明了 $m=2$ 的情况．

虽然在 3.1 节当函数类型 $y=f(x)$为直线时，采用几何方法应用了这一准则，但一般性准则暴露出严重的问题．使用计算解决最优化问题，必须将和式(3-2)对 $f(x)$的参数求导，以找出临界点．然而由于出现了绝对值，这个和式的各种微分不是连续的．所以下面不以这一准则为目标．在第 7 章我们考虑这一准则的另一应用，给出数值近似解的技术．

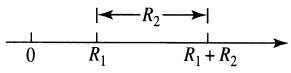


图 3-13 极小化绝对偏差和的一个几何解释

## 最小二乘准则

现在最常用的曲线拟合准则是**最小二乘准则**．使用与前面相同的记号，问题是确定函数类型 $y=f(x)$的参数，极小化和数

$$\sum_{i=1}^{m}|y_i-f(x_i)|^2 \tag{3-3}$$

用此方法解决产生的最优化问题仅需使用几个变量的演算，所以容易普及．然而在现代数学规划技术上的进展(比如解决许多切比雪夫准则应用的单纯形方法)以及准则(3-2)的近似解的数值方法上的进步，削弱了这一优势．当从概率角度加以考虑，假定误差是随机分布时，会提高对使用最小二乘方法的评价，但我们到下一章的最后才会讨论到概率．

现在给出最小二乘准则的几何解释．考虑三个点的情况，以 $R_i=|y_i-f(x_i)|$ 记观测到的和预测的值间的绝对偏差，$i=1,2,3$．将 $R_i$ 考虑为偏差向量的一个数量分量，绘在图 3-14 中．那么向量 $\boldsymbol{R}=R_1\boldsymbol{i}+R_2\boldsymbol{j}+R_3\boldsymbol{k}$ 代表了观测值和预测值间产生的偏离．这一偏离向量的长度给定为

$$|\boldsymbol{R}|=\sqrt{R_1^2+R_2^2+R_3^2}$$

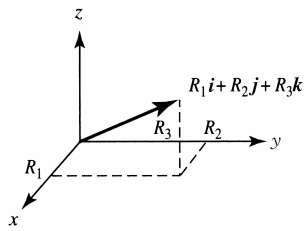


图 3-14 最小二乘准则的几何解释

要极小化 $|\boldsymbol{R}|$ 可以极小化 $|\boldsymbol{R}|^2$(参看问题 1)．所以最小二乘问题是：确定函数类型 $y=f(x)$的参数，以便极小化

$$|\boldsymbol{R}|^2=\sum_{i=1}^{3}R_i^2=\sum_{i=1}^{3}|y_i-f(x_i)|^2$$

也就是说，可以解释最小二乘准则为极小化向量的长度．该向量的坐标代表了观测值和预测值之间的绝对偏差．

对于超过三对的数据集，我们不能再提供一个几何图形，但仍然可以说：我们极小化了其坐标为观测值和预测值间的绝对偏差之向量的长度.

## 谈谈准则

三个曲线拟合准则的几何解释有助于在量化描述方面比较准则．极小化绝对偏差和将赋予每一数据点相等的权值来平均这些偏差．切比雪夫准则对潜在有大偏差的单个点给予更大的权值．最小二乘准则是根据与中间某处的远近来加权，其权与单个点具有的显著偏离有关．由于解析地运用切比雪夫准则和最小二乘准则更方便些，我们现在寻求一个方法来谈谈用这两个准则产生的偏差．

假设用切比雪夫准则，并解出所产生的优化问题，产生函数 $f_1(x)$．拟合产生的绝对偏差定义为

$$|y_i-f_1(x_i)|=c_i,\quad i=1,2,\cdots,m$$

现在定义 $c_{\max}$ 为绝对偏差 $c_i$ 中的最大者．$c_{\max}$ 有一个相连的显著特征．因为函数 $f_1(x)$的参数是按极小化 $c_{\max}$ 的值确定的，$c_{\max}$ 是可获得的最小的极大绝对偏差．

另一方面，假设应用最小二乘准则，解出所产生的优化问题，产生函数 $f_2(x)$．由拟合产生的绝对偏差，如下给定：

$$|y_i-f_2(x_i)|=d_i,\quad i=1,2,\cdots,m$$

定义 $d_{\max}$ 为这些绝对偏差 $d_i$ 的最大者，对于前面讨论的 $c_{\max}$ 的特殊特征，现在还仅能说 $d_{\max}$ 至少比 $c_{\max}$ 大，但可更精确些谈谈 $d_{\max}$ 和 $c_{\max}$.

最小二乘准则涉及 $d_i$ 的特殊特征是它们的平方和是可获得的此类平方和中的最小者．这必然有

$$d_1^2+d_2^2+\cdots+d_m^2 \leqslant c_1^2+c_2^2+\cdots+c_m^2$$

由于对每个 $i$ 有 $c_i \leqslant c_{\max}$，这些不等式推演出

$$d_1^2+d_2^2+\cdots+d_m^2 \leqslant mc_{\max}^2$$

或

$$\sqrt{\frac{d_1^2+d_2^2+\cdots+d_m^2}{m}} \leqslant c_{\max}$$

为了方便讨论，定义

$$D=\frac{\sqrt{d_1^2+d_2^2+\cdots+d_m^2}}{m}$$

那么

$$D \leqslant c_{\max} \leqslant d_{\max}$$

最后的这一关系式是很有启发性的．假设在一个运用最小二乘准则更方便的特殊状态，但又涉及可产生的最大绝对偏差 $c_{\max}$，如果我们计算 $D$，则获得一个 $c_{\max}$ 的下界，而 $d_{\max}$ 给出了一个上界．这样如果 $D$ 和 $d_{\max}$ 之间有巨大的差异，那么建模者应考虑应用切比雪夫准则．

## 习题

1. 用初等演算，说明 $y=f(x)$ 的最小值点和最大值点出现在 $y=f^2(x)$ 的最小值点和最大值点之间．假定 $f(x)\geqslant 0$，为什么能通过极小化 $f^2(x)$ 来极小化 $f(x)$？

2. 对下列每一数据集，构造数学模型写出公式，极小化数据和直线 $y=ax+b$ 间的最大偏差．如果有计算机可用，解出 $a$ 和 $b$ 的估计．

(a)

| $x$ | 1.0 | 2.3 | 3.7 | 4.2 | 6.1 | 7.0 |
|---|---|---|---|---|---|---|
| $y$ | 3.6 | 3.0 | 3.2 | 5.1 | 5.3 | 6.8 |

(b)

| $x$ | 29.1 | 48.2 | 72.7 | 92.0 | 118 | 140 | 165 | 199 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 0.0493 | 0.0821 | 0.123 | 0.154 | 0.197 | 0.234 | 0.274 | 0.328 |

(c)

| $x$ | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 | 5.5 |
|---|---|---|---|---|---|---|---|
| $y$ | 4.32 | 4.83 | 4.27 | 5.74 | 6.26 | 6.79 | 7.23 |

3. 对下列数据，构造数学模型写出公式，极小化数据和模型 $y=c_1x^2+c_2x+c_3$ 间的最大偏差．如果有计算机可用，解出 $c_1$，$c_2$ 和 $c_3$ 的估计．

| $x$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|
| $y$ | 0.06 | 0.12 | 0.36 | 0.65 | 0.95 |

4. 对下列数据，构造数学模型写出公式，极小化数据和模型 $P=ae^{bt}$ 间的最大偏差．如果有计算机可用，解出 $a$ 和 $b$ 的估计．

| $t$ | 7 | 14 | 21 | 28 | 35 | 42 |
|---|---|---|---|---|---|---|
| $P$ | 8 | 41 | 133 | 250 | 280 | 297 |

5. 设变量 $x_1$ 可以是任意实数值．说明下列使用非负变量 $x_2$ 和 $x_3$ 的替换允许 $x_1$ 取任意实数值：

$$x_1 = x_2 - x_3, \quad x_1 \text{ 无限制}$$

且

$$x_2 \geqslant 0 \quad \text{和} \quad x_3 \geqslant 0$$

这样，倘若计算机只允许用非负变量，这一代替允许解出变量 $x_2$ 和 $x_3$ 的线性规划，然后再找出变量 $x_1$ 的值.

## 3.3 应用最小二乘准则

假设我们预想到一个确定形式的模型，并且已经收集了数据并进行了分析. 在这一节用最小二乘准则来估计各种类型曲线的参数.

### 拟合直线

设预期模型的形式为 $y=Ax+B$，并决定用 $m$ 个数据点$(x_i, y_i)(i=1, 2, \cdots, m)$来估计 $A$ 和 $B$. 用 $y=ax+b$ 记作 $y=Ax+B$ 的最小二乘估计. 这时运用最小二乘准则(3-3)，则要求极小化

$$S = \sum_{i=1}^{m}[y_i - f(x_i)]^2 = \sum_{i=1}^{m}(y_i - ax_i - b)^2$$

最优的一个必要条件是两个偏导数 $\partial S/\partial a$ 和 $\partial S/\partial b$ 等于零. 得方程

$$\frac{\partial S}{\partial a} = -2\sum_{i=1}^{m}(y_i - ax_i - b)x_i = 0$$

$$\frac{\partial S}{\partial b} = -2\sum_{i=1}^{m}(y_i - ax_i - b) = 0$$

重写这些方程得出

$$\left.\begin{aligned} a\sum_{i=1}^{m}x_i^2 + b\sum_{i=1}^{m}x_i &= \sum_{i=1}^{m}x_iy_i \\ a\sum_{i=1}^{m}x_i + mb &= \sum_{i=1}^{m}y_i \end{aligned}\right\} \tag{3-4}$$

将 $x_i$ 和 $y_i$ 的全部值带入，从前面的方程可解出 $a$ 和 $b$，用消去法很容易得到参数 $a$ 和 $b$ 的解(参看本节末习题 1). 得出

$$a = \frac{m\sum x_iy_i - \sum x_i\sum y_i}{m\sum x_i^2 - (\sum x_i)^2}, \quad \text{斜率} \tag{3-5}$$

$$b = \frac{\sum x_i^2\sum y_i - \sum x_iy_i\sum x_i}{m\sum x_i^2 - (\sum x_i)^2}, \quad \text{截距} \tag{3-6}$$

对任一数据点集，容易写出计算机计算 $a$ 和 $b$ 值的程序. 方程(3-4)称为**正规方程**.

### 拟合幂曲线

现在对一个给定的数据点集用最小二乘准则拟合 $y=Ax^n$ 形式的曲线，$n$ 为固定数. 研究模型 $f(x)=ax^n$ 的最小二乘估计，应用该准则要求极小化

$$S = \sum_{i=1}^{m}[y_i - f(x_i)]^2 = \sum_{i=1}^{m}[y_i - ax_i^n]^2$$

最优化的必要条件为导数 $\mathrm{d}S/\mathrm{d}a$ 等于零，给出方程

$$\frac{\mathrm{d}S}{\mathrm{d}a} = -2\sum_{i=1}^{m} x_i^n [y_i - ax_i^n] = 0$$

从方程解出 $a$，得

$$a = \frac{\sum x_i^n y_i}{\sum x_i^{2n}} \tag{3-7}$$

记住方程(3-7)中，$n$ 是固定的.

同样可以将最小二乘准则用于其他模型. 应用该方法的限制在于计算最优化过程中要求的各种导数，令这些导数为零，解这些得到的方程，求出模型类型中的参数.

例如，用表 3-3 给出的数据拟合 $y=Ax^2$，并预测 $x=2.25$ 时 $y$ 的值.

**表 3-3 拟合 $y=Ax^2$ 数据集**

| $x$ | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 |
|---|---|---|---|---|---|
| $y$ | 0.7 | 3.4 | 7.2 | 12.4 | 20.1 |

这时，最小二乘估计 $a$ 给定为

$$a = \frac{\sum x_i^2 y_i}{\sum x_i^4}$$

计算出 $\sum x_i^4 = 61.1875$，$\sum x_i^2 y_i = 195.0$. 得到 $a=3.1869$(到小数点后四位). 这一计算给出了一个最小二乘近似模型

$$y = 3.1869x^2$$

当 $x=2.25$ 时，预测 $y$ 值为 16.1337.

## 经变换的最小二乘拟合

在理论上最小二乘准则很易应用，但在实践上可能是有困难的. 例如，用最小二乘准则拟合模型 $y=A\mathrm{e}^{Bx}$. 研究模型的最小二乘估计 $f(x)=a\mathrm{e}^{bx}$，应用该准则极小化

$$S = \sum_{i=1}^{m} [y_i - f(x_i)]^2 = \sum_{i=1}^{m} [y_i - a\mathrm{e}^{bx_i}]^2$$

最优化的必要条件是，$\partial S/\partial a = \partial S/\partial b = 0$. 列出条件式，解这个非线性方程组是不容易的. 许多简单的模型会产生很复杂的求解过程，或者很难解的方程组. 基于这一原因，我们要使用变换，得出近似的最小二乘模型.

在 3.1 节对数据拟合直线，经常发现先变换数据再对变换后的数据拟合直线很方便. 例如，图形拟合 $y=C\mathrm{e}^x$，可以画出 $\ln y$ 对 $x$ 的图，对变换后的数据拟合直线. 同样的想法可用于最小二乘准则，简化拟合过程的计算. 特别地，如果找到一个方便的变换，问题变成在变换后的变量 $X$ 和 $Y$ 间采用 $Y=AX+B$ 的形式，那么方程(3-4)可用来为变换后的变量拟合一条直线. 用上面的例子来说明这一技术.

假设我们想对一数据点集拟合幂曲线 $y=Ax^N$，用 $\alpha$ 记 $A$ 的估计，$n$ 记 $N$ 的估计. 方程 $y=\alpha x^n$ 两边取对数得

$$\ln y = \ln\alpha + n\ln x \tag{3-8}$$

在变量 $\ln y$ 对 $\ln x$ 的图中，方程(3-8)构成一条直线．在图上 $\ln\alpha$ 是 $x=0$ 时的截距，$n$ 是直线的斜率．用变换后变量和 $m=5$ 个数据点，由方程(3-5)和(3-6)解出斜率 $n$ 和截距 $\ln\alpha$，有

$$n=\frac{5\sum(\ln x_i)(\ln y_i)-(\sum\ln x_i)(\sum\ln y_i)}{5\sum(\ln x_i)^2-(\sum\ln x_i)^2}$$

$$\ln\alpha=\frac{\sum(\ln x_i)^2(\ln y_i)-(\sum\ln x_i)(\ln y_i)\sum\ln x_i}{5\sum(\ln x_i)^2-(\sum\ln x_i)^2}$$

从表 3-3 给出的数据得到，$\sum\ln x_i=1.321\ 755\ 8$，$\sum\ln y_i=8.359\ 597\ 801$，$\sum(\ln x_i)^2=1.964\ 896\ 7$，$\sum(\ln x_i)(\ln y_i)=5.542\ 315\ 175$，产生 $n=2.062\ 809\ 314$，$\ln\alpha=1.126\ 613\ 508$ 或 $\alpha=3.085\ 190\ 815$．所以方程(3-8)的最小二乘最佳拟合为(保留小数点后四位)

$$y = 3.0852x^{2.0628}$$

当 $x=2.25$ 时，这一模型预测 $y=16.4348$．注意，这一模型不是我们前面所拟合的平方形式．

假设我们仍然希望对数据集拟合一个二次形 $y=Ax^2$．用 $a_1$ 记 $A$ 的估计，以便与前面计算出的常数 $a$ 和 $\alpha$ 区别．方程 $y=a_1x^2$ 两边取对数，得

$$\ln y = \ln a_1 + 2\ln x$$

这时，$\ln y$ 对 $\ln x$ 的图是一条斜率为 2 的直线，截距为 $\ln a_1$．使用(3-4)式的第二个方程计算截距，有

$$2\sum\ln x_i+5\ln a_1=\sum\ln y_i$$

从表 3-3 列出的数据，得到$\sum\ln x_i=1.321\ 755\ 8$ 和$\sum\ln y_i=8.359\ 597\ 801$．因此，这一方程给出 $\ln a_1=1.143\ 217\ 24$ 或 $a_1=3.136\ 844\ 129$，产生最小二乘最佳拟合(保留小数点后四位)

$$y = 3.1368x^2$$

当 $x=2.25$ 时，这一模型预测 $y=15.8801$，它和第一个预测值 16.1337 有显著差异．第一个预测值是未经数据变换，由 $y=Ax^2$ 的最小二乘最佳拟合得出的二次形 $y=3.1869x^2$ 预测出的．在下一节将比较这两个二次形模型(还有第三个模型)．

前面的例子说明两个事实．第一，如果一个方程进行变换，在变换后的变量间构成一直线方程，方程(3-4)能直接用来解出变换后的图中的斜率和截距．第二，变换后的方程的最小二乘拟合与原方程的最小二乘拟合不是同一个，这个差异的起因是由于所产生的最优化问题是不同的．在原始问题中，寻找曲线时，是极小化原始数据的偏差的平方和，而在变换后的问题中，极小化使用变换后的变量的偏差的平方和．

## 习题

1. 解(3-4)给出的两个方程，得出分别由(3-5)和(3-6)式给出的参数的值．
2. 使用(3-5)和(3-6)式估计直线的系数．使直线和下列数据点间的偏差平方和达到极小．

(a)

| $x$ | 1.0 | 2.3 | 3.7 | 4.2 | 6.1 | 7.0 |
|---|---|---|---|---|---|---|
| $y$ | 3.6 | 3.0 | 3.2 | 5.1 | 5.3 | 6.8 |

(b)

| $x$ | 29.1 | 48.2 | 72.7 | 92.0 | 118 | 140 | 165 | 199 |
|---|---|---|---|---|---|---|---|---|
| $y$ | 0.0493 | 0.0821 | 0.123 | 0.154 | 0.197 | 0.234 | 0.274 | 0.328 |

(c)

| $x$ | 2.5 | 3.0 | 3.5 | 4.0 | 4.5 | 5.0 | 5.5 |
|---|---|---|---|---|---|---|---|
| $y$ | 4.32 | 4.83 | 5.27 | 5.74 | 6.26 | 6.79 | 7.23 |

对每一问，计算 $D$ 和 $d_{\max}$ 以界定 $c_{\max}$. 将你解出的结果与 3.2 节问题进行比较.

3. 求使一数据点集与二次形模型 $y=c_1x^2+c_2x+c_3$ 间偏差平方和极小化的方法. 使用这些方程对下列数据集找出 $c_1$, $c_2$ 和 $c_3$ 的估计.

| $x$ | 0.1 | 0.2 | 0.3 | 0.4 | 0.5 |
|---|---|---|---|---|---|
| $y$ | 0.06 | 0.12 | 0.36 | 0.65 | 0.95 |

计算 $D$ 和 $d_{\max}$ 以界定 $c_{\max}$，将你解出的结果与 3.2 节问题进行比较.

4. 为拟合模型 $P=ae^{bt}$ 做一个适当的变换，使用(3-4)式估计 $a$ 和 $b$.

| $t$ | 7 | 14 | 21 | 28 | 35 | 42 |
|---|---|---|---|---|---|---|
| $P$ | 8 | 41 | 133 | 250 | 280 | 297 |

5. 细心地考虑问题 3 中你拟合二次形时产生的方程组. 假设 $c_2=0$，对应的方程组将会怎样？在 $c_1=0$ 和 $c_3=0$ 的情形，重复这一问题. 提供一个三次的方程组，检查你的结果. 说明如何推广方程(3-4)的系统到拟合一个任意的多项式，如果多项式中有一个或多个系数为零，你将做什么？

6. 计算一个人体重的一般规则如下：对一位女性，用 3.5 乘身高(英寸)，再减 108. 对一位男性，用 4.0 乘身高(英寸)，再减 128. 如果一个人的骨架较小，调整计算结果，削减 10%. 对骨架较大的人加 10%. 对中等体形的人不调整. 收集不同年龄、形体和性别的人的体重对身高的数据，使用(3-4)式和你的数据为男性拟合一条直线，为女性拟合另一条直线. 这些直线的斜率和截距如何，怎样将这些结果与普遍规则进行比较？

在习题 7～10 中按给定模型用最小二乘法拟合数据.

7.

| $x$ | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| $y$ | 1 | 1 | 2 | 2 | 4 |

(a) $y=b+ax$

(b) $y=ax^2$

8. 弹簧拉伸的数据(参看 3.1 节习题 2).

| $x(\times10^{-3})$ | 5 | 10 | 20 | 30 | 40 | 50 | 60 | 70 | 80 | 90 | 100 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $y(\times10^{-5})$ | 0 | 19 | 57 | 94 | 134 | 173 | 216 | 256 | 297 | 343 | 390 |

(a) $y=ax$

(b) $y=b+ax$

(c) $y=ax^2$

9. 美国黄松的数据(参看 3.1 节习题 3).

| $x$ | 17 | 19 | 20 | 22 | 23 | 25 | 28 | 31 | 32 | 33 | 36 | 37 | 39 | 42 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $y$ | 19 | 25 | 32 | 51 | 57 | 71 | 113 | 140 | 153 | 187 | 192 | 205 | 250 | 260 |

(a) $y=ax+b$

(b) $y=ax^2$

(c) $y=ax^3$

(d) $y=ax^3+bx^2+c$

10. 行星的数据.

| 星体 | 周期(秒) | 与太阳的距离(米) |
|---|---|---|
| 水星 | $7.60\times10^6$ | $5.79\times10^{10}$ |
| 金星 | $1.94\times10^7$ | $1.08\times10^{11}$ |
| 地球 | $3.16\times10^7$ | $1.5\times10^{11}$ |
| 火星 | $5.94\times10^7$ | $2.28\times10^{11}$ |
| 木星 | $3.74\times10^8$ | $7.79\times10^{11}$ |
| 土星 | $9.35\times10^8$ | $1.43\times10^{12}$ |
| 天王星 | $2.64\times10^9$ | $2.87\times10^{12}$ |
| 海王星 | $5.22\times10^9$ | $4.5\times10^{12}$ |

拟合模型 $y=ax^{3/2}$.

### 研究课题

1. 建议那些想了解统计相关性度量的初步知识的同学完成教学单元“运用最小二乘准则做曲线拟合”(Curve Fitting via the Criterion of Least-Squares, by John W. Alexander, Jr., UMAP 321)的要求. 这一单元提供了相关、散点图以及直线和曲线回归的简单介绍. 可构造散点图, 为拟合特殊的数据选取合适的函数. 使用计算机程序拟合曲线.
2. 从 2.3 节的研究课题 1～7 中选择一个课题, 用最小二乘法拟合你提议的比例性模型. 将你的最小二乘的结果与 2.3 节使用的模型进行比较, 找出对切比雪夫准则的界定值, 解释这一结果.

### 进一步阅读材料

Burden, Richard L., & J. Douglas Faires. *Numerical Analysis*, 7th ed. Pacific Grove, CA: Brooks/Cole, 2001.

Cheney, E. Ward, & David Kincaid. *Numerical Mathematics and Computing*. Monterey, CA: Brooks/Cole, 1984.

Cheney, E. Ward, & David Kincaid. *Numerical Analysis*, 4th ed. Pacific Grove, CA: Brooks/Cole, 1999.

Hamming, R. W. *Numerical Methods for Scientists and Engineers*. New York: McGraw-Hill, 1973.

Stiefel, Edward L. *An Introduction to Numerical Mathematics*. New York: Academic Press, 1963.

## 3.4 选择一个好模型

现在来考虑形如 $y=Ax^2$ 的各种模型的适当性. 这些模型是我们用前节中最小二乘和经变换的最小二乘准则拟合的. 使用最小二乘准则, 得到模型 $y=3.1869x^2$. 评估模型是否很好地拟合了数据的一个途径是: 计算模型点和实际点间的偏差. 如果计算了偏差的平方和, 我们同时可以界定 $c_{\max}$. 对模型 $y=3.1869x^2$ 和表 3-3 给出的数据, 计算出的偏差列在表 3-4 中.

表 3-4 表 3-3 中的数据和拟合的模型 $y=3.1869x^2$ 间的偏差

| $x_i$ | 0.5 | 1.0 | 1.5 | 2.0 | 2.5 |
|---|---|---|---|---|---|
| $y_i$ | 0.7 | 3.4 | 7.2 | 12.4 | 20.1 |
| $y_i-y(x_i)$ | −0.0967 | 0.2131 | 0.029 98 | −0.3476 | 0.181 875 |

从表 3-4 可算出偏差的平方和为 0.209 54，所以 $D=(0.209\ 54/5)^{1/2}=0.204\ 714$. 由于 $x=2.0$ 时，最大绝对偏差为 0.3476，$c_{\max}$ 能界定如下：

$$D=0.204\ 714\leqslant c_{\max}\leqslant 0.3476=d_{\max}$$

现在来求 $c_{\max}$，因为存在五个数据点，数学问题是最小化五个数 $|r_i|=|y_i-y(x_i)|$ 的最大者. 将其称为 $r$. 我们要极小化 $r$，限制条件为 $r\geqslant r_i$ 和 $r\geqslant -r_i$，$i=1$，2，3，4，5. 将我们的模型记为 $y(x)=a_2x^2$，那么将表 3-3 的观测到的数据代入不等式 $r\geqslant r_i$ 和 $r\geqslant -r_i$，对 $i=1$，2，3，4，5，产生下列线性规划：

最小化 $r$

满足约束条件

$$\begin{aligned}
r-r_1&=r-(0.7-0.25a_2)\geqslant 0\\
r+r_1&=r+(0.7-0.25a_2)\geqslant 0\\
r-r_2&=r-(3.4-a_2)\geqslant 0\\
r+r_2&=r+(3.4-a_2)\geqslant 0\\
r-r_3&=r-(7.2-2.25a_2)\geqslant 0\\
r+r_3&=r+(7.2-2.25a_2)\geqslant 0\\
r-r_4&=r-(12.4-4a_2)\geqslant 0\\
r+r_4&=r+(12.4-4a_2)\geqslant 0\\
r-r_5&=r-(20.1-6.25a_2)\geqslant 0\\
r+r_5&=r+(20.1-6.25a_2)\geqslant 0
\end{aligned}$$

这一线性规划的解产生 $r=0.282\ 93$ 和 $a_2=3.170\ 73$. 那么，我们将最大的偏差从 $d_{\max}=0.3476$ 降到了 $c_{\max}=0.282\ 93$. 注意，我们无法进一步降低模型类型 $y=Ax^2$ 的最大偏差 0.282 93.

现在确定了模型类型 $y=Ax^2$ 的参数 $A$ 的三个估计. 哪个最好呢？表 3-5 中是根据每一个模型从每一个数据点计算出的偏差记录.

对三个模型的每一个可计算偏差平方和与最大绝对偏差，结果列在表 3-6 中.

**表 3-5 每一 $y=Ax^2$ 模型偏差的总结**

| $x_i$ | $y_i$ | $y_i-3.1869x_i^2$ | $y_i-3.1368x_i^2$ | $y_i-3.17073x_i^2$ |
|---|---|---|---|---|
| 0.5 | 0.7 | −0.0967 | −0.0842 | −0.0927 |
| 1.0 | 3.4 | 0.2131 | 0.2632 | 0.2293 |
| 1.5 | 7.2 | 0.0294 75 | 0.1422 | 0.0659 |
| 2.0 | 12.4 | −0.3476 | −0.1472 | −0.2829 |
| 2.5 | 20.1 | 0.181 875 | 0.4950 | 0.282 93 |

**表 3-6 三个模型的结果的总结**

| 准 则 | 模 型 | $\sum[y_i-y(x_i)]^2$ | $\text{Max}\,|y_i-y(x_i)|$ |
|---|---|---|---|
| 最小二乘 | $y=3.1869x^2$ | 0.2095 | 0.3476 |
| 变换后最小二乘 | $y=3.1368x^2$ | 0.3633 | 0.4950 |
| 切比雪夫 | $y=3.170\ 73x^2$ | 0.2256 | 0.282 93 |

如我们所预料，每一个模型都有优势方面．然而，考虑到在变换后的最小二乘模型中偏差平方和的增长，我们将选用简单的规划来选择模型．比如，选择有最小绝对偏差者(同样存在其他的拟合优度的统计指示量，可参看 *Probability and Statistics in Engineering and Management Science*，by William W. Hines and Douglas C. Montgomery，New York：Wiley，1972)，以此去掉明显很差的模型，这些指示量是有用的．但上面的问题用哪个模型最好仍不易回答．具有最小绝对偏差或最小平方和的模型在你最关注的范围中可能拟合得很差．进一步，在第 4 章将看到，不难构造出通过每一数据点的模型，这样偏差平方和与最大偏差都是零．所以，我们回答哪个模型最好要以具体个案为基础，要考虑模型的目的、实际情况要求的精度、数据的准确性以及使用模型时独立变量的值的范围．

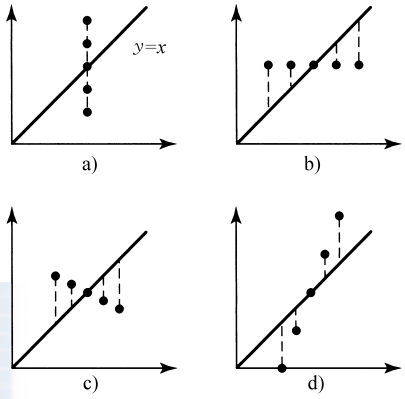


图 3-15　在这些图中，模型 $y=x$ 有同样的偏差平方和

选择模型或评价模型的适当性时，我们可能试图借助所用的最佳拟合准则的值．例如，可能试图使所选模型对给定的数据集有最小的偏差平方和或偏差平方和比一个认定拟合得很好的值要小，然而孤立地用这些指示量可能有严重误导．例如，考虑图 3-15 中显示的数据，四个情形对模型 $y=x$ 产生了完全相同的偏差．因此没有图的帮助，我们可能得出结论，认为每一情况模型拟合数据的结果是相同的．然而如图所示，在追踪数据的倾向时，每一模型的可用性存在显著的不同．下面的例子说明，在决定一个特殊模型的适当性时，如何借助各种指示量．正常情况下，图形有很大帮助．

## 例 1　车辆的停止距离

回顾预测车辆停止距离的问题，将其作为车辆速度的一个函数(问题描述见 2.2 节和 3.3 节)．在 3.3 节的子模型，反应距离 $d_r$ 正比于速度 $v$，并用图形做了检验，比例常数估计为 1.1．类似地，预测刹车距离 $d_b$ 与速度平方 $v^2$ 为正比例的子模型也经过检验，有理由认同该子模型，并估计正比常数为 0.054．因此，停止距离模型定为

$$d = 1.1v + 0.054v^2 \tag{3-9}$$

现在解析地拟合这些子模型，并比较各种拟合．

使用最小二乘准则拟合模型，使用(3-7)式的公式

$$A=\frac{\sum x_i y_i}{\sum x_i^2}$$

其中 $y_i$ 为每一数据点驾驶员的反应距离，$x_i$ 为每一数据点的速度．对表 2-3 给出的 13 个数据点，计算得 $\sum x_i y_i = 409\ 05$ 和 $\sum x_i^2 = 370\ 50$，得出 $A=1.104\ 049$．

对模型 $d_b = Bv^2$ 使用公式

$$B=\frac{\sum x_i^2 y_i}{\sum x_i^4}$$

其中 $y_i$ 为每一个数据点的平均刹车距离，$x_i$ 为每一数据点的速度．对表 2-4 给出的 13 个数据点计算出 $\sum x_i^2 y_i = 8\ 258\ 350$ 和 $\sum x_i^4 = 152\ 343\ 750$，得出 $B = 0.054\ 209$．因为数据相当不精确且是定性地建模，我们归整系数得到模型

$$d = 1.104v + 0.0542v^2 \tag{3-10}$$

模型(3-10)与第 3 章用图示获得的没有显著差异．接下来，我们分析模型拟合得如何．现在已经能够计算表 2-3 中观测的数据点与模型(3-9)和(3-10)预测值间的偏差．这些偏差总结在表 3-7中，两个模型的拟合非常相似．模型(3-9)的最大绝对偏差是 30.4，而模型(3-10)是 28.8．注意，到 70mph 为止，两个模型均过高估计停止距离，而之后低估了停止距离．代替前面个别地拟合的子模型，直接对总的停止距离拟合模型 $d = k_1 v + k_2 v^2$ 应能获得一个更好的模型．个别地拟合子模型并检验子模型的优点是可以检验它们是否很好地说明了状况．

**表 3-7　观测数据点与模型(3-9)和(3-10)预测值间的偏差**

| 速　度 | 图示模型(3-9) | 最小二乘模型(3-10) | 速　度 | 图示模型(3-9) | 最小二乘模型(3-10) |
|---|---|---|---|---|---|
| 20 | 1.6 | 1.76 | 55 | 14.35 | 15.175 |
| 25 | 5.25 | 5.475 | 60 | 12.4 | 13.36 |
| 30 | 8.1 | 8.4 | 65 | 7.15 | 8.255 |
| 35 | 13.15 | 13.535 | 70 | −1.4 | −0.14 |
| 40 | 14.4 | 14.88 | 75 | −14.75 | −13.325 |
| 45 | 16.35 | 16.935 | 80 | −30.4 | −28.8 |
| 50 | 17 | 17.7 | | | |

为确定是否较好地拟合了数据，画一个所用模型和观测数据点的图是有用的．模型(3-10)和观测点画在图 3-16 中．

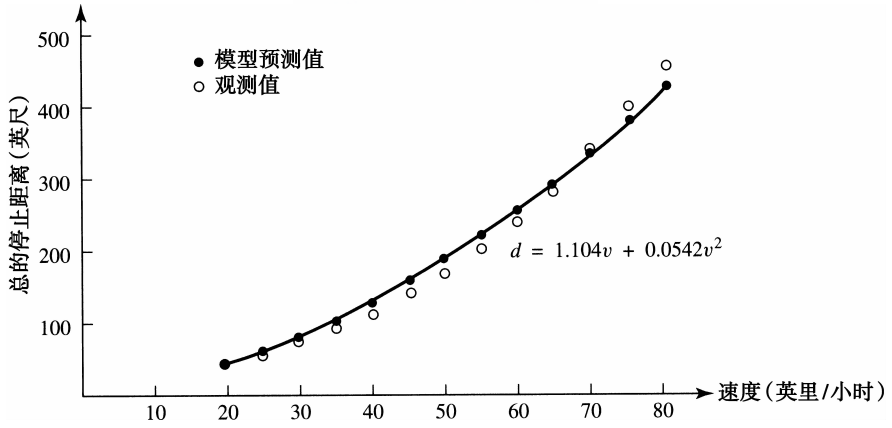


图 3-16　给出的模型和观测数据点的绘图提供了模型适合性的视觉检查

图 3-16 提供了一个模型适合性的视觉检查．看图形可以证实在数据中存在一个确定的倾向．模型(3-10)很好地说明了这种倾向，特别在低速度部分．

有一种很好的办法可以快速确定模型在何处受到破坏，即画出偏差(残差)图，将偏差作为

独立变量的一个函数．模型(3-10)的偏差图绘在图 3-17 中，说明到 70mph 为止，模型的确是可取的．超出 70mph，在预测观测到的状况方面，模型不再适用．

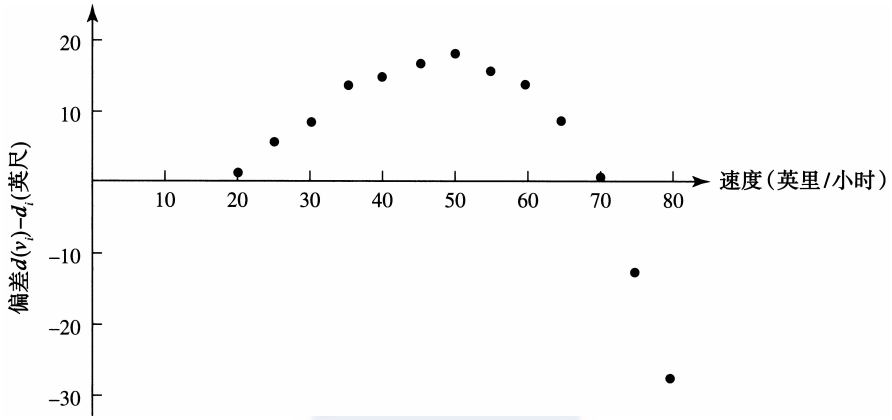


图 3-17　揭示模型拟合得不好的区域的偏差(残差)图

更细心地考察图 3-17，虽然到 70mph 为止偏差相当小，但它们全是正的．如果模型完满地解释了状况，偏差不仅应该小，而且应有正有负．为什么？注意到图 3-17 中偏差的特征中的一个确定模式，可使我们重新考察模型和数据．偏差中模式的特征能提示我们如何进一步精炼模型．在这里，数据收集过程中的不够精确无法为进一步的模型精炼提供保证． ■

## 例 2　比较准则

我们考虑下列涉及直径、高度、体积和直径$^3$ 的数据．观察图 3-18 中的倾向．

| 直径 | 体积 |
|---|---|
| 8.3 | 10.3 |
| 8.8 | 10.2 |
| 10.5 | 16.4 |
| 11.3 | 24.2 |
| 11.4 | 21.4 |
| 12.0 | 19.1 |
| 12.9 | 22.2 |
| 13.7 | 25.7 |
| 14.0 | 34.5 |
| 14.5 | 36.3 |
| 16.0 | 38.3 |
| 17.3 | 55.4 |
| 18.0 | 51.0 |
| 20.6 | 77.0 |

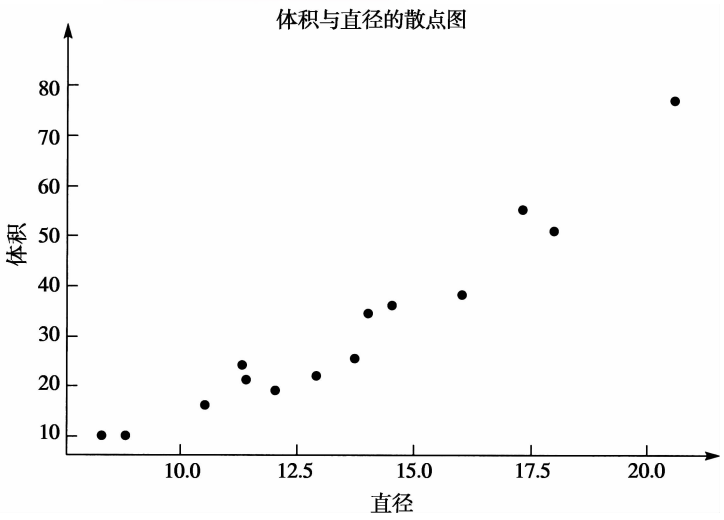


图 3-18　直径对体积的散点图

我们希望拟合模型 $V=kD^3$．比较全部三个准则：(a)最小二乘，(b)绝对偏差和，(c)极小化最

大误差(切比雪夫准则). 虽然这些方法中(b)和(c)的解法未出现在本文中，但我们仍然可以解释这些模型的结果.

(a)**最小二乘法** 使用公式找到最小二乘估计 $k$

$$k=\frac{\sum D_i^3 V}{\sum D_i^6}$$

$$k=1\ 864\ 801/19\ 145\ 566=0.009\ 74$$

这个回归方程是

$$\text{体积}=0.009\ 74\ \text{直径}^3$$

总的 SSE 是 451.

(b)**绝对偏差和** 我们用第 7 章描述的数值优化方法解这个问题.

极小化 $S=\sum|y_i-ax_i^3|$，$i=1$，2，…，14. 这里给出模型值和绝对误差的汇总. $a$ 的最佳系数值为 0.009 995，其相应的绝对误差和为 68.602 55.

| 直径 | 体积 | 模型值 | 绝对误差 | 系数 | 0.009 995 |
|---|---|---|---|---|---|
| 8.3 | 10.3 | 5.714 861 | 4.585 139 | | |
| 8.8 | 10.2 | 6.811 134 | 3.388 866 | | |
| 10.5 | 16.4 | 11.570 16 | 4.829 842 | | |
| 11.3 | 24.2 | 14.421 38 | 9.778 624 | | |
| 11.4 | 21.4 | 14.807 64 | 6.592 357 | | |
| 12 | 19.1 | 17.270 91 | 1.829 094 | | |
| 12.9 | 22.2 | 21.455 59 | 0.744 408 | | |
| 13.7 | 25.7 | 25.7 | 2.45E-06 | | |
| 14 | 34.5 | 27.425 56 | 7.074 441 | | |
| 14.5 | 36.3 | 30.470 21 | 5.829 794 | | |
| 16 | 38.3 | 40.938 44 | 2.638 444 | | |
| 17.3 | 55.4 | 51.744 92 | 3.650 079 | | |
| 18 | 51 | 58.289 31 | 7.289 307 | | |
| 20.6 | 77 | 87.372 15 | 10.372 15 | | |
| 绝对误差和 | | | 68.602 55 | | |

这一模型是

$$\text{体积}=0.009\ 995\ \text{直径}^3$$

应该指出的是，如果使用这一方法并计算 SSE，它应该大于从最小二乘获得的值 451. 另外，如果计算最小二乘法模型的绝对误差和，其值应大于 68.602 55.

(c)**切比雪夫方法** 我们使用第 7 章描述的线性规划方法. 用近似技术解这一规划. 根据切比雪夫方法，规划模型为

$$\text{极小化}\quad R$$

约束为

$$R+10.3-571.787k\geqslant 0$$

$$R-(10.3-571.787k)\geqslant 0$$

$$R+10.2-681.472k \geqslant 0$$
$$R-(10.2-681.472k) \geqslant 0$$
$$\cdots$$
$$R+77-8741.816k \geqslant 0$$
$$R-(77-8741.816k) \geqslant 0$$
$$R,k \geqslant 0$$

其最优解是 $k=0.009\ 936\ 453\ 825$，$R=9.862\ 690\ 776$. 模型为

$$\text{体积}=0.009\ 936\ 453\ 825\ \text{直径}^3$$

目标函数值(极小化的 $R$)即最大误差为 9.862 690 776.

准则就是根据我们追求极小化误差平方和，还是极小化绝对误差和，抑或是极小化最大绝对误差来进行选择.

## 习题

对下面的每一问题，用数据或用经变换的数据(如果适当)使用最小二乘准则求出一个模型. 将你的结果与 3.1 节问题中观测到的图形拟合进行比较，对每一模型计算偏差、极大绝对偏差以及偏差平方和. 如果模型是用最小二乘准则拟合的，求关于 $c_{max}$ 的一个界.

1. 3.1 节问题 3.
2. 3.1 节问题 4$a$.
3. 3.1 节问题 4$b$.
4. 3.1 节问题 5$a$.
5. 3.1 节问题 2.
6. 3.1 节问题 6.
7. (a)在下列数据中，$W$ 表示一条鱼的重量，$l$ 表示它的长度，使用最小二乘准则拟合模型 $W=kl^3$.

| 长度 $l$(英寸) | 14.5 | 12.5 | 17.25 | 14.5 | 12.625 | 17.75 | 14.125 | 12.625 |
|---|---|---|---|---|---|---|---|---|
| 重量 $W$(盎司) | 27 | 17 | 41 | 26 | 17 | 49 | 23 | 16 |

(b)在下列数据中，$g$ 表示一条鱼的身围. 使用最小二乘准则对数据拟合模型 $W=klg^2$.

| 长度 $l$(英寸) | 14.5 | 12.5 | 17.25 | 14.5 | 12.625 | 17.75 | 14.125 | 12.625 |
|---|---|---|---|---|---|---|---|---|
| 身围 $g$(英寸) | 9.75 | 8.375 | 11.0 | 9.75 | 8.5 | 12.5 | 9.0 | 8.5 |
| 重量 $W$(盎司) | 27 | 17 | 41 | 26 | 17 | 49 | 23 | 16 |

(c)两个模型哪个拟合数据较好？全面评判，你更喜欢哪一个模型？为什么？

8. 使用在习题 7(b)中的数据拟合模型 $W=cg^3$ 和 $W=kgl^2$. 解释这些模型，计算适当的指示量并确定哪个模型是最佳的. 做出说明.

## 研究课题

1. 写出一个计算机程序. 求下列模型中系数的最小二乘估计：
   (a) $y=ax^2+bx+c$　　(b) $y=ax^n$
2. 写出一个计算机程序，计算数据点和使用者遇到的任一模型的偏差. 假定模型是用最小二乘准则拟合的，计算 $D$ 和 $d_{\max}$. 输出每一数据点、每一数据点的偏差、$D$、$d_{\max}$ 和偏差平方和.
3. 写出计算机程序，使用(3-4)式和适当的变换后的数据计算下列模型的参数.
   (a) $y=bx^n$　　(b) $y=be^{ax}$　　(c) $y=a\ln x+b$
   (d) $y=ax^2$　　(e) $y=ax^3$