

灵活实用的曲线拟合软件

魏公毅 (中国科学院计算中心)

摘要: 本文介绍一种灵活实用的曲线拟合软件。它对于单自变量 x , 单因变量 y 的一组实验数据, 分别使用线性回归和非线性回归两种算法, 给出 22 种公式的回归系数等结果。能在屏幕上显示结果图形, 或在打印机上快速地绘制精美的图形。从而为用户选择合适的经验公式提供了有效而直观的信息, 并加速了选配经验公式的过程。

一、引言

在生产实际和科学实验中, 人们往往积累许多实验数据及图表等资料, 为了进一步研究其中的内在规律性, 需要建立能近似地表达这些资料的经验公式。

当表达式形式确定后, 通过回归分析方法可以计算表达式中的回归系数, 并给出相应的经验公式。对表达式形式没有什么限制, 而精度要求不高的一类问题采取表达式的半自动挑选方法也是可行的方法。然而, 在实际中往往需要在指定的许多种表达式中选取一种或几种满足要求的表达式。

本曲线拟合软件对于给定的单自变量 x , 单因变量 y 的一组实验数据, 分别使用线性回归和非线性回归两种算法, 给出 22 种公式的回归系数, 以及相应的残差平方和, 残差和, 复相关系数及其临界值和检验结果等信息。并在打印机上快速地绘制精美的图形(也可屏幕显示), 为用户选择合适的函数类型提供有效、直观的信息。

该软件可以在 22 种公式中任意指定所需计算的公式, 绘图的尺寸及坐标注解, 也可以指定文件名的固定或自选输入方式, 因此使用非常灵活。

该软件列出常用的 22 种类型的公式供选用, 并有较多的结果及直观图形, 这对于用户来讲是很实用的。

该软件可以很容易地扩充功能。例如, 增加函数类型, 处理多个自变量的实验数据等。

这个软件由三部分程序组成, 它们是:

线性回归程序 LR

非线性回归程序 NR

绘图程序 GR

该曲线拟合软件以统计软件包 SASD 为基础, 用 FORTRAN 77 语言编写, 源程序约 2000 行, 可以在 IBM PC 机及其兼容机上用 FORTRAN3.0 以上版本运行, 并可在 M2024, M1724, P1350, 3070, M1570 及 LQ 系列等多种打印机上绘图。

二、线性回归程序 LR

1. 线性化方程计算

m 个自变量 x_1, x_2, \dots, x_m 的线性回归模型有如下形式:

$$y = b_1x_1 + b_2x_2 + \dots + b_mx_m \quad (2.1)$$

在本文附录所列 22 种原始方程中, 公式(22)本身属于线性模型。公式(1)-(17), (21)属于非线性模型, 但它们通过变量变换的方法都可以简化为线性模型(2.1)的形式, 所以常称其为广义线性模型。

线性回归程序 LR 使用 SASD 统计包逐步回归分析程序 SRLSTP 可以直接计算公式(1)-(17), (21), (22)线性模型的回归系数, 残差平方和(Q), 残差和, 复相关系数(R)和临界值 $R_{0.05}, R_{0.01}$ 以及检验结果等。

复相关系数按下式计算:

$$R = \sqrt{1 - Q / S[yy]} \quad (2.2)$$

式中 s_{yy} 表示总离差平方和。F 统计量与复相关系数 R 有如下关系式:

$$F = \frac{R^2 / m}{(1 - R^2) / (N - m - 1)} \sim F(m, N - m - 1) \quad (2.3)$$

其中 N 为观测点数, m 为自变量个数。可以推出

$$R = \sqrt{\frac{Fm}{Fm + N - m - 1}} \quad (2.4)$$

只要计算出 F 分布上侧概率分位数

$F_{0.05}(m, N-m-1)$ 及 $F_{0.01}(m, N-m-1)$, 即可按上述公式给出临界相关系数 $R_{0.05} \equiv R_{0.05}(m, N-m-1)$ 及 $R_{0.01} \equiv R_{0.01}(m, N-m-1)$ 的数值, 并可进行下面的线性相关的显著性检验:

若 $R < R_{0.05}$, 则检验结果不显著, 表明线性回归方程无效;

若 $R_{0.05} < R < R_{0.01}$, 则检验结果显著, 表明线性回归方程有意义;

若 $R_{0.01} < R$, 则检验结果极显著, 表明线性回归方程更有意义。

2. 原始方程结果

由线性化方程算得的回归系数应回代到原始方程, 并算出原始方程对应的回归系数, 残差平方和, 残差和, 复相关系数和临界值及检验结果。使用(2.3)或(2.4)计算时, 取 $m = M - 1$, 其中 M 为公式待定参数的总个数。

值得注意的是, 对于公式(7), (13), (14), (21)和(22), 线性化方程与原始方程的计算结果是一致的, 因此, 其复相关系数及其检验结果可以作为回归效果好坏的依据。其余公式算出的复相关系数也可以作为曲线拟合好坏的一种衡量, 但有关的检验见公认的实用方法, 这里参照线性情况给出其检验结果, 仅供参考。

3. 限制条件

为了保证线性化过程变量变换的顺利进行, 要求观测值 $x, y > 0$ 。当 x, y 不满足要求时, 可以在 22 个公式中选取合适的公式进行曲线拟合, 或可选用接近 x, y 的值来替代 x, y , 例如, 当 $x = 0$ 时, 取 $x = 0.01$, 当 $y = 0$ 时, 取 $y = 0.01$ 等。

4. 其它

公式(18)–(20)为不易进行线性化处理的非线性模型, 在线性回归程序 LR 中不作计算, 仅令回归系数全部为 1(或取更合适的数值)。

另外, 数据文件中包括有控制公式进行计算的标志, 绘图尺寸及坐标轴文字注解等内容。程序中提供了固定和自选输入输出名的两种方式, 因而增加了计算的灵活性, 并方便用户使用。

三、非线性回归程序 NR

单自变量 x 的非线性回归模型有如下形式:

$$y = f(x, b_1, b_2, \dots, b_M) \quad (3.1)$$

其中 f 是待定数 b_1, b_2, \dots, b_M 的非线性函数。

非线性回归程序 NR 使用 SASD 统计包非线性回归程序 SRNGMD 计算附录中的回归系数, 残差平方和, 残差和, 复相关系数和临界值以及检验结果。迭代初始值由线性回归程序 LR 的计算结果给出。

值得强调的是, 对于公式(7), (13), (14), (21)和(22), 算出的复相关系数及检验结果可以作为衡量回归好坏的依据。而其余公式的复相关系数也能够作为曲线拟合好坏的一种衡量, 但有关检验未见公认的实用方法, 这里参考线性情况给出其检验结果, 仅供参考。

四、绘图程序 GR

绘图程序 GR 使用 SASD 统计包图形库, 并加上专门为该软件编制的作图子程序 SCGPC 等, 可在屏幕或纸上作出 22 种公式的函数图形。

GR 直接处理 LR 和 NR 产生的结果文件。用户既可以自由选择是屏幕显示还是在纸上作图, 也可以指定打印图形的个数, 图形的高度和宽度。当用户对图形坐标注解有自己的要求时, 可在原始数据文件内容之后, 加上所要注解的内容。如果数据文件中没有这一部分内容, 则程序将按标准方法作图和注解。

绘图程序 GR 为用户选择合适的函数类型提供了直观的信息, 从而加速了选配经验公式的过程。

附录: 22 种函数公式

- | | |
|------------------------------|--|
| (1) $y = ax^b$ | (12) $y = x / a + bx + cx^2$ |
| (2) $y = ae^{bx}$ | (13) $y = a + b / x + c / x^2$ |
| (3) $y = 1 / (a + bx)$ | (14) $y = a + b \log x + c \log^2 x$ |
| (4) $y = x / (a + bx)$ | (15) $y = ax^b e^{cx}$ |
| (5) $y = ax^b + c$ | (16) $y = ae^{bxc}$ |
| (6) $y = ae^{bx} + c$ | (17) $y = ae^{bxcx}$ |
| (7) $y = a + bx + cx^2$ | (18) $y = ae^{bx} + ce^{dx}$ |
| (8) $y = a + bx / c + dx$ | (19) $y = a + bx + ce^{dx}$ |
| (9) $y^2 = a + bx + cx^2$ | (20) $y = ax^b + cx^d$ |
| (10) $y = ae^{bx + cx^2}$ | (21) $y = a + bx + cx^2 + dx^3 + ex^4$ |
| (11) $y = 1 / a + bx + cx^2$ | (22) $y = a + bx$ |