DOI:10.16016/j.1000-5404.1988.02.017

建立辅助诊断方程专用软件

蔡昌启 李爱国 陈 卫 （数学教研室）

计算机辅助鉴别诊断在国内已有不少应用。由于它具有速度快、错分率较小等特点，日益受到医学界重视。目前在处理这类问题时所用的数学方法很多，如最大似然法、贝叶斯公式、回归分析、判别分析等等。但具体处理多限于单一的数学方法，较少提供比较、选择的机会。有的在所用算法上存在一些问题，如用Logistic回归时能否收敛，依赖于初始值的选择，而适宜的初始值较难确定。为解决以上问题，我们设计、编制了一个“建立辅助诊断方程”的软件，它能同时给出四个辅助诊断方程，以便研究者选择使用。而且只要数据符合建立Logistic方方程的条件，即可给出收敛结果，也不需要用户给出初始值。

软件的主要功能及特点如下。

一、给出四个不同的诊断方程和选择的依据

本程序处理的数据是：来自两种须进行鉴别的疾病患者的确诊病例各若干个，一般要求大样本。病例中各种有关的检验指标应有完整记录。

处理结果给出四个辅助诊断方程。一个是基于Bayes准则建立的线性判别方程[1]，一个是用条件极大似然估计—WD迭代算法建立的Logistic判别方程，另外两个都是用解非线性方程组的拟中顿法建立的Logistic判别方程。建立线性判别方程，在理论上要求各预报变量为连续的（计量资料），各总体来自正态分布且有相同的协方差矩阵。在实际问题中，这些前提很难具备。因而仅考虑一种线性判别方程，其判别效果往往不理想。而Logistic判别方程对预报变量的类型没有什么要求，它允许其中使用离散变量（定性分类资料）。我们同时给出这四个方程以供选择，经验表明对减少错分率是有作用的。

程序除给出各诊断方程的回报及验证结果、以及错分率的估计外，还给出了衡量诊断方程优劣的灵敏性和特异性两个指标的值。可以据此做为选取方程及临床诊断的依据。

二、适宜的迭代算法保证Logistic变换收敛

拟合Logistic方程需要用迭代算法。本程序给出三种迭代算法。第一种算法为参数的条件极大似然估计—WD迭代法[2]。它是利用线性判别方程的诸参数的估计值做为初始值而进行迭代计算的。第二、三种算法为解非线性方程组的拟牛顿法中的Broyden秩1算法[3]及Broyden-Fletcher-Shanno秩2算法。只要数据符合Logistic变换的要求（来自两总体的数据要有部分重迭）[4]时，总能保证其收敛。如果数据不符合上述要求，算法可能不收敛，但由程序给出的线性判别方程的判别效果一般较好。同时这三种算法所须初始值都不须人工给出，由程序自动实现。

三、采用了汉字系统及适宜的编程技巧，使用方便。

本程序是在CCDOS系统下用BASIC语言编写的，可在BM-PC/XT机或Apple-Ⅱ机上运行。

对原始数据的输入格式，程序设置了例示，以汉字屏幕提示方式给出了详细说明。输出打印结果用汉字注明。操作过程采用人机汉语对话方式进行。即或无使用手册，或使用者不了解有关统计计算，也都能使用本程序，很便于推广应用。

程序设计使用了模块结构，以覆盖方式运行的技术，节省了内存，适用于一般微机条件。程序由主控制模块，建立线性判别方程及用WD迭代算法拟合Logistic方程的模块，用Broyden秩1算法拟合Logistic方程模块，以及用B-F-S秩2算法拟合Logistic方程模块等四部分组成。只要调出主控制模块，键入执行命令，程序即可运行。它将根据使用者提供的选择，自动进行有关计算而输出所需各项结果。主要输出内容见附录。

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

（本文是在苑锡光教授的帮助、指导下完成的。谨致深切的谢意。）

附　录

建立辅助诊断方程实例

我们曾采用西南医院消化内科实验室收集的肝癌患者44例，肝硬化30例数据。每例病历都测量了七项生化指标：X_1—α_1-AT（α_1-抗胰蛋白酶）；X_2-AFP（甲种胎儿蛋白）；X_3—AKP（碱性磷酸酶）；X_4—LDH（乳酸脱氢酶）；X_5—γ-GT（γ-谷酰转肽酶）；X_6—γ-GT（同工酶），X_7—5'-NPDase（5'—核苷磷酸二酯酶）。y = 0或1（"0"表示肝硬化，"1"表示肝癌）。用本专用软件在IBM-AJ机上进行了例算。其中用36例肝癌、24例肝硬化病历的数据进行迭代计算建立诊断方程。其它14例作为验证诊断方程之用。结果如下：

1. 用Wallker-Duncan（WD）算法建立的"线性诊断方程"和"Logistic诊断方程"分别为：

$$\hat{y} = -18.36009 + 3.320288X_1 + 0.01072052X_2 + 0.01208672X_2 + 0.006365632X_4 - 0.0008906772X_5 + 3.879685X_6 + 5.858294X_7 \cdots\cdots (1)$$

$$\hat{y} = 1/\{1 + EXP[-(-18.35996 + 3.320336X_1 + 0.01072068X_2 + 0.01208706X_3 + 0.0063655655X_4 - 0.0008904999X_5 + 3.879747X_6 + 5.859053X_7)]\} \cdots\cdots (2)$$

两个方程预报和验证结果是相同的。如下两表所示：

预报结果

方程预报	临床诊断		总计
	肝硬化	肝　癌	
肝硬化	24	1	25
肝　癌	0	35	35
总计	24	36	60

R = 0.0166　6灵敏性 = 0.9722　特异性 = 1

验证结果

方程预报	临床诊断		总计
	肝硬化	肝　癌	
肝硬化	0	0	0
肝　癌	6	8	14
总计	6	8	14

R = 0.42857　灵敏性 = 1　特异性 = 0

2. 用拟牛顿法的 Brogden 秩1算法建立的"Logistic诊断方程"为

$$\hat{y} = 1/\{1 + EXP[-(-35.72188 + 5.620884X_1 + 0.02223626X_2 + 0.5983471X_3 + 0.01597772X_4 - 0.0388905X_5 + 14.99827X_6 - 1.456152X_7)]\} \cdots\cdots (3)$$

预报结果

方程预报	临床诊断		总计
	肝硬化	肝　癌	
肝硬化	24	0	24
肝　癌	0	36	36
总计	24	36	60

R = 0　灵敏性 = 1　特异性 = 1

验证结果

方程预报	临床诊断		总计
	肝硬化	肝　癌	
肝硬化	6	0	6
肝　癌	0	8	8
总计	6	8	14

R = 0　灵敏性 = 1　特异性 = 1

3. 用拟牛顿法 B-F-S（Broyden-Fletcher-Shanno）秩2算法建立的"Logistic诊断方程"为

$$\hat{y} = 1/\{1 + EXP[-(-301.4427 + 28.05423X_1 + 0.07320407X_2 + 1.874286X_3 + 0.03217454X_4 + 0.06427406X_5 + 79.92156X_6 + 258.0961X_7)]\} \cdots\cdots (4)$$

（下转150页）

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

2. 实验方法的重复性 批内误差：4份标本同时做3孔测定，CV分别为：5.2%，9.1%，12.5%，3.6%，平均CV7.6%。批间误差：3份标本不同时间测定3次，CV分别为：15.7%，12.8%，6.2%，平均CV11.6%。

3. 回收试验 3份已知AFP含量的标本分别入AFP标准液31ng/ml，62ng/ml，其中一份加入125ng/ml，回收率分别为93.5%、81.7%、96.2%、93.1%、74.2%、98.4%、119.6%，平均回收率为93.8%。

4. 稀释试验 440ng/ml标本稀释两倍后测得AFP含量为200ng/ml，165ng/ml标本稀释两倍后测得AFP含量为66ng/ml。

讨 论

火箭电泳放射试管测量法的原理是：电泳后，自琼脂板加样孔始，不同距离的琼脂块所含放射性与血清标本中的AFP含量相关。测定近加样孔段琼脂（即A管）和远加样孔段琼脂（即B管）的放射性，并计算A管cpm/（A管cpm＋B管cpm）百分比，根据此标准曲线和待测标本所得上述百分比值，即可查出待测标本的AFP含量。

我们体会到，该法具有下列优点：1.流程短，4 h左右可得出结果；2.操作较简便，容易掌握；3.结果便于计算；4.重复性较好，灵敏度与火箭电泳放射自显影法相当，而又克服了自显影法流程长，有时显影模糊不易判断结果等缺点。由于上述优点，我们认为该法可取代放射自显影法来测定AFP含量。

（本文统计学处理曾得到我校医用统计教研室尹全换老师指导，特此致谢。）

（上接148页）

预报结果

方程预报	临床诊断 肝硬化	肝癌	总计
肝硬化	24	0	24
肝癌	0	36	36
总计	24	36	60

R = 0　灵敏性 = 1　特异性 = 1

验证结果

方程预报	临床诊断 肝硬化	肝癌	总计
肝硬化	6	1	7
肝癌	0	7	7
总计	6	8	14

R = 0.071　灵敏性 = 0.875　特异性 = 1

通过以上建立的四个诊断方程的预报和验证的结果，并结合错分率、灵敏性、特异性来看方程（3）的效果最佳，方程（4）次之，若用于临床，看来以方程（3）或（4）为宜。

参 考 文 献

1. 中国科学院计算中心概率统计组。《概率统计计算》。科学出版社。1979

2. Wallker SH, Duncan DB Estimation of the probability of an event as a function of several independent variables. Biometrika 1967; 54:167～179

3. 王德人。《非线性方程组解法与最优化方法》。人民教育出版社。1982

4. Albert A, Anderson JA On the Existance of Maximum Likelihood Estimates in Logistic Regresson Models, Biometrika 1984; 71:1～10

?1994-2016 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net