

文章编号:1005-3085(2006)07-0101-08

## 艾滋病疗法评价及疗效预测问题评析

边馥萍<sup>1</sup>, 姜启源<sup>2</sup>

(1- 天津大学数学系, 天津 300072; 2- 清华大学数学科学系, 北京 100084)

**摘 要:** 本文简述 2006 高教社杯全国大学生数学建模竞赛 B 题的实际背景和一般的解法, 并对同学论文中的一些问题做了简单的评析。

**关键词:** HIV; CD4; 数学模型; 预测

**分类号:** AMS(2000) 90C90; 62J05

**中图分类号:** O221

**文献标识码:** A

### 1 问题的背景

艾滋病的医学全名为“获得性免疫缺陷综合症”(以下简称 AIDS), 是人体感染了人类免疫缺陷型病毒(以下简称 HIV)引起的, 当 HIV 感染者的免疫功能受到病毒的严重破坏时, 感染者便发展为 AIDS 病人。随着免疫力的降低, 人体会越来越频繁和严重地感染上各种致病微生物, 最终会因各种复合感染而死亡。AIDS 已成为当前人类社会最严重的瘟疫之一。联合国艾滋病规划署和世界卫生组织发表的《2005 年度全球艾滋病疫情报告》显示, HIV 感染者总人数已达 4030 万, 自 1981 年以来的死亡总人数为 2500 万。我国 HIV 感染者也已有 100 万人, 专家预测, 如不采取积极有效的措施, 到 2010 年 HIV 感染者将超过 1000 万人。AIDS 的预防及治疗已经成为全社会共同的责任。在 HIV 感染过程中, 病毒以活动、静止两种状态存在于宿主细胞里。含有静止期病毒的细胞一旦被激活, 整合到宿主细胞中的病毒 DNA 开始转录为 RNA, 在病毒结构蛋白及各种粒酶作用下, 经装配病毒 RNA 核心颗粒, 从细胞膜上获得包膜, 成为成熟的 HIV 病毒。成熟的 HIV 再感染新的细胞。随着病毒的复制繁殖, 在机体免疫系统起重要作用的 CD4 细胞数量呈进行性减少。HIV 感染后出现的动态进展过程包含着不同的发展阶段, 如图 1 所示。

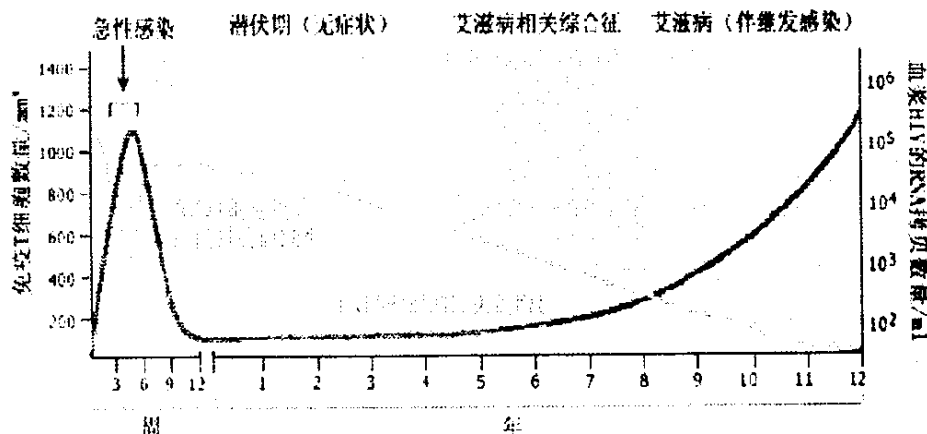


图1: 艾滋病感染的临床过程

对于 HIV 感染的治疗, 现在以针对 HIV 的高效抗逆转录病毒疗法为主。病毒载量是评估治疗方案效果的最重要指标,

(目前认为治疗有效的标志是在开始治疗 8 周后, 血浆病毒载量降低 1 个 log (10 倍), 4 至 6 个月降低到低于检测下限 (< 500 拷贝/mL)。

治疗过程中何时需要改变治疗方案? 由于目前抗逆转录酶病毒药物数量有限, 改变方案意味着感染者将来选择范围缩小了, 所以要非常慎重。临床上需要改变治疗方案的原因有: 最新临床试验结果提示, 感染者正在使用的不是最佳治疗方案; 感染者虽然采用高效的治疗方案, 但 CD4 细胞数量继续下降; 患者有临床进展表现或严重的毒副作用, 使之难以坚持治疗。题目中提出要为患者预测继续治疗的效果, 及确定最佳治疗终止时段, 所采用的 ACTG320 数据与 193A 数据都来源于美国最大的 AIDS 医疗试验机构 ACTG 的报告。

HIV 浓度的测试成本很高, 而 CD4 细胞数量的降低是免疫缺陷进展的直接标志, 被认为是最重要的预见 HIV 感染状态的参考, 也是各类疗法的有效评价指标。在 193A 报告中给出的不同人群、不同疗法的评价指标是 CD4 细胞的数量。

患者的治疗是一个长期的、只能延缓生命的过程, 特别是对于发展中国家的一些患者, 需要考虑治疗的费用, 所以题目中提出, 在考虑治疗成本的前提下, 如何选择最佳疗法与终止时间。这里, 药品价格是从网上查到的。

## 2 问题的解法简述

解决这个问题最常用的方法是统计方法, 这也是大多数同学用的办法, 结合同学的做法可作如下简述。

1. 利用 ACTG320 数据预测继续治疗的效果, 或者确定最佳治疗终止时间

如果随机取若干个病人, 画出他们 CD4 和 HIV 浓度随时间变化的图形 (散点图), 可以看出多数的 CD4 大致有先增后减的趋势, HIV 大致有先减后增的趋势, 启示我们从统计的角度可以建立时间的二次函数模型。

纵向数据分析是生物医学统计中常用的处理这类问题的方法<sup>[1]</sup>。记第  $i$  病人第  $j$  次测量时间为  $t_{ij}$ , 测量的 CD4 (或 HIV) 值为  $y_{ij}$ , 建立回归模型

$$y_{ij} = b_{i0} + b_{i1}t_{ij} + b_{i2}t_{ij}^2 + \varepsilon_{ij}, \quad i = 1, \dots, n, \quad j = 1, \dots, n_i, \quad (1)$$

其中  $b_{i0}, b_{i1}, b_{i2}$  是回归系数,  $\varepsilon_{ij}$  是随机误差, 服从零均值、方差  $\sigma^2$  的正态分布,  $n$  是病人数,  $n_i$  是第  $i$  病人的测量次数。将模型(1)的回归系数分解为

$$b_{ik} = b_k + \eta_{ik}, \quad k = 0, 1, 2, \quad (2)$$

其中  $b_k$  称固定效应参数 (与哪个病人无关),  $\eta_{ik}$  称随机效应参数, 服从零均值、方差为常数的正态分布。将(2)代入(1)得到如下的混合模型

$$\begin{cases} y_{ij} = b_0 + b_1t_{ij} + b_2t_{ij}^2 + \eta_{i0} + \eta_{i1}t_{ij} + \eta_{i2}t_{ij}^2 + \varepsilon_{ij} \\ \eta = (\eta_{i0}, \eta_{i1}, \eta_{i2})^T \sim N(0, D) \\ \varepsilon_{ij} \sim N(0, \sigma^2) \\ \eta, \varepsilon_{ij} \text{相互独立} \end{cases}, \quad D = \begin{pmatrix} d_1^2 & d_{12} & d_{13} \\ d_{12} & d_2^2 & d_{23} \\ d_{13} & d_{23} & d_3^2 \end{pmatrix}, \quad (3)$$

这里  $D$  是随机效应的协方差矩阵, 如果忽略  $\eta_{ik}$  对于  $k$  的相关性, 可以将  $D$  简化为对角阵。根据数据  $t_{ij}$ ,  $y_{ij}$  估计  $b_k$  ( $k = 0, 1, 2$ ) 及  $D$ ,  $\sigma^2$  的方法为

记

$$Y_i = \begin{pmatrix} y_{i1} \\ y_{i2} \\ \dots \\ y_{in_i} \end{pmatrix}, \quad X_i = \begin{pmatrix} 1 & t_{i1} & t_{i1}^2 \\ 1 & t_{i2} & t_{i2}^2 \\ \dots & \dots & \dots \\ 1 & t_{in_i} & t_{in_i}^2 \end{pmatrix}, \quad b = \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix}, \quad (4)$$

则由(3)可知

$$Y_i \sim N(X_i b, X_i D X_i^T + \sigma^2 I_{n_i}), \quad i = 1, 2, \dots, n, \quad (5)$$

记  $V_i = X_i D X_i^T + \sigma^2 I_{n_i}$ , 可以得到(5)的似然函数

$$L(b, D, \sigma^2) = \prod_{i=1}^n \left\{ (2\pi)^{-\frac{n_i}{2}} |V_i|^{-\frac{1}{2}} \exp \left( -\frac{1}{2} (Y_i - X_i b)^T V_i^{-1} (Y_i - X_i b) \right) \right\}, \quad (6)$$

求出(6)的最大值点就得到参数  $b_k (k = 0, 1, 2)$  及  $D, \sigma^2$  的估计值, 用 SAS 软件可以方便地实现。

$b_k$  的大小反映的是按照模型(1)全体病人的总体效应, 在预测继续治疗的效果或确定最佳治疗终止时间时, 利用  $b_k$  得到的是平均意义下的数字结果。而  $d_k^2$  的大小则反映因不同病人而异的分散性, 注意到(2)式, 在  $\eta_{ik}$  服从正态分布的假设下, 容易计算回归系数  $b_{ik}$  属于任意区间的概率, 于是在预测继续治疗的效果或确定最佳治疗终止时间时, 可以给出数字结果在一定置信度下的置信区间。

利用二次函数回归模型, 由 CD4 曲线的最大值点和 HIV 曲线的最小值点可以预测继续治疗的效果或确定最佳治疗终止时间。

纵向数据分析需要较专门的统计知识和软件, 作为它的简化, 不妨只用普通的统计知识(回归分析)和软件(如 MATLAB), 按照模型(1)对每个病人计算出回归系数  $b_{ik}$ , 并对  $i$  求出平均值和方差, 作为上面  $b_k$  和  $d_k^2$  的近似值来应用。

当然, 如果只要求得到全体病人的总体参数, 就可以将模型(1)简化为

$$y_{ij} = b_0 + b_1 t_{ij} + b_2 t_{ij}^2 + \varepsilon_{ij}, \quad i = 1, 2, \dots, n, \quad j = 1, 2, \dots, n_i, \quad (7)$$

由此计算出来的回归系数也作为上面  $b_k$  的近似值来应用, 只不过模型(7)的剩余方差将很大, 因为它相当于模型(3)中  $D$  和  $\sigma^2$  所造成的随机误差的总和。

如果考虑病人初始状态 ( $t = 0$  时的 CD4 和 HIV) 的不同对模型的影响, 可以将(1)和(7)中的  $y_{ij}$  定义为第  $i$  病人第  $j$  次测量的 CD4 (或 HIV) 与初始值之差或之比, 也可以先按照病人初始状态分类, 然后对于每一类建立回归模型。

## 2. 利用 193A 数据评价 4 种疗法的优劣

可以对每种疗法建立如前面(1)或(7)的回归模型, 只是可增加一个年龄变量, 或者先按年龄分类, 再对每一类建立模型。也可以再引入几个 0-1 变量, 表示不同的疗法, 建立统一模型。

利用数据计算出回归模型的系数以后, 要评价 4 种疗法的优劣, 可以进行疗法的两两比较, 即检验它们的回归系数是否有显著差异。

解决这个问题的另一种统计方法是方差分析。疗法当然是必选的因素, 如果还考虑年龄和初始状态, 用双因素分析则比较复杂, 可以先按这些因素分类, 再做疗法的单因素分析。

线性规划是解决这个问题的另一种方法, 简述如下。

为了利用 193A 数据评价 4 种疗法的优劣, 首先进行数据分析, 构造决策单元 (DMUs), 然后构造线性规划模型, 并利用其最优解, 判断决策单元是否达到最优<sup>[2]</sup>。

对 1309 位患者按年龄分组: 如 14 到 25 岁, 25 到 35 岁, 35 到 45 岁, 及 45 岁以上分为 4 组。在每个年龄组中, 按照 4 种疗法和 4 个治疗阶段 (0 到 10 周, 10 到 20 周, 20 到 30 周, 30 到 40 周), 可以得到 16 个决策单元。取 4 种药品的数量作为输入, 取患者在治疗各个阶段末的 CD4 与开始治疗时的 CD4 的比值作为输出。

设 4 种药品在第  $j$  单元的数量分别为  $x_{1j}$  (齐多夫定),  $x_{2j}$  (去羟基苷),  $x_{3j}$  (扎西他滨),  $x_{4j}$  (奈韦拉平),

$$y_{1j} = \frac{e^{\log(CD4+1)_{end}} - 1}{e^{\log(CD4+1)_{beginning}} - 1},$$

并定义决策单元  $DMU_{j_0}$  的效率评价指标为

$$h_{j_0} = \frac{\sum_{i=1}^4 v_i x_{ij_0} + v_0}{u_1 y_{1j_0}}.$$

其中  $v_i$  为输入的权重,  $u_1$  为输出的权重。权重处理成变量的形式, 在求解过程中确定它们。

$h_{j_0}$  越小, 表明  $DMU_{j_0}$  能够用相对较少的输入得到相对较多的输出。

1) 作为评价模型首先建立分式规划模型如下<sup>[2]</sup>

$$\begin{aligned} \min \quad & V_P = \frac{\sum_{i=1}^4 v_i x_{ij_0} + v_0}{u_1 y_{1j_0}} \\ \text{s.t.} \quad & \frac{\sum_{i=1}^4 v_i x_{ij} + v_0}{u_1 y_{1j}} \geq 1, \quad j = 1, 2, \dots, n, \\ & v_i \geq 0, \quad i = 0, 1, \dots, 4, \\ & u_1 \geq 0, \end{aligned} \quad (8)$$

若令  $t = \frac{2}{u_1 y_{1j_0}}$ ,  $\omega_i = tv_i$ ,  $\mu_1 = tu_1$ , (8) 式改写为如下的线性规划模型

$$\begin{aligned} \min \quad & V_P = \sum_{i=1}^4 \omega_i x_{ij_0} + \omega_0 \\ \text{s.t.} \quad & \sum_{i=1}^4 \omega_i x_{ij} - \mu_1 y_{1j} + \omega_0 \geq 0, \quad j = 1, 2, \dots, n, \\ & \mu_1 y_{1j_0} = 1, \\ & \omega_i \geq 0, \quad i = 0, 1, \dots, 4, \\ & \mu_1 \geq 0, \end{aligned} \quad (9)$$

规划(9)的对偶规划模型是

$$\begin{aligned} \max \quad & V_D = \alpha \\ \text{s.t.} \quad & \sum_{j=1}^n \lambda_j x_{ij} + s_i^- = x_{ij_0}, \quad i = 1, 2, 3, 4, \\ & \sum_{j=1}^n \lambda_j y_{1j} - s_1^+ = \alpha y_{1j_0}, \\ & \sum_{j=1}^n \lambda_j = 1, \\ & \lambda_j \geq 0, \quad j = 1, 2, \dots, n, \\ & s_i^- \geq 0, \quad i = 1, 2, 3, 4, \\ & s_1^+ \geq 0, \end{aligned} \quad (10)$$

求解上述模型, 根据决策单元的有效性, 评价疗法。

2) 关于预测模型可参考<sup>[3-5]</sup>

对 DMU 未知输出的走向进行预测的基础, 是假设未知 DMU (用  $DMU_{j_0}$  表示) 相对于已知的 DMU 是有效的。如果已知  $DMU_{j_0}$  的输入向量为  $(x_{1j_0}, x_{2j_0}, x_{3j_0}, x_{4j_0})^T$ , 预测模型(11)得出的输出  $y_{1j_0}^*$ , 对输出的期望是越大越好, 建立预测输出模型(11),

$$\begin{aligned} \max \quad & y_{1j_0} \\ \text{s.t.} \quad & \sum_{j=1}^n \lambda_j x_{ij} \leq x_{ij_0}, \quad i = 1, 2, 3, 4, \\ & \sum_{j=1}^n \lambda_j y_{1j} = y_{1j_0}, \\ & \sum_{j=1}^n \lambda_j = 1, \\ & \lambda_j \geq 0, \quad j = 1, 2, \dots, n. \end{aligned} \quad (11)$$

这样, 不同年龄段的患者就可以根据各种疗法在治疗各个阶段的效率和自己经济情况决定采用哪种疗法、治疗到哪个阶段。

### 3 对同学论文中出现问题的评析

在用统计方法的论文中主要有以下问题。

#### 1. 只做数据拟合, 不做统计检验

已知一组数据  $(x_k, y_k), k = 1, 2, \dots, n$  和一个形式已定、参数  $\beta$  待定的模型(曲线)  $y = f(x, \beta)$ , 确定参数  $\beta$  使数据与模型在误差平方和最小的意义下拟合得最好, 这就是人们熟悉的最小二乘拟合。确定参数  $\beta$  有现成的计算公式和软件, 很容易实现。如在这个题目中假设 CD4 (或 HIV) 与测试时间有一次或二次、三次函数模型, 根据数据可以算出模型的参数。问题是这样得到的模型有无显著意义, 它的置信度多大, 用它作预测时准确程度如何。这些问题只做数据拟合是无法解决的, 需要用回归分析方法做统计检验。

为了简洁地说明统计检验的必要性, 看下面的数字例子。假设有两组数据  $(x, y_1)$  和  $(x, y_2)$ , 见表1。

表1: 两组统计数据

x	0	1	2	3	4	5	6	7	8	9
$y_1$	1.04	1.22	1.38	1.59	1.80	1.99	2.21	2.39	2.65	2.83
$y_2$	0.28	0.61	2.19	2.56	2.49	1.17	3.17	1.29	2.11	3.27

分别拟合两条直线

$$y_1 = a_1 x + b_1, \quad y_2 = a_2 x + b_2,$$

用最小二乘法根据上面两组数据容易计算出  $a_1, b_1, a_2, b_2$  的估计值

$$\begin{aligned} \hat{a}_1 &= 0.2013, & \hat{b}_1 &= 1.0040, \\ \hat{a}_2 &= 0.2025, & \hat{b}_2 &= 1.0025, \end{aligned}$$

即两条直线几乎完全一样, 但是把这两组数据的散点图和两条直线作图对比一下(图2), 直观告诉我们, 用数据  $(x, y_1)$  得到的模型远比用  $(x, y_2)$  得到的模型来得可靠和准确, 而回归分析可以定量地描述模型的可靠和准确程度。

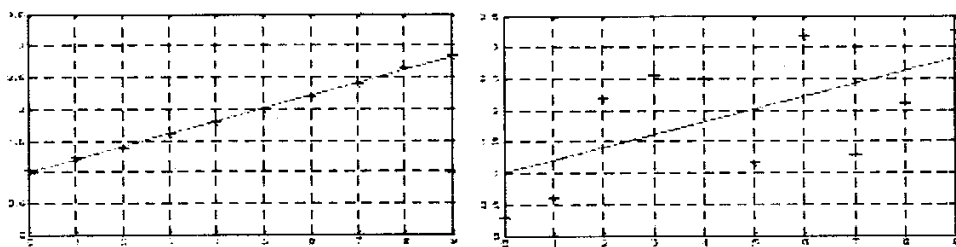


图2: 左图: 数据  $(x, y_1)$  的散点图(+)和拟合直线  $y_1 = a_1x + b_1$ ;  
右图: 数据  $(x, y_2)$  的散点图(+)和拟合直线  $y_2 = a_2x + b_2$

用实现回归分析的软件(常用的 MATLAB 的统计工具箱即可)可以得到结果见表2(置信区间的置信度=5%)

表2: 回归置信区间

	决定系数 $R^2$	F值	P值	剩余方差 $\sigma^2$	a的置信区间	b的置信区间
$y_1 = a_1x + b_1$	0.9985	5294	< 0.0001	0.0006	[0.1950, 0.2077]	[0.9699, 1.0381]
$y_2 = a_2x + b_2$	0.3528	4.361	0.0702	0.7761	[-0.0211, 0.4262]	[-0.1915, 2.1966]

对于第1个模型, 所有指标均表明模型与数据拟合得很好。对于第2个模型,  $R^2$  的数值表明在  $y_2$  的变化中(在与均值偏差平方和的意义下)只有35%是由  $x$  决定的, 其余由随机因素引起;  $F$  值很小, 导致  $p$  值大于0.05, 于是在常用的置信度  $\alpha = 5\%$  下该模型没有显著意义;  $a$  的置信区间包含零点, 即  $a = 0$  (这时模型无任何意义)落在置信水平为95%的区间内。这些指标均表明该模型不能反映数据  $(x, y_2)$  的变化规律。

用这两个模型作预测, 其结果的准确程度也有很大差别。比如对于  $x = x_0 = 5.5$ , 预测值为  $\hat{y}_{10} = \hat{a}_1x_0 + \hat{b}_1 = 2.1113$ ,  $\hat{y}_{20} = 2.1165$ , 二者相差无几, 但是它们的置信区间(置信水平95%)分别为  $[2.0621, 2.1606]$  和  $[0.3899, 3.8432]$ , 显然用第2个模型作预测没有什么价值。

这个例子表明, 用一个模型去拟合一组数据时不能只计算其参数, 应该做统计检验, 对模型(包括得到的参数)的显著性进行分析。如果需要用几个模型(如  $x$  的一次函数、二次函数、...)去拟合一组数据, 比较哪个较好时, 就更要借助统计分析。

## 2. 对原始数据先取平均再用平均值做拟合

在这个题目给的数据中大多数病人的测试时间是0, 4, 8, ... (周), 个别病人是5, 7, ... (周), 有些同学为了得到 CD4 (或 HIV) 的变化趋势, 先在每个时间点上对 CD4 取平均, 再用平均值做拟合, 而不管有的时间点上上百个病人的 CD4, 而有的时间点上只有几个病人的 CD4。

这样做有什么问题呢? 让我们看一个似乎有些极端、可是非常简明的数字例子, 说明这样做拟合可能产生的后果。设有这样的一组数据  $(x, y)$ :  $x = 0, y = 0 : 0.05 : 2$  (以步长0.05从0到2取值);  $x = 1, y = 1 : 0.05 : 3$ ;  $x = 2, y = 0.2$ , 在图3中左图上共84个点。对数据  $(x, y)$  拟合一条直线, 得  $y = 0.7647x + 1.0784$ , 如图3中左图。而如果先在  $x = 0, 1, 2$  对  $y$  取平均, 得到图3中右图的3个点, 再拟合一条直线, 得  $y = 1.3333$ , 如图3中右图, 它显然不能反映原始数据  $(x, y)$  的变化趋势。这是因为取平均后的3个点在做拟合时的权

重相同, 但是  $x = 0, 1$  的点都是 41 个原始数据的平均值, 而  $x = 3$  的点只是 2 个原始数据的平均值。

读者可以对一般情况进行讨论, 看看在什么条件下对原始数据先取平均再用平均值做拟合与直接对原始数据做拟合的结果才是一样的。

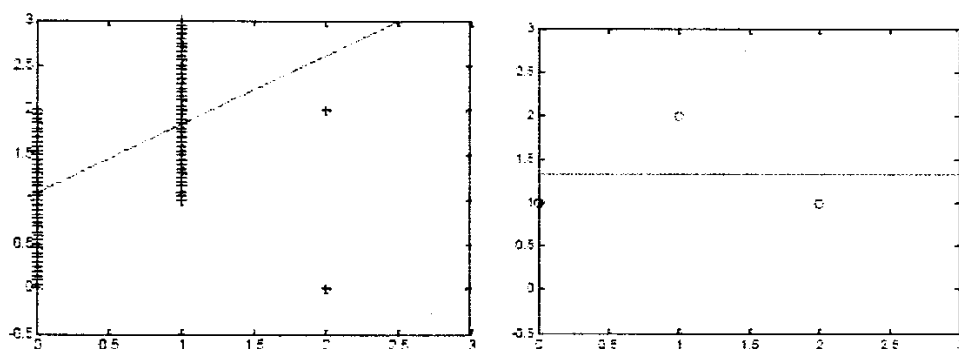


图3: 左图: 数据  $(x, y)$  的散点图(+)和拟合直线;  
右图: 数据  $(x, y)$  对  $y$  取平均后的散点图(o)和拟合直线

### 3. 拟合过度

这个题目给的 CD4 (或 HIV) 随时间的变化虽大体上是先增后降(或先降后增), 但仔细看来却有起伏, 为了拟合这些起伏有些同学用时间的 4 次、5 次甚至更高次数的多项式, 称之为拟合过度。事实上, 这些起伏是由与时间本身无关的其它随机因素引起的, 不应该用增加时间的高次项来拟合。极端的情况是, 对于数据  $(x_k, y_k), k = 1, 2, \dots, n$ , 用  $x$  的  $n-1$  次多项式拟合“最好”, 这条曲线通过全部  $n$  个点, 误差为零, 可是这已经变成多项式插值了。插值一般用于节点数据很精确时插值点的计算问题, 适用本题的显然是拟合而非插值。一般地说, 多项式拟合不要超过 3 次, 对于本题如果用统计检验做显著性分析, 可以发现多数情况是以时间的 2 次多项式为好。

### 4. 只按照图形做直观的定性判断, 不做定量分析

为了比较 4 种疗法的优劣, 有些同学对每种疗法的 CD4 (或取平均) 做散点图, 或折线(散点连线)图, 直观地比较疗法的优劣。对于本题这样做虽然也能得到正确的结果, 但是第一, 这种做法没有普遍的指导意义, 对于其它问题或数据, 这样做不一定能得到结果, 第二, 一种疗法的 CD4 散点图(或折线图)在另一种疗法的上面, 从统计意义上并不能表明其显著性。作为数学建模题目应该给出定量的处理方法。

### 5. 双指标处理不当

对于 CD4 和 HIV 两个指标的处理, 常用的方法是用某种函数形式将二者结合在一起。考虑到 CD4 越大越好, HIV 越小越好, 一些同学取二者之差或二者之商为综合指标, 这样简单处理的问题在于, 这两个指标量纲不同, 数量级差别较大。恰当的办法是, 先分别将其归一化(如将数值变换到 0-1 之间), 然后取加权平均。

除了用统计方法以外, 我们看到一些同学采用微分方程、灰色系统预测方法建模, 根据生态学中描述生灭关系的鲁特尔-伏尔脱拉方程, 建立微分方程组, 而且方程组的形式仅是简单

的一阶线性方程组。由于方程组中各项参数未知,因而将参数视为灰系数。还有的采用一阶一元灰色模型,即 GM(1, 1), 求灰系数。一般的说此方法不大适用于本题, 因为微分方程模型通常求得的是解析解, 或数值解, 因而模型要有一定精度, 求解过程复杂; 它可以求出整体最优解, 对于局部的最优性问题显得束手无策; 需要确定输入、输出之间关系的显式表达式, 过程复杂, 而且需要先确定系数。

不能用灰色预测方法求得这些系数, 采用灰色系统作预测时要知道输入变量的未来值。我们建议采用此类方法建模的同学看文献<sup>[6]</sup>, 从中了解如何采用微分方程最优控制模型, 描述在免疫系统中 CD4 细胞与 HIV 病毒的相互作用过程, 其方程形式比较复杂, 而且其参数是已知的, <sup>[6]</sup>中所研究的 CD4 细胞与 HIV 相互作用过程只是在 50 天左右, 这些参数值不适用于长周期。

某些同学还采用了层次分析法, 模糊评价, 时间序列, 神经网络等方法建立模型, 也不一定合适。简单的说, 层次分析法, 模糊评价用于半定性, 半定量问题转化为定量分析, 本题是定量问题的计算。时间序列模型需要有足够长的样本长度, 这里给出的观测值的个数太少。至于神经网络, 适用于继代分析, 这里不再详述, 希望同学们能对某种数学方法的应用范围能有进一步的理解, 而不是拿来简单套用。

#### 参考文献:

- [1] Geert Molenberghs, Geert Verbeke. Models for Discrete Longitudinal Data[M]. New York: London Springer, 2005.
- [2] 魏权龄. 评价相对有效性的 DEA 方法[M]. 北京: 中国人民大学出版社, 1988.
- [3] Bian F P, Cui L Y, Lawrence Lessner. DEA-based anti-HIV immunotherapy model[J]. Progress in Natural Science, 2004,14(9):828-832
- [4] Bian F P et al. The efficiency valuation of DMUs with undesirable outputs[J]. Transaction of Tianjin University, 2002,8(4):295
- [5] Bian F P, Fan Y. Prediction model of data envelopment analysis with undesirable outputs[J]. Transactions of Tianjin University, 2004,10(1):34-38
- [6] Hem Raj Joshi. Optimal control of an HIV immunology model[J]. Optimal control applications and methods, 2002,23:199-213

## The Comments on the Solution of the Assessment and Prediction of the AIDS Treatments

BIAN Fu-ping<sup>1</sup>, JIANG Qi-yuan<sup>2</sup>

(1- Tianjin University, Tianjin 300072; 2- Tsinghua University, Beijing 100084)

**Abstract:** In this paper the we give the background and general solutions and make the brief comments on the contest problem B for CUMCM-2006.

**Keywords:** AIDS; HIV; CD4; mathematical models; prediction.