乙肝发病高峰季节的预测

朱萌纾

(数理教研室)

摘要 目的 对淮阴市乙肝发病高峰季节进行预测。方法 对原始数据 0－1 化处理及对 0－1 序列分析建模。结果 1995 年第 2 季度和第 3 季度为乙肝发病高峰季节。结论 拟合值的均方根误差 RMSE＝0.3715,平均相对误差为 0.05,精度良好。

关键词 乙型病毒性肝炎 发病 季节性 预测 0－1 时间序列 拟合值 双向差分模型

中图法分类号 O211.67

在实际事物中我们经常遇到呈现两种状态的变量,如股票价格的上涨与下跌,气候变化的冷与暖,传染病患者数的增多与减少,等等。对这类事物,我们可用简单、抽象的“0”和“1”来表征。对原始数据进行 0－1 化处理、分析和建模,然后进行预测。

设等间隔取样的一数量数据序列

$$x(t)=\{x(1),x(2),\cdots\cdots x(N)\}$$

式中 N 为样本量。

1 对原始数据 x(t) 按均值取定标准得到 0－1 序列 $x_1(t)$

$x(t)$ 的均值为

$$\bar{x}=\frac{1}{N}\sum_{t=1}^{N}x(t)$$

以 $\bar{x}$ 为列划分标准,得样本量为 N 的 0－1 序列

不同

$$x_1(t)=\begin{cases}1, & x(t)>\bar{x}\\ 0, & x(t)\leqslant\bar{x}\end{cases}$$

2 将 0－1 序列分别排列出序号序列 $y_0(t)$ 和 $y_1(t)$

将 $x_1(t)$ 中“1”态出现的时间即“1”所对应的序号重新排列成一个新数列

$$y_1(t)=\{y_1(1),y_1(2),\cdots\cdots y_1(N_1)\}$$

也可以把 $x_1(t)$ 中“0”态出现的时间即“0”所对应的序号组成另一个新数列

$$y_0(t)=\{y_0(1),y_0(2),\cdots\cdots y_0(N_0)\}$$

$$N_1+N_0=N$$

3 按问题的具体要求建立 $y_0(t)$ 或 $y_1(t)$ 序列的双向差分预测模型

令 $x_k^{(0)}=y_1(t)\qquad k=1,2,\cdots,N(N=N_1)$

或 $x_k^{(0)}=y_0(t)\qquad k=1,2,\cdots,N(N=N_0)$

$$x_k^{(1)}=\sum_{j=1}^{k}x_j^{(0)}\qquad k=1,2,\cdots,N$$

$$Z_0=\frac{1}{2(N-2)}\left[\sum_{k=2}^{N-1}x_k^{(0)}+\sum_{k=2}^{N-1}x_{k+1}^{(0)}\right]$$

$$Z_1=\frac{1}{2(N-2)}\left[\sum_{k=2}^{N-1}x_k^{(0)}x_k^{(1)}+\sum_{k=2}^{N-1}x_{k+1}^{(0)}x_k^{(1)}\right]$$

$$S_1=\frac{1}{(N-2)}\sum_{k=2}^{N-1}x_k^{(1)}$$

$$V_0=\frac{1}{(N-2)}\sum_{k=2}^{N-1}1^2=1$$

$$V_1=\frac{1}{(N-2)}\sum_{k=2}^{N-1}\left[x_k^{(1)}\right]^2$$

$$Z=\begin{bmatrix}Z_0\\ Z_1\end{bmatrix},\quad V=\begin{bmatrix}V_0 & S_1\\ S_1 & V_1\end{bmatrix}$$

$$B=\begin{bmatrix}b_0\\ b_1\end{bmatrix}$$

表 1 淮阴市 1992～1994 年每季度乙肝发病人数

年份	季度			
	Ⅰ	Ⅱ	Ⅲ	Ⅳ
1992	75	117	147	219
1993	235	310	320	180
1994	211	215	215	180

将数据按年份季度顺序排列得序列

$x(t)=\{75,117,147,219,235,310,320,180,211,215,215,180\}$

$$\bar{x}=\frac{1}{12}\sum_{t=1}^{12}x(t)=202$$

按均值 202 作为标准使序列 0—1 化，见表 2。)

表 2 序列 0—1 化

序号	1	2	3	4	5	6	7	8	9	10	11	12
$x_1(t)$	0	0	0	1	1	1	1	0	1	1	1	0

将表 2 中数码为"1"的序号列为一新序列

$y_1(t)=\{4,5,6,7,9,10,11\}$

令 $x^{(0)}=\{4,5,6,7,9,10,11\}$

$N=7$，一次累加

$x^{(1)}=\{4,9,15,22,31,41,52\}$

$$Z_0=\frac{1}{10}[(5+6+7+9+10)+(6+7+9+10+11)]=8$$

$$Z_1=\frac{1}{10}[(5\times9+6\times15+7\times22+9\times31+10\times41)+(6\times9+7\times15+9\times22+10\times31+11\times41)]$$

$$S_1=\frac{1}{5}(9+15+22+31+41)=23.6$$

$$V_0=1$$

$$V_1=\frac{1}{5}(9^2+15^2+22^2+31^2+41^2)=686.4$$

$$Z=\begin{bmatrix}8\\209.6\end{bmatrix},\quad V\begin{bmatrix}1 & 23.6\\236.6 & 686.4\end{bmatrix}$$

相应的时间响应方程为：

$$X_{k+1}^{(1)}=\left(X_1^{(1)}+\frac{4.2096}{0.1607126}\ e^{0.1607126}-\frac{4.20967}{0.1607126}\right.$$

把 $X_1^{(1)}=4$ 代入上式得

$$X_{k+1}^{(1)}=30.19377697e^{016707126k}-26.19377697$$

$$X_K^{(1)}=X_K^{(1)}-X_{K-1}^{(1)}\quad K=N,\cdots2$$

计算结果见表 3。

表 3 计算结果

序号	1	2	3	4	5	6	7	8	9
$x^{(0)}$	4	5	6	7	9	10	11		
$\hat{x}^{(0)}$	4	5.26	6.18	7.26	8.53	10.01	11.76	13.81	16.21
残差	0	−0.26	−0.18	−0.26	0.47	−0.01	−0.76		

表中序号 8、9 的 $\hat{x}^{(0)}$ 值为两步预测值。拟合值的均方根误差 RMSE＝0.3715，平均相对误差为 0.05。由此可见精度良好。

预测结果

$\hat{x}_8^{(0)}=13.81\approx14$

$\hat{x}_9^{(0)}=16.21\approx16$

14 表示 1995 年第 2 季度，16 表示 1995 年第 4 季度，即预测 1995 年 2、4 季度乙肝发病人数较多。实况 1995 年第 2 季度发病人数 475 人，第 4 季度 395 人，是发病人数最多的两个季度，预则正确。

参 考 文 献

1 魏风英，曹鸿兴. 长期预测的数学模型及其应用. 北京：气象出版社，1990

(收稿：1996—12—20 修回：1997—04—20)

(本文编辑：吴 进)