

· 技术方法 ·

应用时间序列统计方法分析广东省 1984 ~ 1993 年乙型脑炎季节性分布特性

李廷杰 陈秀山 李燕芬

摘要 对构成双向有序列联表时间序列数据的广东 1984 ~ 1993 年 10 年乙型脑炎(乙脑)逐月发病资料,应用游程检验和有序对数比检验两种不同统计方法进行季节性分布分析,获一致的统计结论,乙脑发病呈明显的一年一峰(6 月或 7 月)季节性周期,这对流行病学工作者掌握两种统计方法的基本原理和方法有一定帮助。

关键词 流行性乙型脑炎 游程检验 有序对数比检验

Application of the Time-Series Method to Analyse the Seasonal Distribution of Epidemic Encephalitis B Incidence in Guangdong Province in the Years of 1984-1993 Li Tingjie, Chen Xiushen, Li Yanfen. *Guangdong College of Pharmacy, Guangzhou 510224*

Abstract This paper analysed the data of epidemic encephalitis B incidence that was constituted to time-series and the two-ordered contingency table in Guangdong province from 1984 to 1993. Two different kinds of statistical methods of run-test and ordered-log-ratio test were applied to study the characteristics of seasonal distribution. Consistent conclusion was obtained to show that there appeared an obvious seasonal cyclic characteristic of a peak every year (June or July) for the incidence of epidemic encephalitis B. It is helpful for the epidemiologists to understand the principle and method of these statistical tests.

Key words Epidemic encephalitis B Run test Ordered-log-ratio test

时间序列是随机过程论的重要组成部分。疾病在一定时间内发生的频率是随机变量,常呈规律性变化,这在传染病的发生和发展中尤为明显。借助适当的统计方法进行分析,了解其变化的规律性,如升降趋势、季节性升高、周期性变化等时序分布特点,对于探讨疾病的流行动态、研究致病因素、考核或评价防治措施提供科学依据。笔者对广东省 1984 ~ 1993 年 10 年间乙型脑炎(乙脑)逐月发病资料,应用时间序列游程检验和有序对数比检验进行分析,探讨该病是否存在季节性周期的分布特征,获得满意结果。

结果与分析

一、乙脑发病时间分布的特点:从表 1 可见,广东省 1984 ~ 1993 年 10 年来,虽年间略有小波动,但总的发病水平下降趋势明显,且具有季节性传染病的特征,发病局限于夏秋季,一年一峰很有规律,除个别年份(1990 年)在 7 月以外,都以每年的 6 月为高峰。

二、乙脑季节性周期的统计分析方法:

1. 时间序列季节性分布的游程检验:

(1) 基本思想:游程检验是用以检验只取元素 A 或 B(非 A)的样本是否来自概率为 P 的二项分布总体,从包含有元素 A 或 B 的总体中抽样,抽得 m 个 A、n 个 B,按抽得先后次序把 A 和 B 统一排列起来,具有相同的事

件(即符号)的连续称为一个游程。游程检验的统计量为游程的个数 r 和长度 l , 如 B、AA、BBB、A, 此处 $m=3, n=4, m+n=7, r=4$, 长度 l 分别为 1、2、3、1。检验的无效假设,

$$P\{r=r_0\} = \begin{cases} 2 \binom{m-1}{k-1} \binom{n-1}{k-1} \binom{m+n}{m} & \text{当 } r_0 = 2k \\ \binom{m-1}{k-1} \binom{n-1}{k-2} + \binom{m-1}{k-2} \binom{n-1}{k-1} \binom{m+n}{m} & \text{当 } r_0 = 2k - 1 \end{cases} \quad (1)$$

表 1 广东省 1984~1993 年乙脑逐月发病率(/10 万)

年份	月 份												年发病率 (/10 万)
	1	2	3	4	5	6	7	8	9	10	11	12	
1984	0.002	0.002	0.008	0.007	0.094	0.858	0.443	0.067	0.026	0.003	0.002	0.000	1.502
1985	0.003	0.000	0.000	0.006	0.034	0.938	0.763	0.165	0.021	0.005	0.003	0.006	1.947
1986	0.002	0.000	0.000	0.009	0.090	0.551	0.392	0.114	0.022	0.015	0.003	0.000	1.200
1987	0.000	0.000	0.002	0.012	0.111	0.730	0.360	0.050	0.011	0.000	0.000	0.002	1.277
1988	0.000	0.000	0.000	0.000	0.012	0.737	0.661	0.126	0.010	0.003	0.001	0.000	1.551
1989	0.000	0.000	0.000	0.003	0.040	0.458	0.195	0.026	0.016	0.000	0.000	0.000	0.741
1990	0.000	0.000	0.003	0.000	0.014	0.308	0.460	0.075	0.004	0.001	0.000	0.001	0.869
1991	0.002	0.000	0.000	0.002	0.036	0.504	0.228	0.037	0.006	0.000	0.000	0.000	0.816
1992	0.000	0.000	0.000	0.003	0.022	0.204	0.105	0.047	0.011	0.000	0.002	0.000	0.395
1993	0.000	0.000	0.002	0.009	0.007	0.106	0.054	0.006	0.011	0.002	0.002	0.000	0.199

据式(1)可制作 r 的频数分布表或累计频数表, 显著性检验时, 可直接查概率:

$$P\{r \leq r_0\}$$

或 $P\{r \geq r_\alpha\} = \alpha, P\{r \leq r_\alpha\} = 1 - \alpha$

且 r 的平均值

$$E(r) = \frac{2mn}{m+n} + 1$$

$$V(r) = \frac{2mn(2mn - m - n)}{(m+n)^2(m+n-1)}$$

$$u = \frac{r - E(r)}{\sqrt{V(r)}}$$

u 服从标准正态分布 $N(0, 1)$ 。

在实际应用中, m 和 n 较小, 直接查游程个数检验用 r 界值表或 l 界值表甚为方便。

不探究较严谨的数理推导, 更为直观的理解是, 当 m 和 n 一定时, r 值不太大或不太小, 游程的长度 l 不会太长, 则样本来自随机分布的总体; 反之, 当游程个数超过一定范围, 或长度超过一定界限, 可判断某种动态趋势的存在, 如长度检验, 可判断季节性升高是否有显著性意义。

(2) 检验目的与检验结果: 每年按月发病

$$\lambda = \ln \left(\frac{n_1}{n_2} \right) + \frac{1}{2} \ln \left(\frac{n_1}{n_3} \right) + \frac{1}{3} \ln \left(\frac{n_1}{n_4} \right) + \dots + \frac{1}{c-1} \ln \left(\frac{n_1}{n_c} \right)$$

即每一次抽得 A 的概率为 P , 抽得 B 的概率为 $1-P$; 备择假设, 抽得 A 或 B 的概率无规律性。无效假设若成立, 当 m 和 n 不变时, 根据 A、B 的组合情况, r 可为偶数或奇数, 得:

水平变动是随机的, 抑或存在季节性周期升高的特征。

从首年 1 月份开始, 逐一比较相邻两个月发病率的大小, 凡后一个比前一个升高者记以“+”号, 降低者记以“-”号, 相同者记以“0”号。10 年共 120 个月发病率, 可得符号 119 个。于符号变更处划一条斜线, 则两条斜线间作为一游程, 得游程个数为 35, 与游程长度分别为 1、2、3、4、6、8 对应的游程个数分别是 14、3、8、6、3、1, 即共有 4 个游程的长度达到或超过 6, 检验结果显示长游程连续地、有规律地出现, 说明乙脑在广东的时间分布的季节性周期有显著意义($P < 0.05$)。

2. 时间序列季节性分布的有序对数比检验:

(1) 基本思想:

有序对数比的概念: 设有一按序次分类变量 C 类, 各类的观察频数依次为 n_1, n_2, \dots, n_c , 定义有序对数比(λ)为:

$$\begin{aligned}
 & + \ln\left(\frac{n_2}{n_3}\right) + \frac{1}{2} \ln\left(\frac{n_2}{n_4}\right) + \dots + \frac{1}{c-2} \ln\left(\frac{n_2}{n_c}\right) \\
 & + \ln\left(\frac{n_3}{n_4}\right) + \dots + \frac{1}{c-3} \ln\left(\frac{n_3}{n_c}\right) \\
 & + \dots + \ln\left(\frac{n_{c-1}}{n_c}\right)
 \end{aligned} \tag{2}$$

有序对数比达到度量有序变量的变化趋势,各项中乘上一个系数是为了消去因水平数的不同,而由间隔度(如 \$n_1\$ 到 \$n_3\$, 间隔度为 2)所产生的影响。

④有序对数比的系数(\$K_i\$)及方差 \$V(\lambda)\$: 当 \$C\$ 为奇数(\$C = 2k + 1, K = 1, 2, \dots\$) 时:

$$\lambda = K_1 \ln\left(\frac{n_1}{n_c}\right) + K_2 \ln\left(\frac{n_2}{n_{c-1}}\right) + \dots + K_{\frac{c-1}{2}} \ln\left(\frac{n_{\frac{c-1}{2}}}{n_{\frac{c+1}{2}}}\right) \tag{3}$$

其中

$$K_1 = 1 + \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{c-1}$$

$$K_2 = \frac{1}{2} + \frac{1}{3} + \dots + \frac{1}{c-2}$$

$$K_3 = \frac{1}{3} + \dots + \frac{1}{c-3}$$

$$\dots \dots \dots$$

$$K_{\frac{c-1}{2}} = \frac{1}{\frac{c-1}{2}} + \frac{1}{\frac{c+1}{2}} \left(\frac{c-1}{2} \text{ 和 } \frac{c+1}{2} \text{ 均为自然数} \right)$$

式(3)中 \$\lambda\$ 的方差 \$V_{(\lambda)}\$ (常用 \$S_\lambda^2\$ 代替) 为

$$\begin{aligned}
 V_{(\lambda)} = & K_1^2 \left(\frac{1}{n_1} + \frac{1}{n_c} \right) + K_2^2 \left(\frac{1}{n_2} + \frac{1}{n_{c-1}} \right) + \dots + \\
 & K_{\frac{c-1}{2}}^2 \left(\frac{1}{n_{\frac{c-1}{2}}} + \frac{1}{n_{\frac{c+1}{2}}} \right)
 \end{aligned} \tag{4}$$

当 \$C\$ 为偶数时, 则式(3)最末一项系数为 \$K_{\frac{c}{2}} \left(= \frac{1}{\frac{c}{2}} \right)\$ 代替 \$K_{\frac{c-1}{2}} \left(= \frac{1}{\frac{c-1}{2}} + \frac{1}{\frac{c+1}{2}} \right)\$, 并且相应于式(4) \$\lambda\$ 的方差为:

$$\begin{aligned}
 V_{(\lambda)} = & K_1^2 \left(\frac{1}{n_1} + \frac{1}{n_c} \right) + K_2^2 \left(\frac{1}{n_2} + \frac{1}{n_{c-1}} \right) + \dots \\
 & + K_{\frac{c}{2}}^2 \left(\frac{1}{n_{\frac{c}{2}}} + \frac{1}{n_{\frac{c}{2}+1}} \right)
 \end{aligned} \tag{5}$$

可将不同水平数有序对数比的系数(\$K_i\$) 求出并列成附表, 可简化计算。

表3 各行有序对数比 \$\lambda_i\$ 及 \$S_\lambda^2\$

	1	2	3	4	5	6	7	8	9	10
\$\lambda_i\$	0.0783	-0.1343	-0.0499	0.1102	-0.1079	0.0451	-0.0943	0.0599	-0.0196	0.0064
\$S_\lambda^2\$	62.2646	62.2028	62.2374	62.4266	62.5338	62.5791	62.5574	62.9185	62.5991	62.6359

据式(6) 求得 \$\bar{\lambda}\$ 为 7.8406, 进而按式(7) 求得统计量 \$Y^2 = 9.8636\$, 自由度 \$df = 10 - 1 = 9\$, 所以 \$Y^2 < \chi_{0.05(9)}^2 = 16.91\$, 故 \$P > 0.05\$。

(四)利用有序对数比检验季节性周期, 积累多年的逐月发病率资料构成双向有序 \$r \times c\$ 表。按上述公式求出 \$\lambda_i\$ 和 \$S_\lambda^2\$ 后, 进而求出 \$\bar{\lambda}\$:

$$\bar{\lambda} = \left(\sum_{i=1}^r \lambda / S_\lambda^2 \right) \sqrt{\sum_{i=1}^r \frac{1}{S_\lambda^2}} \tag{6}$$

再按下式求出统计量 \$Y^2\$

$$Y^2 = \sum_{i=1}^r (\lambda_i - \bar{\lambda})^2 / S_\lambda^2 \tag{7}$$

无效假设为各年的按月发病季节性周期性升高趋势相同。若此假设成立, \$Y^2\$ 服从自由度为 \$r-1\$ 的卡方分布。容易理解如果按月发病率(即各行数据) 都是按照一定的规律变化, 比如都有相同的季节性升高特征, 则接受无效假设(\$P > 0.05\$)。反之, 若各年逐月发病趋势不一致, 不存在相同的季节性升高的特征, 则拒绝无效假设(\$P = 0.05\$)。

(2) 检验目的与检验结果: 仍然是检验广东省 1984 ~ 1993 年乙脑发病季节性分布的特征。

本例 \$r = 10, c = 12\$, 按照上述公式求得系数 \$K_i\$ (也可查附表), 列于表 2。各行的有序对数比 \$\lambda\$ 及其方差 \$S_i^2\$ 亦分别求出, 列于表 3。

表2 有序对数比 \$\lambda\$ 的系数 \$K_i\$ 值

\$K_1\$	\$K_2\$	\$K_3\$	\$K_4\$	\$K_5\$	\$K_6\$
3.02	1.93	1.33	0.84	0.51	0.17

说明广东省 1984 ~ 1993 年 10 年间每年都存在一致性的季节性周期升高的特征。

讨 论

相对于初等概率论是研究静态性质的数学方法, 随机过程论则是研究随机现象的动态特性, 研究随机现象发生、发展过程的一个数学分支, 并已发展成为应用极广泛的数学工具之一, 而时间序列研究又是随机过程论中应用性特别强的组成部分。

研究疾病的流行病学特征时, 首先要研究它的分布, 这是描述性流行病学的基本任务。正确地描述疾病的时间分布, 往往能反映或提示某些因素与发病的关系, 更为我们深入调查指明途径和方向, 因此又是分析性流行病学的基础。基于这些原因, 本文主要对累积 10 年的广东乙脑发病规律是否存在季节性周期升高进行了研究。这种资料是一组时

间序列, 构成 $r \times c$ 表是双向有序的列联表, 诚然分析时间序列的方法较多, 随数据的性质和分析目的而异。流行病学者调查收集所得资料仅是一样本资料, 虽知当以样本数据推论总体的流行趋势时, 应考虑抽样误差的影响, 这往往易被某些人忽视, 单纯从罗列出来的发病数据就断然下结论, 正确的做法, 应寻找相应的数理统计方法进行检验, 检验的结论才具科学性和可靠性。固然一种统计检验方法往往有多种用途, 但分析同一问题, 也可以应用多种统计方法。本文着重针对同一资料, 应用两种不同的检验方法, 对于获得广东乙脑发病存在季节性周期这一结论却是殊途同归的。

(收稿: 1997-05-14 修回: 1997-06-19)

漳平市象湖村饮水型地方性氟中毒流行病学分析

陈文金 郑清源 杨寿钦

为了解我市象湖村饮水型氟中毒流行情况, 1982 年我站组织力量对该村开展了氟中毒流行病学调查。结果报告如下。

一、调查内容与方法: 采用氟离子选择电极法测定该村饮水和人群尿样氟含量。按中共中央地方病办公室制定的《地方性氟中毒防治工作标准(试行)》, 开展现场人群氟中毒调查, 对有氟骨症临床表现者作 X 线摄片诊断。

二、结果与分析:

1. 氟中毒流行程度: 该村饮水点共有 20 处, 水氟含量在 $0.2 \sim 7.0 \text{mg/L}$, 中位数为 3.2mg/L , 超出 1.0mg/L 饮水占 70%, 饮高氟水人数占 65.8% (部分居民有混饮高低氟水现象); 人群氟斑牙率为 75.7% (502/663), 其中 8~15 岁人群患病率为 73.9%, 以着色型、中度病变为主; 氟骨症率为 3.2% (21/663), 临床表现以 I 度为主 (62%), 无 II 度

病人, X 线检查以 I 度 (含早期改变) 为主 (81%), 16 岁以上人群 II 度氟骨症率为 0.8%; 病区与低氟区人群尿氟含量的几何均值分别为 3.6mg/L (37 份) 和 0.6mg/L (51 份)。该病区经模糊综合评判判定为中等病区。

2. 氟中毒流行病学关系: 水氟含量与氟斑牙、氟骨症率有较好的对应上升趋势 ($P < 0.001$); 氟斑牙率与年龄呈余弦曲线关系下降 ($P < 0.001$), 氟骨症率与年龄则为指数曲线上升 ($P < 0.001$), 女性的氟斑牙、氟骨症率均显著高于男性 ($P < 0.05$); 氟斑牙、氟骨症率均随着饮水年限的增加而上升 ($P < 0.001$), 但表现出中毒程度与年限长短的此高彼低或此低彼高现象。此外, 外地迁入者的氟斑牙率明显低于当地生长者 ($P < 0.001$), 而前者的氟骨症率显著高于后者 ($P < 0.05$); 未生育女性的氟骨症率显著低于生育者, 且生育胎数多者 II 度以上患者所占比例大 ($P < 0.001$)。