

纵向生命质量资料分析的 Markov 过程法及其应用

万崇华¹ 方积乾² 周文清³ 秦学军²

【摘要】 目的 提出一种分析纵向生命质量资料的方法并探讨有关的应用问题。方法 基于 Markov 过程思想提出 Markov 过程法, 并且用于 212 例吸毒者在戒毒期间生命质量的变化及影响因素分析。结果 职业对状态 1 \rightarrow 2 和状态 2 \rightarrow 3 的转移强度系数分别为 0.7912 和 0.6797; 戒毒时间对状态 1 \rightarrow 2 的转移强度系数为 0.5663, 戒毒方式对状态 2 \rightarrow 3 的转移强度系数为 0.2765。结论 本文方法是一较好的纵向资料分析法, 既能分析生命质量的状态转移又能分析转移的影响因素。职业、戒毒方式和戒毒时间能影响生命质量的变化。

【关键词】 Markov 过程 纵向资料 生命质量 吸毒

Markov process method in analyzing longitudinal quality of life data WAN Chonghua^{*}, FANG Jiqian, ZHOU Wenqing, et al. ^{*} Department of Health Statistics, Kunming Medical College, Kunming 650031

【Abstract】 Objective To present a method to analyze longitudinal quality of life data and discuss some relevant questions. **Method** Based on the theories of Markov process, the Markov method was presented and used to analyze the change of quality of life status for 212 cases being detoxified drug addicts and their influence factors. **Results** The intensity coefficients of status transient for 1 \rightarrow 2 and 2 \rightarrow 3 for occupation were 0.7912 and 0.6797 respectively; the coefficients for status 1 \rightarrow 2 for the detoxification time was 0.5663 and that for 2 \rightarrow 3 for the detoxification method was 0.2765. **Conclusions** The Markov method is good to analyze longitudinal quality of life data for it can analyze both the transient probabilities between any two status of quality of life and their influence factors. The results showed that the occupation of drug addicts, the detoxification time and method can affect the change of quality of life.

【Key words】 Markov process Longitudinal data Quality of life Drug addiction

随着生命质量(quality of life, QOL)研究的兴起,涌现了大量的生命质量资料。鉴于生命质量资料的复杂性,一些学者^[1~3]对其统计分析方法进行了专门的研究讨论。由于

生命质量的时间依赖性,纵向的重复多次的生命质量测评才真正体现生命质量研究的精髓。但纵向资料的处理即使在其它领域也尚未很好解决,在生命质量领域更是鲜为人知。由此严重阻碍了生命质量测评的开展及深入,以致于目前多数仅开展横切面的测定或者前后两次的测定。

如果将生命质量在时间上的变化视为 Markov 过程,即可试用 Markov 过程思想来处理。实际上,已有一些研究者^[1,2]注意到

1 昆明医学院卫生统计学教研室 云南 650031
2 中山大学卫生统计学教研室
3 苏州工业园区管委会
高等学校博士点学科专项科研基金资助(基金编号

了这一点,但具体方法尚未见报道。为此,笔者应用 Markov 过程思想提出一种分析方法,并应用于吸毒者生命质量资料的分析。

方法及原理

由于生命质量的变化与时间有关,采用非齐性 Markov 过程来处理。生命质量在时间上的变化是一个连续过程,其状态是连续的。鉴于随机过程分析的复杂性,我们将其离散化后按纯不连续 Markov 过程处理。方法是将生命质量综合为一个指标(生命质量总分),然后定出一些界点,从而分为不同的状态。根据各状态间的关系可得到不同的变化模式,但基本的模式如下:

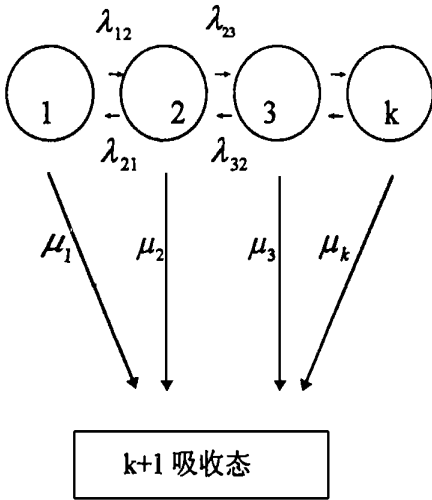


图 1 生命质量变化的基本模式

设生命质量的变化为随机过程 $X(t)$, 其转移概率定义为:

$$p_{ij}(t, \tau) = \Pr\{X(t) = j | X(\tau) = i\} \quad (1)$$

$i = 1, \dots, k, j = 1, \dots, k+1$ 。

其状态间的转移强度定义为:

$$\lambda_{ij}(t) = \lim_{dt \rightarrow 0} \frac{\Pr\{X(t+dt) = j | X(t) = i\}}{dt} \quad (2)$$

$$\mu_i = \lambda_{i, k+1}(t) =$$

$$\lim_{dt \rightarrow 0} \frac{\Pr\{X(t+dt) = k+1 | X(t) = i\}}{dt} \quad (3)$$

$i, j = 1, \dots, k, i \neq j$ 。

转移概率与转移强度的关系由 Kolmogorov 方程得到:

$$\frac{dP(t)}{dt} = P(t)Q \quad (4)$$

式中 Q 为转移强度矩阵,按上述基本模型有:

$$Q = \begin{pmatrix} -(\mu_1 + \lambda_{12}) & \lambda_{12} & 0 & \dots & 0 & \mu_1 \\ \lambda_{21} & -(\mu_2 + \lambda_{21} + \lambda_{23}) & \lambda_{23} & \dots & 0 & \mu_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -(\mu_k + \lambda_{k-1}) & \mu_k \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

经求解式(4)得:

$$P(t) = A \text{diag}\{e^{\rho_1 t}, \dots, e^{\rho_{k+1} t}\} A^{-1} \quad (5)$$

式中, ρ 为 Q 的特征根, A 为一方阵,其各列元素为各特征根 ρ 对应的特征向量。

在上述基础上即可构造模型并估计参数。

1. 构造模型: 类似于 Cox 回归模型,我们将每个转移强度定为 n 个伴随变量 $Z_m(t)$ 的线性模型^[4] ($m = 1, \dots, n$):

$$\lambda_{ij}(t) = \beta_{ij,0} + \sum_m \beta_{ij,m} Z_m(t) \quad (6)$$

2. 参数估计: 按上述设置的模型,将有大量的参数需要估计,任两状态转移间均有 $n+1$ 个参数。除非从专业上判断某些因素对某转移无作用,从而从模型中剔除之。

我们采用最大似然法来估计参数。设第 i 个个体在第 j 次测定时的时刻为 t_{ij} , 所处的状态为 $S(t_{ij})$, 参加测定的次数为 m_i , 则似然函数定义为各转移概率的乘积:

$$L = \prod_{i=1}^N \prod_{j=1}^{m_i} P_{S(t_{ij})S(t_{i,j+1})}(t_{ij}, t_{i,j+1}) \quad (7)$$

N 为参与测定的总例数。

可见,通过(5)、(6)和(7)式将 β 和 L 联系起来,使 L 达到最大即可估计出一组参数 β 。具体计算时可采用 Hooke-Jeeve 直接搜

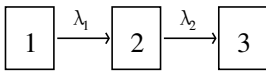
索法来求解参数^[5]。

3. 假设检验: 伴随变量对各状态转移的作用大小有无统计学意义可通过假设检验来实现。这与 logistic 或 Cox 模型中检验各参数的方法相同, 多用似然比检验进行。这里从略。

实例分析

为了对吸毒者在戒毒期间生命质量的变化规律及影响因素进行分析, 用我们研制出的吸毒者生命质量测定量表 QOL-DA^[6], 对 212 例吸毒者的生命质量进行了 5 次纵向测定(每次间隔一周), 同时对一般人口特征(如性别、文化程度)、吸毒情况(如吸毒方式、每日吸毒量、吸食种类)和社会家庭情况的 41 个因素进行了调查。这里用前述的 Markov 过程法来分析其生命质量的变化及其影响因素。

首先, 根据各时点生命质量测定情况将其分为三个渐进状态: 1: QOL < 120, 2: 120 ≤ QOL ≤ 139, 3: QOL ≥ 140。



其次, 考虑戒毒方式 Z₁, 性别 Z₂, 文化程度 Z₃, 职业 Z₄ 和戒毒时间 Z₅ 五个因素, 其中前四个不随时间变化, 从而有〔为简便计, 省去 λ(t) 中的 t〕:

$$\begin{cases} \lambda_1 = A_0 + A_1Z_1 + A_2Z_2 + A_3Z_3 + A_4Z_4 + A_5Z_5(t) \\ \lambda_2 = B_0 + B_1Z_1 + B_2Z_2 + B_3Z_3 + B_4Z_4 + B_5Z_5(t) \end{cases}$$

可推得转移概率与转移强度的关系为(未列出者均为 0):

$$\begin{aligned} p_{11} &= e^{-\lambda_1} \\ p_{12} &= \frac{\lambda_1}{\lambda_1 - \lambda_2} (e^{-\lambda_2} - e^{-\lambda_1}) \\ p_{13} &= \frac{\lambda_2 e^{-\lambda_1} - \lambda_1 e^{-\lambda_2}}{\lambda_1 - \lambda_2} + 1 \\ p_{22} &= e^{-\lambda_2} \\ p_{23} &= 1 - e^{-\lambda_2} \\ p_{33} &= 1 \end{aligned}$$

根据 5 个时点的测定资料, 按(7)式并用直接搜索法即可求得各参数值(见表 1)。

经检验, 职业、戒毒时间对转移 1 → 2 有显著性, 戒毒方式和职业对转移 2 → 3 作用有显著性。

表 1 吸毒者生命质量状态转移强度系数估计值

因素	转移 1 → 2 (A)	转移 2 → 3 (B)
常数项	0.2409	0.4029
戒毒方式	0.1267	0.2765
性别	0.1507	0.0900
文化程度	0.2817	0.3200
职业	-0.7912	0.6797
戒毒时间	0.5663	0.0917

讨 论

作者将连续型生命质量划分为离散状态, 从而可大大简化处理方法。状态划分的关键是定出分界点, 这往往从专业角度来考虑而且带有一定的主观性。如 Gentleman^[7]在研究 AIDS 病的进程时分别按 CD₄ T 淋巴细胞计数和 IgA 浓度来划分状态。前者按 ≥500(1), 200~499(2), <200(3) 和 AIDS(4) 分为四个状态; 后者按 <136(1), 136~200(2), 201~291(3), ≥292(4) 和 AIDS(5) 分为五个状态。Kay^[8]在研究肝细胞癌的进程时按某时刻与初始时刻甲胎蛋白(AFP)含量的对数值的比值来划分状态。≤1.05 为正常状态, >1.05 为升高状态, 随后就是吸收态: 死亡。其取对数变换的目的主要是使资料分布对称化。显然, 取 1.05 为界值点就是比初始时增加 5% 以上即认为进入另一状态。在生命质量评定中, 笔者认为可采用下面三种分法:

1. 按条目界点来划分: 用各条目理论上的界点分的总和为界点。比如由 20 个 5 级式条目构成的量表, 每个条目按等级 1~5 分, 则理论上的量表粗分为 100 分, 除两个极端点外其余三个界点的理论得分的极大值分别为 40、60 和 80, 用此三个界点即可将生命

质量状态分为 4 个部分。

2. 按增长的比率划分: 此即类似 Kay 的方法, 只是生命质量一般均呈正态分布, 没有必要作对数变换。增长率的确定也不一定按 5%、10%, 可根据实际资料变化的大小来确定。

3. 结合某些特征来划分: 可结合频数分布、假设检验和某些症状体征等来划分。比如当按某界点化分没有频数或频数太小时, 说明此状态的实际意义可能不大。按某界点化分差异无显著性时意义可能也不大。

本文介绍的 Markov 过程处理生命质量状态变化的方法, 不但可分析各状态的变化情况(转移概率), 而且可分析任意两状态间转移的影响因素。这对深入的分析生命质量纵向变化及影响因素是非常有用的。其关键是状态的合理划分及建立伴随变量与转移强度的关系模型。

在实际应用中, 可根据实际情况对基本模式进行删减, 略去一些状态间的转移通路, 得到不同的特殊模式。如规定不能进行逆转, 各状态只能渐进, 则处理起来就非常方便。本文实例正是如此。在癌症和某些严重的慢性病治疗或随访观察过程中, 其生命质量愈来愈差, 直到到达终点(死亡)。在另一些情形下(如吸毒者的戒毒过程, 疾病治疗后的康复过程等)生命质量逐渐变好, 并达到终点(脱毒、痊愈等)。这些都可简化为渐进模型。

本文中转移强度模型设置为线性关系主要是为了处理方便, 实际上也可设为指数等其它模型^[9]。各模型中伴随变量的多少可根据专业知识来取舍, 舍弃一些不重要的因素, 处理起来就方便一些。如有必要, 尚可估计出各状态的平均逗留时间^[10]和任意状态间的转移概率。

此外, 本文方法不限于生命质量资料的

分析, 凡可根据某特征划分为多个状态的资料均可采用。

尽管 Markov 过程法分析纵向生命质量具有诸多优点, 但也有一些局限性。首先是前述的状态划分具有一定的任意性; 其次, 计算问题也是一大难题, 实际工作者很难编制相应的计算程序, 而且由于状态划分和建模的多样性, 也难以编制通用的计算程序, 这在一定程度上限制了其推广使用。

参 考 文 献

- 1 Olschewski M, Schumacher M. Statistical analysis of quality of life data in cancer clinical trials. *Statistics in Medicine*, 1990, 9: 749- 763.
- 2 Cox DR, Fitzpatrick R, Fletcher AE, et al. Quality-of-life assessment: can we keep it simple? *JR. Statist Soc A*. 1992, 155: 353- 393.
- 3 万崇华, 方积乾. 生命质量资料的统计分析方法. *中华预防医学*, 1996, 30: 172- 174.
- 4 方积乾, 伍超标, 毛建华, 等. 肿瘤潜隐期的二阶段模型(I) — 非齐性 Markov 模型. *应用概率统计*, 1995, 11: 202- 212.
- 5 王德人编. 非线性方程组解法与最优化方法. 北京: 高等教育出版社, 1979, 208- 220.
- 6 万崇华, 方积乾, 陈丽影, 等. 海洛因成瘾者生存质量测定量表的制定及其考评. *中国行为医学科学杂志*, 1997, 6: 169- 171.
- 7 Gentleman RC. Multi - state Markov models for analysing incomplete history data with illustrations for HIV disease. *Statistics in Medicine*, 1994, 13: 805- 821.
- 8 Kay R. A Markov model for analysing cancer markers and disease states in survival studies. *Biometrics*, 1986, 42: 855- 865.
- 9 Anderson PK. Multistate models in survival analysis: a study of nephropathy and mortality in diabetes. *Statistics in Medicine*, 1988, 7: 661- 670.
- 10 Duffy SW, Chen HH, Tabar L, et al. Estimation of mean sojourn time in breast cancer screening using a markov chain model of both entry to and from the preclinical defectable phase. *Statistics in Medicine*, 1995, 14: 1531- 1543.

(收稿: 1998- 08- 29 修回: 1998- 09- 10)