

## 在馆读者的实时学习状态抽样分析与估算

——以南京师范大学社科专业图书馆为例

邓盛<sup>1</sup>, 杨宇舟<sup>2</sup>, 朱全新<sup>3</sup>

(1. 南京师范大学 图书馆, 江苏 南京, 210023; 2. 厦门大学 经济学院, 福建 厦门, 361000; 3. 南京师范大学 数学科学学院, 江苏 南京, 210023)

**摘要:** 根据南京师范大学社科专业图书馆 2016 年 4.22—5.22 的在馆读者信息, 以可反映学习状态的参数——在位人数、总占位数及吃零食、玩手机、用电脑、睡觉、交谈、学习等的人数, 对该馆内读者的实时学习状态进行了抽样分析。首先, 通过读者出入馆信息确定了状态参数的抽样时间为上午 10—11 时, 并定义学风指标为学习人数除以实际在位人数; 再对抽样时间点的学习状态参数及天气情况建立模型。通过建立负二项回归模型、广义线性模型, 并应用稳健性检验方法得到了在馆读者的实时学习状态。结果表明: 学风指标区间为 60%~80%, 处于良好状态; 8—11 时的进馆人数与学习氛围呈正相关关系, 天气因素对馆内读书人数有较大影响, 其是无法改变的外生变量, 可通过改善图书馆的硬件设施与服务质量提升进馆人数; 模型估计的学风指标与实际情况基本一致。

**关键词:** 图书馆; 实时学风; 抽样调查; 稳健性检验方法

中图分类号: G 251

文献标志码: A

文章编号: 1672-6146(2017)04-0018-05

## Analysis and estimation of learning style in Library

Deng Sheng<sup>1</sup>, Yang Yuzhou<sup>2</sup>, Zhu Quanxin<sup>3</sup>

(1. Library of Nanjing Normal University, Nanjing 210023, China; 2. School of Economics, Xiamen University, Xiamen 361000, China; 3. School of Mathematical Sciences, Nanjing Normal University, Nanjing 210023, China)

**Abstract:** Based on the present readers' information from Nanjing Normal University professional library between Apr 22th and May 22th in 2016, the readers' real-time learning style are analyzed by some parameters reflecting learning status such as the present number, the total number of occupied seats, and the numbers of who are eating snack, playing phones, using computer, sleeping, talking and learning. Firstly, the readers' entrance information are used to ensure that the sampling time of status parameters is 10—11 a.m. each day, and we define the learning style index as the number of who are learning divided by the real number of who are present seated. Secondly, a model is put forward based on the learning status parameters and weather conditions from sampling time point. By the Negative Binomial Regression, Generalized Linear Model and robust test, the real-time learning status of present readers is put forward. The results show that the interval of the learning style is 60%~80%, which is good; there is a positive correlation between the entrance number of 8—11 o'clock and the learning atmosphere. The weather condition is an unchanged exogenous variable that affects the number of readers. The improved library's hardware facilities and service quality can increase the entrance number. The index made by the model is basically consistence with the actual circumstance.

**Key words:** Learning style; real-time learning style; sampling survey; robustness test method ; sample interval

通信作者: 朱全新, 05401@njnu.edu.cn。收稿日期: 2017-09-10

基金项目: 国家自然科学基金(61374080); 江苏省自然科学基金(BR20161552)。

学风是高校办学水平评估的重要指标<sup>[1]</sup>,其可以通过学生读者在图书馆实时学习状态部分体现出来。目前,我国对学风建设的研究还处于一个较低的水平,主要集中在学风建设的加强及对策<sup>[2]</sup>,也见于通过加强图书馆建设干预学风<sup>[3-6]</sup>。但对于学风的定量分析与研究鲜见。图书馆作为公共学习重要场所,其在内的自然的学习状态体现了读者的自我学习欲望<sup>[7]</sup>,实时地反映了某一群体的学风状态。本文将以南师范大学社科专业图书馆为例,用能反映学习状态的参数对馆内读者的实时学习状态进行抽样,并建立模型对图书馆内的实时学风进行定量分析。以期得到一些参量间的相互关系及对学风的影响,并为加强学风建设及制定提升入馆人数措施提供参考。

## 1 实时学风抽样统计与分析

本文拟从3个方面讨论馆内读者的实时学习状态。首先,筛选出入馆信息数据,选取非功利学习时间段在馆读者量相对稳定且最高的时间点进行持续抽样统计,得到图书馆读者的实时学习行为数据;其次,通过每天进馆流量实时数据对稳定时间点的实时学风进行统计调查,得到实时学习状态的样本区间;最后,通过模型估算学习氛围的范围值。

### 1.1 择取 10—11 时作为抽样时间的依据

当图书馆进出流量基本平衡且最小时,就可以认定其为当天数据最稳定的时间点,这个时点的数据最能体现读者在图书馆的实时学习状态。在抽样时间 2016.4.22—2016.5.22 中,图书馆进馆人数的分布如图 1 所示。图 1 显示,每天每小时平均进馆人数有明显波动;图中曲线代表了均值点的波动轨迹,而 10—11 时的均值正好处于全天中较平稳的低进出值。其它时间数据起伏波动较大(另外 20—21 时数据因 21 点闭馆,呈自然闭合时间),不具备数据分析上的普遍意义,因此,本文对 10—11 时的数据进行了抽样调查。

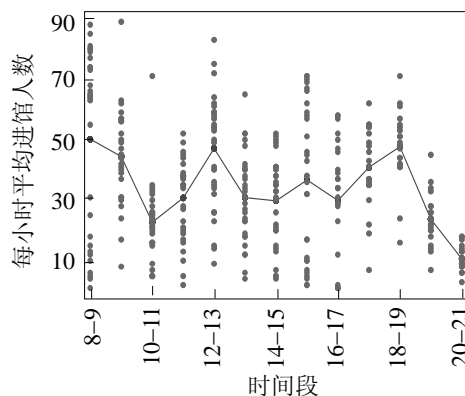


图 1 2016 年 4 月 22 日—5 月 22 日 8—21 时平均进馆人数分布状况

### 1.2 10—11 时学习状态抽样数据统计

在对 10—11 时学习状态的抽样中,统计了从 2016 年 4 月 22 日到 5 月 22 日为期 1 个月的数据。

表 1 不同行为读者的人数统计结果

变量名	均值	标准差	中位数	最小值	最大值
seat	70.13	30.13	72.50	21.00	132.00
occup	108.70	40.27	102.50	39.00	198.00
snack	3.30	1.95	3.00	0.00	7.00
phone	10.13	3.70	10.00	2.00	19.00
computer	3.60	2.19	3.50	0.00	10.00
asleep	1.10	1.06	1.00	0.00	3.00
chat	3.20	1.95	3.00	0.00	7.00
learn	48.63	24.98	51.50	8.00	98.00

这个时间段正处于学生期中考试完结,而期末考试尚远,且无大规模行业认证考试的相对非功利性时期。本文以在位人数(seat)、总占位数(occup)及吃零食(snack)、玩手机(phone)、用电脑(computer)、睡觉(asleep)、交谈(chat)、学习(learn)等读者人数为学习状态的统计指标,统计结果如表 1 所示。由表 1 可知,实际在座读者的平均数量

(70.13 人)少于总占位的平均数量(108.7 人),这可能是由于跨坐、暂时不在座位、占座等现象增加了总占位的数量。因此,实际在馆读者应该大于在位人数,但小于总占位数。交谈、吃零食、玩手机、用电脑、睡觉的人都较少,在座学习的平均人数为 48.63,超过总在位人数的一半。

用学习人数除以实际在位人数作为学风指标( $I_s$ )。则由表 1 可得到学风指标变化如图 2 所示。图 2 中剔除了节假日时间的数据,但为了数据在时间上的连贯性,保留了节假日

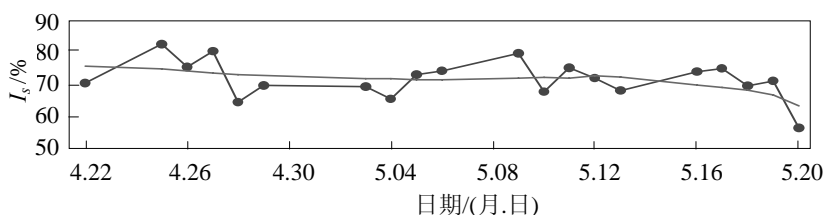


图 2 2016 年 4 月 22 日—5 月 22 日 10—11 时学风指标变化趋势

的时间刻度,而没有显示相应数值。事实上即便考虑节假日时间的数据,趋势也不会发生明显的变化。由图2可知,1个月的实时学风指标稳定在60%~80%之间。

## 2 模型选择与估值

### 2.1 负二项回归模型

由表1可知,平均在馆人数方差均值比为12.83,远大于1。那么学习人数数据存在过度分散(over dispersion)的特征,不满足进行泊松回归(Poisson Regression)的条件,因此选择负二项回归(Negative Binomial Regression)对数据进行分析。由文献[8]有负二项分布的密度函数

$$f(y|\mu, \alpha) = \frac{\Gamma(y + \alpha^{-1})}{\Gamma(y+1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\alpha^{-1} + \mu}\right)^y, \quad \alpha \geq 0, \quad y = 0, 1, 2, \dots \quad (1)$$

其中: $y$ 是被解释变量,也是计数数据; $\mu$ 是指数化的解释变量的条件均值,且 $E(y_i | x_i) = \mu_i = \exp(x_i' \beta)$ ,  $x_i$ 是影响被解释变量 $y$ 的第 $i$ 种可能因素,包括图书馆在馆人数,天气与气温,是否为周末等等; $\alpha$ 表示数据分散程度; $\Gamma(\cdot)$ 是伽马函数,即当 $y$ 为整数时, $\Gamma(y+a)/\Gamma(a) = \prod_{j=0}^{y-1} (j+a)$ 。因此

$$\ln\left(\frac{\Gamma(y + \alpha^{-1})}{\Gamma(\alpha^{-1})}\right) = \sum_{j=0}^{y-1} \ln(j + \alpha^{-1}) \quad (2)$$

将式(2)代入式(1),考虑指数均值 $\mu_i = \exp(x_i' \beta)$ ,有 $\ln L(\alpha, \beta) = \ln \sum_{i=1}^n \{[\sum_{j=0}^{y_i-1} \ln(j + \frac{1}{\alpha})] - \ln y_i! - (y_i + \frac{1}{\alpha}) \cdot \ln(1 + \alpha \exp(x_i' \beta)) + y_i \ln \alpha + y_i x_i' \beta\}$ 。对此求关于 $(\beta, \alpha)$ 的一阶导,令其等于0,有 $\sum_{i=1}^n \frac{y_i - \mu_i}{1 + \alpha \mu_i} x_i = 0, \sum_{i=1}^n (\frac{1}{\alpha^2} \cdot (\ln(1 + \alpha \mu_i) - \sum_{j=0}^{y_i-1} \frac{1}{j + \alpha^{-1}}) + \frac{y_i - \mu_i}{\alpha(1 + \alpha \mu_i)}) = 0$ 。

### 2.2 广义线性模型

因为负二项模型与泊松模型假定了数据分布的具体形式,为了防止结果过于依赖于假设,因此使用更一般的广义线性模型(GLM)。其概率密度函数为 $f_{\text{GLM}}(y|\theta, \phi) = \exp(\frac{a(\phi)}{\theta y - b(\theta)} + c(y, \phi))$ ,其中: $b(\cdot)$ 满足 $E(y) = b'(\theta)$ ;函数 $a(\phi) = V(y) / b''(\theta)$ ;函数 $c(\cdot)$ 则是归一化常数项。不同的 $a(\cdot)$ 与 $b(\cdot)$ 形式导致不同的GLM形式。如对于泊松模型,相当于 $b(\theta) = \exp(\theta), a(\phi) = 1, c(y, \phi) = \ln y!$ ,因此 $E(y) = V(y) = \exp(\theta)$ 。通常情况下, $\theta = x' \beta$ 。因此,广义线性模型的极大似然函数满足 $L_{\text{GLM}} = \sum_{i=1}^n ((\theta(x_i' \beta) y_i - b(\theta(x_i' \beta))) \frac{1}{a(\phi)} + c(y_i, \phi))$ 。故一阶条件为 $\sum_{i=1}^n \frac{1}{\omega_i} (y_i - \mu_i) \frac{\partial \mu_i}{\partial \beta} = 0$ ,其中波动函数为 $\omega_i = a(\phi) v(\mu(x_i' \beta))$ ,以及 $v(\mu_i) = b''(\beta_i)$ 。

### 2.3 结果与分析

模型选择10—11时的实时学习状态指标进行分析。首先,从开馆时间到11时前的进出人数会对10—11时的学习人数产生影响;其次,图书馆的人数会受到天气与气温的影响;最后,周末与节假日也会影响图书馆人数,为此将周末时间剔除。

根据以上分析,本文构建模型 $E(I_{st} | x_t) = \mu_t = \exp(x_t \beta) = \exp(\beta_1 n_8 + \beta_2 n_9 + \beta_3 n_{10} + \beta_4 T_{\min} + \beta_5 r_t + c)$ 。其中: $I_{st}$ 为 $t$ 时刻的学风指标, $n_8, n_9, n_{10}$ 分别表示8—9, 9—10, 10—11时进馆人数, $r_t$ 为 $t$ 时刻天气虚拟变量(雨天为1,否则为0), $T_{\min}$ 是最低气温, $c$ 为常数项。为了防止节假日数据产生误差项,在以上模型中去掉了节假日的时间段。

为了防止因为模型选择而产生特定的偏误,故采用了3种不同的模型进行分析,即基准模型负二项回归、GLM框架下的泊松回归(泊松GLM)与负二项回归(负二项GLM)。

本文采用Stata12.0作为数据分析工具,根据文献[9]的方法,可以得到如表2所示的分析结果。由表2可知,前3个小时的进馆人数与图书馆的学风呈正向关系,其中8—9时进馆人数每增加1人,11时

在馆的学习人数增加 1.1%~1.4%; 9-10 时进馆人数每增加 1 人, 11 时在馆的学习人数增加 0.75%~0.80%; 而到了 10-11 时的进馆人数, 对在馆学习人数的影响只有 0.66% 左右。因此, 随着进馆时间越靠近抽样时间, 进馆人数与学习氛围的关联越小。从上述回归结果

**表 2 学习人数的负二项回归与 2 种 GLM 回归分析结果**

参数	负二项回归	泊松 GLM	负二项 GLM
$n_8$	0.011 4***(3.499 8)	0.011 4***(3.412 0)	0.011 1***(3.701 3)
$n_9$	0.007 7**(2.466 6)	0.007 5**(2.468 1)	0.008 0**(2.412 3)
$n_{10}$	0.006 6**(2.083 7)	0.006 5**(1.996 9)	0.006 7**(2.139 7)
$T_{min}$	0.083 7***(3.367 3)	0.081 3***(3.554 7)	0.086 4***(3.337 3)
$r_t$	-0.3845***(-2.765 9)	-0.367 1***(-2.756 3)	-0.402 9***(-2.939 9)
$c$	1.504 2**(2.511 2)	1.555 4***(2.746 4)	1.466 9**(2.375 9)
Lalpha C	-4.310 5***(-5.838 4)		
$N$	20	20	20

注: 括号中的数据为 t 统计量; \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ 。

可以得出, 最低气温每上升 1℃, 会使图书馆读书人数上升 8.5%~9.0%; 下雨天要比不下雨的时候少 30.7%~33.2% 的学习人数。因此, 外界环境对实时学习状态的影响要远超图书馆的内部环境影响。天气与气温都是无法改变的外生变量, 进馆人数可通过改善图书馆的硬件设施与服务质量而得到提升。在考虑周末节假日的虚拟变量情况下, 是否下雨对被解释变量的影响不再显著, 这说明周末节假日的天气情况对被解释变量的影响有所不同。与其他自然因素比较(是否下雨、最低气温等), 进馆人数对实时学习状态的影响要小得多, 但它在评估学习氛围算法中的作用不可忽视, 因为它是唯一能够人为调整的因素。

为了方便不同模型之间的比较, 同时为了更直观地判断不同模型的拟合效果, 给出实际值与估计值的协方差矩阵如表 3 所示。对拟合值与真实数据的相关系数矩阵进行分析, 可以发现泊松 GLM 的拟合效果最佳, 但与其他 2 种方法的结果相差很小。

**表 3 实际值与估计值的协方差矩阵**

变量	实时学习	负二项回归	泊松 GLM	负二项 GLM
实时学习	1.000 0			
负二项回归	0.888 2	1.000 0		
泊松 GLM	0.890 0	0.999 9	1.000 0	
负二项	0.885 1	0.999 8	0.999 3	1.000 0

不同模型存在着系数波动, 但不同模型中解释变量系数的正负性并没有发生变化, 模型的显著性也没有发生变化, 同一变量的不同系数大小也没有很明显的波动, 因此, 可以认为不同模型的拟合效果相似。

### 2.4 稳健性检验

为了保证数据的稳健性, 利用图书馆在位人数再次对上述模型进行分析。由于在馆学习人数必然属于图书馆在座人数, 因此若用在座人数评估图书馆学习氛围, 则会稀释真实数据, 降低模型的解释力。在这种情况下, 只要模型主要拟合系数的定性结论不发生变化, 就可以判定模型对数据的解释力是比较稳健的。

在座人数的分析结果如表 4 所示。由表 4 可知, 对于不同被解释变量, 解释变量所拟合的参数性质(正负性)与学习人数的分析结果相比没有发生变化, 只是对被解释变量(在座人数)的影响程度降低了。这种影响程度的降低, 与降低的解释力是一致的: 前 3 h 的进馆人数与图书馆学习氛围呈正向关系, 其中 8—9 时进馆人数每增

**表 4 在座人数的负二项回归与 2 种 GLM 回归分析结果**

参数	负二项回归	泊松 GLM	负二项 GLM
$n_8$	0.008 7***(3.944 8)	0.008 5***(3.644 6)	0.008 7***(4.232 1)
$n_9$	0.006 6**(2.299 1)	0.006 4**(2.205 5)	0.006 9**(2.363 0)
$n_{10}$	0.004 7*(1.665 9)	0.004 5(1.506 8)	0.004 8*(1.806 0)
$T_{min}$	0.081 6***(3.996 1)	0.080 1***(4.160 2)	0.083 2***(3.939 9)
$r_t$	-0.271 3***(-3.291 1)	-0.266 1***(-3.241 5)	-0.275 1***(-3.426 6)
$c$	2.157 6***(4.504 4)	2.211 3***(4.850 2)	2.108 1***(4.286 2)
Lalpha C	-4.540 1***(-6.833 2)		
$N$	20	20	20

注: 括号中的数据是 t 统计量; \*  $p < 0.1$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ 。

加 1 人, 图书馆学习人数增加 0.85%~0.87%; 9—10 时进馆人数每增加 1 人, 图书馆学习人数增加 0.64%~0.69%; 10—11 时进馆人数每增加 1 人, 图书馆学习人数增加 0.45%~0.48%。最低气温每上升 1℃, 图书馆读书人数上升 8.3%~8.7%。对于是否下雨的虚拟变量, 可以发现在正常时间段内, 雨天的在座人数比不下雨时少 23.4%~24.1%。

表 5 在座人数实际值与估计值的协方差矩阵

变量	实时学习	在座人数	负二项回归	泊松 GLM	负二项 GLM
实时学习	1.000 0				
在座人数	0.983 4	1.000 0			
负二项回归	0.873 6	0.873 5	1.000 0		
泊松 GLM	0.873 9	0.873 8	1.000 0	1.000 0	
负二项 GLM	0.872 1	0.872 2	0.999 9	0.999 9	1.000 0

同理, 在座人数分析的协方差矩阵如表 5 所示。表 5 与表 3 相比, 拟合效果也要差一些, 但泊松 GLM 的拟合效果仍然要比其他 2 个模型高一些。

### 2.5 学习人数估计

通过以上不同模型对参数的估计结果, 以及得到的估计值与实际值之间的相关系数, 能粗略地描述学习氛围的波动情况。这为估算或者预测其他时间段的实时学习状态提供了参数。在利用社科图书馆 2014 年至 2015 年的相关数据, 对实时学习状态进行估算时, 首要是找到所有可能的解释变量, 包括 2014 年至 2015 年对应时期的天气状况、8—11 时进馆人数等。在此基础上, 得到了图 3 所示的学习氛围估计值。

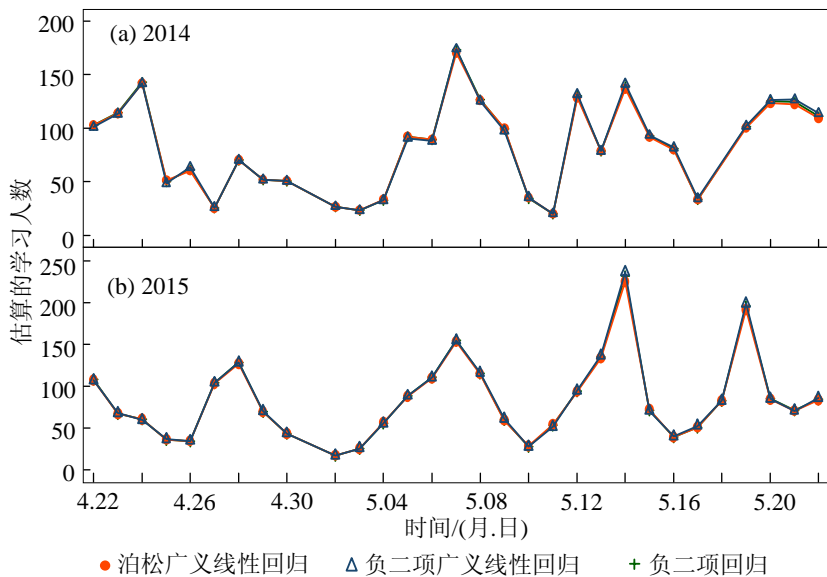


图 3 2014 与 2015 年 4 月 22 日—5 月 22 日学习人数的不同模型估值

用不同模型估算在不同时间内的学习人数分布如图 3 所示。由图 3 可知, 在估算的学习人数中, 最多接近满员, 出现在五一劳动节之后的 2 个星期中。最少人数小于 50 人, 往往出现在周末。用同样的方法, 也可以估算出不同年份相同时间段的图书馆在座人数, 从而得到实时学风的估值。

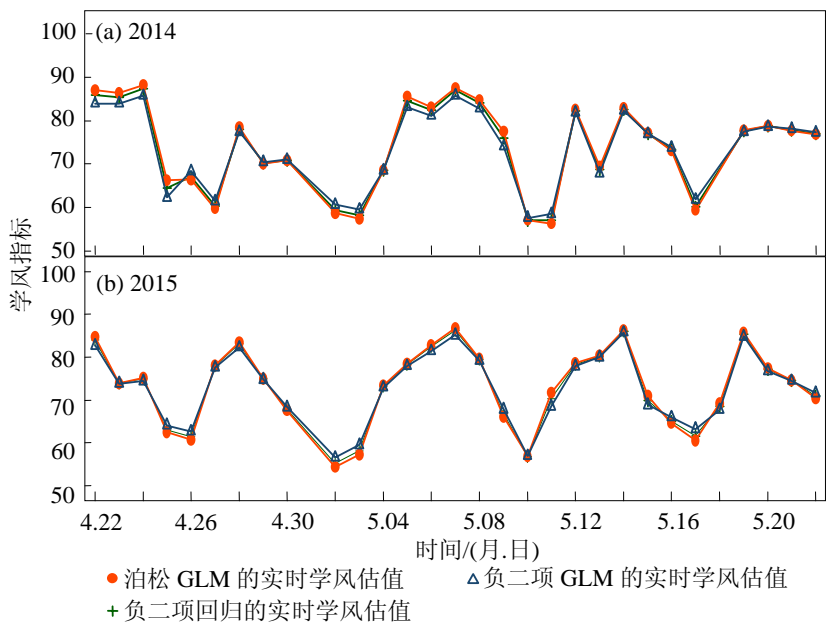


图 4 2014 与 2015 年 4 月 22 日—5 月 22 日实时学风百分比估计值

2014 与 2015 年 4 月 22 日—5 月 22 日的实时学习状态百分比估计值分布如图 4 所示。由图 4 可知, 学风百分比主要集中在 60%~80% 的区间内, 并且学习状态随着周末与节假日出现呈周期性的上下波动。通常在节假日, 图书馆的实时学习状态也到达最低值。

3 种不同估算方法的结果及拟合效果都非常接近, 说明研究方法是可行的。

(下转第 44 页)

- [3] 古伟良. 高频等离子体气相反应法制备氮化硅超微细粉[J]. 稀有金属, 1991, 13(6): 52-55.
- [4] 管绍茂, 王迅. 半导体表面钝化技术及其应用[M]. 北京: 国防工业出版社, 1981: 134-138.
- [5] 田民波, 刘德令. 膜科学与技术手册[M]. 北京: 机械工业出版社, 1991: 5.
- [6] 于映, 陈抗生. 用 PECVD 法在金属衬底上沉积氮化硅薄膜[J]. 电子元件与材料, 1996, 15(1): 27-29.
- [7] 赵永军, 王民娟, 杨拥军, 等. PECVD SiN<sub>x</sub>薄膜应力的研究[J]. 半导体学报, 1999, 20(3): 183-187.
- [8] 于映, 陈跃. 氮化硅介质薄膜内应力实验研究[J]. 真空科学与技术, 2001, 21(1): 51-54.
- [9] Armin G. Overview on SiN surface passivation of crystalline silicon solar cells [J]. Solar Energy Materials & Solar Cells, 2001, 65: 239-248.
- [10] 吴永汉, 窦菊芳. 椭圆法测膜厚的直接计算方法[J]. 物理实验, 1996, 18(1): 11-13.
- [11] Yelundur V, Rohatgi A, Jeong J W, et al. PECVD SiN<sub>x</sub> induced hydrogen passivation in string ribbon silicon [C]// 28th IEEE photovoltaic specialists conference, 2000: 91.

(责任编辑: 刘刚毅)

(上接第 22 页)

### 3 结论

本文从出入馆信息数据入手, 首先确定了上午 10—11 时作为抽样调查的基准时间, 然后针对 2016 年 4 月 22 日到 5 月 22 日的图书馆读者行为进行了为期 1 个月的抽样统计。发现多数读者是在认真学习, 学习状态指数维持在 60%~80%之间, 并且随着周末与节假日而发生波动。随后, 本文根据天气、气温、图书馆显示的读者进入人数等数据, 对图书馆学习状态进行了估计: 不仅发现外在因素更大程度上影响读者的学习状态, 而且发现每天的读者来馆数量也与图书馆学习状态息息相关。在此基础上, 回顾 2014—2015 年的历史数据, 拟合得到了图书馆在 2014—2015 年对应时间的学习人数与学习状态的状况。2 年对应时间的估值结果显示, 图书馆学风仍然处于 60%~80%的区间内, 并且随周末节假日的周期性变化明显。

#### 参考文献:

- [1] 任波. 论高校“四风”建设[J]. 重庆城市管理职业学院学报, 2013, 13(4): 10-15.
- [2] 张勇, 荣翠琴, 王玲. 试论高职院校图书馆的阅读推广模式—以成都航空职业技术学院图书馆为例[J]. 大学图书馆学报, 2014, 32(2): 64-67.
- [3] 杨剑瑜, 杨文奎. 图书馆嵌入式服务信息资源整合平台的模型构建[J]. 湖南文理学院学报(自然科学版), 2014, 26(2): 91-94.
- [4] 段汶君. 试论高校图书馆导读的必要性和具体实施[J]. 图书馆工作与研究, 2006(4): 54-55.
- [5] 马莉, 惠调艳. 高校教学质量影响因素分析及其对策建议[J]. 西安电子科技大学学报(社会科学版), 2011, 21(2): 98-101.
- [6] 倪佳. 阅读推广与大学图书馆管理服务创新[J]. 湖南文理学院学报(自然科学版), 2016, 28(1): 79-82.
- [7] 刘秋林. 教育现代化进程中的高校图书馆[J]. 湖南文理学院学报(社会科学版), 2005, 30(1): 107-109.
- [8] Cameron A C, Trivedi P K. Regression analysis of count data [M]. Cambridge: Cambridge university press, 2013.
- [9] Long J S, Freese J. Regression models for categorical dependent variables using Stata [M]. College Station, Texas: Stata press, 2006.

(责任编辑: 张红)