

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会  
电话：0471-4343756

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email：2010@tzmcm.cn

## 第三届“ScienceWord 杯”数学中国

### 数学建模网络挑战赛 承 诺 书

我们仔细阅读了第三届“ScienceWord 杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛报名号为：

参赛队员（签名）：

队员 1：冯亚男

队员 2：张红冉

队员 3：李家宝

参赛队教练员(签名)： 无

参赛队伍组别：大学

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会  
电话：0471-4343756

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email：2010@tzmcm.cn

## 第三届“ScienceWord 杯”数学中国

### 数学建模网络挑战赛

### 编号专用页

参赛队伍的参赛号码：（请各个参赛队提前填写好）：

1259

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会  
电话：0471-4343756

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email：2010@tzmcm.cn

## 2010 年第三届“ScienceWord 杯”数学中国 数学建模网络挑战赛

题 目 高校图书馆的智能服务

关 键 词 统计 正态分布 曲线拟合 区间分段 信誉限借

### 摘 要：

高校图书馆的设立满足了广大师生汲取知识的需要。然而，即使各图书馆实行超期罚款的政策，仍然有一部分人拖欠图书，从而影响了图书流通。

本文对题目中给定数据的统计分析，关注了拖欠图书人数与超期天数的对应关系，并经 K-S 检验验证了拖欠本数与超期时间满足正态分布关系。对二者关系的曲线拟合，我们在曲线拐点处对其进行分区。从总体和区间两个角度分别讨论了何类书易造成借阅超期，以及发生超期的主要原因。在分析统计的过程中我们主要借助了 EXCEL 和 SPSS 软件。

通过分析我们发现：

超期在30—45天的书占到拖欠总数目的25%，主要拖欠原因为无意识错过还书期限；

45-60天人群只占到1.5%，主要拖欠原因为罚款较少；

60-1400天的拖欠人群占50%，主要拖欠原因为教学科研需要，长期使用；

超期1400-5000天且尚未归还的图书，认为属于历史累积数据，可能由于忘记、丢失等其他原因，归还图书馆概率较低。因此本类数据虽然总体数量较大（占总数量25%），但数据相对分散，造成其数据庞大原因为历史累积，即每次统计由于其归还的概率很小，总会划归到拖欠书籍行列中，且拖欠时间越来越长，拖欠数量越来越多。所以本超期时间段内数据不计入统计分析。

从总体分析，教材工具类科研教学用书的拖欠情况最为严重，

在建立模型之后，我们对题目第二部分中所给定的高校借阅制度做了评价并提出了改进方案。我们对改进后的借阅方案做了数值模拟，可以看出，拖欠书数量减少 43%，改善了原方案。

参赛队号 1259

所选题目 C

参赛密码 \_\_\_\_\_  
(由组委会填写)

# 第三届数学中国数学建模网络挑战赛

地址：内蒙古数学会  
电话：0471-4343756

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2010@tzmcm.cn

## 英文摘要

### Summary

Library has expanded our outlook in a large degree. However, there always exist some people who fail to return the books in time though they can be punished for doing that. This phenomenon has caused great trouble to the books' circulation.

In this article, by analysing the given data we would like to discuss the possible reasons for the returning delay and try to find out which kind of books is easy to be kept overtime. With the help of SPSS, we complete the work by statistics analysing of the given data.

We criticized the current book lending system according to our new model.

At last, we propose a solution to solve this problem and numerical results reveal the method is practical and available.

## 一、问题重述

高校图书馆存在着许多不良的现象, 借阅超期现象很常见。对于借阅图书超期, 原因众多。考虑到高校的特点, 大体上可以把超期原因分为三类: 1、无意错过还书期限; 2、由于教学科研需要, 长期使用; 3、由于罚款数目较少, 所以故意拖延。请通过给出的数据或自行查找相关数据, 建立合理的模型, 分析哪些书容易发生超期现象, 以及当发生超期时, 分析其可能的原因。以便图书馆采取相应的措施改进服务。

假设某高校的借书规则为: 借书时限为1 个月, 可以续借一次, 续借时限为15 天, 超期一天罚款0.1 元。试根据上述问题中的模型评价这种借书规则的优劣。请设计一种更加合理的借阅规则。

## 二、模型

### 一. 图书的分类

参考中图法, 我们依据不同图书的内容和时效性将其分为十类:

1. 文学类 (如《尼罗河上的惨案》)
2. 家居生活类 (如《微波烹调食谱》)
3. 哲学历史类 (如《社会发展史》)
4. 艺术体育类 (如《现代西方插花艺术设计沙龙》)
5. 经济政治类 (如《国际货币》)
6. 语言类 (如《大学英语语法过关训练》)
7. 工程建筑类 (如《工程机械》)
8. 计算机及电子信息类 (如《数字通信=原理、技术及其应用》)
9. 数理化类 (如《数学手册》)
10. 其它类 (如《本领测试》)

其中, 语言类、工程建筑类、计算机及电子信息类、数理化类、经济政治类图书属于教材范畴。在这五类教研工具书籍中, 计算机及电子类图书更新快、时效短, 而相比之下数理化类的更新慢、时效长, 较长时间内都可以使用。

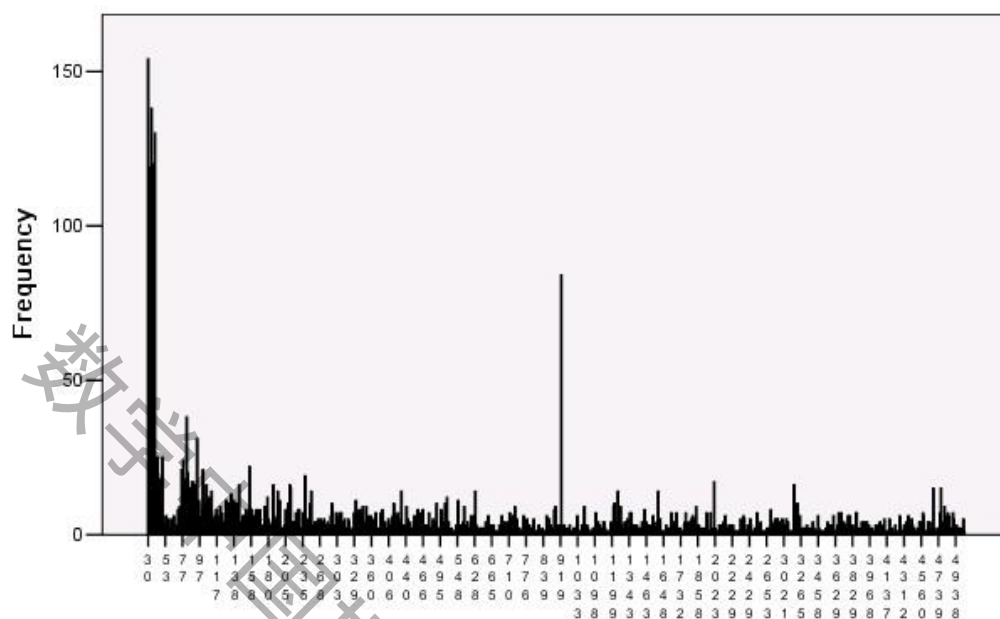
更细致的区分如地质类、宗教类图书, 因其在总体中所占比例很小 (低于 0.1%), 为方便研究, 本模型将其归为“其它类”图书。

### 二. 对书籍拖欠时间区间的划分

本模型对题目中所给的 4386 项超期 30 天以上图书做了统计分析。首先借助 SPSS 软件, 我们得到了欠书时间与欠书本数的对应关系图。如下:

#1259

欠书时间



欠书时间

经过 K—S 检验，我们发现数据基本符合正态分布[1]。

观察图像分析可知：

(1)横轴时间跨度比较大(30—4938天)，而且由于单独一天偶然性比较大，不适合进行分析检验。所以必须对上述时间跨度按一定标准进行分组，具体分组方法如下：

观察图像和数据，可得如下结论：

①拖欠时间在 30—60 天区间的书比较集中，共计 1084 本，占总数的 25%，故以此百分比为粗略标准进行划分；

②拖欠时间在 60—240 天区间的书较 30-60 天呈现递减趋势，共计 1099 本，占总数 25%；

③拖欠时间在 240—1000 天欠书数目较少且趋于平缓。虽然在第 919 天异常突起，但从分布看与两侧不连续，认为是由突发状况引起，属于偶然事件，不做过多考虑。本时间区间内书目共计 1054 本，占总数 25%；

④拖欠时间在 1000 天以上者，共计 1149，占总数 25%左右。

(2)于是将所有数据按照超期天数进行划分：

①对于第一组，即 30—60 天，由数据知，30-45 天区间内共计 1015 本，远远大于 45-60 天，从图像亦可观察到此特点。于是分为  $[30, 45)$ ， $[45, 60)$  两区间；

②对于第二组，即 60—240 天，数据量占总数 25%，考虑到时间跨度不大(180 天，相对总样本很小)，因此将其间隔 30 天进行划分如下： $[60, 90)$ ， $[90, 120)$ ， $[120, 150)$ ， $[150, 180)$ ， $[180, 210)$ ， $[210, 240)$ ，共计 5 区间。

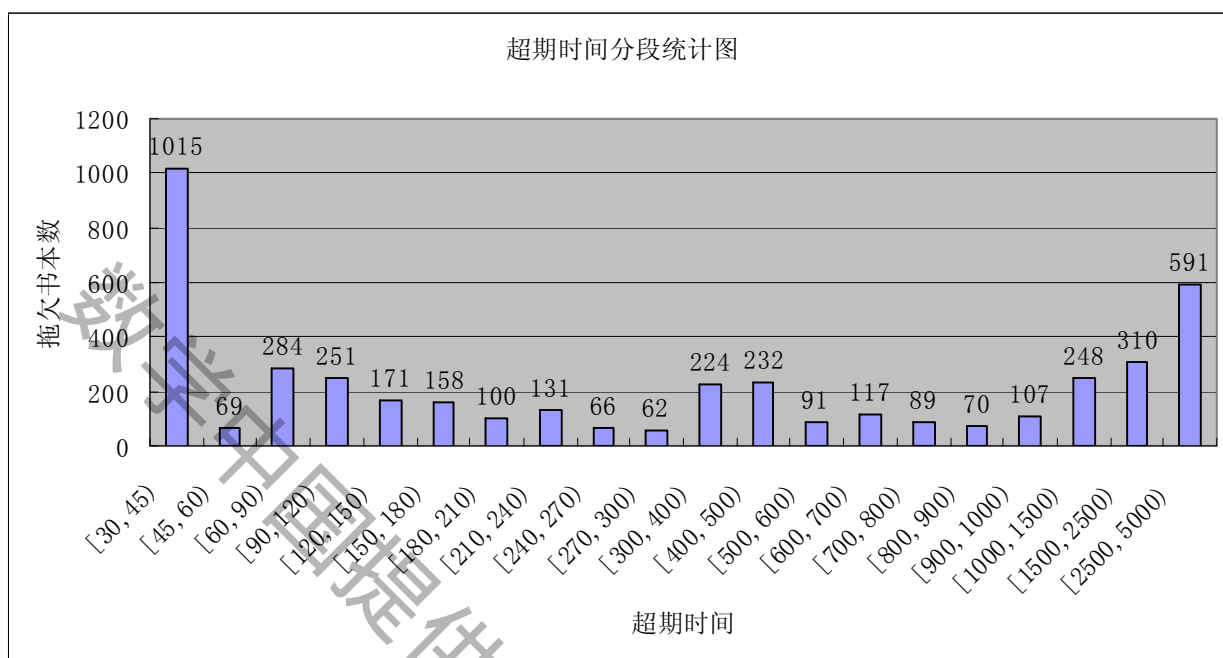
③对于第三组，即 240—1000 天，数据量仍占总数的 25%，但考虑到时间跨度相对较大(760 天)，因此将其间隔 100 天进行划分如下(为和前一部分比较，前两组仍按 30 间隔划分)： $[240, 270)$ ， $[270, 300)$ ， $[300, 400)$ ， $[400, 500)$ ， $[500, 600)$ ， $[600, 700)$ ， $[700, 800)$ ， $[800, 900)$ ， $[900, 1000)$ ，共计 9 区间

④对于第四组，即 1000 天以上，数据量占总数 25%左右，但时间跨度非常大，相

#1259

对每天数据量很少,因此将其间隔 500 天以上划分如下:  $[1000, 1500)$ ,  $[1500, 2500)$ ,  $[2500, 5000)$ , 共计 3 区间。

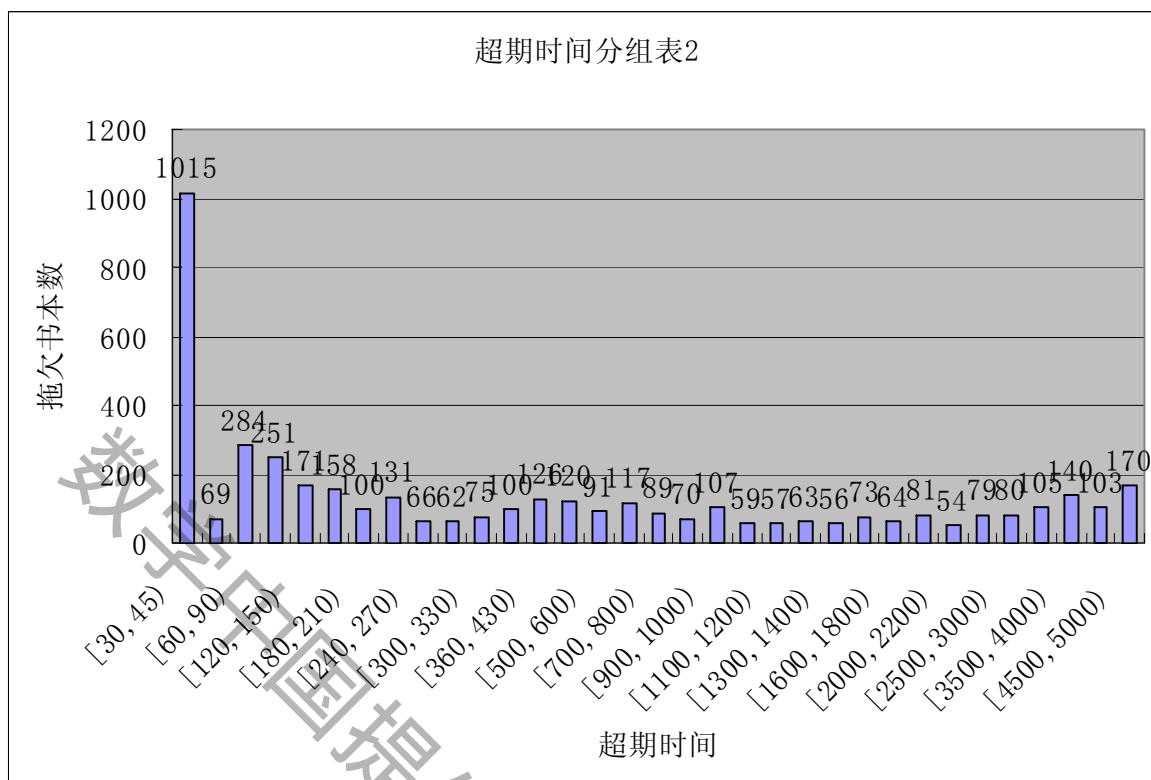
按照分组及各组拖欠书本数进行统计, 统计图如下:



观察分析上图, 发现:

- 1、 $[30, 45)$ ,  $[45, 60)$ 两区间有其特殊性, 需要单独进行研究。
- 2、 $[60, 90)$  区间开始, 数据基本满足一定规律递减, 但是在 $[300, 400)$ ,  $[400, 500)$ 两区间反常的大。分析原因, 发现是因为时间间距过大引起的, 因此应该对这两个区间进行细化, 重新划分小区间, 为了比较, 仍采用 30 天为时间间隔重新划分为 $[300, 330)$ ,  $[330, 360)$ ,  $[360, 390)$ ,  $[390, 420)$ ,  $[420, 450)$ ,  $[450, 480)$ ,  $[480, 500)$ 共 7 区间。重新统计数据得出自 $[360, 390)$ 区间开始, 数量跳跃性减少, 因此将后续 $[360, 390)$ ,  $[390, 420)$ ,  $[420, 450)$ ,  $[450, 480)$ ,  $[480, 500)$ 区间合并为 $[360, 430)$ ,  $[430, 500)$ 两区间。
- 3、 $[600, 700)$ 区间开始, 数据比较平稳, 但从 $[1000, 1500)$  区间开始, 数据突然增大。分析原因, 发现是因为时间间距过大引起的, 因此应该对后续区间进行细化, 具体操作步骤同上步, 重新统计数据, 作图如下:

#1259



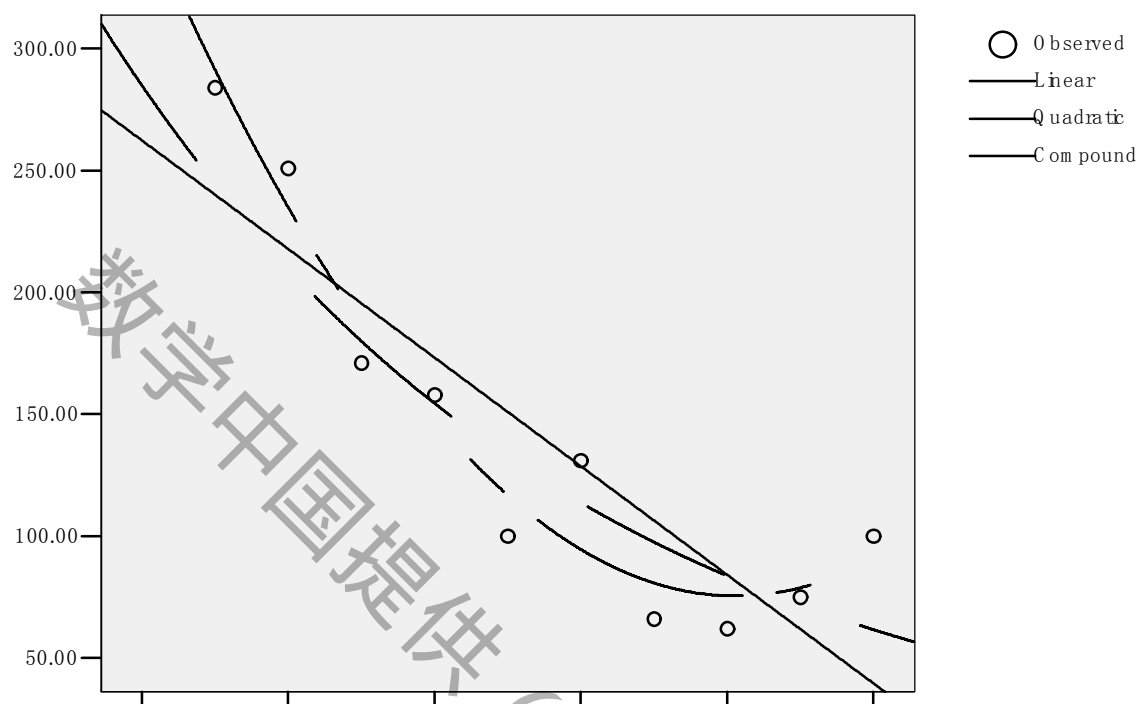
观察此图，发现分组情况比较良好。具体可分为五大组如下：

- 1、[30, 45) 区间，因为数据非常集中
- 2、[45, 60) 区间，数据锐减
- 3、[60, 360) 区间，数据满足一定规律。从此区间之后，虽然图示数据高度基本持平，但区间间隔增大至其 3 倍，所以数据减至前一组三分之一。
- 4、[360, 1400) 区间，数据较为平缓。从此区间之后，虽然图示数据高度基本持平，但区间间隔增大至其 2 倍，所以数据减至前一组二分之一。
- 5、[1400, 5000) 区间，数据较为平缓。虽然图示数据高度起伏，但区间间隔增大至 500 天，所以数据较前一组减小。

运用 SPSS 软件[2]拟合验证，结果如下：

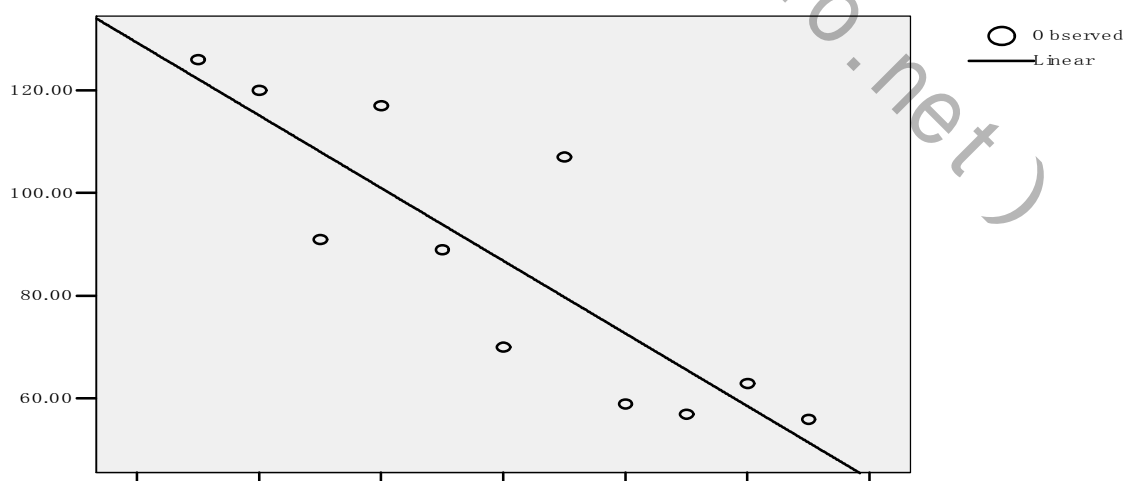
- (1)对于前两组，单区间数据各有其特点，单区间自成一组，无需验证
- (2)对第三组 10 个区间分别进行线性、二次方、混合拟合，三种拟合结果如下图示：



**VAR00009**

其中混合拟合近似效果最好。(图中用点划线表示混合拟合，圆圈表示原始抽样数据)。

(3)第四组 11 个区间基本满足线性递减关系，线性拟合结果如下：

**VAR00012**

(4)第五组 10 个区间各数据独立性较大，具有较大的随机性。其方差为 43.028.

本组超期图书都在1400天以上，认为属于历史累积数据，可能由于忘记、丢失等其

#1259

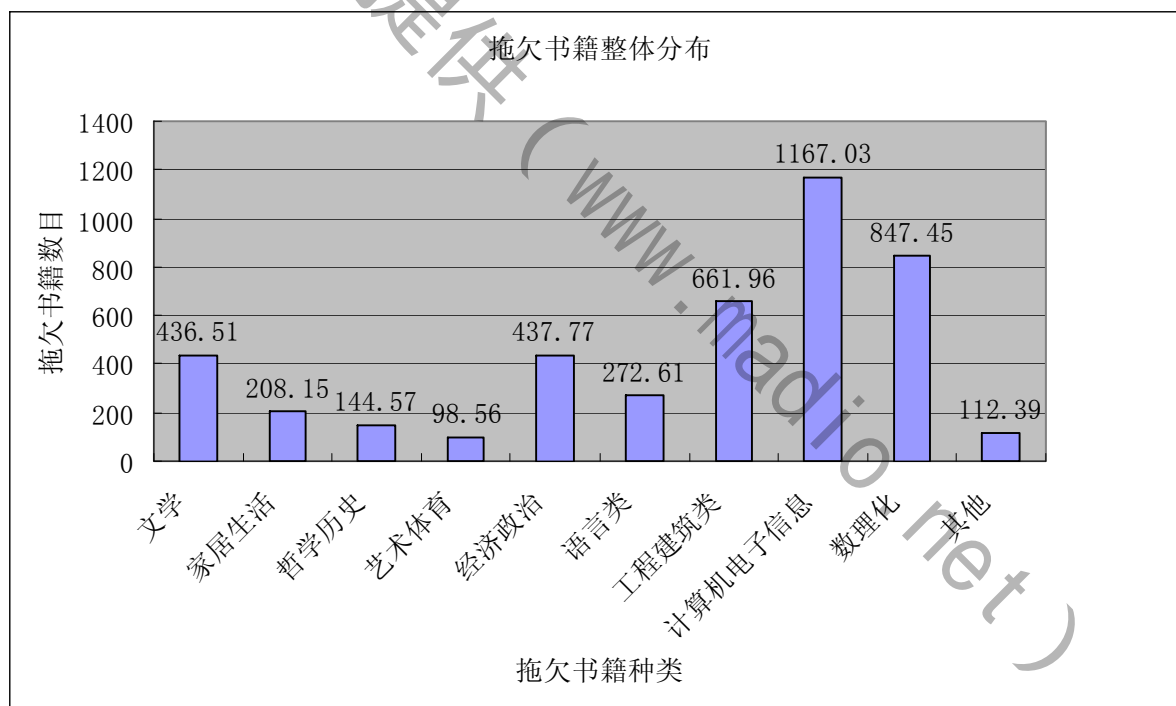
他原因，归还图书馆概率较低。因此本类数据虽然总体数量较大（占总数量25%），但数据相对分散，造成其数据庞大原因为历史累积，即每次统计由于其归还的概率很小，总会划归到拖欠书籍行列中，且拖欠时间越来越长，拖欠数量越来越多。所以本超期时间段内数据不计入统计分析。

### 三、易超期书种类研究

若想对整体时间段上书籍归还情况进行分析，由于数据庞大，首先要对数据进行随机抽样处理。在本文中，利用 EXCEL 中的随机抽样函数完成抽样功能。因为在按照时间划分的上述五个区间段内，书籍被拖欠的原因不同。因此，若直接对 4386 个数据进行抽样，则对不同种类的书影响不同。所以从所分五组中进行分层抽样，更能准确的反应被拖欠书籍的种类，而且方便分析书籍被拖欠的原因。

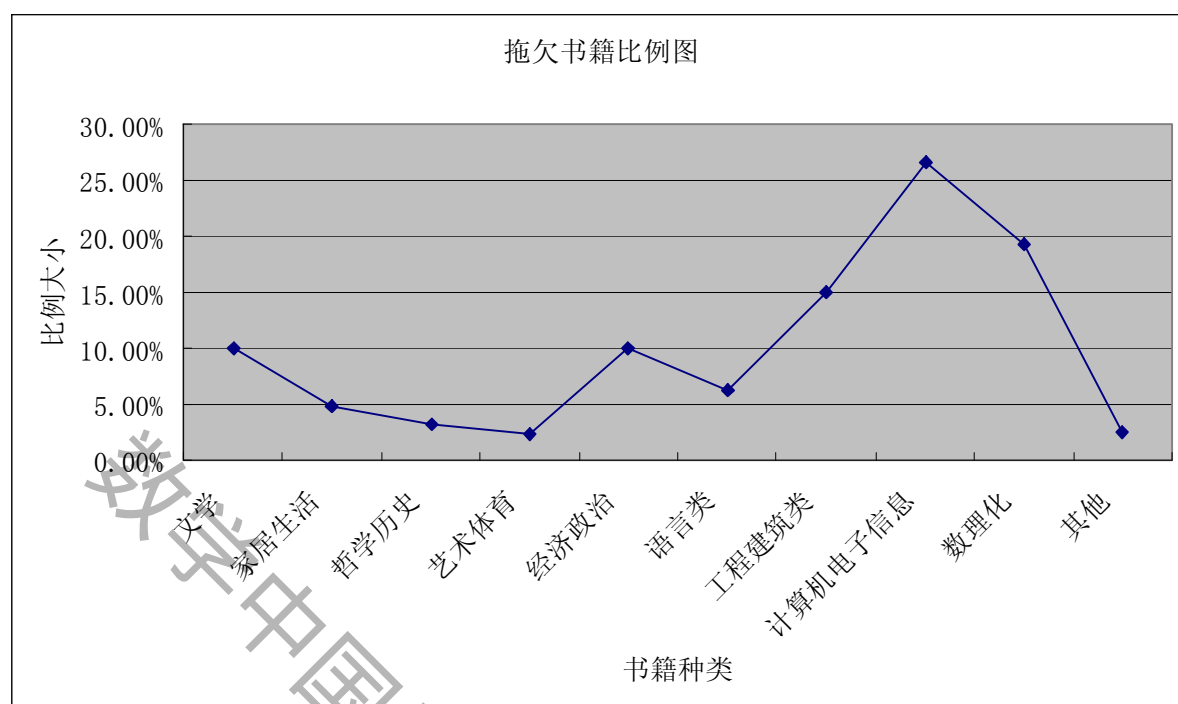
由上，本文先对划分的每个区间进行 20%比例随机抽样，分析不同种类的书籍的归还情况。不过，对于样本空间数据较少的情况，要酌情多抽取一些（25%）。之后比较各个区间不同种类的书的分布情况，在此基础上，测定整体的不同种类的书籍的归还比例。

下面为分析的整体不同种类书籍的比例图。



为方便比较比例大小，下面给出其比例图：

#1259



从上面拖欠书籍的整体情况来看，可以发现，不同种类的书所占比例差别很大，其中，计算机电子信息类书籍所占比重最多，其次为数理化类、工程建筑类。艺术体育类所占比重最少。比重较大说明在超期归还的书籍中数目较多，直观说明了这种书“容易”发生超期现象。观察上图，计算机电子信息类书籍容易发生超期现象，其次为数理化类、工程建筑类。上述都属于教学研究类书籍，而非教学类书籍诸如文学、家居生活之类一次性阅读较强的书籍发生拖欠的现象则较少。

这也是高校图书馆本身特性所决定。因为高校图书馆藏书教学类书籍远远多于其他类书籍，面向借阅对象也多为教师、学生，对教学工具类书需求量较大，由此造成的拖欠现象自然也较多。而非教材类书籍借阅次数相对较少，发生拖欠现象也较少。

此外，这也符合不同种类书籍自身使用特点。对于教学工具类书籍，使用周期相对较长，短时间内很难用完，需要长期使用，所以发生超期拖欠现象概率较大；而文学等非教材书籍，一次性阅读较强，使用周期短，读完即还，则发生拖欠现象概率较小。

综上所述，对于本题中高校图书馆，教学工具类书籍易发生超期现象，尤其以计算机电子等实效性较强的教学工具类书籍最易发生超期现象；而非教学类书籍发生超期现象较少，不易发生超期拖欠。

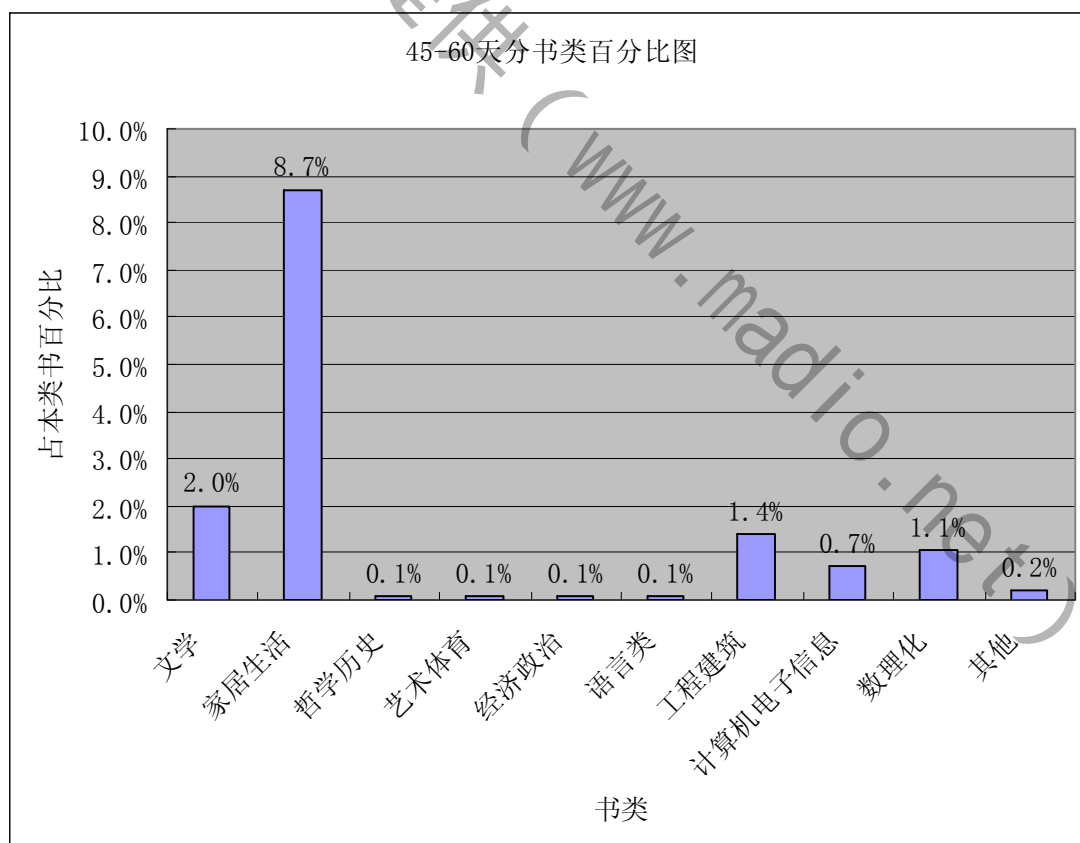
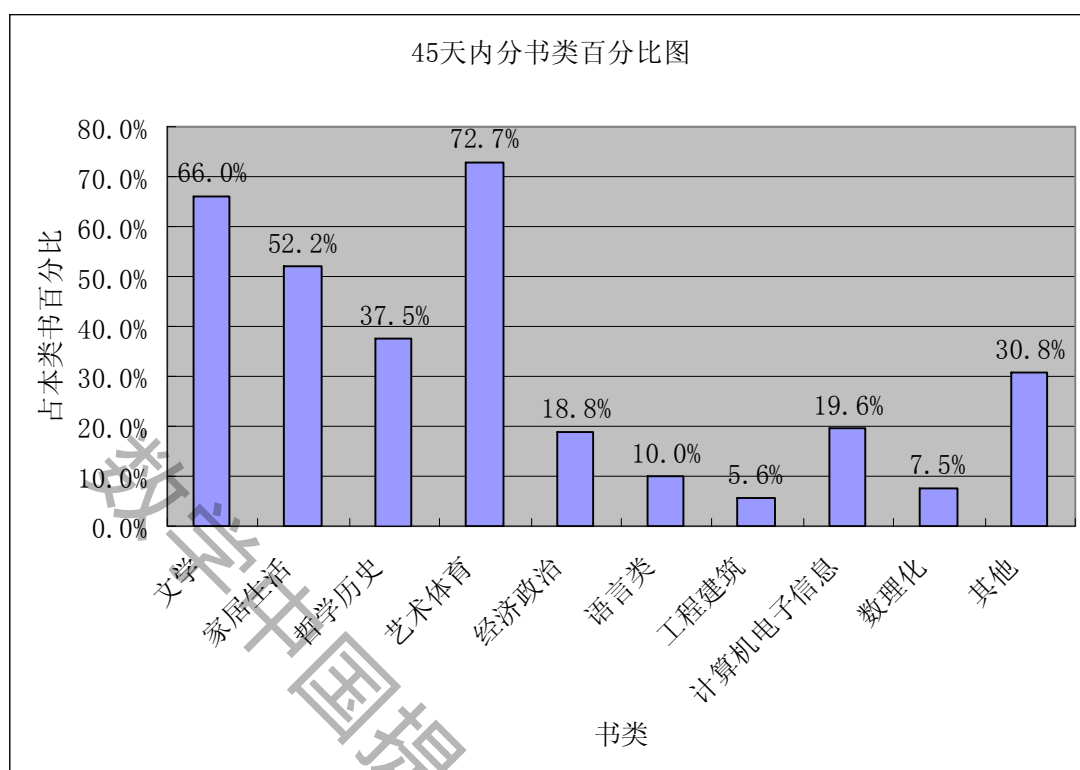
#### 四、超期原因分析

根据题中所给条件，超期原因分为三种：

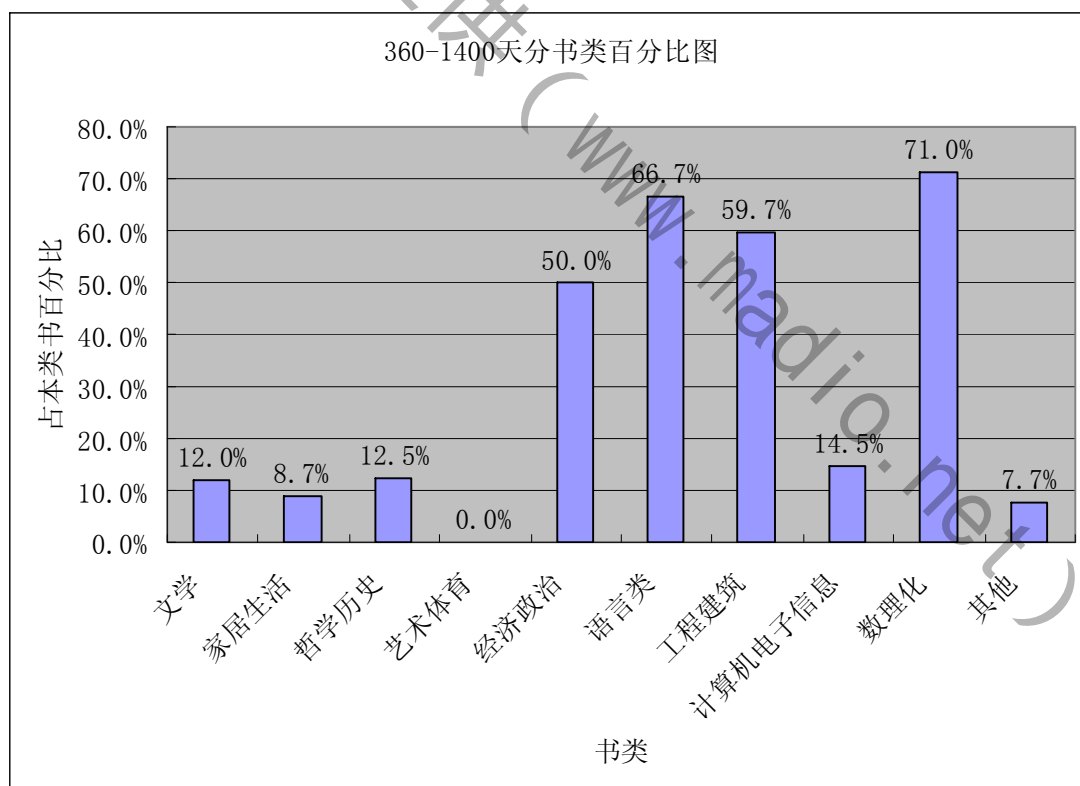
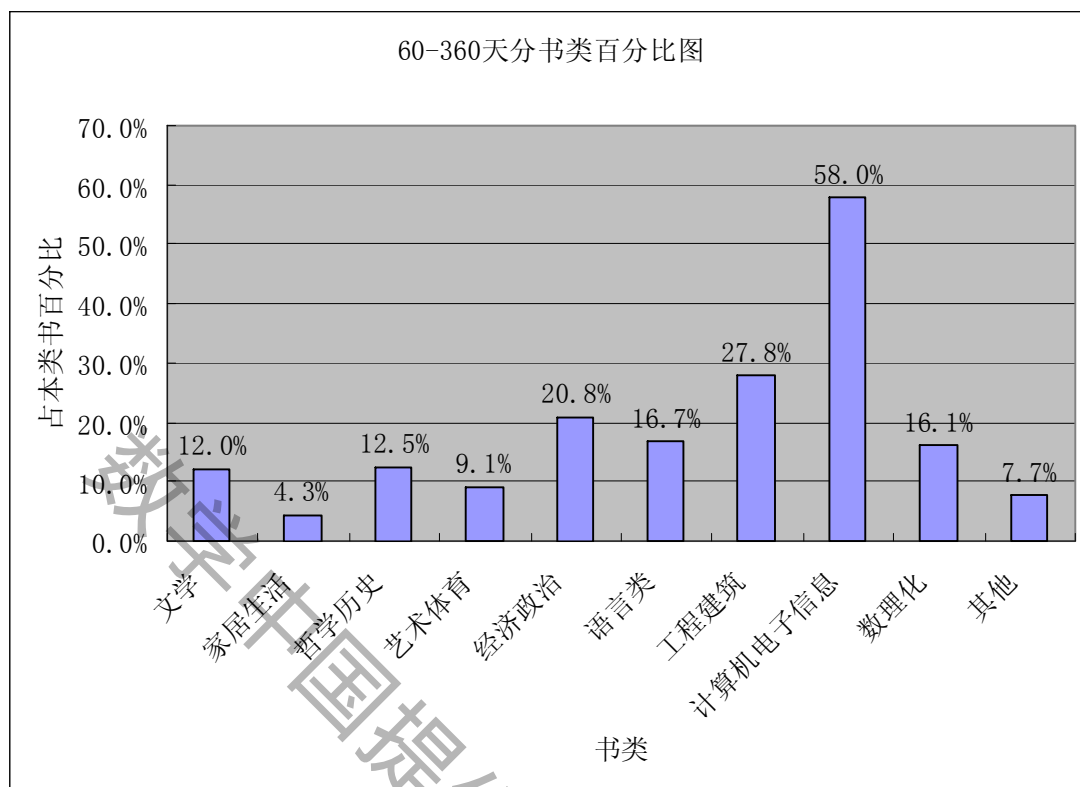
- 1、无意错过还书期限；
- 2、由于教学科研需要，长期使用；
- 3、由于罚款数目较少，所以故意拖延。

对于不同类型图书，不同时间段造成拖欠现象的原因也不同。研究数据取自高校图书馆，则借阅图书中教材类（包括上述5-9类图书）占大多数。分析原因是必须要考虑到不同种类图书本身数量差距很大的情况。因此，研究对象应该定义为每一时间段，每一类书所还数目占该类书总数目的百分比。对此进行抽样统计（每一时间段抽样率20%），所得结果如下图：

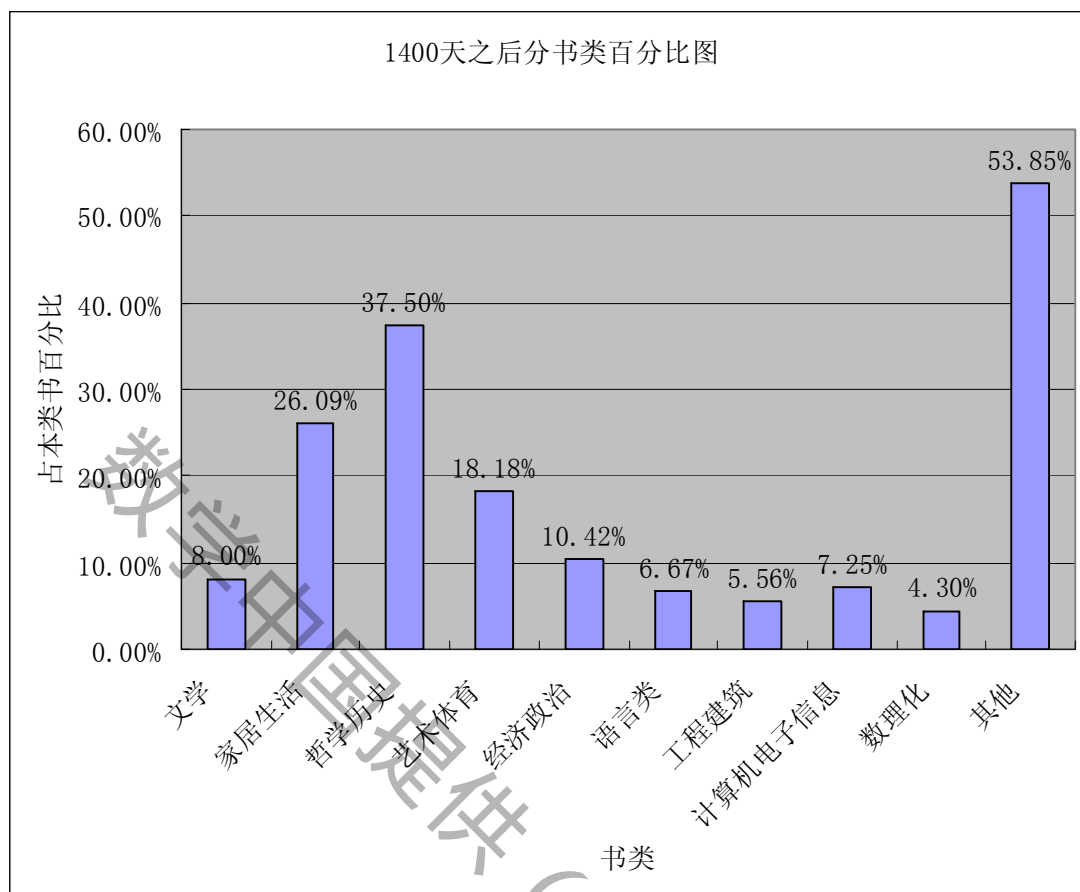
#1259



#1259



#1259



### 1、从时间段角度分析

对于超期在30—45天的图书，观察发现下一时间段，即超期45—60天的图书数目锐减，说明在本时间段内拖欠原因多为无意识错过还书期限。当期发现超期时，多能立即归还，不会进一步拖欠，造成拖欠时间延长。

对于超期在45—60天的图书，多为由于罚款数目较少，所以故意拖延的原因。因为如果无意识遗忘，一般不会超过60天，即可归还。而如果教学需要，也不会在90天之内归还。下一时间区间60—360天内，拖欠书籍增多就是因为这个原因。

对于超期在60—360天的图书，拖欠原因是教学科研需要，长期使用。因为拖欠如此长时间，欠款数目不小，基本不可能由于罚款数目较少，所以故意拖延的原因。也不可能无意识错过还书期限三个月以上。而且这一时间段内拖欠最多的书籍是时效性比较强的，专业知识更新比较快的计算机电子类书籍，这类书籍使用周期较非教材类书籍长，而较数理化类教材短，因此在60—360天时间段内，大部分拖欠都是此类书籍教学科研需要所致。

对于超期在360—1400天的图书，拖欠原因也是教学科研需要，长期使用。同样是因为拖欠如此长时间，欠款数目不小，基本不可能由于罚款数目较少，所以故意拖延的原因。也不可能无意识错过还书期限一年以上。和上一时间段相比，此时间段内教学科研所需多为使用周期较长的，专业更新较慢的数理化类、建筑工程类等教材辅导用书。

对于超期在1400天以上的图书，认为属于历史累积数据，可能由于忘记、丢失等其他原因，归还图书馆概率较低。因此本类数据虽然总体数量较大（占总数量25%），但数据相对分散，造成其数据庞大原因为历史累积，即每次统计由于其归还的概率很小，总会划归到拖欠书籍行列中，且拖欠时间越来越长，拖欠数量越来越多。所以本超期时间段内数据不计入统计分析。

## 2、从书类角度分析

对于文学类书籍，分析上面前四幅图（第五幅图为超期1400天以上的情况，已经阐述过不具有统计分析性）比较可知，该类书籍拖欠时间多为30—45天，这也是由于该类书籍的特点所决定的。文学类书籍多为一次性阅读，读后大部分都会归还，拖欠原因可能为无意识错过还书期限。

对于家居生活类书籍，分析上面前四幅图比较可知，该类书籍虽然拖欠时间多为30—45天，一部分拖欠时间较长，这是由于对于某些此类图书，使用周期较短，看后即可归还。也有一部分使用时间较长，因此此类图书拖欠原因是两方面的，可能为无意识错过还书期限，也可能是学习研究需要长期使用，研究结束后归还。

对于哲学历史类、艺术体育书籍，拖欠时间分布类似于家居生活类书籍，原因亦类同，不重复叙述。

对于政治经济类、语言类、建筑工程类、数理化类教材辅导工具书等，因为其使用周期一般相对很长，可以当做工具书使用，知识更新换代相对较慢，因此拖欠时间多在60天以上，拖欠原因是学习研究需要长期使用，研究结束后归还。

对于计算机电子类教材辅导工具书，虽然同样为教材类书籍，但由于其使用周期较短，专业知识更新较快，因此拖欠时间多在60—360天，拖欠原因是学习研究需要长期使用，研究结束后归还。

其他类书籍，高校借阅较少，暂不多作讨论。

## 五、所给借书规则评价

### 1. 借书时限为1个月，可以续借一次，时间为15天

所加模型中可知，在 $[30, 45)$ 天时间内，有25%的拖欠的书。如果续借成功，这些拖欠书籍将按时归还，不再属于拖欠。但存在问题是处于本区间内拖欠原因主要是无一错过还书时间，所以即使有续借制度，也很可能错过续借时间，使续借制度形同虚设。

### 2、超期一天罚款0.1 元

本条规则不是很合理，超期罚款应该随时间增长而逐渐增多，不应该不管超期多少天都罚款0.1元。尤其在 $[45, 60)$ 区间内，主要原因是罚款数目较少，所以故意拖延，因此在此区间可以适当提高罚款额。

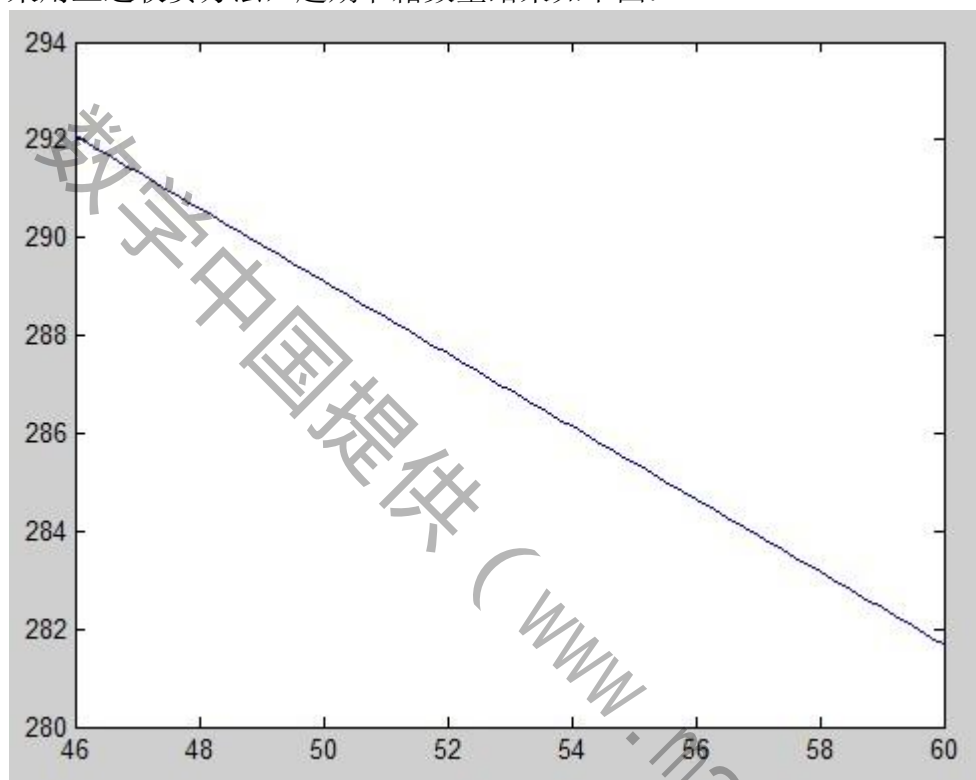
## 六、更合理的借阅规则

### 1、有超期未还书者，不允许继续借阅图书。（信誉限借）

分析所给数据，发现有超期未还，又继续借阅图书者，共有316人，786本书实行此规则，可避免此现象发生。这部分书占拖欠总数的18%，可以大大减少拖欠书数量。

## 2、节约超期超过45天，超期罚款额将随超期时间递增

对于超期时间在 $[45, 60)$ 区间内，主要原因是罚款数目较少，所以故意拖延，因此可适当提高罚款金额。由所给数据知，这部分拖欠罚款最高为 $60 \times 0.1 = 6$ 元，因此可在 $[45, 52)$ 天每天0.2元， $[52, 60)$ 天每天0.3元，如此此部分拖欠者将提前归还拖欠图书。而图书馆不是以营利为目的，所以罚款不宜定的过高，而且60天之后拖欠原因主要为科研教学需要长期使用，不会因为罚款价格上升而提前归还拖欠图书，所以可以适当考虑60天以后将罚款额降低。但要保证罚款总额随超期时间递增。采用上述收费方法，超期书籍数量结果如下图：



## 3. 图书到期前邮件提醒

由于大量还书者集中在超期45天之内还书，拖欠原因经分析主要无意错过书时间，因此图书到期前可以采取发邮件等手段提醒借阅者，续借或者还书。这在 $[30, 45)$ 之内的拖欠书籍将不会存在，大大减少拖欠数量。

## 4. 时效性强书籍限借制度

对于教材工具书类，尤其是需求量大馆藏不多的电子计算机类书籍，应该没人限借1本，还后再借。给数据统计，拖欠计算机类书籍超过一本的人有52人，共计286本该类图书，占总拖欠图书的7%。有此制度，这些书将不会被拖欠。而对于阅读周期较长，馆藏较丰，时效性不是很强的数理化等类教学科研用书，可以适当放宽，但也应限定个人借书数量。



## 七、模型缺陷

1. 本模型对图书的分类不够细致，由于所给数据图书序列号没有统一的标准，所以对图书分类主观性较强。
2. 数据只提供应还时间，没提供实际还书时间。模型构建建立在所给数据动态变化假设基础上，即数据表中图书不会继续出现大量拖欠现象。
3. 新借阅规则中图书馆提醒比较困难，并不一定每个借阅者都能看到，因此还是会有读者由于无意错过还书时间而造成拖欠。

## 八、参考文献

- [1] 冯敬海，沈玉波，宋立新，杨剑.《概率论与数理统计》，大连：大连理工大学出版社，2008。 119 页
- [2] 米红，张文璋.《实用现代统计分析与 SPSS 应用》，当代中国出版社，2000。 35 页