Operations Research and Fuzziology 运筹与模糊学, 2023, 13(3), 1778-1790
Published Online June 2023 in Hans. https://www.hanspub.org/journal/orf
https://doi.org/10.12677/orf.2023.133178

Hans汉斯

基于定序回归的医生预约火爆程度的影响因素分析

周婷婷

贵州大学数学与统计学院，贵州 贵阳

收稿日期：2023年3月30日；录用日期：2023年6月11日；发布日期：2023年6月19日

摘　要

近年来，随着信息技术和互联网的火速发展，传统医疗模式正在向“互联网 + 医疗健康”转型，在线问诊、网络预约等医疗服务也逐渐普及。为提高患者看病时医生预约成功率，本文采集500条医生在线诊断的相关数据，共选取了11个影响医生预约火爆程度的解释变量，即：专家团队、职称、所属科室、挂号支持、图文问诊费用、视话问诊费用、在线问诊评价数、医院就诊评价数、评分、问诊量、关注数，运用定序回归建立医生预约火爆程度模型，探究不同因素对医生预约火爆程度的影响大小。最后研究结果表明：患者一旦预约“图文问诊费用偏低、医院就诊评价数较高、关注数较多”的医生，就会发现该医生的预约量较大，预约火爆程度偏高，预约成功率偏低。

关键词

医生预约火爆程度，在线问诊，描述性分析，定序回归

An Analysis of Factors Influencing the Popularity of Doctor’s Appointment Based on Sequential Regression

Tingting Zhou

School of Mathematics and Statistics, Guizhou University, Guiyang Guizhou

Received: Mar. 30th, 2023; accepted: Jun. 11th, 2023; published: Jun. 19th, 2023

Abstract

In recent years, with the rapid development of information technology and the Internet, the tradi-

文章引用: 周婷婷. 基于定序回归的医生预约火爆程度的影响因素分析[J]. 运筹与模糊学, 2023, 13(3): 1778-1790.
DOI: 10.12677/orf.2023.133178

tional medical model is transforming to "Internet + medical and health", and medical services such as online consultation and online appointment are gradually popularized. In order to improve the success rate of doctor's appointment when patients see a doctor, 500 data related to doctor's online diagnosis were collected in this paper, and 11 explanatory variables affecting the popularity of doctor's appointment were selected, namely: expert team, professional title, department, registration support, graphic consultation cost, visual consultation cost, online consultation evaluation number, hospital treatment evaluation number, score, consultation volume, attention number. Sequential regression was used to establish the doctor's appointment popularity model to explore the influence of different factors on the doctor's appointment popularity. Finally, the results show that once patients make an appointment with a doctor with "low cost of graphic consultation, high evaluation number of hospital visits and high number of attention", they will find that the doctor has a large amount of appointment, a high degree of appointment popularity and a low success rate of appointment.

Keywords

Doctor's Appointment Popularity, Online Consultation, Descriptive Analysis, Sequential Regression

1. 引言

随着生活水平的不断提高，人们日益增长的医疗需求与我国医疗资源不足、配置不合理之间的矛盾更加突出，为扭转当前不合理的医疗资源配置格局，解决资源配置不均衡问题，国家出台了《国务院办公厅关于推进分级诊疗制度建设的指导意见》[1]，为全面推进分级诊疗制度建设提供政策框架和机制保障。在此背景下，互联网结合当前社会发展现状和网络普及程度，发挥自身突破时间、空间限制方面的优势，为分级诊疗制度的建立提供了一条开拓性的解决思路——在线诊断平台[2]。它将互联网技术、网络对话技术、评论机制等其它领域的成熟技术和方法引进健康医疗领域，患者们通过电脑手机等各种终端设备就可以在平台上以图文、语音、电话等多种方式进行疾病诊断。该平台不仅降低了患者就医的时间、空间和金钱成本，还合理配置了医疗资源，对构建分级诊疗服务体系，提升医生预约量具有重要研究意义。

医生的预约量大小体现了医生的预约火爆程度，而影响医生的预约火爆程度的因素有很多。例如：刘笑笑[3]和薛书峰[4]分别基于好大夫在线问诊数据发现，医生的在线努力和声誉对其预约量和咨询量有显著影响，服务价格也在其中发挥中介作用。李勇[5]等提出一种混合医生推荐模型，并实证验证了患者更倾向于选取相似患者推荐的医生。范晓妞等[6]结合医患问诊数据，发现医患双方的知识交换量、信任关系、患者收益、沟通成本等均会对医生的预约量产生影响。谭博仁[7]通过对医生图片和患者评论对医生问诊量的影响进行实证分析，发现医生图片和患者评论都对医生预约量有显著影响。因此站在患者角度，为探究什么样的医生他的预约量更高，本文提出在不考虑短期不可变因素的前提下，利用定序回归建立医生预约火爆程度模型，以探究医生职称，所属科室，评分及视话问诊费用等 11 个因素对患者预约医生火爆程度的影响大小。

2. 定序回归模型

传统的线性回归模型预测的因变量取值范围为任意实数，在实际应用中我们常常需要对非连续型数据建模，其中一类典型的数据即是定序数据[8]，也就是没有数值意义但是有顺序意义的数据。最常见的例子就是问卷调查给出的选项：非常满意、满意、一般、不满意、非常不满意，它们就是一类定序数据。定序变量介于连续变量和定类变量之间，是在测量层次上被分为相对次序的不同类别，但并不连续。以定序数据作为因变量的线性回归模型被称为定序回归模型，即：解释变量 $X=\left(1,X_{11},X_{12},\cdots,X_{nn}\right)'$，对应的的回归系数为 $\beta=\left(\beta_0,\beta_{11},\beta_{12},\cdots,\beta_{nn}\right)'$，其中 β_0 是截距项。然后再定义：$X'\beta=\beta_0+\beta_{11}X_{11}+\beta_{12}X_{12}+\cdots+\beta_{nn}X_{nn}$。与 0-1 逻辑回归一样，直接定义 $Y=X'\beta+\varepsilon$ 是不适用的，因为 $X'\beta+\varepsilon$ 为任意取值的数值，而 Y 为离散型的定性的指标，因变量 Y 为消费者对于商品的偏好程度，而在消费者对产品进行打分时，内心会有一个更加精确的产品的偏好，该偏好在进行偏好度调查时没有被直观显示，是潜在的一种喜好程度，而且这种喜好是连续的。消费者对于不同类型的相近产品喜好度是相近的，当消费者进行产品的选择购买时就会出现左右为难的情况。最后假设用 M 来表示消费者潜在的偏好，可以由分数显示出来。当消费者对某一产品更加喜欢时，累计得分就会越高。对应地，当 M 取值特别低时，消费者对某一产品的偏好度就会低。在数学上，可以假设

$$Y=\begin{cases}1, & M<a_1\\ 2, & a_1\le M<a_2\\ 3, & a_2\le M<a_3\\ 4, & a_3\le M\end{cases} \tag{1}$$

式中：a_1-a_3 为 4 个喜好程度划分的的阈值，需要根据具体数据进行分析。

3. 数据来源与指标设计

作为一个初步的探索性研究，本文的数据是基于狗熊会精品案例库模拟生成的 500 条医生在线诊断的相关数据，一行数据对应一位在线医生。数据中的因变量为预约火爆程度：是指在线医生的预约量大小，有 1，2，3，4 四个级别，其中：1 表示预约量小、2 表示预约量较小、3 表示预约量较大、4 表示预约量大。此外，还选取了 11 个解释变量，这些变量主要以医生的预约量大小为特征，具体见表 1。

Table 1. Data variable description table
表 1. 数据变量说明表

变量类型	变量名	详细说明	取值范围	备注
因变量	预约火爆程度	定序变量(4 种)	1 = 预约量小； 2 = 预约量较小； 3 = 预约量较大； 4 = 预约量大；	
自变量	专家团队	分类变量	是、否	是专家团队：86.2%
	职称	分类变量	主任医师、副主任医师	主任医师：70%
	所属科室	分类变量	包括心外科、呼吸内科、整形科等科室	
	在线挂号支持	分类变量	是、否	支持在线挂号：79.6%

Continued

图文问诊费用	数值型变量	10~600	单位：元
视话问诊费用	数值型变量	10~1200	单位：元
在线问诊评价数	数值型变量	1~2502	单位：条
医院就诊评价数	数值型变量	1~4361	单位：条
评分	数值型变量	8.2~10	评分大于等于 9：97.4%
问诊量	数值型变量	1~2647	
关注数	数值型变量	12~17,083	

4. 描述性分析

本文数据共涉及 11 个解释变量：专家团队、职称、所属科室、挂号支持、图文问诊费用、视话问诊费用、在线问诊评价数、医院就诊评价数、评分、问诊量、关注数。接下来，本文将进行描述分析检查数据质量，并初步探索医生预约火爆程度与各潜在影响因素之间的相关关系，为后续建模做准备。首先，本文对医生的预约火爆程度(1、2、3、4)绘制柱状图，见图 1。从图 1 中可以看出，分布是比较均匀的，其中预约量较小(2)的人数最多，预约量大(4)的人数最少。

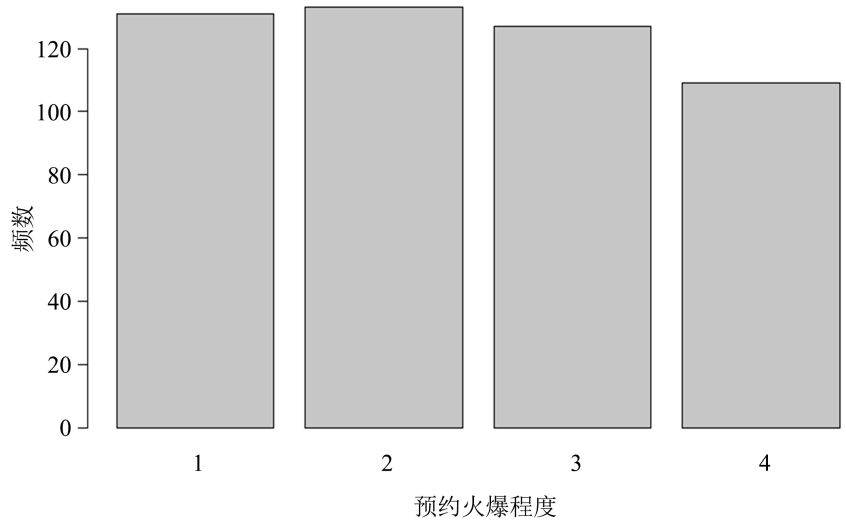

Figure 1. Barchartofappointmentpopularity
图 1. 预约火爆程度柱状图

接下来本文分别对 11 个解释变量做描述分析。专家团队变量有两个不同取值(是专家团队、不是专家团队)，其中是专家团队的医生有 431 人(占比 86.2%)，不是专家团队的医生有 69 人(占比 13.8%)。职称是一个 0-1 型解释变量，它有两个不同取值(主任医师或副主任医师)，其中是主任医师的医生有 350 人(占比 70%)，是副主任医师的医生有 150 人(占比 30%)。

所属科室是一个多分类解释变量，包括心外科、呼吸内科、整形科等科室，其柱状图分布见图 2。从图 2 中可以看出，科室类型太多导致柱状图太分散，因此查阅资料决定合并科室。把神经外科，心外科，普通外科等科室归为外科，把呼吸内科，内分泌科，消化内科等科室归为内科，把整形科，医学心

理科，放射治疗科等归为辅助科室，此外还有中医科和门诊科，合并后一共有 5 个大的科室。这样做的好处是让数据变得更简单，模型更加稳健。合并 5 个科室的柱状图分布见图 3。

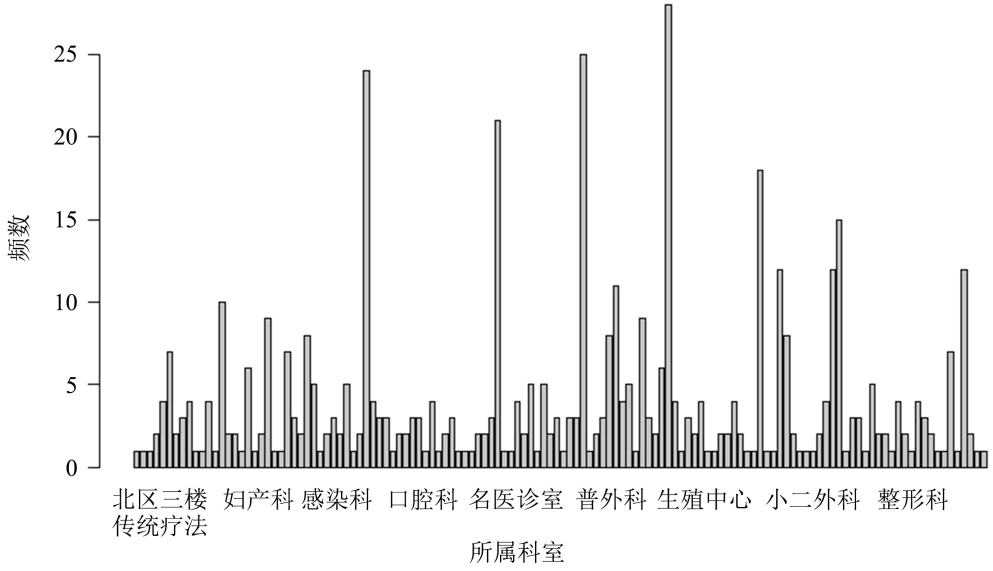

Figure 2. Department bar chart
图 2. 所属科室柱状图

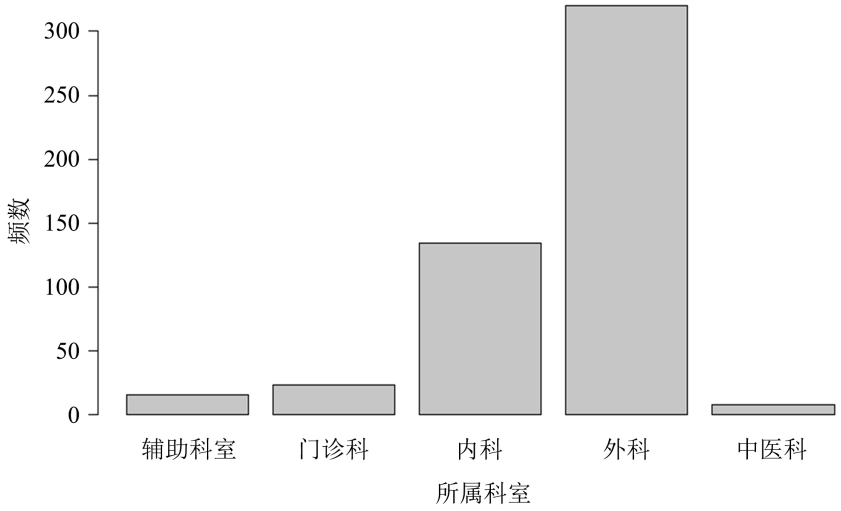

Figure 3. Bar chart of the department after the merger
图 3. 合并后所属科室柱状图

支持在线挂号是一个 0-1 型解释变量，有两个不同的取值(是或否)，其中支持在线挂号的有 398 人(占比 79.6%)，不支持在线挂号的有 102 人(占比 20.4%)。图文问诊费用的直方图见图 4 (左)，从图中可以看出，图文问诊费用的直方图不是非常连续并且严重右偏，故先对它做对数变换，再画直方图，见图 4 (右)，从图 4 (右)中我们发现分布情况好了很多。视话问诊费用是一个连续型的数值变量，在本文中，最高视话问诊费用是 1200 元，而最低视话问诊费用是 10 元。它的直方图不是非常连续，并且也是一个严重右偏的数据。因此，与图文问诊费用类似，对该变量进行对数变换后，再做直方图，具体分布见图 5。

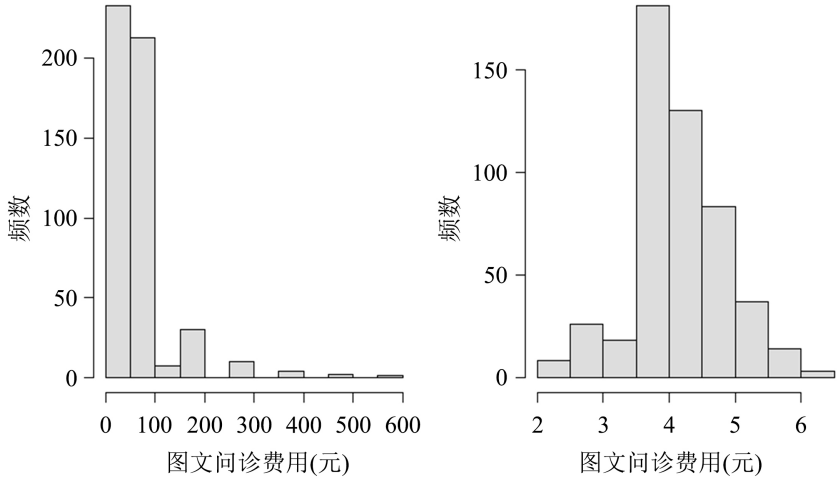

Figure 4. Histogram of consultation cost
图 4. 图文问诊费用直方图

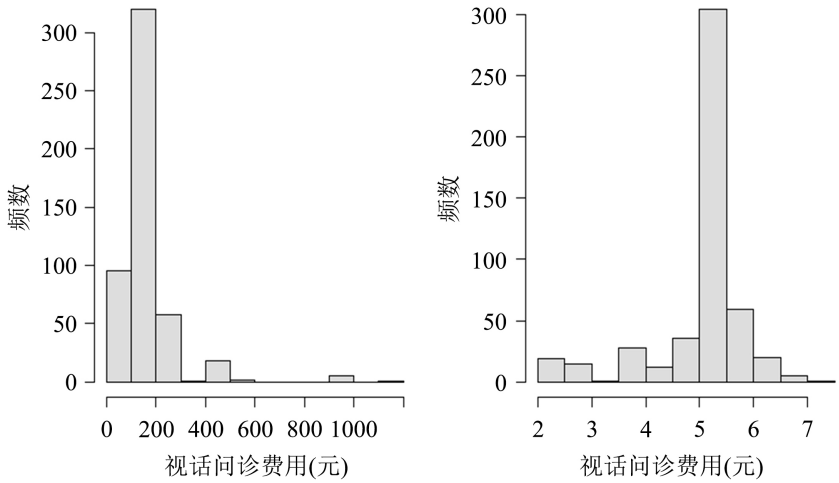

Figure 5. Histogram of the cost of visual consultation
图 5. 视话问诊费用直方图

在线问诊评价数是一个连续型变量，它的直方图见图 6 (左)。从图 6 (左)中我们发现其直方图不是非常连续且严重右偏，故决定对其先做对数变换，再画直方图，具体见图 6 (右)。在图 6 (右)中可以看到经过对数变换后的直方图的分布近似于正态分布。同理，与在线问诊评价数的类似，对医院就诊评价数也进行相应的对数变换，具体见图 7。

评分是一个连续型变量，评分越高，表示病人对该医生越满意，反之，则表示病人对该医生越不满意。

在本文中，评分大于等于 9 的占比为 97.4%，评分的直方图见图 8 (左)，可以看出它是一个左偏的数据，因此本文决定对该变量先进行对数变换，再做直方图。见图 8 (右)，我们发现直方图的分布均匀了很多。然后，对问诊量的各个取值计算频数，并画出其柱状图，见图 9 (左)。从图 9 (左)中我们可以看出数据严重右偏，因此决定对该变量进行对数变换后，再做直方图，见图 9 (右)。从图 9 (右)中可见，对数变换后的直方图分布情况好了很多，近似于正态分布。对关注数做同样的分析，结果见图 10。从图 10 (右)中可以看出，变量经过对数变换后，直方图的分布均匀了很多。

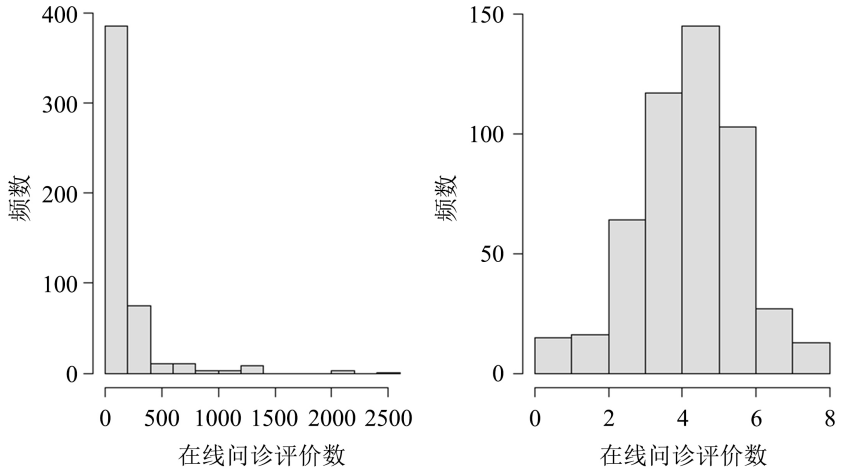

Figure 6. Histogram of evaluation number of online consultation
图 6. 在线问诊评价数直方图

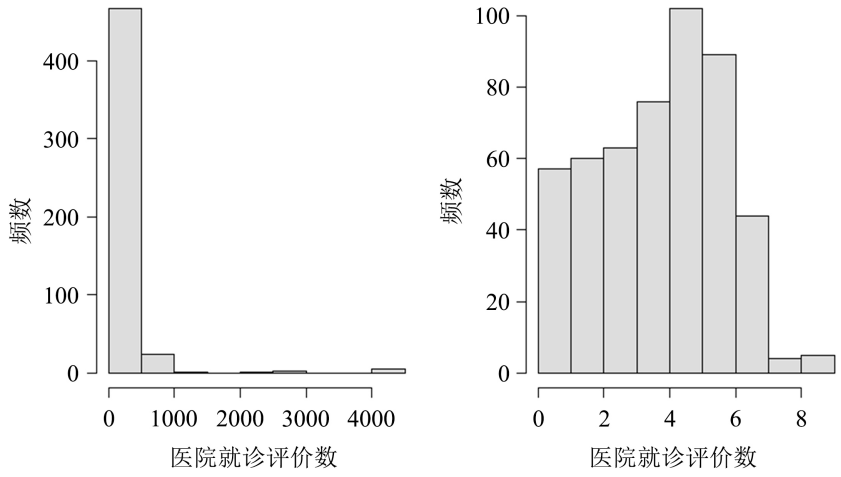

Figure 7. Histogram of evaluation number of hospital visits
图 7. 医院就诊评价数直方图

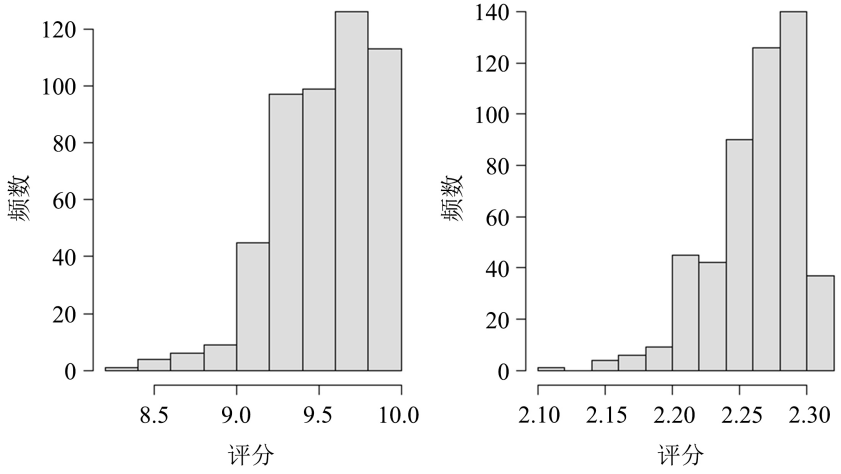

Figure 8. Score histogram
图 8. 评分直方图

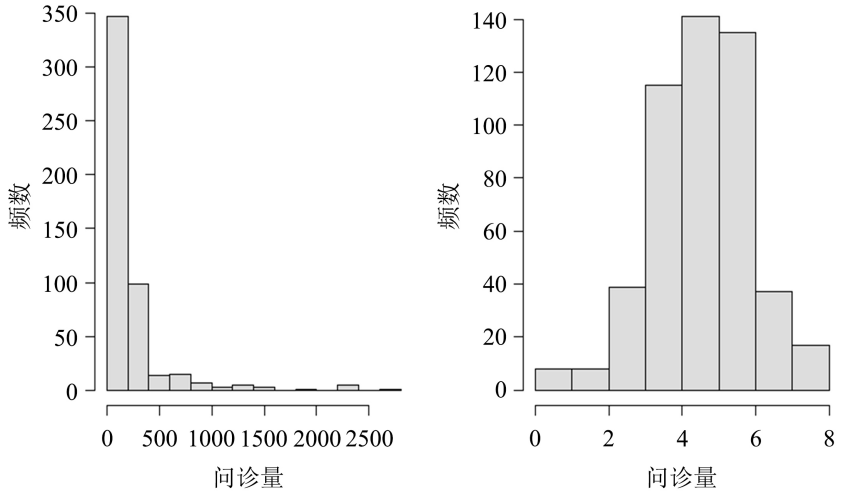

Figure 9. Histogram of consultation volume
图 9. 问诊量直方图

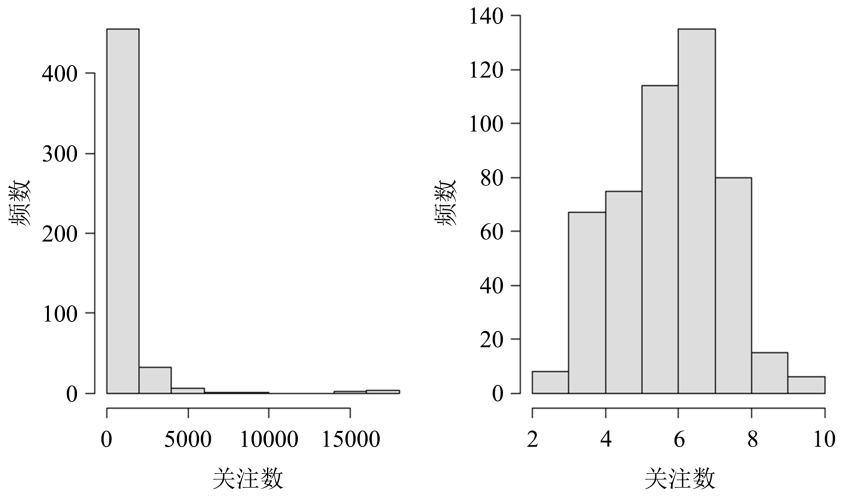

Figure 10. Histogram of concerns
图 10. 关注数直方图

接下来，对因变量(预约火爆程度)和所有的解释变量做相应的描述性统计分析。首先，对专家团队、职称、支持在线挂号这 3 个分类变量与因变量(预约火爆程度)之间的相关关系做进一步描述统计分析。具体步骤为：首先根据预约火爆程度的不同取值(1~4)把样本分成 4 组。再对每组的解释变量中的某一个取值(如是专家团队)计算占比，并通过折线图进行可视化展示，具体见图 11。

从图 11 中可以看出，预约火爆程度与是专家团队占比、主治医师占比及支持在线挂号的关系比较混乱，缺乏清晰规律。接下来对评分这个连续型变量离散化。由前面的分析可知，报告中评分大于等于 9 的医生为 387 人，占比为 97.4%，而评分小于 9 的医生只有 13 人，占总人数的 2.6%。因此，我们考虑根据评分这个解释变量的取值，把样本分成两组，即：一组是评分正常(评分大于等于 9)，另一组是评分不正常(评分小于 9)。对此，计算变量中评分正常的医生人数所占比率，再利用折线图进行可视化展示，见图 12。从图 12 中我们可以看出，随着预约量增大，评分正常占比越低，变化范围较大(大致从 100%下降到 92%)。

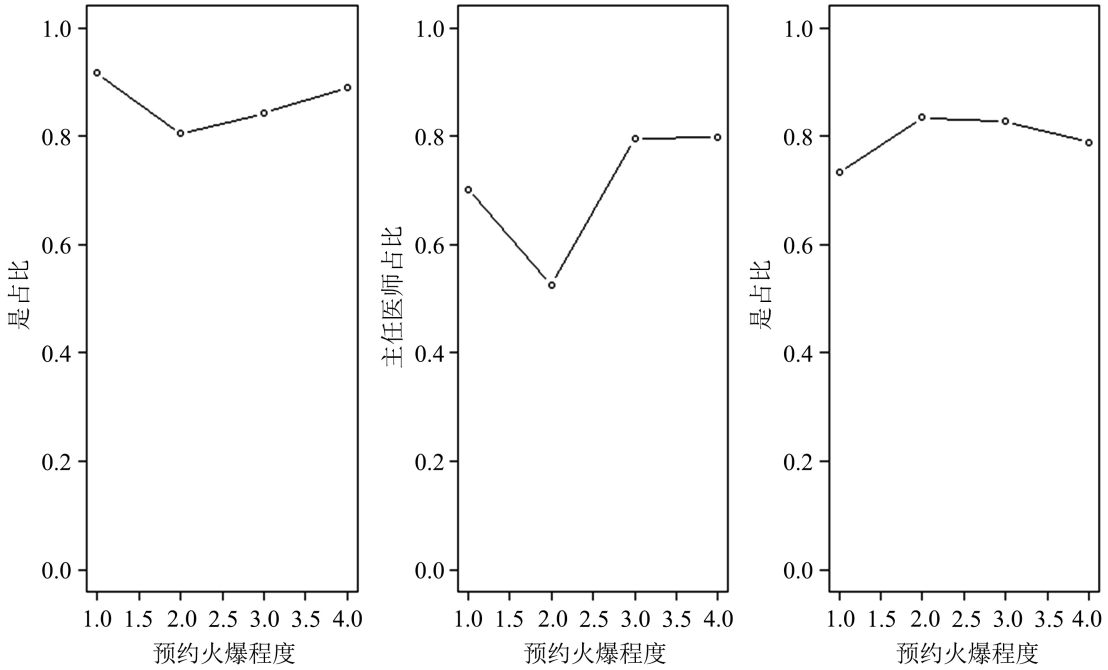

Figure 11. Line chart of the popularity of reservation and three categorical
图 11. 预约火爆程度与 3 个分类变量折线图

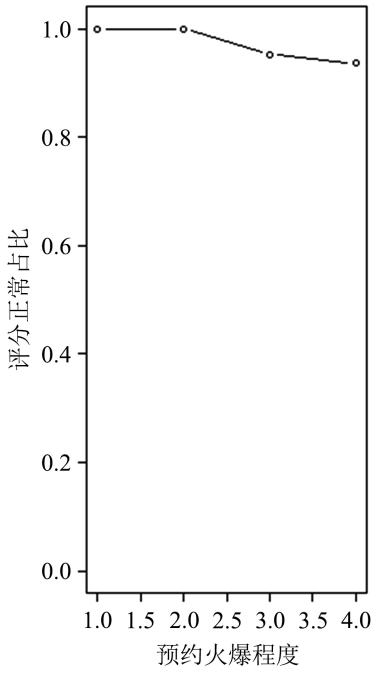

Figure 12. Line chart of popularity of variables reservation and score
图 12. 预约火爆程度与评分折线图

然后，本文分析预约火爆程度与 6 个数值型变量(图文问诊费用、视话问诊费用、在线问诊评价数、医院就诊评价数、问诊量、关注数)之间的关系，观察 6 个数值型变量的分布情况，并利用箱线图进行可视化展示，具体见图 13。然而，值得注意的是，此处的 6 个数值型变量都是经过对数变换后的。从图 13

中可以看出，对数图文问诊费用越高的组预约火爆程度似乎越高；就对数视话问诊费用而言，预约火爆程度 4 组间的费用似乎都差不多；对于对数医院就诊评价数，预约量较大的一组(3 组)所对应的对数医院就诊评价数似乎最多，而对数关注数也有类似的规律；关于对数问诊量，预约量较小的一组(2 组)所对应的对数问诊量似乎最低，对于其余的 3 组，对数问诊量似乎都相差不大；对于对数在线问诊评价数，预约火爆程度同对数在线问诊评价数之间缺乏清晰规律。

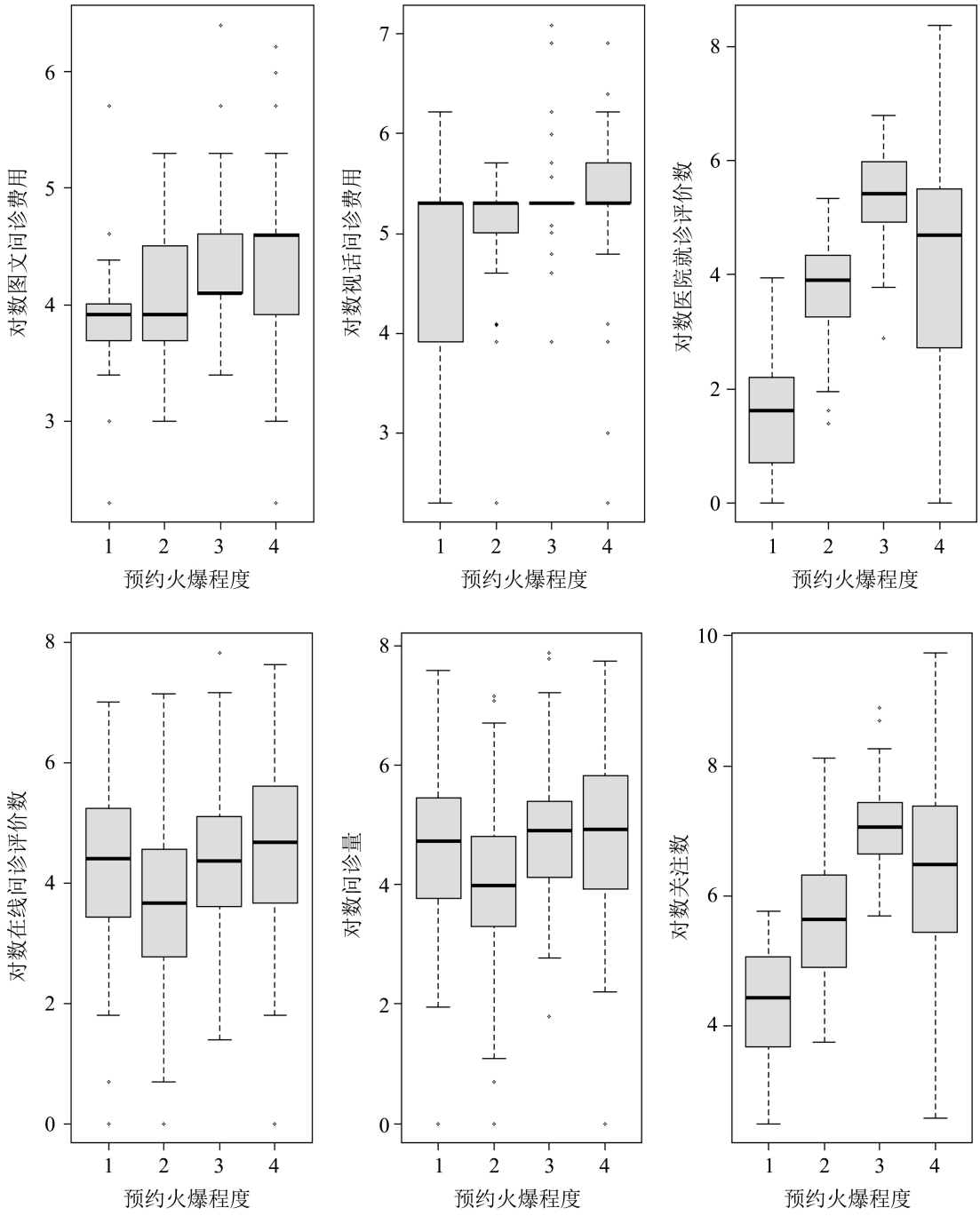

Figure 13. Boxplot between reservation popularity and six numerical variables
图 13. 预约火爆程度与 6 个数值型变量之间箱线图

5. 模型建立

为了探究影响医生预约量大小的显著因素，本文把医生预约火爆程度作为因变量建立相关的定序回归模型。首先对 11 个解释变量建立定序回归全模型，并对它做相应的变量选择，其中 AIC 选中了 8 个解释变量：专家团队、挂号支持、图文问诊费用(对数变换后)、视话问诊费用(对数变换后)、医院就诊评价数(对数变换后)、在线问诊评价数(对数变换后)、问诊量(对数变换后)以及关注数(对数变换后)，具体情况见表 2。

Table 2. Results of three sequential regression models
表 2. 三种定序回归模型结果

变量名称	全模型回归系数	P 值	AIC 回归系数	P 值	BIC 回归系数	P 值
截距项 1/2	3.381	0.000	3.499	0.000	2.474	0.000
截距项 2/3	4.477	0.000	4.592	0.000	3.544	0.000
截距项 3/4	5.356	0.000	5.461	0.000	4.398	0.000
专家团队——是	0.296	0.048	0.298	0.040		
职称——主任医师	0.079	0.522				
所属科室——外科	0.314	0.350				
所属科室——内科	0.448	0.190				
所属科室——门诊科	0.318	0.438				
所属科室——中医科	0.483	0.344				
所属科室——辅助科室	0.382	0.392				
挂号支持——是	0.184	0.185	0.218	0.100		
对数图文问诊费用	0.176	0.092	0.207	0.033	0.223	0.012
对数视话问诊费用	0.135	0.079	0.138	0.066		
对数医院就诊评价数	0.235	0.000	0.235	0.000	0.226	0.000
对数在线问诊评价数	0.277	0.030	0.265	0.040		
正常评分	−0.346	0.338				
对数问诊量	−0.269	0.008	−0.258	0.010		
对数关注数	0.277	0.001	0.278	0.001	0.274	0.000
模型全局检验	P 值小于 0.001		P 值小于 0.001		P 值小于 0.001	

由表 2 可知，整个模型是高度显著的。在 10%显著性水平下，AIC 选中的 8 个变量都显著，其中问诊量(对数变换后)参数估计显著为负，这说明问诊量多的医生如果被预约，预约火爆程度偏低；专家团队(是)的参数估计显著为正，说明和不是专家团队的医生相比，是专家团队的医生如果被预约，预约火爆程度偏高；挂号支持(是)的参数估计显著为正，说明是支持挂号的医生如果被预约，预约火爆程度偏高；图文问诊费用和视话问诊费用(对数变换后)参数估计显著为正，说明图文问诊费用和视话问诊费用越高的医

生如果被预约，预约火爆程度偏高；医院就诊评价数和在线问诊评价数(对数变换后)参数估计显著为正，说明评价条数越多的医生如果被预约，预约火爆程度偏高；关注数(对数变换后)参数估计显著为正，说明关注人数越多的医生如果被预约，预约火爆程度偏高；因为预约火爆程度由 1，2，3，4 组成，所以模型产生了 3 个不同的截距项，它们的估计分别是 3.499 (1|2)、4.592 (2|3)以及 5.461 (3|4)。相较于 AIC，BIC 又去掉了 5 个显著性水平相对较低的解释变量，它们分别是：专家团队、挂号支持、图文问诊费用(对数变换后)、在线问诊评价数(对数变换后)、问诊量(对数变换后)。此外，BIC 模型中也有 3 个截距项，其估计值分别是 2.474 (1|2)、3.544 (2|3)和 4.398 (3|4)。

6. 模型应用

在建立的预约火爆程度模型中，BIC 模型都比 AIC 模型简单，并且表现良好。因此，本小节将以 BIC 模型为例，探究其应用思路，比如帮助医院建立预约看病系统，预测医院里每个医生被预约看病的预约量大小，定义其预约火爆程度，以期能增强病人预约成功率。假设某天某医院的预约系统收到来自某个病人的预约，该病人预约的医生是在专家团队里的并且支持挂号，他的图文问诊费用为 50 元，医院就诊评价数 500 条，问诊量和关注数分别为 900 人和 1520 人，请问该病人预约的这个医生的预约火爆程度是偏低还是偏高？经分析，我们可该病人预约的这名医生是在专家团队里的并且支持挂号的，他的图文问诊费用低于预约系统里医生图文问诊费用的平均水平 78.8 元，而医院就诊评价数远远多于预约系统里医生的医院就诊评价数的平均评价数 183 条，同样地，这名医生的问诊量和关注数也是远远多于预约系统里医生问诊量的平均水平 213 人和关注数的平均水平 870。因此通过初步判断，这名医生的预约火爆程度似乎是偏高的。为了获得更精准的预判，我们将该名医生的相关数据放入所建立的预约火爆程度模型中得到：这名医生的预约火爆程度是 4 (预约量大)。这名医生的预约火爆程度高的原因如下：第一，他是专家团队里的医生，因此病人认为他的医术好，更信任他。第二，这名医生在线支持挂号，对病人来说这更加省时省力。第三，他的图文问诊费用为 50 元，低于预约系统里医生图文问诊费用的平均水平 78.8 元，相较于预约系统里的大部分医生，他更便宜一点。第四，这名医生的医院就诊评价数，问诊量和关注数也是远远多于预约系统里医生的平均评价数 183 条，问诊量的平均水平 213 人和关注数的平均水平 870，这说明他与预约系统里的大部分医生相比，更能得到病人的喜爱和信任。

7. 结论与展望

本文以医生“预约火爆程度”为因变量，利用 500 条医生在线诊断的相关数据，共采集了 11 个影响医生“预约火爆程度”的解释变量，即：专家团队(是否属于)、职称(主任医师或副主任医师)、所属科室、挂号支持(是否支持)、图文问诊费用(对数变换后)、视话问诊费用(对数变换后)、在线问诊评价数(对数变换后)、医院就诊评价数(对数变换后)、评分(对数变换后)、问诊量(对数变换后)、关注数(对数变换后)构建了预约火爆程度模型，以便探究不同因素对医生预约火爆不同程度的影响。最后通过预约火爆程度模型，发现病人一旦预约“图文问诊费用偏低、医院就诊评价数较高、关注数较多”的医生，就会发现该医生的预约量较大，约火爆程度偏高，预约成功率较低。

本文受数据限制，只考虑了 11 个特征指标。而在生活中的预约经验启示我们还可以有很多相关的解释变量，包括医生的岁数，医生所属医院是否是三甲医院，医院所在城市，是否有多个工作地点等。怎样在不违反法律并且在医院和医生同意的情况下获得更好的解释变量，是我们将来研究的重要方向。

参考文献

[1] 国务院办公厅关于推进分级诊疗制度建设的指导意见国办发[2015]70 号[J]. 中国制药信息, 2016(4): 23-26.

[2] 王若佳, 张璐, 王继民. 基于机器学习的在线问诊平台智能分诊研究[J]. 数据分析与知识发现, 2019, 3(9): 88-97.
[3] 刘笑笑. 在线医生信誉和医生努力对咨询量的影响研究[D]: [硕士学位论文]. 哈尔滨: 哈尔滨工业大学, 2014.
[4] 薛书峰. 互联网医疗的定价影响因素研究[D]: [硕士学位论文]. 南京: 南京大学, 2016.
[5] 李勇, 黄俊. 一种混合医生推荐算法的研究[J]. 信息通信, 2018(2): 67-70.
[6] 范晓妞, 艾时钟. 在线医疗社区参与双方行为对知识交换效果影响的实证研究[J]. 情报杂志, 2016, 35(7): 173-178.
[7] 谭博仁. 在线问诊平台中患者对医生选择意愿的影响因素研究[D]: [博士学位论文]. 北京: 北京邮电大学, 2019.
[8] 郭梦笛, 赵小山. 葡萄酒消费者偏好度调查研究[J]. 天津职业技术师范大学学报, 2020, 30(3): 51-56.