小组成员：

016年5月3日

游乐园客流疏导方案及酒店入住量预测

摘要

随着城市的发展与人们对生活质量的追求，越来越多的娱乐场所应运而生。其中，为保证每位游客的游园体验达到最佳，游乐园等具有大型器械且对人数限制要求苛刻的场所，做好关于客流量的预估与疏导方面的工作显得尤为重要。

本文针对Youth乐园的规划图与园内皇冠假日酒店2015年的全年预定数据，进行了下列问题的探讨：

首先针对第一个问题，在已预估出的高峰期每日客流总量的基础上，对于每一个游乐项目，将游客从到达该项目位置时到可以进行游戏的等待时间与放弃排队走到另一个项目在路上所用时间与排队等待的时间之和作比较。其中，每分钟到达任一个项目的游客数量满足泊松分布，并先将问题一般化，在没有疏导提示时游客到达该项目，即初始时刻时所有项目的排队人数相同，再进一步对模型优化，按照疏导提示进行人流量分流。

其次针对第二个问题，先将数据按照每天的预定房间数进行整理，选择考虑以下四个变量：季节、工作日/周末、法定假日、暑期。通过日期数据整理出各个日期对应的前面所述的四个变量的取值。对附件2中的数据做柱状图并对变量进行初步的定性分析，进而得到考虑的变量与因变量预定房间数之间的可能关系。然后将数据进行回归分析，对拟合度较低的模型进行逐步回归，对于无规律的时间段进行随机数预测。

但问题一的现有方案仍然建立在每个项目的吸引度相同和人流量平稳且随意游走的基础上。为了解决这一弊端，本文根据不同项目到达人数所满足的泊松分布提供了调查方案和模型改进。

关键词：提醒疏导 时间成本 人流量 回归分析 平均随机数

一、问题重述

随着社会的发展以及人们由满足基本生活到享受生活的态度的转变，选择到娱乐场所游玩的人越来越多。特别是具有大型设施的游乐场，每个设备对人数都有限定，故不能同时满足所有人的需求。在如此高的客流量的基础上尽可能满足游客的游园体验即游客可以在较少的时间里玩较多的游戏项目，引导提示显得尤为重要。

Youth游乐园即将盛大开园，作为本市建有最多过山车的游乐园，受到了青少年的热捧。预计届时园区将迎来每天1万的大客流。为此提出以下问题：

问题一：在保障每位游客体验游乐设施的前提下，建立对每个游乐项目的等候游客进行游览提醒和疏导的模型，以达到游园体验最优。



表一



图一

问题二：皇冠假日酒店是游乐园内的酒店，目前已开业，为有需要的游客提供住宿便利。请根据该酒店历史预订数据信息,综合考虑影响房间预定量的主要因素建立数学模型。并根据酒店2015年全年预定数据, 预测2016年1月至3月每天预定房间数。

二、问题分析

2.1问题一

问题一中，在客流量较大的情况下，由于每个游乐项目对游客进行疏导的目的是尽量让每位游客花费在排队等候的时间上最少，即在每一个项目中，游客选择等待的时间与步行去其他项目并排队等待的时间之和的大小关系，本文以此为目标建立模型，并考虑到达该项目时已有排队人数以及步行到另一个项目的时间里新增的排队人数，最终得到在不同的排队人数情况下可以选择的游览方向。由于每个项目的分析方式相同，故在建立模型时可以选取一个地点进行分析。

2.2问题二

问题二中，首先可以定性分析房间预定量与四个因素的关系，得出一个定性结论，之后借助数学软件对数据进行回归分析，得出最后的模型。在预测房间预定量时，观察到一二月数据之间的相似性，三月数据的特殊性，所以单独将一二月数据进行多元回归得出回归模型，再进行预测。观察到三月数据具有分两层的性质，所以选取合适的模型来预测三月数据。

三、模型假设

1. 假设每一个游乐项目对游客的吸引度相同
2. 假设一天中游乐园的游客络绎不绝，符合泊松分布
3. 假设游客在步行中没有停留与折返
4. 假设游乐园每个项目的器械连续不间断的工作
5. 假设游乐园每日营业时间为8小时
6. 假设游乐园中的道路足够宽，即路上不存在拥堵情况。

四、符号说明

|  |  |  |
| --- | --- | --- |
| di→j |  | 从i到j的最短距离（i,j=A,B,C,D,E,F,G,H,I,J，单位:m） |
| v |  | 一般情况下人散步时的速度，设为50(单位:m/s) |
| Ni |  | 游乐项目的每场容纳游客数(i=A,B,C,D,E,F,G,H,I,J，单位:个) |
| n |  | 游客到达某项目时初始的排队人数(单位:个) |
| △nij |  | 游客从i项目走到j项目的时间里j项目新增的排队人数(i=A,B,C,D,E,F,G,H,I,J单位:个) |
| λ |  | 单位时间内到达各项目的总人数(单位:个) |
| λ’ |  | 每个项目单位时间到达的人数(单位:个) |
| ti |  | 选择在原项目等待时所需要的时间(i=A,B,C,D,E,F,G,H,I,J,单位:min) |
| tij |  | 选择离开去其他项目到进行其他项目的时间(i,j=A,B,C,D,E,F,G,H,I,J单位:min) |
| Ti |  | 游乐项目的每场持续时间(i=A,B,C,D,E,F,G,H,I,J，单位:min) |

表二 符号表

五、数据处理

1. 对时间单位统一到分钟，便于进行时间比较。
2. 对附件2中的数据进行补充，如：2015/1/3这一天没有预定记录，则将这一天的房间预定数记为0。

六、模型的建立与求解

6.1问题一

6.1.1模型的建立

根据题目中给出的预计每日人流量，且客流到达数量符合泊松分布P(λ)，其中λ为已出现排队的情况下单位分钟内到达游乐园的人数。根据泊松分布的期望E(P)=λ，表示游客在单位分钟内到达的数量，又因为当日的预计人流量共有1万人，且在排队情况下每个项目的承载量达到满员，且每场间断时间不计，即任一时刻每个项目都能达到满员，故只需考虑除去所有项目的总人数时的游客到达量。依据假设中的营业时间可以得到平均每分钟到达游乐园的游客数量，即λ：



又由假设1，每个项目对游客的吸引度相同，在任意时刻每个项目对于宏观的不同游览程度的人来说地位相同，所以对每一个项目的每分钟游客到达量为



不失一般性，选择A项目进行分析，取单独的一位游客作为时间比较对象，设游客在到达A时为零时刻（即此时A游戏项目刚好开始），则如果选择在A项目等待所需要的时间为

(其中，n未知，[ ]为取整命令)

如果该游客放弃排队A项目去其他的项目，假设去B项目，则从A到B的时间为



在该时间段内同时有其他人陆续到达B项目，所以B项目新增的人数为

且在这个过程中项目B一直不停的运作，则在游客到达B项目时，项目B正在进行的这一轮还需要时间才能结束，与新增人数和原排队人数构成的新的排队人数之和，得到:



一般情况为：

设一游客到达i项目，则该游客选择在i项目等待所需要的时间为：



若游客选择去j项目，则在路上所花的时间内j项目新增的人数为



选择离开i项目去j项目路上所需要的时间与在j项目排队所需要的时间之和为：



为得出在i项目的引导路线，我们需要比较游客在i项目等待所花费的时间与去其他各个项目在路上和等待花费的总时间进行比较（i=A,B,C,D,E,F,G,H,I,J）。即改变n值来判断

的正负。下面我们将会对每种情况做出讨论和求解。

6.1.2模型的求解

根据图一所示项目之间距离，我们计算出了各个项目之间的最短距离，如下所示

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | A | B | C | D | E | F | G | H | I | J |
| A | 0 |  |  |  |  |  |  |  |  |  |
| B | 300 | 0 |  |  |  |  |  |  |  |  |
| C | 600 | 300 | 0 |  |  |  |  |  |  |  |
| D | 1050 | 750 | 450 | 0 |  |  |  |  |  |  |
| E | 350 | 350 | 500 | 950 | 0 |  |  |  |  |  |
| F | 1550 | 1250 | 950 | 500 | 1200 | 0 |  |  |  |  |
| G | 900 | 900 | 1050 | 1150 | 550 | 650 | 0 |  |  |  |
| H | 1050 | 1150 | 1400 | 1550 | 900 | 1050 | 400 | 0 |  |  |
| I | 600 | 800 | 950 | 1400 | 450 | 1500 | 850 | 450 | 0 |  |
| J | 250 | 550 | 850 | 1300 | 600 | 1800 | 1150 | 800 | 350 | 0 |

表三（单位：米）

先利用Excel求出（具体数据见附录），再利用MATLAB，首先求出当游客在i项目的等待时间与去其他项目所花费的总时间之差等于零处的n值，若无零点则适当改变n的取值使得时间之差的绝对值最小，从而得出当游客处于任意一个项目，通过排队人数的多少来确定继续等待还是去其他项目所在地。当然，游乐园负责人也可以设置LED显示屏来显示此时的排队人数以及推荐路线。

1. 当游客到达A项目时，我们得出 ******

推荐去项目B,C,D,E,G,H,I,J;

推荐去项目B,C,E,G,H,I,J;

推荐去项目B,C,E,G,H,I,J;

推荐去项目B,C,E ,I,J;

推荐去项目B,C,E ,J;

推荐去项目B,C,E,F,J;

1. 当游客到达B项目时，我们得出

推荐去项目A,C;

推荐去项目C;

推荐游客原地等待

1. 当游客到达C项目时，我们得出

无论n为何值，推荐游客原地等待

1. 当游客到达D项目时，我们得出

推荐游客原地等待

推荐去项目C或者原地等待;

推荐去项目C,F或者原地等待;

推荐去项目B,C,F或者原地等待

1. 当游客到达E项目时，我们得出

推荐游客原地等待

推荐去项目A或者原地等待;

推荐去项目A,B,C或者原地等待

f. 当游客到达F项目时，我们得出

无论n为何值，推荐游客原地等待

g. 当游客到达G项目时，我们得出

推荐游客原地等待

推荐去项目H或者原地等待

h. 当游客到达H项目时，我们得出

无论n为何值，推荐游客原地等待

1. 当游客到达I项目时，我们得出

推荐游客原地等待

推荐去项目J或者原地等待;

推荐去项目A,J或者原地等待;

推荐去项目A,H,J或者原地等待;

推荐去项目A,C,E,H,J或者原地等待;

推荐去项目A,B,C,E,H,J或者原地等待

1. 当游客到达J项目时，我们得出

推荐游客原地等待;

推荐去项目A,B,C或者原地等待

以上仅为理想情形下的结果，考虑实际情况，推荐游客去上述结果中距离游客所在地最近的前三个项目。

6.2问题二

6.2.1模型的建立

首先，根据题目，先要分析影响房间预定量的主要因素如季节，工作日/周末，法定假日，暑假，令其为x1，x2，x3，x4。在此之前可以先用Excel对数据进行处理，使得影响因素可以数字化如在节假日则将此变量x3取值为1，不在节假日就将该变量取值为2等。对于同一天房间预定量的整合，可以使用Excel中的数据透视表来计算。数据处理完毕之后运用SAS软件来分别画出房间预定量与这四个因素的柱状图，定性分析各个因素与房间预定量的关系，得出数据分布规律再给出合理模型。预测2016年一月到三月的数据时，通过数据观察可以发现一二月份的数据分布较均匀没有显著的差别与波动，三月份数据在三月后半月出现明显上升且一直保持在一个比较大的预定量的状态，所以将一二月份数据放在一起处理，三月份数据单独处理。运用SAS软件对一二月数据进行回归分析，然后预测2016年一二月数据。三月数据分为上半月和下半月，上半月数据范围在1到15之间波动，且数据并无显著规律性，下半月数据基本在15到35之间波动，数据具有随机性且并无明显特征，所以采用产生均匀随机数的方法来预测2016年三月数据。

6.2.2模型的求解

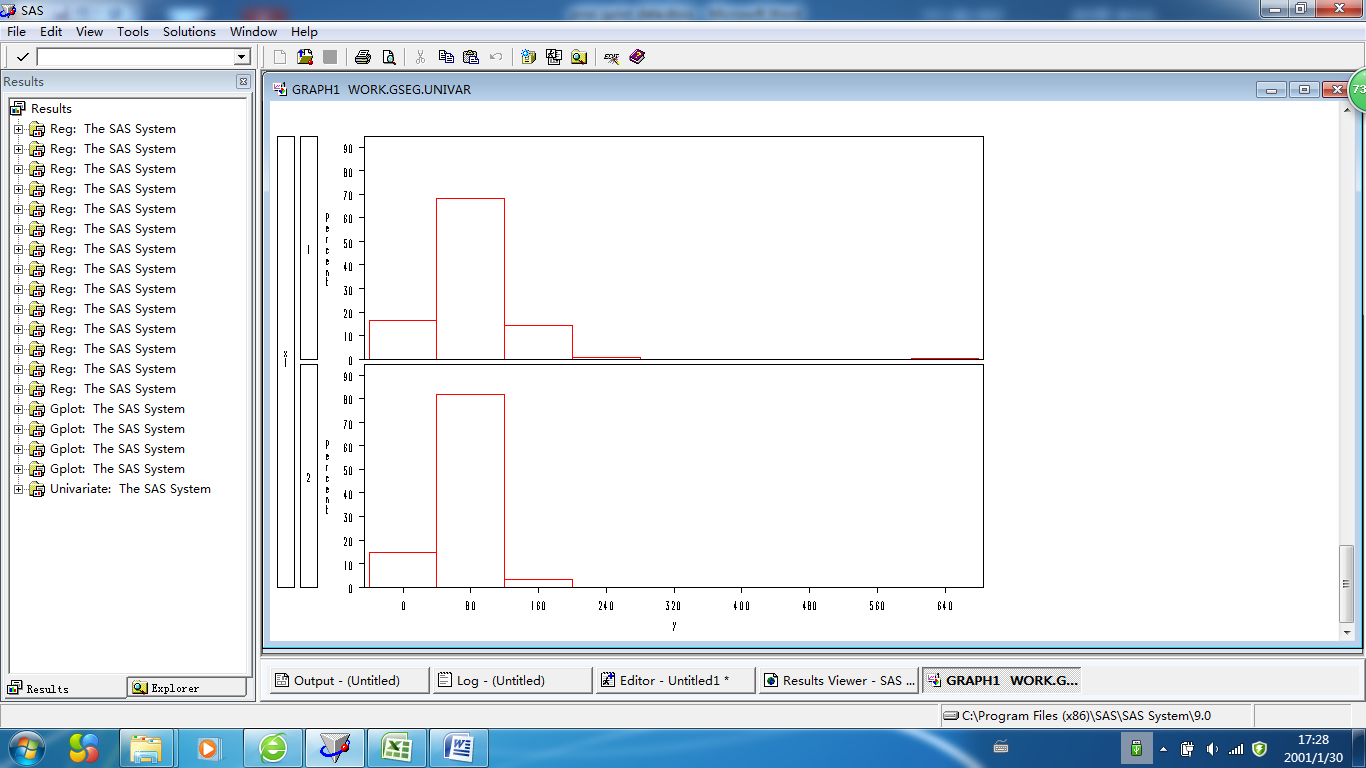
首先运用Excel对数据进行处理如下图表四所示（因数据较多所以只展示与部分数据）。

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| day | y | x1（工作日/周末） | x2（季节） | x3（节假日） | x4（寒暑假） |
| 2015/1/2 | 10 | 1 | 4 | 1 | 2 |
| 2015/1/4 | 1 | 2 | 4 | 2 | 2 |
| 2015/1/5 | 5 | 1 | 4 | 2 | 2 |
| 2015/1/8 | 2 | 1 | 4 | 2 | 2 |
| 2015/1/9 | 3 | 1 | 4 | 2 | 2 |
| 2015/1/11 | 2 | 2 | 4 | 2 | 2 |
| 2015/1/14 | 4 | 1 | 4 | 2 | 2 |
| 2015/1/16 | 2 | 1 | 4 | 2 | 2 |
| 2015/1/17 | 14 | 2 | 4 | 2 | 2 |
| 2015/1/20 | 3 | 1 | 4 | 2 | 2 |
| 2015/2/2 | 1 | 1 | 4 | 2 | 1 |
| 2015/2/4 | 1 | 1 | 4 | 2 | 1 |
| 2015/2/6 | 1 | 1 | 4 | 2 | 1 |
| 2015/2/9 | 2 | 1 | 4 | 2 | 1 |
| 2015/2/25 | 2 | 1 | 4 | 2 | 1 |
| 2015/2/26 | 3 | 1 | 4 | 2 | 1 |
| 2015/2/27 | 1 | 1 | 4 | 2 | 1 |
| 2015/3/1 | 3 | 2 | 1 | 2 | 1 |
| 2015/3/2 | 6 | 1 | 1 | 2 | 1 |
| 2015/3/3 | 5 | 1 | 1 | 2 | 1 |
| 2015/3/4 | 2 | 1 | 1 | 2 | 1 |
| 2015/3/5 | 2 | 1 | 1 | 2 | 1 |
| 2015/3/6 | 5 | 1 | 1 | 2 | 1 |
| 2015/3/7 | 9 | 2 | 1 | 2 | 1 |

表三

运用SAS得出柱状图如下（程序见附录）。

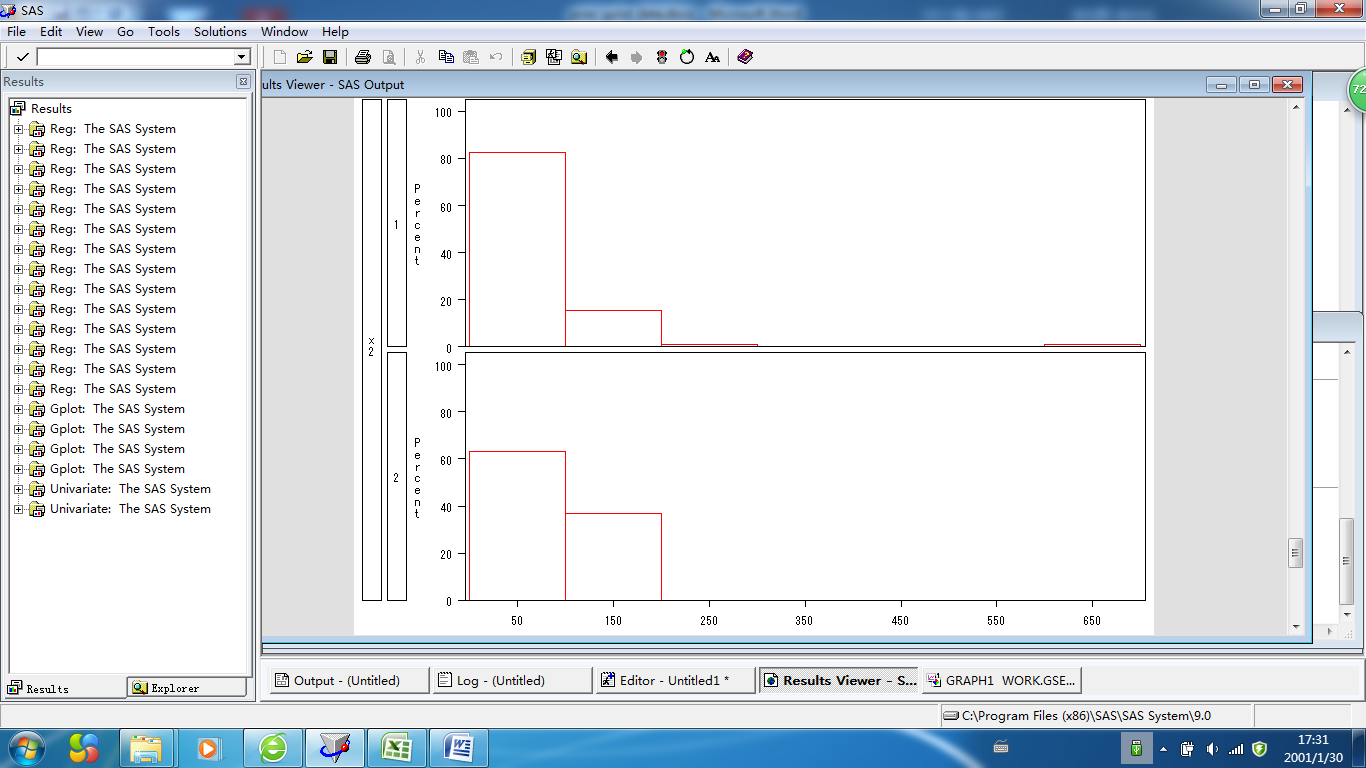
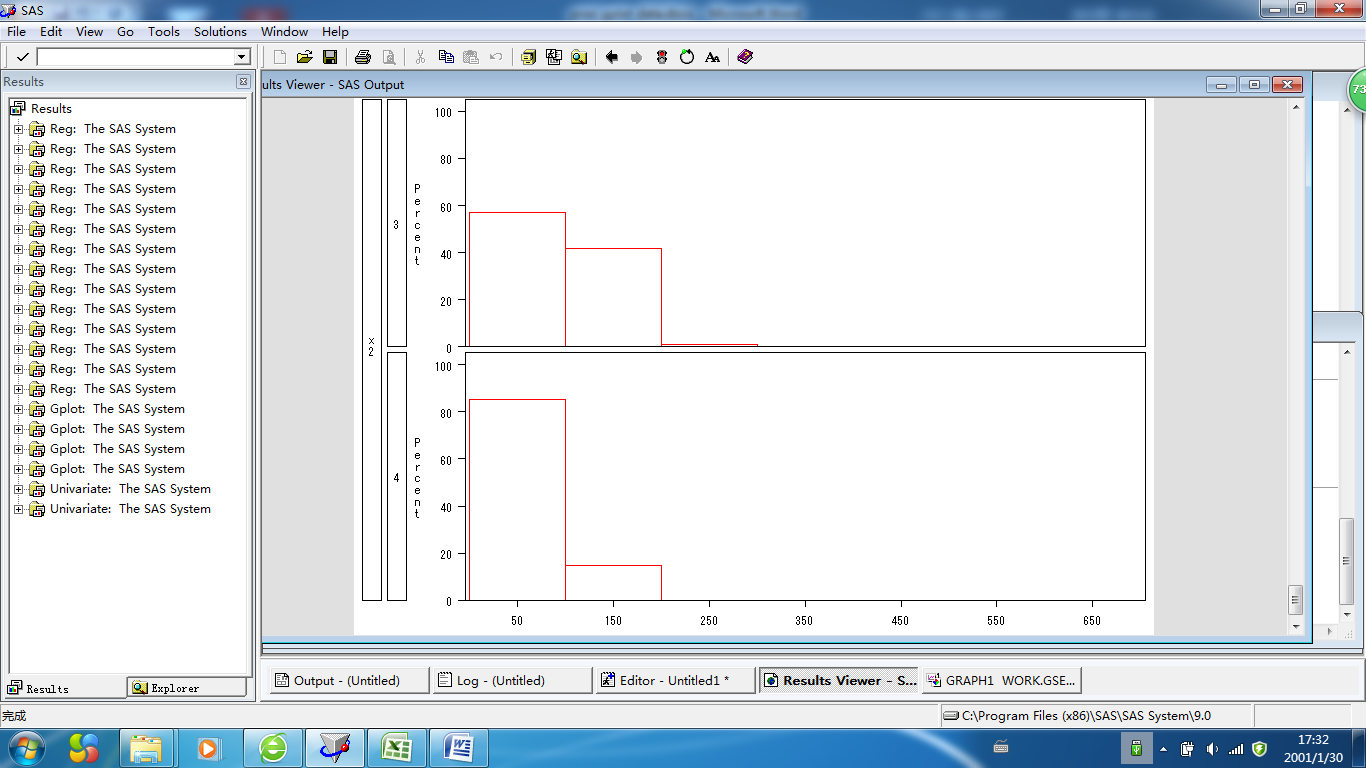
X1与y的柱状图：



图二

由图一观察到，工作日预定量总数大于周末预定量，但是预定量的趋势基本呈现正态分布，所以预定量与工作日和周末的关系并不明显。

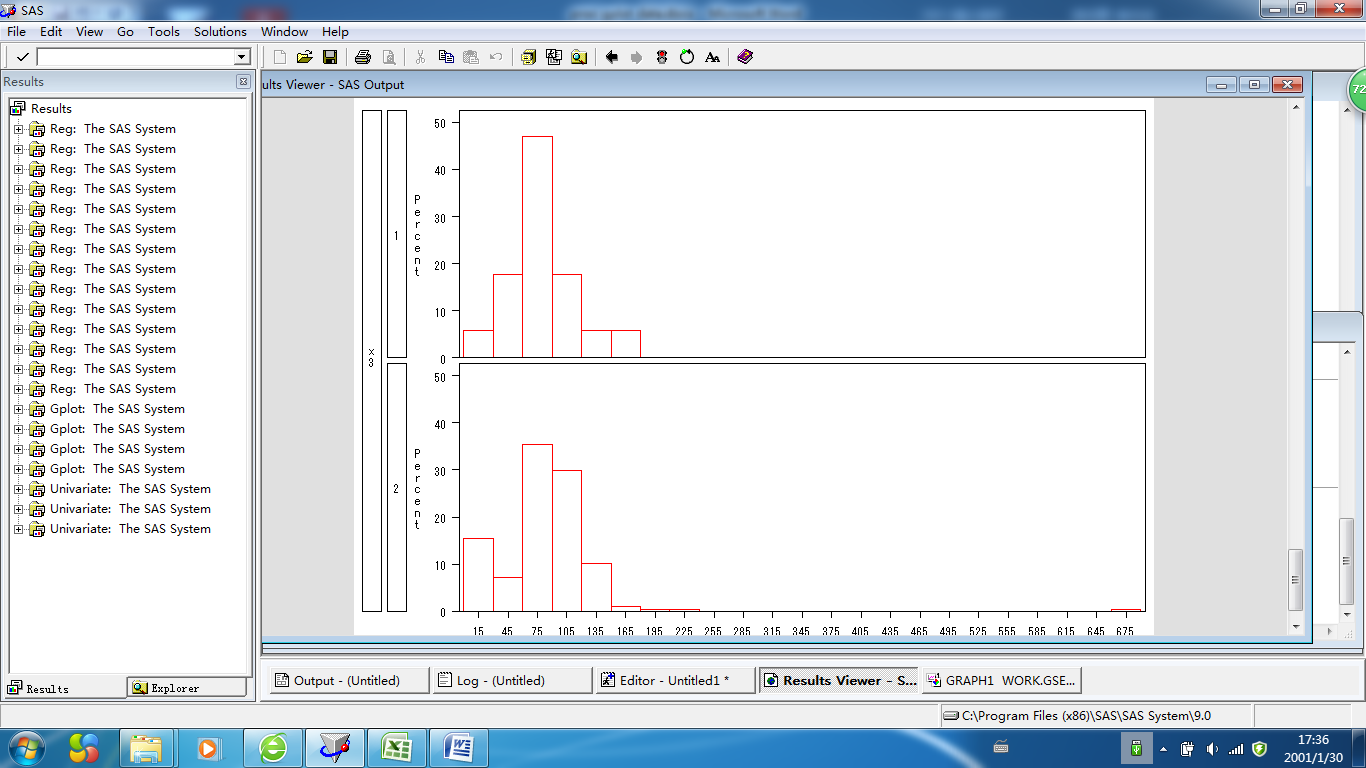
X2与y的柱状图：

图三 图四

由图二与图三分析得知四个季节整体总预定量是基本相同的，但春冬季节预定量基本上都在小的数据范围里，夏秋季节预定量则有很多位于大的数据范围里。所以季节对预定量影响是显著的。

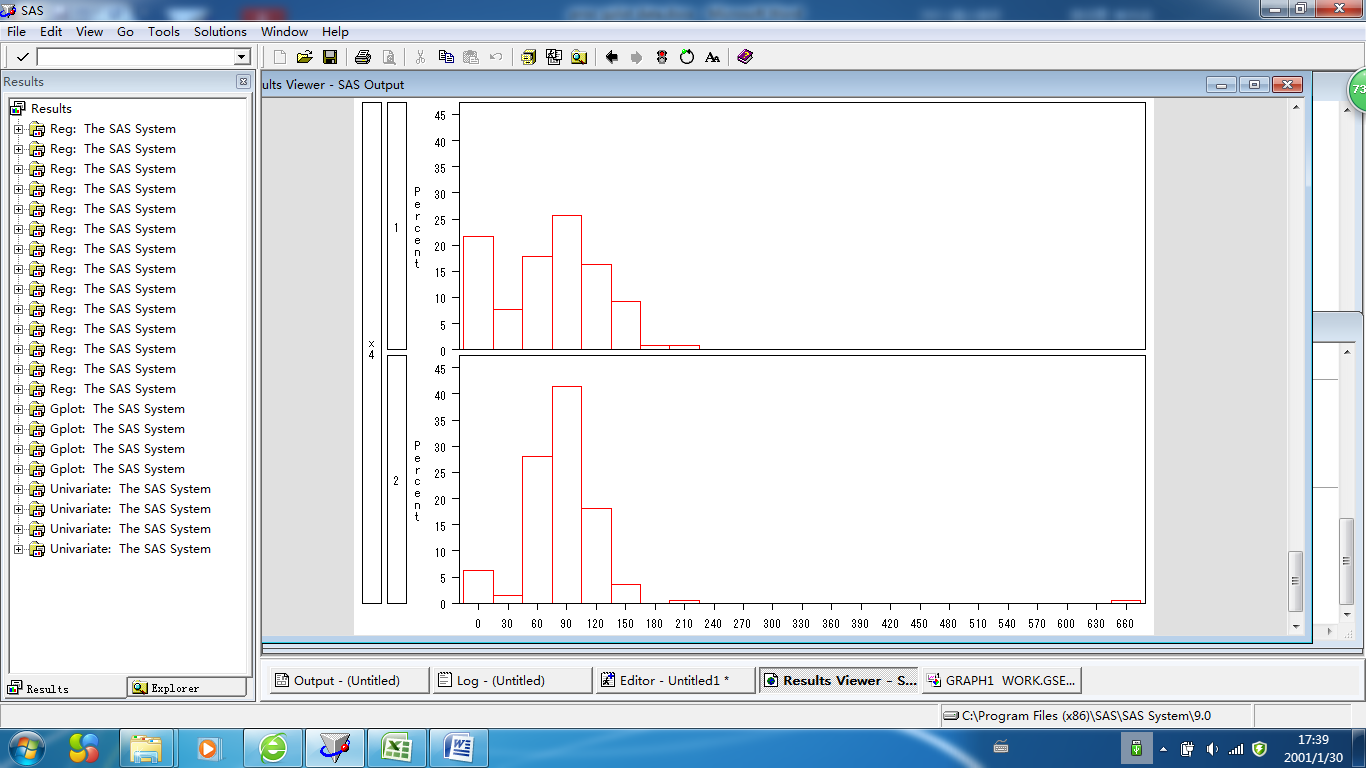
X3与y的柱状图：



图五

由图四分析，在节假日时期房间预定量呈现正态分布趋势，在非节假日时期房间预定量也具有正态分布趋势，由于在节假日房间预订会在节假日之前显著增加，具有后效性，所以预定量大的数据会出现在节假日前的一段时期，

X4与y的柱状图：



图六

由图五分析得，在寒暑假时期因为有较长一段时间的假期，所以在这段时间房间预定量基本是均匀分布，在非寒暑假时期房间预定量呈正态分布趋势。

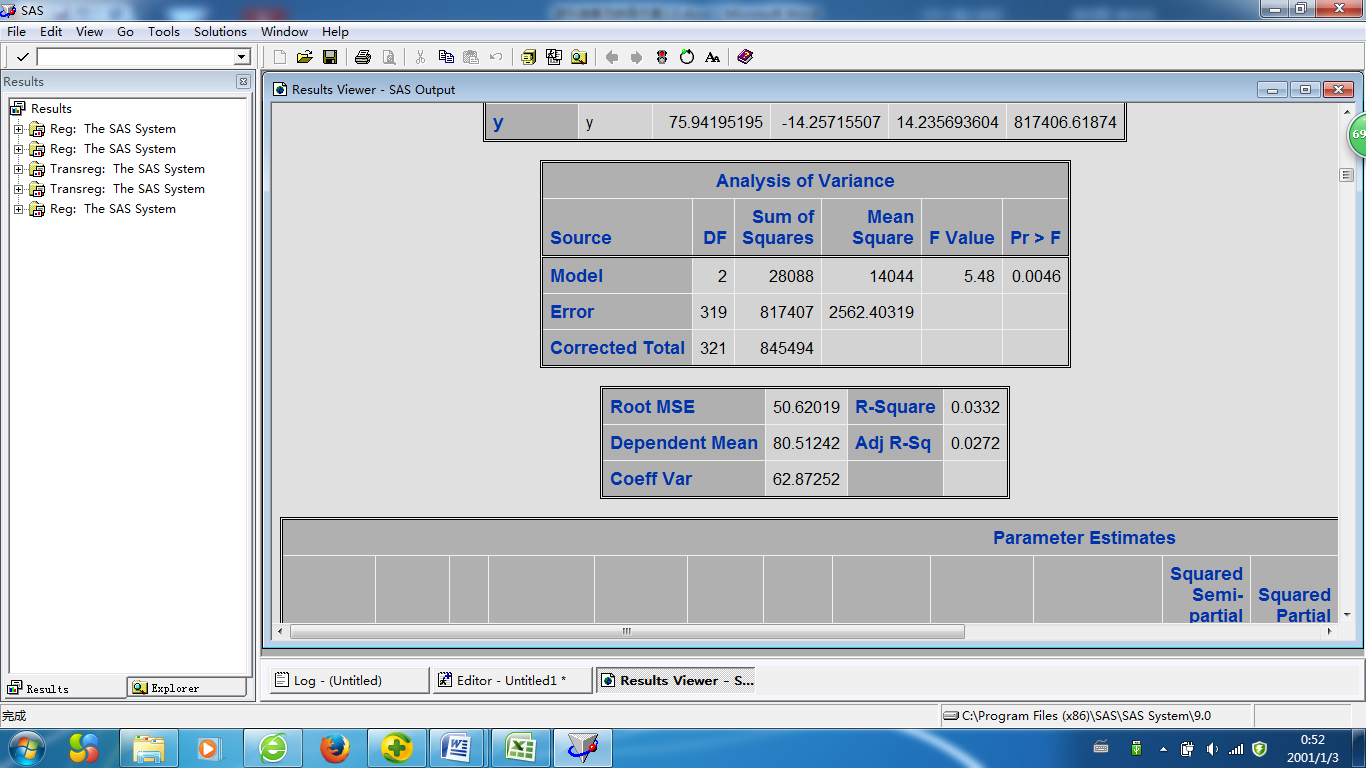
多元回归预测模型：

X1=1（在工作日）/2(在周末) x2=1（在春季）/2（在夏季）/3（在秋季）/4（在冬季）

X3=1（在节假日）/2（在非节假日） x4=1（在寒暑假）/2（在非寒暑假）。

Y=b0+b1\*x1+b2\*x2+b3\*x3+b4\*x4,其中b0,b1,b2,b3,b4皆为常数项。

回归结果如下(程序见附录)：



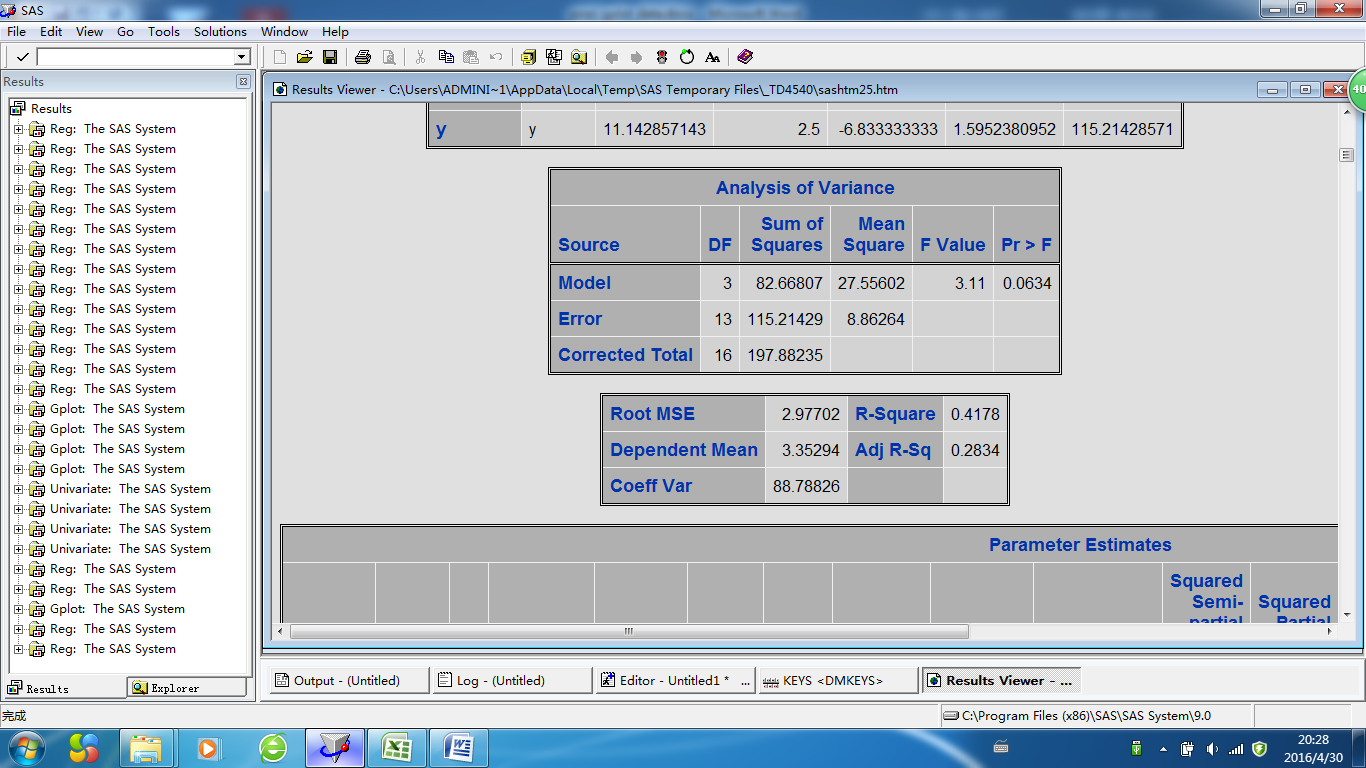
图七

得出回归预测模型为Y=90.15618-14.25716x1-1423569x4

在预测2016年的预定量时先将2015年一二月份数据提取之后建立回归模型：

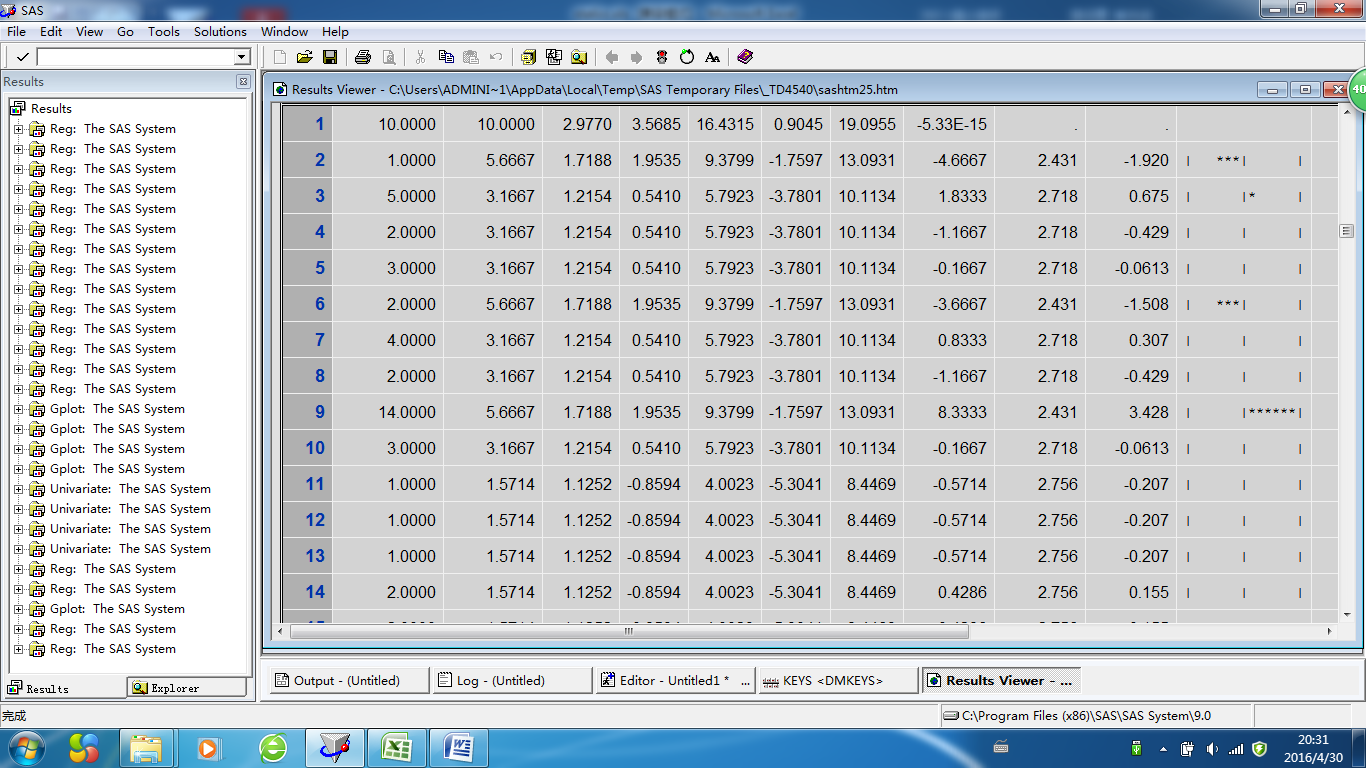
Y=a0+a1\*x1+a3\*x3+a4\*x4,其中a0,a1,a3,a4都为常数。因为一二月份都为冬季，所以将季节变量x2剔除，只考虑其余三个变量。

得到回归结果如下（程序见附录）：



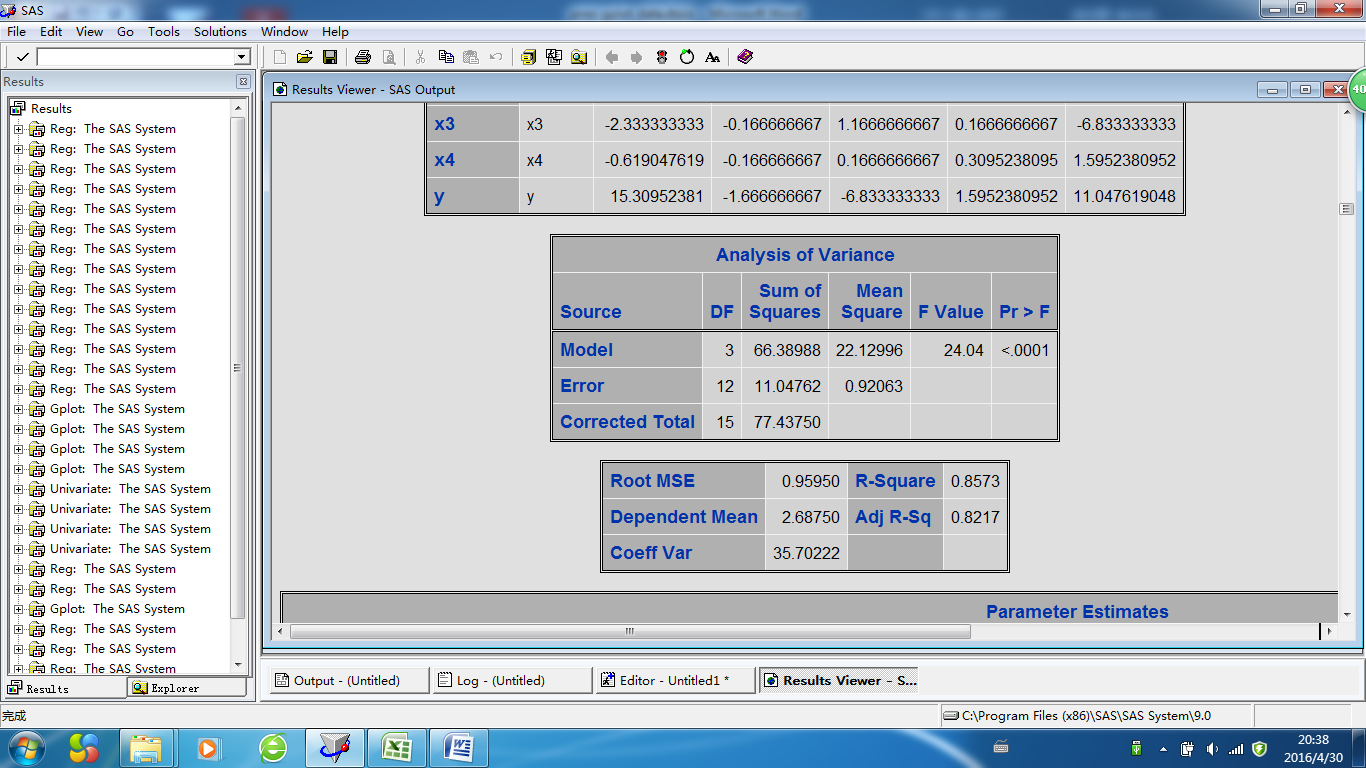
图八

在观察预测值时发现异常值：



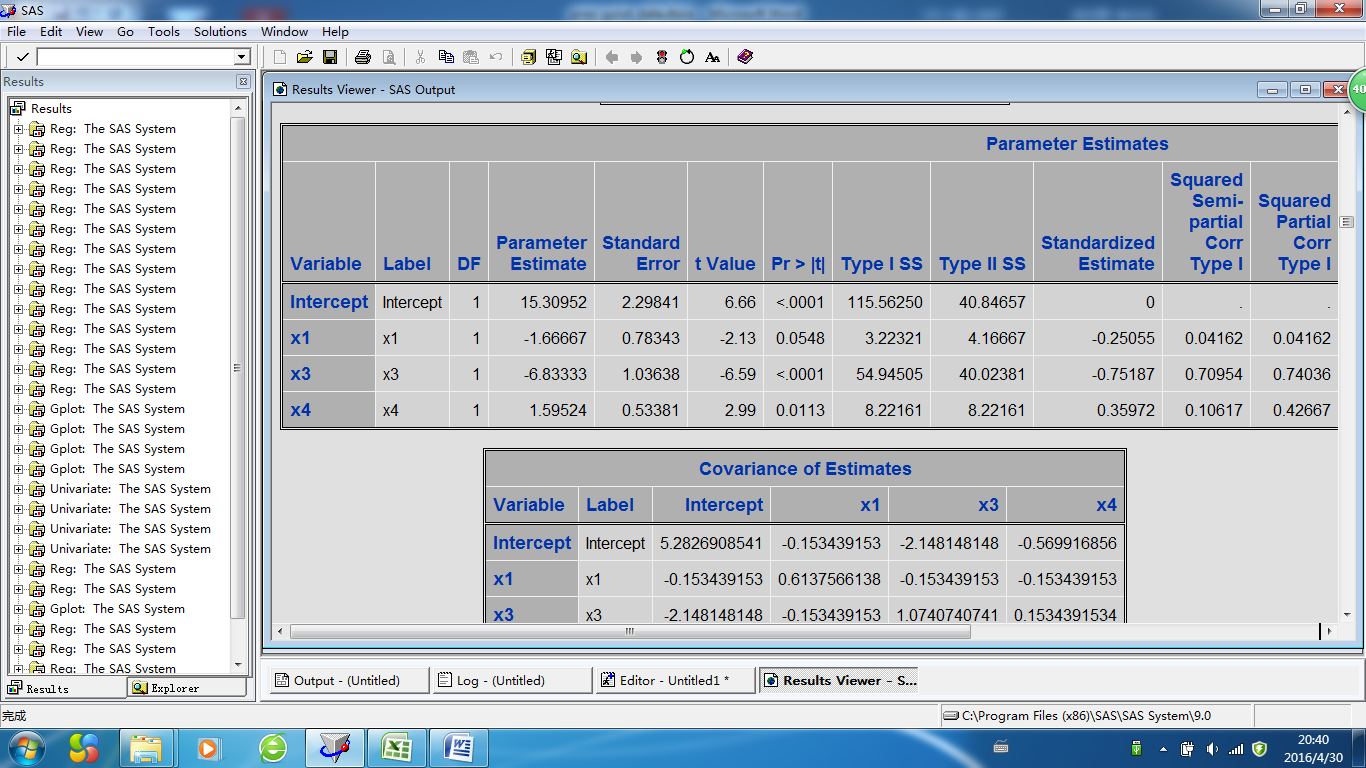
图九

所以将该异常值剔除再次回归得到如下结果：



图十

r-square显著提升



图十一

截距项以及x1，x2，x3的p检验也都通过。

回归方程：Y=15.309-1.6667x1-6.8333x3+1.5952x4

观察数据得知，三月数据与一二月数据有显著差别，在气温上升之后，酒店预订数量显著升高，所以将三月单独预测。可以发现三月数据在15号之前是一到十五之间的均匀随机数，在15号之后是5到30之间的均匀随机数，因为在三月后半月气温开始回升，春季季节性明显，所以可以通过Excel中randbetween（）产生不同范围内的均匀随机数的方法来预测。

最后得到的预测结果：

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| day | x1 | x3 | x4 | y^ |
| 1月1日 | 1 | 1 | 2 | 9.9994 |
| 1月2日 | 2 | 1 | 2 | 8.3327 |
| 1月3日 | 2 | 1 | 2 | 8.3327 |
| 1月4日 | 1 | 2 | 2 | 3.1661 |
| 1月5日 | 1 | 2 | 2 | 3.1661 |
| 1月6日 | 1 | 2 | 2 | 3.1661 |
| 1月7日 | 1 | 2 | 2 | 3.1661 |
| 1月8日 | 1 | 2 | 2 | 3.1661 |
| 1月9日 | 2 | 2 | 2 | 1.4994 |
| 1月10日 | 2 | 2 | 2 | 1.4994 |
| 1月11日 | 1 | 2 | 2 | 3.1661 |
| 1月12日 | 1 | 2 | 2 | 3.1661 |
| 1月13日 | 1 | 2 | 2 | 3.1661 |
| 1月14日 | 1 | 2 | 2 | 3.1661 |
| 1月15日 | 1 | 2 | 2 | 3.1661 |
| 1月16日 | 2 | 2 | 2 | 1.4994 |
| 1月17日 | 2 | 2 | 2 | 1.4994 |
| 1月18日 | 1 | 2 | 2 | 3.1661 |
| 1月19日 | 1 | 2 | 2 | 3.1661 |
| 1月20日 | 1 | 2 | 2 | 3.1661 |
| 1月21日 | 1 | 2 | 2 | 3.1661 |
| 1月22日 | 1 | 2 | 2 | 3.1661 |
| 1月23日 | 2 | 2 | 2 | 1.4994 |
| 1月24日 | 2 | 2 | 2 | 1.4994 |
| 1月25日 | 1 | 2 | 2 | 3.1661 |
| 1月26日 | 1 | 2 | 2 | 3.1661 |
| 1月27日 | 1 | 2 | 2 | 3.1661 |
| 1月28日 | 1 | 2 | 2 | 3.1661 |
| 1月29日 | 1 | 2 | 2 | 3.1661 |
| 1月30日 | 2 | 2 | 2 | 1.4994 |
| 1月31日 | 2 | 2 | 2 | 1.4994 |
| 2月1日 | 1 | 2 | 1 | 1.5709 |
| 2月2日 | 1 | 2 | 1 | 1.5709 |
| 2月3日 | 1 | 2 | 1 | 1.5709 |
| 2月4日 | 1 | 2 | 1 | 1.5709 |
| 2月5日 | 1 | 2 | 1 | 1.5709 |
| 2月6日 | 2 | 2 | 1 | -0.0958 |
| 2月7日 | 2 | 1 | 1 | 6.7375 |
| 2月8日 | 1 | 1 | 1 | 8.4042 |
| 2月9日 | 1 | 1 | 1 | 8.4042 |
| 2月10日 | 1 | 2 | 1 | 1.5709 |
| 2月11日 | 1 | 2 | 1 | 1.5709 |
| 2月12日 | 1 | 2 | 1 | 1.5709 |
| 2月13日 | 2 | 2 | 1 | -0.0958 |
| 2月14日 | 2 | 2 | 1 | -0.0958 |
| 2月15日 | 1 | 2 | 1 | 1.5709 |
| 2月16日 | 1 | 2 | 1 | 1.5709 |
| 2月17日 | 1 | 2 | 1 | 1.5709 |
| 2月18日 | 1 | 2 | 1 | 1.5709 |
| 2月19日 | 1 | 2 | 1 | 1.5709 |
| 2月20日 | 2 | 2 | 1 | -0.0958 |
| 2月21日 | 2 | 2 | 1 | -0.0958 |
| 2月22日 | 1 | 1 | 1 | 8.4042 |
| 2月23日 | 1 | 2 | 1 | 1.5709 |
| 2月24日 | 1 | 2 | 1 | 1.5709 |
| 2月25日 | 1 | 2 | 1 | 1.5709 |
| 2月26日 | 1 | 2 | 1 | 1.5709 |
| 2月27日 | 2 | 2 | 1 | -0.0958 |
| 2月28日 | 2 | 2 | 1 | -0.0958 |
| 2月29日 | 1 | 2 | 1 | 1.5709 |

表四

表四中y^为预测值，预测值可以采取取整的方法来得到预定数，负值就将该天的预测值取为0。

|  |  |
| --- | --- |
| Day | y^ |
| 3月1日 | 10 |
| 3月2日 | 13 |
| 3月3日 | 4 |
| 3月4日 | 8 |
| 3月5日 | 3 |
| 3月6日 | 7 |
| 3月7日 | 6 |
| 3月8日 | 1 |
| 3月9日 | 10 |
| 3月10日 | 5 |
| 3月11日 | 9 |
| 3月12日 | 2 |
| 3月13日 | 7 |
| 3月14日 | 3 |
| 3月15日 | 2 |
| 3月16日 | 23 |
| 3月17日 | 20 |
| 3月18日 | 7 |
| 3月19日 | 6 |
| 3月20日 | 5 |
| 3月21日 | 10 |
| 3月22日 | 29 |
| 3月23日 | 25 |
| 3月24日 | 30 |
| 3月25日 | 14 |
| 3月26日 | 5 |
| 3月27日 | 5 |
| 3月28日 | 30 |
| 3月29日 | 7 |
| 3月30日 | 12 |
| 3月31日 | 24 |

表五

其中y^为预测值。

七、模型的评价

**7.1模型的优点**

问题一中充分考虑了游客实际的用时情况，时间成本的比较方法较为精准。结果给出了在不同排队人数情况下的游览项目提醒，使得选择的范围更加准确。基于假定的情况将游客游览的动态过程简化为一个个体在不同选择时的时间比较。

问题二中分析预定量与四个因素的关系时，使用了定性分析来进行提前预测，使得分析结果更加合理。多元回归的预测模型具有很好的拟合度，使得最后的预测数据较为准确。对于三月数据，发现其特殊性，灵活使用了随机数模型来预测数据，使得最后的预测结果并无大的偏差。

**7.2模型的缺点**

在问题一的模型中，我们假定各个时刻进园的人数是相同的，未考虑实际情况是否会有高峰期的波动。并且未考虑到游客游览的随意性。对于项目持续时间较长的项目选择上有可能出现游客等待该项目的时间总是大于去其他项目的时间。

在问题二中，分析房间预定量与四个因素之间的关系的回归模型，预测精度并没有达到很高，使得模型具有一定偏差。

**7.3模型的改进与推广**

在已得到的结果的基础之上，游览提醒对游客的走向有一定的导向性，因此被推荐项目的游客单位时间到达人数多于其他未被推荐的项目，所以应该按照被推荐频率的比例分配人流量。

本文上述结果中A项目被其他4个项目推荐，因为每个项目推荐项目取距离最近的前3个，因此总的被推荐次数为30，则A项目的单位时间游客到达量为单位时间总的到达游园游客的到达量的。另外，因为A项目持续时间较长，所以可以令当在A项目排队的人数小于每场容纳的人数的时候就推荐游客在A项目前排队等待。

本文的模型可用于其他营业时常与其他客流量情况下公共娱乐场所的游览指示方案。

在目前得出的房间预定量与四个因素之间的关系使模型精度有一定欠缺，所以可以通过后续增加变量等方式提高模型精度。

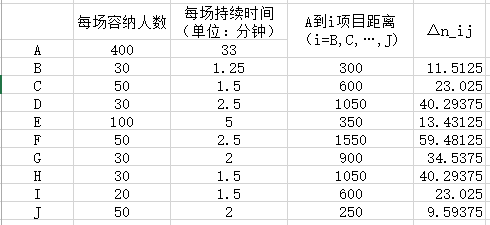
预订房间模型可以应用于酒店管理房间数量方面，使得酒店管理更有效率。

参考文献

何晓群 刘文卿，《应用回归分析》，北京，中国人民大学出版社，2001

附录1：

当游客刚到达A项目时



当游客刚到达B项目时



当游客刚到达C项目时



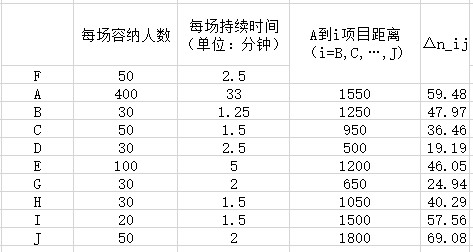
当游客刚到达D项目时



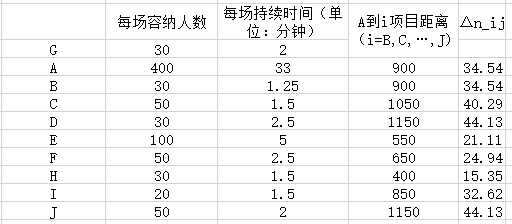
当游客刚到达E项目时



当游客刚到达F项目时



当游客到达G项目时



当游客到达H项目时



当游客到达I项目时



当游客刚到达J项目时



附录2

SAS画柱形图的程序：

**proc** **univariate** data=statistic;

var y;

class x1;

histogram;

**run**;

**quit**;

**proc** **univariate** data=statistic;

var y;

class x2;

histogram;

**run**;

**quit**;

**proc** **univariate** data=statistic;

var y;

class x3;

histogram;

**run**;

**quit**;

**proc** **univariate** data=statistic;

var y;

class x4;

histogram;

**run**;

**quit**;

回归预测模型程序：

**proc** **reg** data=statistic;

model y=x1 x2 x3 x4/all seletion=stepwise;

**run**;

**quit**;

一二月数据回归程序：

**proc** **reg** data=sss1;

model y=x1 x3 x4/all selection=stepwise;

**run**;

**quit**