

引用格式: 敖长林, 王菁霞, 孙宝生. 基于大数据的空气质量对公众外出游玩影响研究[J]. 资源科学, 2020, 42(6): 1199-1209.  
[Ao C L, Wang J X, Sun B S. Impact of air quality on public outings based on big data[J]. Resources Science, 2020, 42(6): 1199-1209.] DOI: 10.18402/resci.2020.06.16

# 基于大数据的空气质量对公众外出游玩影响研究

敖长林, 王菁霞, 孙宝生

(东北农业大学管理科学与工程系, 哈尔滨 150030)

**摘要:** 基于大数据探究空气质量对公众外出游玩的影响, 不仅可以丰富空气质量议题的实践研究, 还对加强空气污染治理、促进旅游产业发展具有重要意义。本文以哈尔滨市为研究对象, 采用携程旅行网 16554 条在线评论数据、环保部空气质量指数数据以及历史天气数据, 基于负二项回归与有序 Probit 回归分别构建公众外出游玩次数与游玩满意度模型, 定量评估空气质量对公众外出游玩的影响及影响程度。研究结果表明: ①在控制温度、风速、节假日等影响因素不变的基础上, 空气质量会显著影响公众外出游玩次数和满意度; ②公众外出游玩次数会随着空气污染程度的增加而显著减少, 与空气质量为优相比, 游玩次数在“严重空气污染”时下降 40.1%, 在“空气质量为良”时下降 15.1%; ③针对哈尔滨市旅游景区, 严重空气污染会显著降低公众外出游玩满意度, 而较低程度的空气污染对公众外出游玩满意度不具有显著影响。研究不仅为准确评估空气质量对公众外出行为的影响提供新视角, 也为相关空气污染防治及旅游政策的制定提供参考。

**关键词:** 空气质量; 旅游大数据; 外出游玩; 负二项回归; 有序 Probit 回归; 哈尔滨市

DOI: 10.18402/resci.2020.06.16

## 1 引言

空气污染作为当前严峻的环境问题, 威胁着世界各国的可持续发展, 尤其是发展中国家<sup>[1]</sup>。近年来, 随着城镇化和现代化的不断推进, 中国的空气污染问题愈发凸显, 特别是 2013 年所爆发的严重雾霾事件, 范围波及了 25 个省份, 100 多个大中型城市。大范围的空气污染对公众的健康造成了严重危害, 据 2016 年世界卫生组织 (WHO) 报道, 世界 92% 的人口遭受着空气污染的威胁, 其中中国因空气污染导致的死亡人数居世界前列<sup>[2]</sup>。空气污染不仅直接影响公众的健康和寿命<sup>[3,4]</sup>, 还对公众的外出活动产生负面效应<sup>[5]</sup>, 如消费出行选择、外出活动次数和满意度等, 进而降低社会经济活力和城市旅游形象, 成为阻碍旅游业发展的重要因素之一<sup>[6-8]</sup>。外出游玩作为公众外出活动的重要形式, 同样会受到

空气污染的影响<sup>[9]</sup>。探究空气质量对公众外出游玩活动的影响机制有助于评估公众规避空气污染的意愿强度, 明晰空气污染对公众生活质量和满意度的影响程度, 促进区域的可持续发展。

现有研究主要通过问卷调查、访谈和实地调研等方法探究空气质量对公众外出活动次数、出行方式和旅游体验满意度等的影响<sup>[10-12]</sup>。但问卷调查等传统方法存在数据收集成本较高、样本有限以及受访人员理解偏差等问题<sup>[13,14]</sup>。此外, Lu 等<sup>[15]</sup>的研究指出, 来自几个封闭问题的有限信息无法为更复杂的分析提供丰富的数据。随着 Web2.0 技术的快速发展, 各种类型的大数据被广泛应用于科学、工程、商业、管理、旅游等领域, 为信息时代下的科学研究作出了重要贡献<sup>[16]</sup>。其中, 基于大数据的旅游研究就是一个新兴的典型例子。作为旅游大数据的一

收稿日期: 2019-07-12 修订日期: 2019-10-28

基金项目: 国家自然科学基金项目 (71874026)。

作者简介: 敖长林, 男, 黑龙江杜蒙县人, 教授, 研究方向为生态环境管理、评价理论与方法。E-mail: aochanglin2002@126.com

种重要类型,海量在线旅游评论(包括酒店评论、餐馆评论和景区评论)数据因具有易获取、范围广、成本低和客观性等特征,受到众多学者的关注<sup>[17-19]</sup>。如 Bakhshi 等<sup>[20]</sup>采用在线旅游评论数据研究发现,气温、降雪、降雨和季节等因素会显著影响公众对餐馆的线上评分。应用在线旅游评论探究游客满意度等旅游问题已成为当前研究的热点<sup>[21]</sup>。然而,鲜有基于在线旅游评论大数据分析空气质量对公众外出游玩影响的研究。

鉴于此,本文以长期遭受空气污染影响的哈尔滨市为例,结合携程旅行网的在线旅游评论数据、生态环境部发布的空气质量指数数据以及历史天气数据,通过构建基于负二项回归和有序 Probit 回归的公众外出游玩次数与游玩满意度模型,揭示空气质量对公众外出游玩的影响,从而为相关空气污染防治和旅游政策的制定提供参考。

## 2 研究区概况、数据来源与研究方法

### 2.1 研究区概况

哈尔滨市位于黑龙江省南部,地理位置介于 125°42'E—130°10'E、44°04'N—46°40'N 之间,属于热门旅游城市和国际冰雪文化名城,素有“东方小巴黎”“东方莫斯科”的美名,1998 年被国家旅游局评选为首批中国优秀旅游城市,2017 年 12 月当选中国十佳冰雪旅游城市,2018 年 10 月获得全球首批“国际湿地城市”称号。辖区内有圣索菲亚教堂、中央大街、松花江和太阳岛等多处人文自然景观,吸引了大批游客。

然而,进入 21 世纪以来,哈尔滨市空气污染状况愈发严重的问题日益凸显<sup>[22]</sup>。哈尔滨市统计局数据显示,2017 年哈尔滨平均 PM<sub>2.5</sub> 浓度约为 87 μg/m<sup>3</sup>,远超过世界卫生组织(WHO)所建议的 10 μg/m<sup>3</sup> 标准。因此,本文选取哈尔滨市作为研究空气质量对公众外出游玩影响的案例地具有一定的典型性和代表性。

### 2.2 数据来源及处理

#### 2.2.1 数据来源

探究空气质量对公众外出游玩次数与满意度的影响,需获取两类关键的数据:公众外出游玩数据、空气质量指数数据;已有研究表明天气因素也会对公众的外出行为产生影响<sup>[20]</sup>,因此本文在模型

构建过程中加入天气条件等控制变量以排除其对结果的干扰,进而提高空气质量在模型中的解释能力。

游玩次数、游玩评级、景区门票价格和景区等级数据来源于携程旅行网(ctrip.com),携程旅行网是中国领先的在线旅行服务公司,提供机票预订、度假预订和旅游资讯等全方位的旅行服务,拥有超过 300 万注册用户,月活跃用户数超过 68 万,其在线评论功能受到了广大游玩者的认可,庞大的用户数量可以确保网络数据的代表性<sup>[23]</sup>,作为数据来源,已多次应用于基于大数据的旅游研究中<sup>[24,25]</sup>。

消费者在景区游玩后,到携程旅行网上该景区的页面对景区进行评分(1~5 分),并对此次游玩作出整体评价,可以为研究提供各景区的评论数量和总体评价情况。虽然以每日的评论数量作为公众外出游玩次数有一定的局限性:由于并不是所有的外出游玩者都在网上进行评论,评论数量会小于真实的游玩次数,但线上评论数量与真实游玩次数具有高度相关性<sup>[5]</sup>。因此,日评论数量可作为真实游玩次数的替代变量,且能够较好地反映公众外出游玩次数随时间的变化趋势。

数据采集根据以下 3 个原则选定研究景区:①在哈尔滨市 9 个城区范围内;②自 2016 年以来全年开放;③在携程旅行网上评论数量不少于 100 条。由于在携程旅行网上未找到双城区内符合上述原则的景区,因此仅考虑南岗区、道里区等 8 个城区的景区评论和评级,最终选取了哈尔滨市 8 个城区的 33 个景区,如中央大街、圣索菲亚教堂、太阳岛等,于 2018 年 7 月利用 Python 语言编程获取 28626 条景区原始研究数据。由于携程旅行网对每个景区仅保留 300 页的评论,且每个景区的日评论数量各不相同,因此以评论时间跨度最短的景区的起始时间(2016 年 9 月)作为研究的起始时间。从采集得到的研究数据中筛选出时间维度在 2016 年 9 月至 2018 年 7 月的数据,共计 16554 条。在此基础上,由于景区的自身属性会显著影响公众满意度<sup>[26]</sup>,所以还需采集各景区的门票价格和景区等级数据,以控制景区自身属性对公众游玩满意度的影响。

空气质量指数(AQI)是 2012 年 3 月国家发布的新空气质量评价标准,污染监测包含二氧化硫、二

2020年6月

氧化氮、PM10、PM2.5、一氧化碳和臭氧6项,数据每小时更新一次。由于依据3个原则而选取的旅游景区主要集中在分布于松花江沿岸,各监测站的AQI数据整体差距并不明显,且存在公众跨区域出行的情况,因此本文未选用哈尔滨市各空气质量监测站的数据,而是选取中华人民共和国生态环境部发布的哈尔滨市AQI作为全市的统一取值。公众外出游玩后并非立即评论,以小时为单位进行研究并无必要,因此选择以天为单位的空气质量指数数据<sup>[5]</sup>,时间区间为2016年9月—2018年7月。

控制变量中的天气条件数据,如:日均温度(*temperature*)、日均风速(*wind speed*)、日均能见度(*visibility*)、日降水量(*precipitation*)、是否降雪(*snow*)以及是否降雨(*rain*)等来源于网站“<https://www.wun->

[derground.com/](http://derground.com/)”。公众外出游玩次数、空气质量指数与部分天气条件数据的变化趋势如图1所示。

## 2.2.2 数据处理

### (1) 空气质量指数

空气质量指数(AQI)是定量描述空气质量状况的无量纲指数,描述空气清洁或者污染的程度,以及对健康的影响<sup>[27]</sup>。AQI将空气质量分为6个等级,AQI值在0~50为优,51~100为良,101~150为轻度污染,151~200为中度污染,201~300为重度污染,大于300为严重污染。

### (2) 评论数量与评分等级

将采集数据中各景区每日评论数量的总和作为日评论数量,即每日公众外出游玩次数(以下简称游玩次数)。对于评分等级,即公众外出游玩满

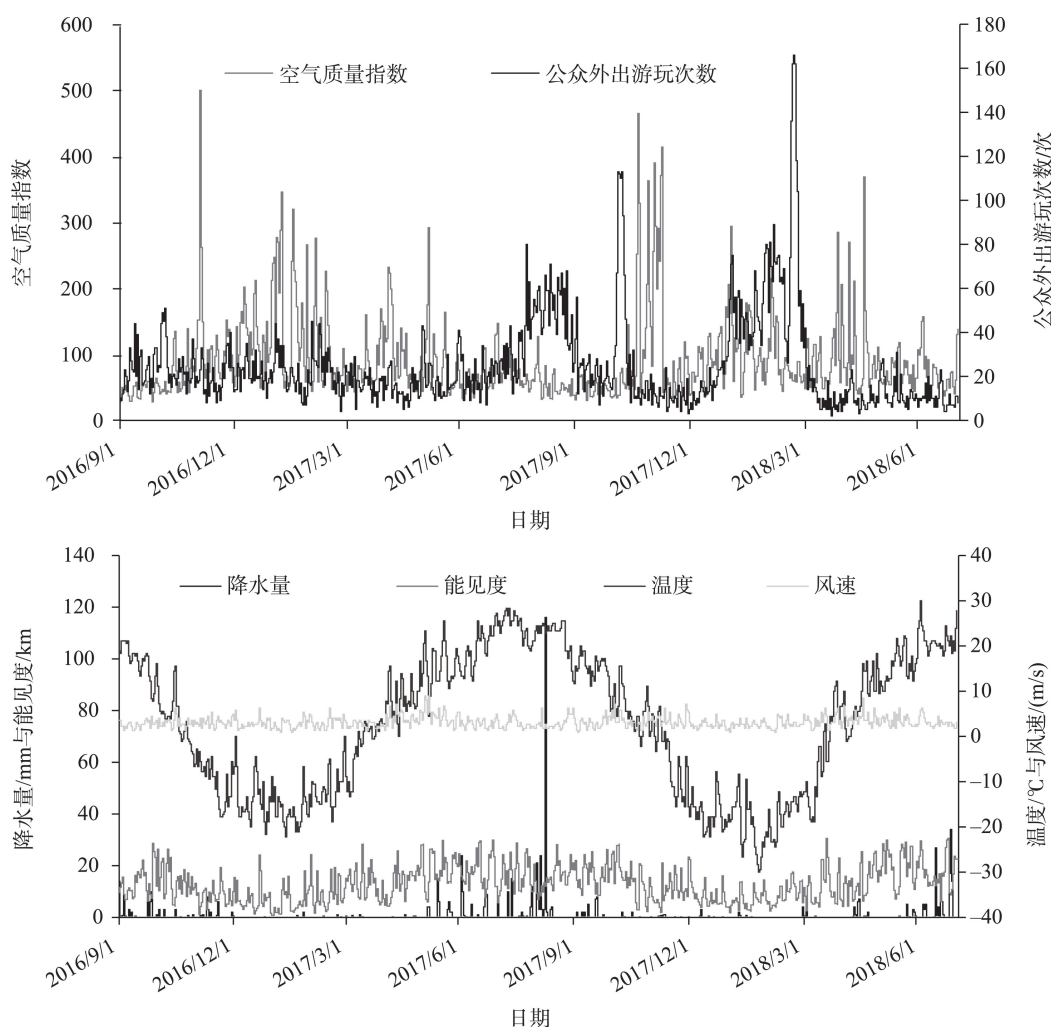


图1 公众外出游玩次数、空气质量指数、降水量、温度、能见度、风速变化趋势

Figure 1 Trends of trip frequency, air quality index (AQI), precipitation, temperature, visibility, and wind speed

意度(以下简称游玩满意度),将每日各景区评论的评分总和除以各景区的评论数量作为各景区每日的公众外出游玩满意度。

(3)其他控制变量

利用收集的湿度数据来计算评论当天的平均温度估计值  $\bar{T} = \left( \sum_{i=1}^{24} T_i \right) / 24$ ,  $T_i$  为第  $i$  小时的温度,将温度划分为 4 个类别,描述公众对温度的感知<sup>[20]</sup>。同时,参考中国气象局天气变量的划分标准,将日均风速划分为 5 个类别,日均能见度划分为 3 个类别,日降水量划分为 3 个类别,如表 1 所示。

已有文献<sup>[26]</sup>表明研究对象的自身属性会显著影响公众满意度,因此在研究空气质量对哈尔滨市各景区满意度的影响时,将景区自身属性(景区门票价格与景区等级)纳入自变量中作为控制变量,用于控制景区质量对公众游玩满意度的影响。依据哈尔滨市各景区的实际价格情况,将门票价格分为 3 个类别,如表 1 所示。

由于本文探究的是空气质量对哈尔滨市整体的游玩次数影响,而不是空气质量对每个景区单独游玩次数的影响。加之,不是所有景区每天都有评

论,若将各景区的价格和等级加入游玩次数模型中,会破坏哈尔滨市整体游玩次数的时间连续性。因此,本文在构建公众外出游玩次数模型时没有考虑景区自身属性因素。

由于哈尔滨市供暖时间从 10 月持续至次年 4 月底,因此将 10 月 1 日至次年 4 月 30 日划分为供暖季,其他时间划分为非供暖季。因为法定假日对公众外出游玩的选择存在明显差异,所以将中国法定节假日分为节假日与非节假日 2 类。同时,由于在周一至周日期间公众的外出游玩机会存在着明显不同,所以将一周七天分为周末与非周末 2 个类别。

2.3 研究方法

2.3.1 公众外出游玩次数模型构建

分析自变量与因变量关系时,常采用多元线性回归,并用最小二乘估计回归参数。但由于多元线性回归要求因变量是连续型变量,而本文的研究数据为离散型,因此需要考虑其他的分析方法。泊松回归与负二项回归模型已广泛应用于交通事故<sup>[28]</sup>、医学<sup>[29]</sup>、流行病学<sup>[30]</sup>等方面的研究。通过对数据的观察与分析,公众外出游玩次数  $y$  的方差  $\sigma^2 = 436.510$ , 均值  $\mu = 24.802$ , 负二项回归更适合因

表 1 空气质量指数及其他控制变量分类结果

Table 1 Classification results of air quality index (AQI) and other control variables

属性	分类	属性	分类
空气质量指数 AQI	[0, 50]	降水量 precipitation/mm	=0
	[51, 100]		(0, 25]
	[101, 150]		(25, ∞)
	[151, 200]	雪 snow	=0 无雪
	[201, 300]		=1 有雪
温度 temperature/°C	(300, ∞)	雨 rain	=0 无雨
	(-∞, -7]		=1 有雨
	(-7, 5]	供暖季 season	=0 非供暖季
风速 wind speed/(m/s)	(5, 21]		=1 供暖季
	(21, 38]	节假日 holiday	=0 非节假日
	(-∞, 1.5]		=1 节假日
	(1.5, 3.3]	周末 weekend	=0 非周末
	(3.3, 5.4]		=1 周末
能见度 visibility/km	(5.4, 7.9]	门票价格 price/元	[0, 100]
	(7.9, 10.7]		(100, 200]
	(-∞, 2]		(200, ∞)
	(2, 10]		
	(10, ∞)		

2020年6月

变量过度离散的情况,即因变量的方差大于均值<sup>[31]</sup>。依据数据的属性特点及统计特征,本文基于负二项回归构建公众外出游玩次数 $y$ 与 $AQI$ 之间的函数模型,并加入控制变量,模型可表示为如下形式:

$$\ln(y) = I + \beta_{AQI}AQI + \beta_{temperature}temperature + \beta_{windspeed}windspeed + \beta_{visibility}visibility + \beta_{precipitation}precipitation + \beta_{snow}snow + \beta_{rain}rain + \beta_{season}season + \beta_{holiday}holiday + \beta_{week}week \quad (1)$$

式中: $I$ 是截距; $AQI$ 是哈尔滨市空气质量指数数据,反映哈尔滨市每日的空气质量情况,也是主要解释变量;在模型的控制变量中, $temperature$ 、 $windspeed$ 、 $visibility$ 、 $precipitation$ 、 $snow$ 、 $rain$ 等为天气变量,分别为每日的温度、风速、能见度、降水量、是否降雪、是否降雨等, $season$ 为供暖季变量, $holiday$ 为节假日变量, $weekend$ 为周末变量。 $\beta$ 为游玩次数模型的边际效应,是指在控制其他影响因素不变的条件下, $\beta$ 对应的自变量变化时游玩次数对数的变化率<sup>[31]</sup>。传统的负二项回归模型采用最大似然法估计参数,在负二项回归的基础上,采用贝叶斯参数估计方法即为贝叶斯负二项回归。相比于传统的负二项回归模型,贝叶斯负二项回归模型的拟合参数不再是固定值,而是具有某种先验分布的随机变量。本文采用马尔可夫链蒙特卡罗方法(MCMC)对贝叶斯负二项回归参数进行抽样,分别采用不同的先验分布进行贝叶斯建模,即假设负二项回归参数分别服从正态分布、均匀分布和拉普拉斯分布。

### 2.3.2 公众外出游玩满意度模型构建

公众外出游玩满意度指定为1到5的整数,此时最小二乘回归模型不再适用。有序Probit回归模型适用于序数因变量和二元因变量,因此本文采用有序Probit回归模型对公众外出游玩满意度进行建模。

有序型选择模型是有序离散数据分析与预测的常用模型之一,将反映变量映射成具有明确顺序的有序变量。模型的一般形式为:

$$z_i^* = \beta'_{AQI}AQI + \beta'_{temperature}temperature + \beta'_{windspeed}windspeed + \beta'_{visibility}visibility + \beta'_{precipitation}precipitation + \beta'_{snow}snow + \beta'_{rain}rain + \beta'_{season}season + \beta'_{holiday}holiday + \beta'_{week}week + \beta'_{price}price + \beta'_{level}level + \varepsilon_i \quad (2)$$

式中: $z_i^*$ 表示游玩满意度 $z$ 的内在趋势,它并不能被直接测量,游玩满意度变量有5个类别,相应取值为 $z=1,2,\dots,5$ ,共有4个分界点(即 $\mu_j(j=1,2,3,4)$ ,且 $\mu_1 < \mu_2 < \mu_3 < \mu_4$ )将各相邻类别分开,即若 $z^* \leq \mu_1$ ,则 $z=1$ ;若 $\mu_1 < z^* \leq \mu_2$ ,则 $z=2$ ;以此类推,若 $z^* > \mu_4$ ,则 $z=5$ 。 $\varepsilon_i$ 为误差项,是难以观测但对反应变量有影响的其他因素的总和,当 $\varepsilon_i$ 服从正态分布时,模型即为有序Probit回归模型<sup>[32]</sup>。在式(1)解释变量基础上,将景区门票价格 $price$ 与等级 $level$ 变量加入控制变量中,用于控制景区自身条件对游玩满意度的影响。 $\beta$ 为满意度模型的边际效应,是指在其他条件不变的情况下, $\beta$ 对应的自变量变化时游玩满意度的变化率。

## 3 结果与分析

### 3.1 公众外出游玩次数模型结果及分析

在对连续型变量进行赋值的基础上,使用R语言编程对负二项回归模型和贝叶斯负二项回归模型进行参数估计,结果如表2所示。其中,通过对比分别服从正态分布、均匀分布和拉普拉斯分布的贝叶斯负二项回归模型结果,发现服从3种先验分布的模型拟合结果基本一致,且正态分布的拟合效果最优,因此本文选取服从正态先验分布的贝叶斯负二项回归模型展开进一步分析。模型的主要解释变量为 $AQI$ ,并加入节假日、雪、雨、周末、供暖季、温度、风速等控制变量。

由于 $AIC$ 指标会忽略贝叶斯负二项回归模型估计参数的先验分布信息,因此本文使用均方误差( $MSE$ )评估模型拟合精度的高低。由表2可知,负二项回归模型的 $MSE$ 值较低,说明该模型的拟合效果更优。同时,负二项回归模型的McFadden  $R^2$ 值为0.289,介于0.2到0.4区间范围内<sup>[33]</sup>,说明该模型的解释能力较好。

负二项回归模型的结果表明, $AQI$ 和控制变量中的温度、节假日、供暖季、风速等因素均会对公众外出游玩次数产生显著影响。

公众外出游玩次数与空气质量指数( $AQI$ )呈显著负相关,在控制变量不变的条件下,以空气质量为优的类别作为参考,其他类别的 $AQI$ 均会导致游玩次数的减少。当 $AQI$ 大于300时,公众外出游玩

表2 参数估计及检验结果

Table 2 Parameter estimation and test results

变量	负二项回归		贝叶斯负二项回归(正态先验)	
	系数	标准误	系数	标准误
常数项	3.875***	0.234	3.872***	0.239
节假日	0.900***	0.086	0.902***	0.090
雪	0.006	0.090	0.009	0.090
雨	0.044	0.071	0.043	0.071
周末	0.064	0.051	0.063	0.052
供暖季	-0.188**	0.087	-0.187**	0.090
$AQI [51, 100]$	-0.164***	0.062	-0.162**	0.063
$AQI [101, 150]$	-0.285***	0.089	-0.281***	0.092
$AQI [151, 200]$	-0.269**	0.131	-0.263*	0.135
$AQI [201, 300]$	-0.449***	0.153	-0.434***	0.153
$AQI (300, \infty)$	-0.513**	0.227	-0.491**	0.235
温度(-7, 5]	-0.503***	0.074	-0.502***	0.075
温度(5, 21]	-0.592***	0.088	-0.590***	0.088
温度(21, 38]	-0.024	0.121	-0.018	0.121
风速(1.5, 3.3]	-0.147	0.094	-0.147	0.095
风速(3.3, 5.4]	-0.292***	0.100	-0.292***	0.102
风速(5.4, 7.9]	-0.174	0.143	-0.171	0.146
风速(7.9, 10.7]	-0.351	0.323	-0.315	0.324
能见度(2, 10]	0.030	0.205	0.029	0.210
能见度(10, $\infty$ )	-0.076	0.219	-0.075	0.224
降水量(0, 25]	-0.101	0.069	-0.100	0.070
降水量(25, $\infty$ )	-0.382	0.353	-0.332	0.358
$AIC$		5276.6		——
$McFadden R^2$		0.289		——
$MSE$		1.059		333.462

注: \*、\*\*、\*\*\*分别表示在10%、5%、1%显著性水平下显著。

次数最少,是空气质量为优时的59.9% ( $\beta_{AQI(300, \infty)} = -0.513$ ),即下降了40.1%; $AQI$ 在201~300时,游玩次数是空气质量为优时的63.8% ( $\beta_{AQI[201, 300]} = -0.449$ ); $AQI$ 在101~150和151~200时的游玩次数差别较小,分别是空气质量为优时游玩次数的75.2%和76.4%;而 $AQI$ 在51~100时,游玩次数是空气质量为优时的84.9% ( $\beta_{AQI[51, 100]} = -0.164$ ),下降了15.1%,说明空气质量良好时,公众更愿意外出游玩,而当空气质量处于严重污染或重度污染时,外出游玩次数会明显减少,表明公众会产生规避空气污染的行为,这与郑思齐等<sup>[5]</sup>探究空气污染

对北京市居民外出就餐行为影响的研究结果基本一致。

在控制变量中,以温度低于-7℃的类别作为参考,温度在-7℃~5℃和5℃~21℃时,游玩次数分别是温度低于-7℃的60.5% ( $\beta_{temperature(-7, 5]} = -0.503$ )和55.3% ( $\beta_{temperature(5, 21]} = -0.592$ ),表明哈尔滨市温度在低于-7℃时能够促进公众外出游玩次数增加。与1.5 m/s的风速相比,3.3~5.4 m/s的风速对游玩次数有显著的负向影响,其游玩次数是1.5 m/s风速条件下游玩次数的74.7% ( $\beta_{windspeed(3.3, 5.4]} = -0.292$ ),说明低风速为公众提供了良好的外出游玩条件,增加了

2020年6月

公众外出游玩的可能。哈尔滨市景区在节假日期间的游玩次数是非节假日游玩次数的2.46倍,由于公众在节假日有大量自由安排的时间,因此会在节假日期间放松身心,进行外出游玩,这与事实情况相符。供暖季的游玩次数是非供暖季游玩次数的82.9% ( $\beta_{season} = -0.188$ ),其原因可能是在供暖季公众更愿意处于温和的室内,从而减少了外出游玩次数。

### 3.2 公众外出游玩满意度模型结果及分析

将AQI作为主要解释变量,并在模型中加入节假日、雪、雨、周末、供暖季、温度、风速、能见度、降水量、景区价格和景区等级等控制变量。使用R语言编程对有序Probit回归模型进行参数估计,结果如表3所示。

表3 有序Probit回归参数估计及检验结果

Table 3 Parameter estimation of Ordered Probit regression and test results

变量	系数	标准误
AQI [51, 100]	0.031	0.032
AQI [101, 150]	0.014	0.048
AQI [151, 200]	-0.039	0.072
AQI [201, 300]	-0.085	0.087
AQI (300, $\infty$ )	-0.679***	0.121
温度(-7, 5]	-0.073	0.046
温度(5, 21]	-0.105**	0.051
温度(21, 38]	-0.047	0.065
风速(1.5, 3.3]	-0.027	0.051
风速(3.3, 5.4]	-0.015	0.055
风速(5.4, 7.9]	-0.019	0.078
风速(7.9, 10.7]	-0.045	0.170
能见度(2, 10]	-0.079	0.120
能见度(10, $\infty$ )	-0.073	0.128
降水量(0, 25]	0.057	0.038
降水量(25, $\infty$ )	-0.305**	0.125
门票价格(100, 200]	0.025	0.037
门票价格(200, $\infty$ )	-0.182***	0.051
景区等级	0.008	0.007
节假日	-0.008	0.046
雪	-0.054	0.052
雨	0.031	0.038
周末	0.034	0.029
供暖季	0.091*	0.050
AIC		16447

利用偏差检验有序Probit回归模型的拟合优度,偏差渐近地遵循 $\chi^2$ 分布,模型的拟合结果如表3所示。由拟合结果可以得出,有序Probit回归模型的P值约为 $5.477e-12$ ,在1%的显著性水平下通过显著性检验。因此,可采用有序Probit回归模型探究空气质量对公众外出游玩满意度的影响。

有序Probit回归模型的结果表明,AQI和控制变量中的温度、降水量、门票价格、供暖季等因素对公众外出游玩满意度有显著影响。在温度、风速、节假日等控制变量不变的情况下,以空气质量为优的类别作为参考,空气质量为严重污染时,会显著降低公众外出游玩满意度,而在空气污染不严重的情况下,游玩满意度与空气质量为优时的游玩满意度没有显著差异。说明公众在外出游玩时会忽视较低程度空气污染产生的有害影响,与Yan等<sup>[34]</sup>基于地理标记的微博数据和空气污染数据评估中国居民的暴露风险所得结论基本一致。

在模型的控制变量中,以温度低于 $-7^{\circ}\text{C}$ 的类别作为参考,温度在 $5^{\circ}\text{C}\sim 21^{\circ}\text{C}$ 时,对游玩满意度有显著的负向影响,其原因可能是在温度介于 $5^{\circ}\text{C}\sim 21^{\circ}\text{C}$ 时哈尔滨市特色旅游项目相对较少,导致了游玩满意度的降低;以降水量为0 mm作为参考,降水量超过25 mm时,公众外出游玩满意度会下降,原因在于降水量过大时,会严重影响景区的观赏性和道路交通,从而降低了游玩满意度;以门票价格不大于100元的类别作为参考,门票价格高于200元的系数显著为负,说明与门票价格不大于100元相比,门票价格高于200元时的游玩满意度会降低,因为高门票价格使公众对景区产生较高期望值,当游玩体验未达到心理预期时,公众的满意度会下降。供暖季的系数显著为正,表明与非供暖季相比,供暖季对外出游玩满意度有显著的正向影响。

## 4 结论、建议与展望

### 4.1 结论

本文采用携程旅行网的在线评论、空气质量指数和天气状况等相关数据,通过构建公众外出游玩次数与游玩满意度模型,基于大数据定量分析了空气质量对公众外出游玩次数和满意度的影响,得到如下结论:

(1)基于传统负二项回归和以正态分布、均匀分布、拉普拉斯分布为先验分布的贝叶斯负二项回归分别构建公众外出游玩次数模型,通过对比上述4种模型的检验结果,得出传统负二项回归模型的拟合程度表现更优,能更好地解释空气质量对公众外出游玩次数的影响。此外,基于有序Probit建立的公众外出游玩满意度模型拟合效果较好,表明该模型能够满足数据解释的需要。

(2)在温度、风速、节假日等控制变量不变的基础上,空气质量会显著影响公众外出游玩次数和满意度。其中,公众外出游玩次数会随着空气污染程度的增加而显著减少,表明公众为了减少在空气污染环境中的暴露风险会采取规避行为。具体而言,与空气质量为优相比,严重空气污染的影响最大,会导致外出游玩次数下降40.1%;而空气质量为良时影响最小,游玩次数会下降15.1%。此外,严重空气污染会显著降低公众外出游玩满意度,但较低程度的空气污染对公众外出游玩满意度不具有显著影响,表明公众对哈尔滨市景区的游玩体验质量并不会因较低程度的空气污染问题而下降。

#### 4.2 建议

根据上述研究结论,为了降低哈尔滨市空气污染对公众外出游玩的影响,提出以下建议:

(1)本文的研究结果证实了空气质量是影响公众外出游玩次数的主要因素,由于工业生产、加工等过程中大量煤炭燃烧增加了城市空气污染物的排放,加之秸秆焚烧过程中大量有害气体和烟雾的形成,导致公众会为了降低空气污染的暴露风险而采取一定的规避行为,更倾向于选择在空气质量优良时外出游玩。这种因空气污染而导致的旅游景区游玩人次减少会影响到第三产业的经济效益,进而减缓城市的整体发展。因此,为促进哈尔滨市旅游产业的蓬勃发展,为公众提供良好的生活环境,政府应加强《哈尔滨市燃煤污染防治条例》<sup>[35]</sup>、《黑龙江省禁止秸秆露天焚烧工作奖惩暂行规定》<sup>[36]</sup>等政策的正确引导,加大宣传力度,提高企业与广大群众对空气污染问题的认知程度,激发公众的空气质量保护意识。

(2)尽管较低程度的空气污染并未显著降低公

众外出游玩满意度,但在严重空气污染条件下,公众的外出游玩满意度已遭受严重的负面影响。因此,为遏制空气污染,切实改善空气质量,提高社会经济活力,政府应长期关注空气污染问题,认真贯彻执行国务院和地方政府印发的大气污染防治行动计划,加大执法监管力度。在未来的政策制定、实施和管理过程中,结合城市实际情况,出台更具针对性和操作性的大气污染防治地方性法规或规章。同时应加大生态环境的保护力度,提高公众的环境保护意识,减少空气污染行为,提倡文明、节约、绿色、低碳的生活方式,进而促进地区的健康可持续发展。

#### 4.3 不足与展望

互联网大数据为空气质量影响公众外出行为的研究提供了新的数据来源。相较于传统基于问卷和访谈的方法,旅游在线评论数据更加充足<sup>[51]</sup>、更具有时间连续性和客观性,可有效避免因调查时间间断而造成数据缺失,以及问卷引导方式、问卷质量等引起的数据偏差等问题。但本文仍存在以下不足:

(1)在运用网络评论时,评论覆盖的用户大部分是受过教育、年轻和可接触到网络的人群<sup>[25]</sup>,可能会产生样本偏差。同时,本文数据来源仅限于携程旅行网,尚未对全网数据(去哪儿网、途牛网等)进行采集,研究结果可能与真实情况存在差异,未来研究可以做进一步拓展。此外,虽然旅游网站可能存在虚假评论的现象,但由于携程旅行网已经制定了相关规则且采取了一定措施以有效避免虚假评论,同时考虑到虚假评论检测的困难性<sup>[38]</sup>,因此本文并未对评论的真实性进行进一步的检测。

(2)与现有文献一样,研究面临无法排除公众因某天的空气污染而取消未来数天的外出游玩活动和游玩后未在当天评论的情况,即本文无法为可能存在的时间滞后效应提供证据,未来研究需采用更为严谨的研究设计对此情况进行弥补和解释。而且,研究对象只针对哈尔滨市,研究结果是否适用于其他地区还有待验证。此外,针对公众外出游玩满意度影响因素的研究仅采用了网络评分数据,由于个体间的评分标准并不完全相同,因此如何将



2020年6月

网络评论中文本型数据与公众外出游玩满意度模型相结合,有待在今后研究中进一步探索。

### 参考文献(References):

- [1] Li Y, Guan D, Tao S, et al. A review of air pollution impact on subjective well-being: Survey versus visual psychophysics[J]. *Journal of Cleaner Production*, 2018, 184: 959–968.
- [2] Dupont A. Improving and monitoring air quality[J]. *Environmental Science and Pollution Research*, 2018, 25: 15253–15263.
- [3] Holgate S T. ‘Every breath we take: The lifelong impact of air pollution’ – a call for action[J]. *Clinical Medicine*, 2017, 17(1): 8–12.
- [4] Martinez G S, Spadaro J V, Chapizanis D, et al. Health impacts and economic costs of air pollution in the metropolitan area of Skopje[J]. *International Journal of Environmental Research and Public Health*, 2018, 15(4): 626.
- [5] 郑思齐, 张晓楠, 宋志达, 等. 空气污染对城市居民户外活动的影响机制: 利用点评网外出就餐数据的实证研究[J]. *清华大学学报(自然科学版)*, 2016, 56(1): 89–96. [Zheng S Q, Zhang X N, Song Z D, et al. Influence of air pollution on urban residents’ outdoor activity: Empirical study based on dining-out data from the Dianping website[J]. *Journal of Tsinghua University (Science and Technology)*, 2016, 56(1): 89–96.]
- [6] Chen C M, Lin Y L, Hsu C L. Does air pollution drive away tourists? A case study of Sun Moon Lake National Scenic Area, Taiwan [J]. *Transportation Research Part D-Transport and Environment*, 2017, 53: 398–402.
- [7] 刘嘉毅, 陈玉萍, 夏鑫. 中国空气污染对入境旅游发展的影响[J]. *资源科学*, 2018, 40(7): 1473–1482. [Liu J Y, Chen Y P, Xia X. Research on the effect of air pollution on the development of inbound tourism in China[J]. *Resources Science*, 2018, 40(7): 1473–1482.]
- [8] 徐冬, 黄震方, 黄睿. 基于空间面板计量模型的雾霾对中国城市旅游流影响的空间效应[J]. *地理学报*, 2019, 74(4): 814–830. [Xu D, Huang Z F, Huang R. The spatial effects of haze on tourism flows of Chinese cities: Empirical research based on the spatial panel econometric model[J]. *Acta Geographica Sinica*, 2019, 74(4): 814–830.]
- [9] 刁贝娣. 雾霾污染及其对城镇居民游憩活动的影响[D]. 武汉: 中国地质大学, 2017. [Diao B D. Fog Haze and Its Impact on Urban Residents’ Recreational Activities: A Case Study of Wuhan [D]. Wuhan: China University of Geosciences, 2017.]
- [10] Zhao P, Li S, Li P, et al. How does air pollution influence cycling behaviour? Evidence from Beijing[J]. *Transportation Research Part D: Transport and Environment*, 2018, 63: 826–838.
- [11] Jiang Y Q, Huang G L, Fisher B. Air quality, human behavior and urban park visit: A case study in Beijing[J]. *Journal of Cleaner Production*, 2019, 240.
- [12] 徐戈, 冯项楠, 李宜威, 等. 雾霾感知风险与公众应对行为的实证分析[J]. *管理科学学报*, 2017, 20(9): 1–14. [Xu G, Feng X N, Li Y W, et al. Empirical study on the perceived risk of smog and public coping behavior[J]. *Journal of Management Sciences in China*, 2017, 20(9): 1–14.]
- [13] Yang X, Pan B, Evans J A, et al. Forecasting Chinese tourist volume with search engine data[J]. *Tourism Management*, 2015, 46: 386–397.
- [14] Li X, Pan B, Law R, et al. Forecasting tourism demand with composite search index[J]. *Tourism Management*, 2017, 59: 57–66.
- [15] Lu W, Stepchenkova S. Ecotourism experiences reported online: Classification of satisfaction attributes[J]. *Tourism Management*, 2012, 33(3): 702–712.
- [16] 明雨佳, 刘勇, 周佳松. 基于大数据的山地城市活力评价: 以重庆主城区为例[J]. *资源科学*, 2020, 42(4): 710–722. [Ming Y J, Liu Y, Zhou J S. Vitality assessment of mountainous cities based on multi-source data: A case of Chongqing Municipality, China[J]. *Resources Science*, 2020, 42(4): 710–722.]
- [17] 任武军, 李新. 基于互联网大数据的旅游需求分析: 以北京怀柔为例[J]. *系统工程理论与实践*, 2018, 38(2): 437–443. [Ren W J, Li X. Tourism demand analysis on Internet big data: The case of Huairou, Beijing[J]. *Systems Engineering- Theory & Practice*, 2018, 38(2): 437–443.]
- [18] Li J J, Xu L Z, Tang L, et al. Big data in tourism research: A literature review[J]. *Tourism Management*, 2018, 68: 301–323.
- [19] 缪秀梅, 陈烨天, 米传民. 基于ISM和在线评论的汤山温泉顾客满意度研究[J]. *中国管理科学*, 2019, 27(7): 186–194. [Miao X M, Chen Y T, Mi C M. Study on consumer satisfaction of Tangshan hot springs based on ISM and online reviews[J]. *Chinese Journal of Management Science*, 2019, 27(7): 186–194.]
- [20] Bakhshi S, Kanuparth P, Gilbert E. Demographics, Weather and Online Reviews: A study of Restaurant Recommendations[C]. Seoul, Korea: International Conference on World Wide Web, 2014.
- [21] Sun B, Ao C, Wang J, et al. Listen to the voices from tourists: Evaluation of wetland ecotourism satisfaction using an online reviews mining approach[J]. *Wetlands*, 2020.
- [22] Mao B, Ao C, Wang J, et al. Does regret matter in public choices for air quality improvement policies? A comparison of regret-based and utility-based discrete choice modelling[J]. *Journal of Cleaner Production*, 2020, 254: 120052.
- [23] Hou Z P, Cui F S, Meng Y H, et al. Opinion mining from online travel reviews: A comparative analysis of Chinese major OTAs using semantic association analysis[J]. *Tourism Management*, 2019, 74: 276–289.
- [24] Ye Q, Law R, Gu B, et al. The influence of user-generated content

- on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings[J]. *Computers in Human Behavior*, 2011, 27: 634–639.
- [25] 敖长林, 李凤佼, 许荔珊, 等. 基于网络文本挖掘的冰雪旅游形象感知研究: 以哈尔滨市为例[J]. *数学的实践与认识*, 2020, 50(1): 44–54. [Ao C L, Li F J, Xu L S, et al. Image perception of ice and snow tourism based on online text mining: A case study of Harbin[J]. *Mathematics in Practice and Theory*, 2020, 50(1): 44–54.]
- [26] Guo Y, Barnes S J, Jia Q. Mining meaning from online ratings and reviews: Tourist satisfaction analysis using latent dirichlet al location[J]. *Tourism Management*, 2017, 59: 467–483.
- [27] 康恒元, 刘玉莲, 李涛. 黑龙江省重点城市 AQI 指数特征及其与气象要素之关系[J]. *自然资源学报*, 2017, 32(4): 692–703. [Kang H Y, Liu Y L, Li T. Characteristics of air quality index and its relationship with meteorological factors in key cities of Heilongjiang province[J]. *Journal of Natural Resources*, 2017, 32(4): 692–703.]
- [28] Zou Y J, Ash J E, Park B J, et al. Empirical bayes estimates of finite mixture of negative binomial regression models and its application to highway safety[J]. *Journal of Applied Statistics*, 2017, 45(9): 1652–1669.
- [29] Castrejón-Pérez R C, Borges-Yáñez S A, Irigoyen-Camacho M E, et al. Negative impact of oral health conditions on oral health related quality of life of community dwelling elders in Mexico city, a population based study[J]. *Geriatrics & Gerontology International*, 2017, 17(5): 744–752.
- [30] Luo J X, Qu Y M. Analysis of hypoglycemic events using negative binomial models[J]. *Pharmaceutical Statistics*, 2013, 12(4): 233–242.
- [31] Hilbe J M. *Negative Binomial Regression*[M]. Cambridge: Cambridge University Press, 2011.
- [32] Variyam J N, Blaylock J, Smallwood D. A probit latent variable model of nutrition information and dietary fiber intake[J]. *American Journal of Agricultural Economics*, 1996, 78(3): 628–639.
- [33] Mcfadden D. Quantitative methods for analyzing travel behaviour of individuals: Some recent developments[J]. *Cowles Foundation Discussion Papers*, 1978.
- [34] Yan L X, Duarte F, Wang D, et al. Exploring the effect of air pollution on social activity in China using geotagged social media check-in data[J]. *Cities*, 2019, 91: 116–125.
- [35] 哈尔滨市人民政府. 哈尔滨市燃煤污染防治条例[EB/OL]. (2016–06–11) [2020–03–11]. [http://www.harbin.gov.cn/art/2016/6/1/art\\_222\\_203.html](http://www.harbin.gov.cn/art/2016/6/1/art_222_203.html). [Harbin Municipal People's Government. Regulations on prevention and control of coal pollution in Harbin [EB/OL]. (2016–06–11) [2020–03–11]. [http://www.harbin.gov.cn/art/2016/6/1/art\\_222\\_203.html](http://www.harbin.gov.cn/art/2016/6/1/art_222_203.html).]
- [36] 黑龙江省生态环境厅. 黑龙江省禁止秸秆露天焚烧工作奖惩暂行规定[EB/OL]. (2018–09–17) [2020–03–11]. <http://www.hljdep.gov.cn/hbzl/jgjsqjz/mbfb/2018/09/20511.html>. [Department of Ecology and Environment of Heilongjiang Province. Interim provisions on awards and punishments for prohibiting open burning of straw in Heilongjiang province[EB/OL]. (2018–09–17) [2020–03–11]. <http://www.hljdep.gov.cn/hbzl/jgjsqjz/mbfb/2018/09/20511.html>.]
- [37] 刘逸, 保继刚, 陈凯琪. 中国赴澳大利亚游客的情感特征研究: 基于大数据的文本分析[J]. *旅游学刊*, 2017, 32(5): 46–58. [Liu Y, Bao J G, Chen K Q. Sentimental features of Chinese outbound tourists in Australia: Big-data based content analysis[J]. *Tourism Tribune*, 2017, 32(5): 46–58.]
- [38] Liu B. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*[M]. Cambridge: Cambridge University Press, 2015.

## Impact of air quality on public outings based on big data

AO Changlin, WANG Jingxia, SUN Baosheng

(Department of Management Science and Engineering, Northeast Agricultural University, Harbin 150030, China)

**Abstract:** Exploring the impact of air quality on public outings based on big data not only can enrich the practical research on air quality issues, but also have important significance for strengthening air pollution control and promoting the development of tourism industry. Taking Harbin City as the research object, this study used the data of 16,554 Ctrip online reviews, the Ministry of Environmental Protection's air quality index data, and historical weather data and constructed the number of public outings and satisfaction models based on negative binomial regression and ordered Probit regression to quantitatively assess the influences and degrees of air quality on public outings. The findings are as follows. (1) Controlling for other influencing factors such as temperature, wind speed, and holidays, air quality significantly affected the number of public outings and the satisfaction of outings. (2) The number of public outings decreased significantly with the increase of air pollution. Compared with excellent air quality, the number dropped by 40.1% when the air quality was severe, and by 15.1% when the air quality was good. (3) For the tourist attractions in Harbin City, severe air pollution significantly reduced satisfaction of public outings, while lower levels of air pollution did not significantly affect satisfaction of public outings. This research not only provides a new perspective with an accurate assessment of the influences of air quality on public outing behavior, but also provides references for relevant air pollution prevention and tourism policy formulation.

**key words:** air quality; tourism big data; outings; negative binomial regression; ordered probit regression; Harbin City