

空间抽样模型及应用对比分析

——以杭州市老年人口抽样调查为例

何晨昊, 俞垠含, 刘 干*

杭州电子科技大学经济学院, 浙江 杭州

收稿日期: 2023年1月17日; 录用日期: 2023年2月13日; 发布日期: 2023年2月28日

摘 要

日前数字化浪潮席卷全国, 人口老龄化程度不断加深, 数字鸿沟问题更加尖锐, 基于杭州成为数字化优势城市的考虑, 本文对杭州市老年人口刷脸服务的使用现状开展抽样调查, 希望以刷脸服务为切入点探究数字鸿沟问题。考虑到空间异质性和空间自相关性, 利用三明治空间模型进行抽样布点, 布点因素有老年人口密度和县级行政区人均GDP, 并与简单随机抽样和空间分层抽样进行比较, 最终得出以下三点结论: 1) 在缺乏先验信息的情况下, 使用简单随机抽样得出的结果较为合理。当具有一定量的先验信息时, 空间分层抽样可以提高抽样布点的精度。2) 从结果来看, 综合因素分层抽样和三明治空间抽样较为相似, 影响因素的分层很大程度影响了抽样的布点, 重要程度越高的层级得到的样本点越多, 得到的调查信息越丰富。3) 人的经验知识等先验信息不同会导致分层不同, 可以使样本集中在更加有调查意义的区域, 故分层前的先验信息尤为重要。

关键词

多目标抽样, 空间抽样, 刷脸服务

Spatial Sampling Model and Application Comparative Analysis

—Taking the Sample Survey of the Elderly Population in Hangzhou as an Example

Chenhao He, Yinhan Yu, Gan Liu*

School of Economics, Hangzhou Dianzi University, Hangzhou Zhejiang

Received: Jan. 17th, 2023; accepted: Feb. 13th, 2023; published: Feb. 28th, 2023

*通讯作者。

Abstract

Recently, the digital wave has swept the country, the aging of the population has been deepening, and the digital divide is more acute. Based on the consideration of Hangzhou's digital advantage, this paper conducts a sample survey on the use status of face brushing services for the elderly population in Hangzhou, hoping to explore the digital divide problem from all the starting points of face brushing services. Considering the spatial heterogeneity and spatial autocorrelation, the sandwich spatial model was used to sample the distribution points, and the distribution factors were the density of the elderly population and the per capita GDP of the county-level administrative districts, and compared with simple random sampling and spatial stratified sampling, the following three conclusions were finally obtained: 1) In the absence of prior information, the results obtained by simple random sampling are reasonable. When there is a certain amount of prior information, spatial layered sampling can improve the accuracy of sampling distribution points. 2) From the results, the stratified sampling of comprehensive factors and sandwich space sampling are similar, and the stratification of influencing factors greatly affects the distribution of sampling, and the higher the importance level, the more sample points are obtained, and the richer the survey information obtained. 3) Different prior information such as human experience knowledge will lead to different stratification, which can concentrate the sample in a more investigative area, so the prior information before stratification is particularly important.

Keywords

Multi-Objective Sampling, Spatial Sampling, Face Brushing Service

Copyright © 2023 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

当前社会老龄化问题日益严重,根据第七次全国人口普查数据,截止2020年中国65周岁及以上的老年人口总人口比重达13.5%。老龄化问题给我国经济社会带来了深刻的影响[1][2],其中老年人“数字鸿沟”问题是众多问题中的难点[3]。基于此,将老年人作为研究主体,通过调查老年人的刷脸服务现状[4],可以作为探究解决数字鸿沟问题的突破口。目前,在数字基础设施、数字产业、公共服务数字化变革和数字经济生态环境的四大领域中杭州优势持续领先,位列第一[5],故本次抽样调查地点选取杭州市[6],并采取问卷调查方式进行调查。

传统的抽样调查一般只针对单个目标,而社会问题的调查研究往往受到多个方面因素的影响,单目标抽样是指在一次调查中估计目标只有一个,即利用一套样本只估计总体的一个指标。传统的单目标抽样难以精确满足调查者的需求,此时就需要进行多目标抽样[7]。多目标抽样,也称多主题抽样、多指标抽样,它是在一次调查中估计目标有两个或两个以上,即利用一套样本估计总体的两个或两个以上的指标[8]。本次调查涉及老年人口的年龄、退休前收入及职业、刷脸服务的参与程度和满意度等指标。

要得到多目标的调查信息,样本的布点就是需要首先解决的问题。本次调查采用综合因素分层的方法,使用空间分层抽样和三明治空间抽样进行样本布点[9],并与简单随机抽样进行对比,不断改善现实中遇到的抽样问题。因此,开展对社会多目标抽样调查,在空间抽样理论上,解决抽样过程中遇到

的问题具有重要的实际应用价值[10] [11] [12] [13]。

2. 研究区域与数据来源

2.1. 研究区域

杭州市位于长江三角洲南沿和钱塘江流域，地形复杂多样。杭州市西部属于浙西丘陵区，东部属于浙北平原地带，地势低平，河流交织，湖泊密布，物产资源丰富，具有典型的“江南水乡”特征。截至2020年，杭州市总面积16850平方千米，常住人口1196.5万人。本文以杭州市作为研究范围，行政区划界见图1。

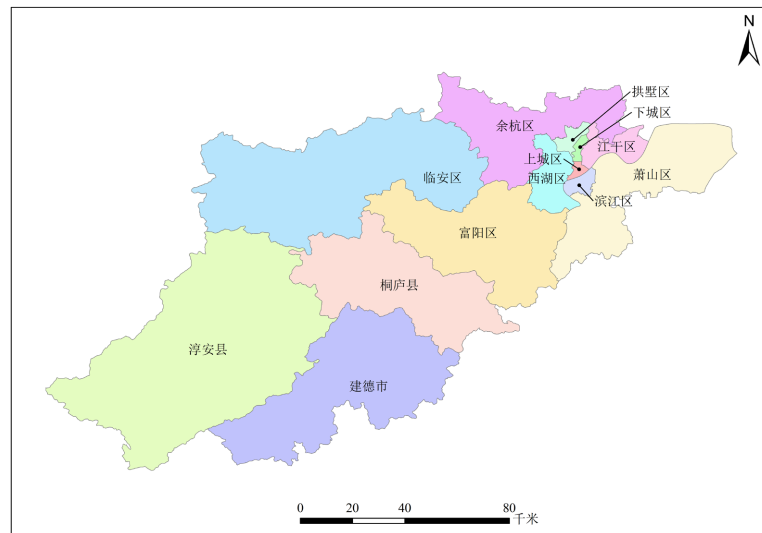


Figure 1. Diagram of overview of the study area
图 1. 研究区域概况图

2.2. 数据来源

研究项目开始时间为2020年，涉及到的杭州市人均GDP、老年人口数据、杭州市各县级行政区土地面积来自于2020年杭州统计年鉴，空间单元数据采取于杭州市2020年行政区划图。

3. 空间抽样方法与应用

本次研究以杭州市行政区划图为基础，以2020年杭州统计年鉴的老年人口数据制作出杭州市的老年人口密度分级图，同时考虑杭州市人均GDP数据制作出人均GDP分级图，利用ArcGIS软件制作出综合因素分层图[14]。

ArcGIS是目前世界上使用最广泛的GIS系统，功能有空间分析、制图可视化、3D GIS、实时GIS、遥感图像管理、处理、分析及分享，空间数据存储及管理。根据杭州市行政区划图，导入杭州市老年人口分级数据、杭州市人均GDP分级数据及综合因素分层数据，将结果可视化，展示分级结果。

空间抽样使用王劲峰的三明治空间抽样模型，该模型由报告层、知识层和样本层相互独立的3层组成[15]，对各层进行综合压缩后，根据用户需求进行样本点的布设，并利用了空间上抽样对象的变异特征，最终根据抽样对象的特征完成抽样，并与空间简单随机抽样和空间分层抽样进行对比。空间抽样的流程图如图2所示。

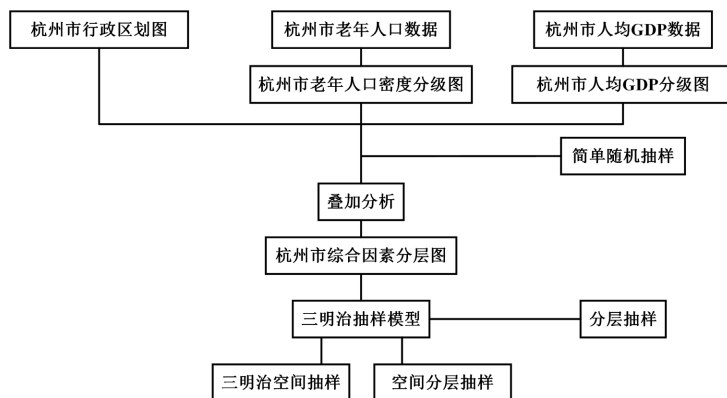


Figure 2. Flow chart of spatial sampling
图 2. 空间抽样流程图

由上图，本次研究以杭州市作为布点区域，通过杭州市老年人口数据制作出老年人口密度分级图，通过杭州市人均 GDP 数据制作出人均 GDP 分级图，结合以上两个因素进行叠加分析，制作出综合因素分层图，以分层抽样作为理论依据分别进行空间分层抽样和三明治空间抽样，并结合空间简单随机抽样进行对比。

3.1. 抽样布点各因素分析

根据杭州市老年人口对刷脸服务的调查内容，选择两个因素：老年人口密度和人均 GDP。抽样的主体对象是老年群体，故每一个地区的老年人口数量则是反映空间特征的关键因素之一。同时各地区的人均 GDP 会导致刷脸设备的齐全与否，从而成为影响老年人口使用刷脸服务的客观限制因素。故将二者作为抽样调查布点因素。

3.1.1. 空间抽样布点因素

1) 老年人口密度分级

通过 2020 年杭州市统计年鉴的数据统计，各县级行政区的老年人口密度从 23.89 人/km² 到 3963.24 人/km²，使用 ArcGIS 软件制作出杭州市老年人口密度分级图，如图 3 所示，可以反映各地区人口密度的

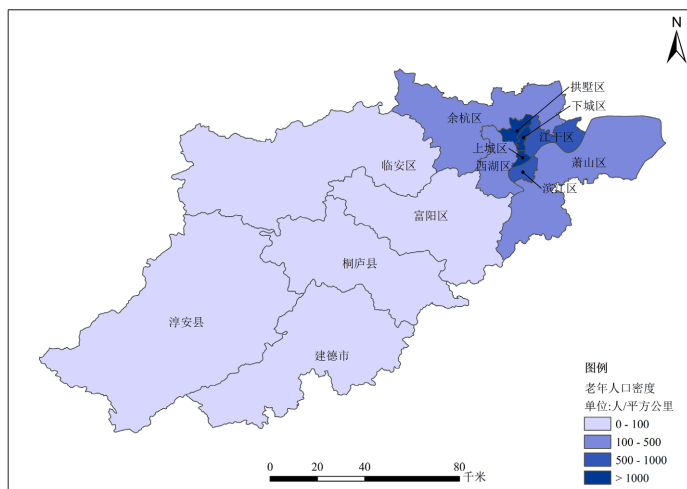


Figure 3. Diagram of division of elderly population density in Hangzhou
图 3. 杭州市老年人口密度分级图

变化,人口密度经常用来研究人口的分布。由于各国家、各地区的人口密度相差比较大,目前尚未有统一的人口密度划分标准,故按照数据的信息将其分为4级,如图3所示,颜色越深表明老年人口密度越大,杭州市的老年人口主要集中在拱墅区、上城区和下城区附近。老年人口密度小于100的区域则是淳安县和临安区等地。

2) 人均 GDP 分级

人均 GDP,可以反映各地区的居民收入和生活水平。老年人口的刷脸服务使用频率,与外部的科技因素有相关关系,而科技的发展程度又与人均 GDP 有关联。根据2020年杭州市统计年鉴计算出各县级行政区的人均 GDP,从5.55万元/人到36.72万元/人不等。对于具体人均 GDP 数据进行相应分级,共4个等级,如图4所示,滨江区的居民收入和生活水平较高,而在淳安县和建德市等县级行政区的人均 GDP 则较低,颜色越深代表人均 GDP 的数值越高。

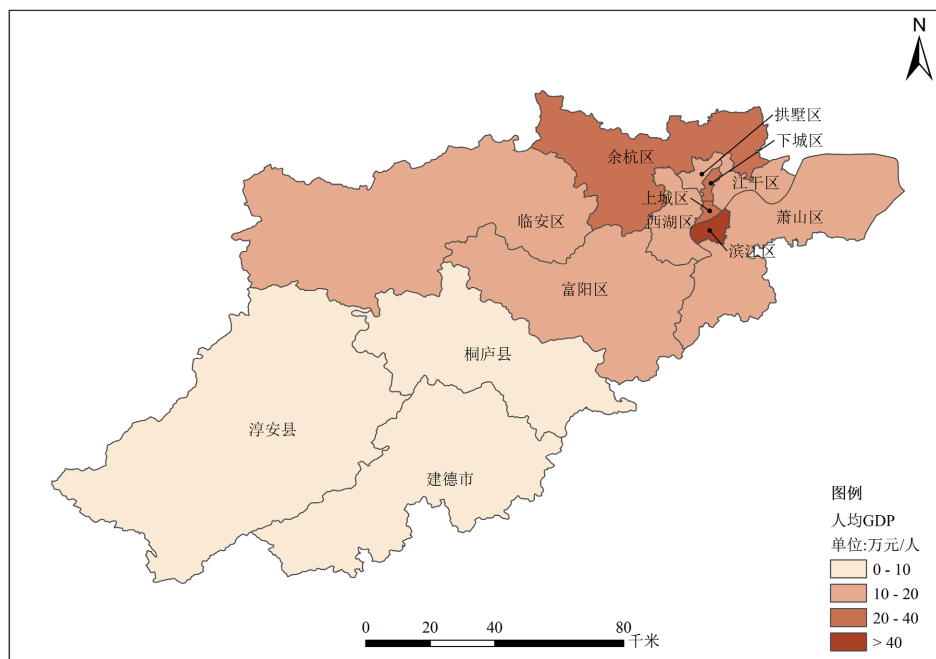


Figure 4. Diagram of division of GDP per capita in Hangzhou

图4. 杭州市人均 GDP 分级图

3) 综合因素分层

在 ArcGIS 软件的支持下,我们将杭州市各行政区的老年人口密度和人均 GDP 进行叠加,根据区域的复杂程度进行重新分类。首先,由于老年人口密度分布图与人均 GDP 分布图都是以杭州市行政区划图作底图进行分级的,所以非常利于设定重要程度参数进行两个因素的不同重要程度的叠加,这里的重要程度是指同一个区老年人口密度和人均 GDP 的分布状况,我们会给老年人口密度越高、人均 GDP 越高的区分别赋予越大的重要程度参数,最后通过叠加重要程度,进行重新分层。同时考虑到重要程度越高的地方更有利于我们得到完整的信息,所以在保留样本代表性的前提下,当两个重要程度不同的区进行叠加的时候,我们将会倾向于将它的新的重要程度定为两个重要程度中较高的一个重要程度,具体重要程度叠加如表1所示。最终制作出综合因素分层分布图,如图5,滨江区、拱墅区、上城区和下城区的重要程度高,而桐庐县、淳安县和建德市的重要程度相对较低。

Table 1. Hierarchical combination of the integrated factors

表 1. 综合因素分层表

老年人口密度 \ 人均 GDP	1	2	3	4
1	1	-	-	-
2	2	2	3	4
3	-	3	-	4
4	-	-	4	-

注：数字越大代表重要程度越高，数字 2 代表地区重要程度比数字 1 代表地区重要程度高，以此类推；“-”表示在地图上未出现这两种因素等级同时存在的地块。

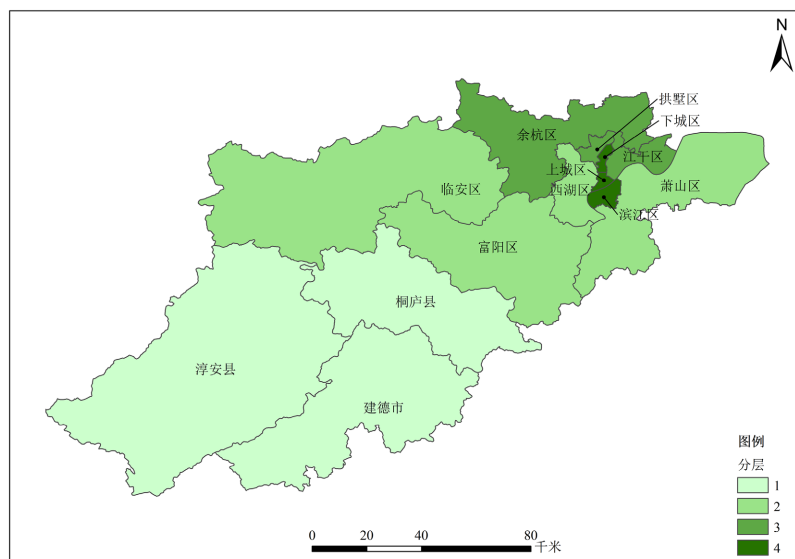


Figure 5. Diagram of hierarchical combination of the integrated factors

图 5. 综合因素分层图

3.2. 空间抽样理论方法

3.2.1. 空间简单随机抽样

根据空间简单随机抽样就是在地理空间上等概率地抽取若干样本单元。样本单元既可以是一个点，也可以是一个行政单元，或是一个样方。

3.2.2. 空间分层抽样

根据 Tobler 第一定律，距离越近的对象，其相似程度就越高。传统的分层抽样方法不考虑样本的空间坐标信息，可能会使同一层的对象相距较远，甚至被空间上其他层隔开。故空间分层抽样不仅要求层内方差小、层间方差大，还要求同一层的样本在空间上连在一起。研究区域均值和方差如下：

$$\bar{Z} = (1/n) \sum_{h=1}^L n_h \bar{Z}_h \tag{1}$$

$$V(\bar{Z}) = \sum_{h=1}^L W_h^2 V(\bar{Z}_h) \tag{2}$$

式中， n_h 是第 h 层的样本个数； \bar{Z}_h 是第 h 层的样本均值； L 为分层个数； W_h 为权重； $V(\bar{Z}_h)$ 是每个层内部的均值方差。

3.2.3. 三明治空间抽样

为了消除对象的空间异质性,将研究区域划分为多个分区,构成知识层,这里的层不仅限于地理的分类或分区,也可以按属性划分。再根据用户对抽样精度的要求计算各层的样本量,并分配到各知识层,计算每个知识层的样本均值和方差。最后将报告层与知识层叠加相切,将均值和方差传递到报告层,完成样本层、知识层和报告层的信息及误差传递。

3.3. 空间抽样方法的结果分析

在实际的抽样调查中我们需要权衡经济成本和抽样精度这两个因素,经济成本一般表达为:

$$\text{Cost} = c_0 + n \times c_1 \quad (3)$$

其中, c_0 为固定成本(或基本花费), n 是样本量, c_1 为平均每个样本的花费[16]; 样本量越大, 抽样精度越高[17]。根据实际经济条件, 设置每一个调查点为 100 元预算, 最终确定最大样本容量为 36。我们通过各层之间的方差相对关系最终确定方差, 从而完成抽样。如图 6~8 所示, 分别进行简单随机抽样、综合因素分层抽样和三明治空间抽样, 进行样本点对比:

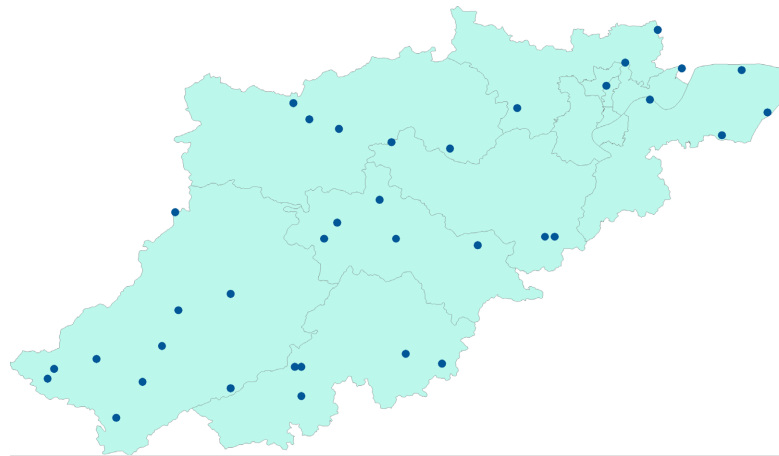


Figure 6. Diagram of simple random sampling
图 6. 简单随机抽样

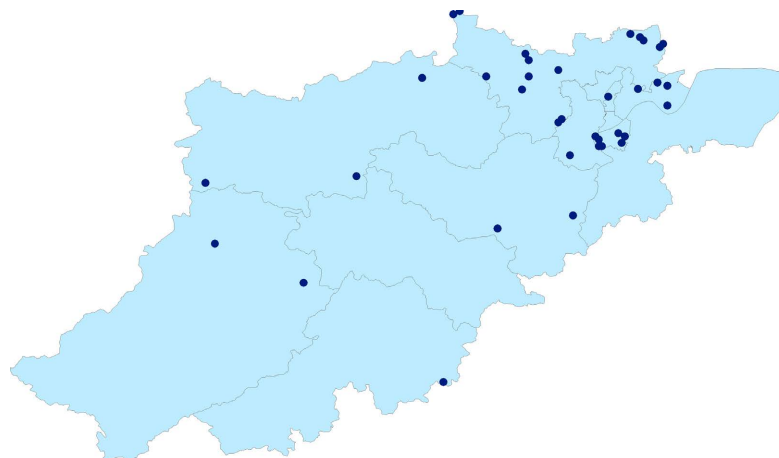


Figure 7. Diagram of hierarchical sampling combination of the integrated factor
图 7. 综合因素分层抽样

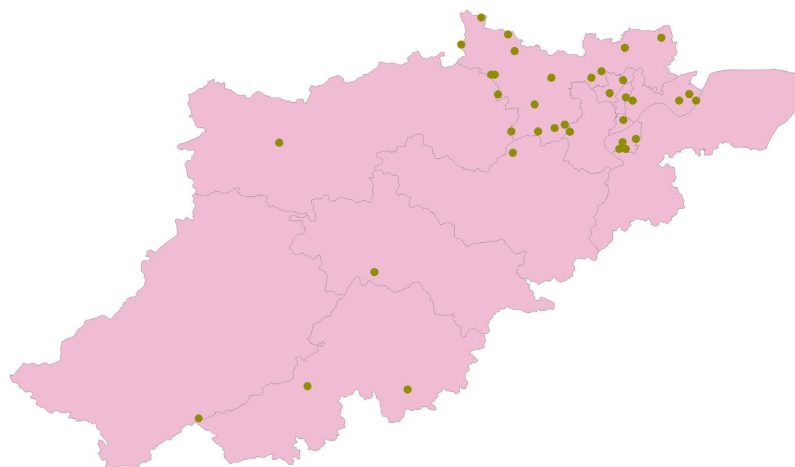


Figure 8. Diagram of sandwich space sampling
图 8. 三明治空间抽样

1) 简单随机抽样的样本量为 36, 其结果如图 6 所示, 抽样的过程中不涉及到分层, 总体而言, 样本点的分布较为均匀, 在杭州市的西南地域和中部地区布点较为密集。在分级为 1 的地区, 样本量为 19, 可见在没有分层的先验信息约束下, 样本的随机性较大, 在重要程度最低的地区布点超过 50%, 如具体实施抽样调查会产生较大误差。

2) 通过综合因素分层抽样, 样本量为 36, 用户给定抽样的方差为 0.01, 其结果如图 7 所示, 样本点的分布较不均匀, 集中在杭州市的东北地区, 而在分级为 1 的部分, 重要程度较低, 样本点的分布少, 共有 3 个样本点, 重要程度分级为 3 及以上的地区共占比 75%, 可见拥有先验信息与没有先验信息抽样有明显区别。整体而言, 样本点的分布与分层的关联程度大, 分层级数高的样本点较多, 分层级数较低的地域样本点较少。

3) 三明治空间抽样是在分层抽样的基础上进行的, 显著性水平 α 为 0.05, 知识层均值的空间标准差为 9.65, 其结果如图 8 所示。因三明治空间抽样同样用到综合因素分层图, 故样本点的分布与综合因素分层抽样较为类似, 主要集中在杭州市的东北区域, 重要程度越高, 相应的样本点就越多, 在分级为 3 及以上的地区, 共有 29 个样本点, 占比 80.55%, 略高于综合因素分层抽样。

总的来说, 简单随机抽样的样本点分布较为均匀, 随机性强, 综合因素分层抽样和三明治空间抽样的结果较为相似, 都受综合因素分层影响较大。分层赋予的重要程度越高也就是可调查性越强, 区域内的样本点越多。

4. 结论

本文对杭州市老年人口刷脸服务使用现状的抽样调查, 阐释了多目标抽样在社会调查中的应用, 选择空间抽样方法中的三明治空间抽样。根据先验知识和实际情况, 分析对比简单随机抽样、综合因素分层抽样和三明治空间抽样三种抽样方式, 得到以下结论:

1) 在缺乏先验信息的情况下, 使用简单随机抽样得出的结果较为合理。当具有一定量的先验信息时, 再使用简单随机抽样会存在较大误差, 而空间分层抽样和三明治空间抽样方法可以提高抽样布点的精度, 使得样本点的代表性增强。

2) 对于此次杭州市老年人口刷脸服务的调查问卷布点问题, 从综合因素分层抽样和三明治空间抽样的结果来看, 两种空间抽样方法较为相似, 分层很大程度影响了抽样的布点, 重要程度越高的层级得到

的样本点越多,得到的调查信息越丰富。相较而言,三明治空间抽样更能消除空间异质特性,在重要程度高的布点数高于空间分层抽样。

3) 空间分层抽样,是地理调查经常采用的一种方法。人的经验知识等先验信息不同会导致分层不同,分层的结果会极大影响抽样的结果,可以使样本集中在更加有调查意义的区域,故分层前的先验信息尤为重要。

在本次的社会问卷调查中,抽样分层仍带有部分主观先验知识,涉及到的社会经济问题通过设置参数控制。综合运用各种目标,建立一个新模型来衡量各层之间的关系,得出更为精确的抽样布点方法,是我们下一步研究的方向。

参考文献

- [1] 王志宝,孙铁山,李国平. 近 20 年来中国人口老龄化的区域差异及其演化[J]. 人口研究, 2013, 37(1): 66-77.
- [2] 穆光宗,张团. 我国人口老龄化的发展趋势及其战略应对[J]. 华中师范大学学报(人文社会科学版), 2011, 50(5): 29-36.
- [3] 邱泽奇,张树沁,刘世定,许英康. 从数字鸿沟到红利差异——互联网资本的视角[J]. 中国社会科学, 2016(10): 93-115+203-204.
- [4] 平安养老险推“智能养老”服务“刷脸”让养老金领取更简单[J]. 山东人力资源和社会保障, 2017(10): 63.
- [5] 张一博. 数字经济背景下长春市人才发展对策研究——基于上海、杭州数字经济人才发展的分析[J]. 中国市场, 2022(6): 174-175.
- [6] 邵雨莲. 杭州市老年人口空间分布及演化规律研究[D]: [硕士学位论文]. 杭州: 浙江大学, 2014.
- [7] 冯士雍. 抽样调查应用与理论中的若干前沿问题[J]. 统计与信息论坛, 2007, 22(1): 5-13.
- [8] 杨光辉. 对多目标抽样的再认识[J]. 统计与决策, 2013(8): 32-35.
- [9] 郝一炜,刘晓宇,金勇进. 空间双重平衡抽样设计及其应用? [J]. 调研世界, 2019(3): 26-31.
- [10] 高丽玲,李新虎,王翠平,邱全毅,崔胜辉,赵千钧. 空间抽样的理论方法与应用分析——以厦门岛问卷调查为例[J]. 地球信息科学学报, 2010, 12(3): 358-364.
- [11] 冯希莹,王星. 空间定位抽样技术:民意调查中的新方法[J]. 社会工作(下半月), 2010(7): 59-61.
- [12] 姜成晟,王劲峰,曹志冬. 地理空间抽样理论研究综述[J]. 地理学报, 2009, 64(3): 368-380.
- [13] 曹志冬,王劲峰,李连发,姜成晟. 地理空间中不同分层抽样方式的分层效率与优化策略[J]. 地理科学进展, 2008(3): 152-160.
- [14] 闫庆武,卞正富. 基于 GIS 的社会统计数据空间化处理方法[J]. 云南地理环境研究, 2007(2): 92-97.
- [15] 刘铁军,王劲峰,邹艳丽. 三明治抽样模型在中国农村住区现状调查中的应用[C]//中国地理学会. 地理学核心问题与主线——中国地理学会 2011 年学术年会暨中国科学院新疆生态与地理研究所建所五十年庆典论文摘要集. 2011: 210.
- [16] 李连发,王劲峰,刘纪远. 国土遥感调查的空间抽样优化决策[J]. 中国科学(D 辑: 地球科学), 2004(10): 975-982.
- [17] 刘建红,朱文泉. 耕地变化空间抽样调查方案的精度与效率分析[J]. 农业工程学报, 2010, 26(10): 331-336.