

## INVESTIGATING POPULATION BALANCE IN THE STATE OF GEORGIA USING SPATIAL CLUSTERING

ALI ABOLHASSANI<sup>id</sup>\* AND SOMAYYEH TARI<sup>id</sup>

**ABSTRACT.** Identification of population ratio disruption in the population structure is one of the challenges that every country faces. Population aging is a kind of demographic abnormality that lack of attention causes population problems. A timely warning about the aging of society can be useful for planning about having children on the one hand and providing suitable facilities for the elderly on the other hand. For example, the capital of Japan is a good example of an urban environment suitable for the elderly.

One of the anomaly detection tools is spatial clustering using scan statistics. In the last three decades, the scan statistic method has been a very important and active field in statistical research. Identifying areas on geographic maps, where the concentration of points (elderly, sick, criminals, certain animal species, etc.) is significant, is important in many fields such as epidemiology, politics, criminology, zoology, and so on. With the help of scan statistic method, spatial clusters can be identified. In this article, we introduce the scan statistic method based on Poisson distribution. Using simulation, we investigate the efficiency of this method in identifying spatial clusters. Based on the results obtained from the simulation, the Poisson scan statistic method is a suitable method for detecting anomalies in the count spatial data. As an application of spatial clustering, we consider the population structure of the state of Georgia and identify areas where the elderly population is significantly high. These areas should be prioritized in the implementation of population reform programs.

---

Keywords: Spatial clustering, Scan statistic, Monte-Carlo hypothesis testing, Poisson distribution, Likelihood ratio

Article Type: Research Paper.

Communicated by Afshin Parvardeh.

\*Corresponding author.

Received: 12-06-2023, Accepted: 27-04-2024, Published Online: 26-11-2024.

This work has been financially supported by Azarbaijan Shahid Madani University under the grant number 906/1402.

**Cite this article:** A. Abolhassani and S. Tari, Investigating population balance in the state of Georgia using spatial clustering, *Journal of Mathematics and Society*, **9** no. 3 (2024) 107–124.

<http://dx.doi.org/10.22108/msci.2024.138014.1585> .



## 1. Introduction

The outbreak of diseases such as COVID-19 has significantly impacted communities, with some cities experiencing widespread transmission while others have lower prevalence rates. Identifying regions with a high density of infected individuals that hold statistical significance is crucial. Timely alerts to governmental and health authorities about the existence of such areas can effectively prevent human casualties and alleviate the burden of additional healthcare costs on the government. Pinpointing these areas may even yield insights into the causative factors of the disease. For instance, during the cholera epidemic, Dr. John Snow mapped the geographical locations of individuals who succumbed to the disease in London and discerned a correlation between mortality due to cholera and contamination of drinking water sources, as most individuals affected were in proximity to these water sources [30]. This paper presents the map of cholera in London in Figure 1 which is generated using R software. Additionally, interested readers can refer to [20] to view the original map designed by Snow.

While mortality rates may be higher around water sources, such concentrations may occur randomly. Hence, investigating the significance of these clusters is imperative. When population density with a specific characteristic (such as being infected, elderly, or obese) in an area is statistically significant, we can infer spatial clustering has occurred. One statistical tool for detecting these areas is the scan statistic method, introduced by Kulldorff [18].

Identifying spatial clusters is not only vital in epidemiology but also applicable across various research domains: astronomy [19], [5], [9], [16], image analysis [12], [34], [4], criminology [22], [29], [14], [8], [10], ecology [31], [23], geography [24], [6], [21], [28], pattern recognition [13], biology [11], [25], epidemiology [27], [17], [7], [33], public transportation [32], [3], [15]. A comprehensive review of the scan statistic method is provided in [2].

## 2. Main Results

Kulldorff and Nagarwalla [18] introduced the Likelihood Ratio Test (LRT) method for spatial cluster detection. In this method, a two-dimensional window scans the study area map, and the window size changes during scanning. After scanning the map, spatial clusters of *points* or *cases* are examined based on a statistical hypothesis test introduced in equation 2.1. It is worth mentioning that each member of the population with a specific attribute is called a *point* or *case* and the term *control* is applied to members of the population that lack that specific attribute.

Let's assume that the study map has  $I$  cities or cells, and for the  $i$ th city,  $i = 1, \dots, I$ , the number of cases and the population are denoted by  $x_i$  and  $n_i$ , respectively. Also, suppose the number of cases in each city follows a Poisson distribution, i.e.,  $X_i \sim Pois(n_i\theta_i)$ . Each sub-area  $Z$  of the map can be considered as a candidate cluster. Thus, the class of candidates will have  $2^{|I|} - 1$  members, where  $|I|$  is the cardinality of the set  $I$ .

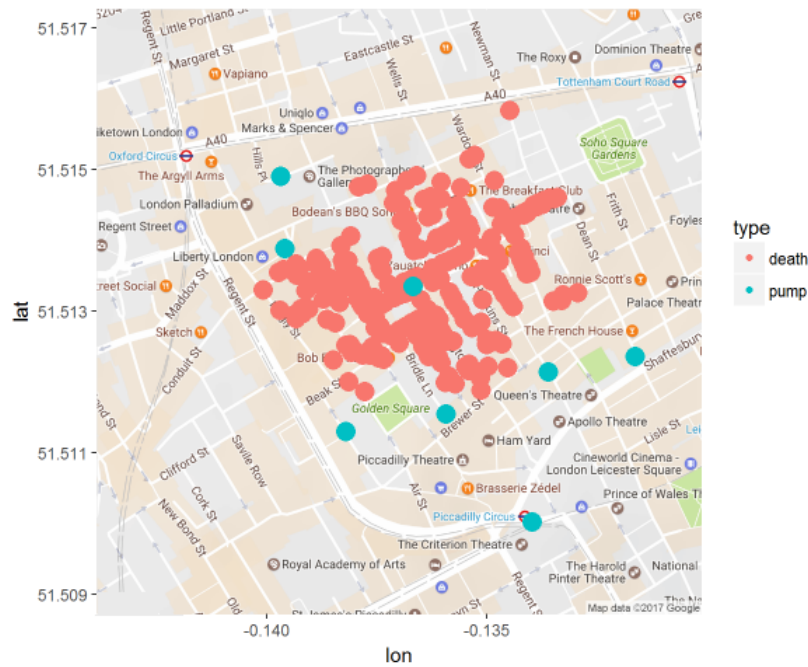


FIGURE 1. The geographic location of water pumps (blue) and cholera patient fatalities (red) on the map of London in 1856.

The class containing all possible candidates is denoted by  $\mathcal{Z}$ . The total number of cases in  $Z$  is  $X_Z = \sum_{i \in Z} X_i$ . Similarly,  $X_{\bar{Z}} = \sum_{i \in \bar{Z}} X_i$  represents the total number of cases in  $\bar{Z}$ , the complement of  $Z$ . The total population in areas  $Z$  and  $\bar{Z}$  are  $n_Z = \sum_{i \in Z} n_i$  and  $n_{\bar{Z}} = \sum_{i \in \bar{Z}} n_i$ , respectively. Thus, the total population on the map is  $n = n_Z + n_{\bar{Z}}$ . Suppose  $\theta_i = \theta_Z$  for each area  $i \in Z$ , and similarly,  $\theta_i = \theta_0$  for each area  $i \in \bar{Z}$ . To find the cluster location on the map, a hypothesis test

$$(2.1) \quad H_0 : \theta_Z = \theta_0, \forall Z \in \mathcal{Z} \quad v.s. \quad H_1 : \exists Z \in \mathcal{Z}, \theta_Z > \theta_0$$

is conducted.

Each member of  $\mathcal{Z}$  can be considered a potential cluster. Since the actual cluster location is unknown,  $Z$  is treated as a parameter. The likelihood function is

$$(2.2) \quad L(Z) = L(Z, \theta_0, \theta_Z) = \left[ \prod_{i \in Z} \frac{e^{-n_i \theta_Z} (n_i \theta_Z)^{x_i}}{x_i!} \right] \left[ \prod_{j \notin Z} \frac{e^{-n_j \theta_0} (n_j \theta_0)^{x_j}}{x_j!} \right].$$

Hence, the likelihood ratio is

$$(2.3) \quad \lambda(Z) = \frac{\sup_{\theta_Z > \theta_0} L(Z, \theta_Z, \theta_0)}{\sup_{\theta_Z = \theta_0} L(Z, \theta_Z, \theta_0)}.$$

The scan statistic is defined as  $\lambda = \sup_{Z \in \mathcal{Z}} \lambda(Z)$ .

To find the spatial cluster,  $\lambda(Z)$  is calculated for each  $Z \in \mathcal{Z}$ . The next step is to determine the region  $Z$  that maximizes  $\lambda(Z)$ , known as the Most Likely Cluster (MLC). Since the distribution of



the scan statistic is unknown, hypothesis testing is performed using the Monte Carlo method. This method is described as follows:

To perform the test using the Monte Carlo method, the number of cases under the null hypothesis in hypothesis testing 2.1 is first generated for each city. The value of  $\lambda(Z)$  is then calculated based on the simulated data to determine the MLC. This process is repeated 499 times, and the scan statistic is calculated in each iteration. Let the computed values be  $\lambda_i, i = 1, 2, \dots, 499$ . These values are sorted in ascending order, i.e.,  $\lambda_{(1)} \leq \lambda_{(2)} \leq \dots \leq \lambda_{(499)}$ . The value of  $\lambda_R$  for the actual data is placed among the sorted values. The  $p$ -value is then calculated as  $\frac{1 + \sum_{i=1}^{499} I(\lambda_i > \lambda_R)}{500}$ , where  $I()$  is the indicator function. It should be noted that the detection of circular and non-circular cluster are discussed in [1].

Five criteria are introduced to measure the efficiency of spatial clustering methods: Recall, Precision, F1 score, Biasness, and Power:

$$\text{Recall} = \frac{\#(\text{Identified Cluster} \cap \text{True Cluster})}{\#(\text{True Cluster})}$$

$$\text{Precision} = \frac{\#(\text{Identified Cluster} \cap \text{True Cluster})}{\#(\text{Identified Cluster})}$$

where  $\#(A)$  denotes the number of members of set  $A$ . The F1 score is the harmonic mean of Precision and Recall and is defined as:

$$F1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Biasness can also be a measure for comparing clustering methods. Prates et al. [26] discussed Biasness in clustering methods. It is noteworthy that the closer these metrics are to one, the better the method has identified the true cluster with fewer errors. Therefore, one is the optimal value for these metrics.

Another criterion for assessing the effectiveness of clustering methods is the power of the test. It is the probability of correctly rejecting the null hypothesis. If a method used for clustering has lower power, it means that it is less likely to detect a cluster in a map containing a cluster.

### 3. Simulation, real application, and conclusions

In this section, we delve into simulation studies considering various scenarios. Initially, we consider a map similar to Figure 2. Each cell of the map includes the number of patients, the total number of individuals in that cell, and the geographical coordinates of its center. We utilize a Poisson distribution to simulate the number of patients in each cell. In the first scenario, the number of patients inside the purple region (cluster location) is generated from a  $Pois(24)$ . Outside this region, the number of patients is generated from a  $Pois(20)$ . We perform cluster detection according to the method described in Section 2 and compute the bias, power, precision, recall, and F1 score. We repeat the simulation of maps under the null hypothesis 499 times. In each iteration, we calculate the introduced metrics.

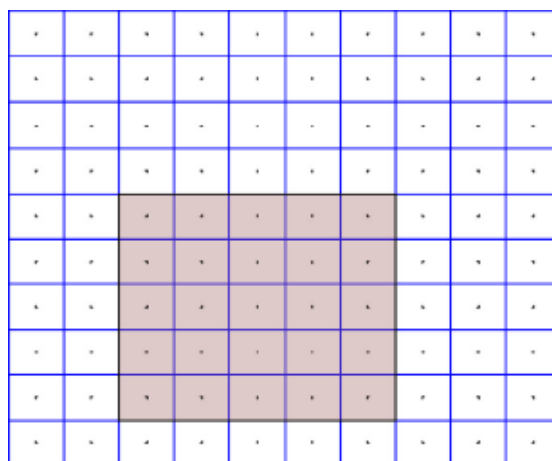


FIGURE 2. A map with 100 cells and a spatial cluster (purple area).

The results related to this scenario are presented in Table 1, in the first column from the right. As observed, the test power is low, around 0.19. This means that in this scenario, the presence of a cluster is only alerted by the scan statistic in 19% of the repetitions. On the other hand, bias, recall, precision, and F1 score deviate from the optimal value of 1, indicating that the identified clusters track the true cluster with more errors.

In the second scenario, the parameter inside the cluster is set to 30, and the parameter outside the cluster is set to 20. Similar to the previous stage, we perform cluster detection and calculate the five introduced metrics. We expect these metrics to be closer to the optimal value compared to the previous scenario. The results are shown in Table 1, in the middle column. Not only has the test power increased from 0.19 to 0.93, but also the values of other metrics have become closer to the optimal value. In other words, the presence of a cluster is alerted in 93% of the repetitions, and the actual cluster location is more accurately tracked.

In the final scenario, we consider the parameter value of 60 for the Poisson distribution inside the cluster and 20 for the distribution outside the cluster. The results of cluster detection are presented in Table 1, in the leftmost column. In this scenario, in all iterations, a cluster presence alert is issued (test power equals 1), and the identified cluster exactly matches the true cluster.

To work on real data, we use the `georgia` dataset available in the `GISTools` package in the R software. This dataset includes 14 variables, including the geographical longitude and latitude of each region, its population, the elderly population in each region, and so on. The map of the state of Georgia, which consists of 159 regions is in Figure 3.

TABLE 1. Simulation results for three different scenarios

	<i>Pois</i> (20) vs <i>Pois</i> (60)	<i>Pois</i> (20) vs <i>Pois</i> (30)	<i>Pois</i> (20) vs <i>Pois</i> (24)
Bias	1.01	1.05	1.1752
Recall	1	1	0.81
Precision	1	1	1.9
Power	1	0.93	0.19
F1	1	1	0.83

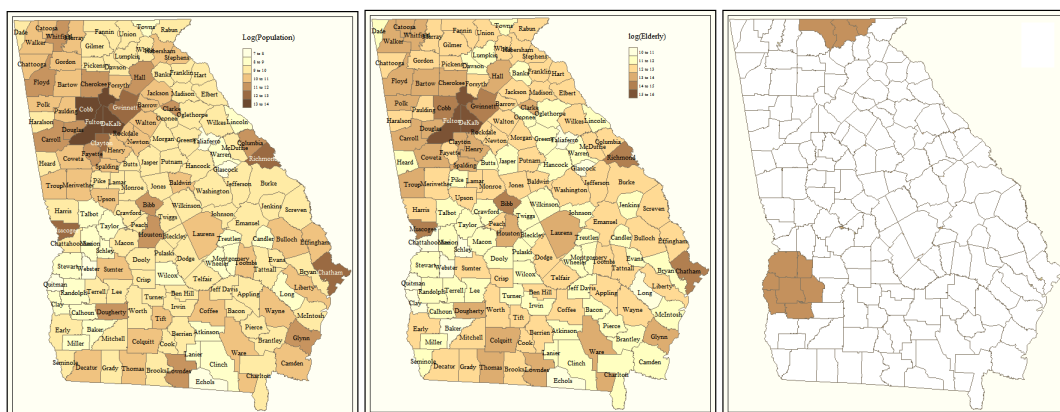


FIGURE 3. The map of the state of Georgia. Left figure: Color coded based on population count, middle figure: Color coded based on the elderly rate, right figure: Location of identified clusters.

We intend to examine which regions have a higher population density of the elderly and whether this density is statistically significant. Using scan statistic method, two clusters are identified in the western and northern regions of Georgia. In the identified areas, the elderly population is higher than in other areas. More precisely, these areas are relatively older compared to other areas. Therefore, it is recommended that these areas be prioritized in population planning, and appropriate measures be taken in terms of urban infrastructure for the elderly and providing conditions for fertility for young people.

The final conclusion of the article is as follows: population aging is a demographic anomaly that can lead to population-related problems if not addressed. Timely warnings about the aging of society can be useful for planning both in terms of fertility and providing suitable facilities for the elderly. Currently, Iran is also experiencing the transition phase of the age structure of the population from youth to elderly, and it seems that by the beginning of the fifteenth century (Solar Hijri), we are facing an increase in the elderly population in the country. Therefore, it is necessary to issue timely warnings

to control population aging and provide appropriate urban and welfare facilities for this segment of society through thoughtful measures.

**Ali Abolhassani**

Department of Mathematics, Azarbaijan Shahid Madani University, Tabriz, Iran

Email: [ali.abolhassani@azaruniv.ac.ir](mailto:ali.abolhassani@azaruniv.ac.ir)

**Somayyeh Tari**

Department of Mathematics, Azarbaijan Shahid Madani University, Tabriz, Iran

Email: [s\\_tari@azaruniv.ac.ir](mailto:s_tari@azaruniv.ac.ir)

## بررسی توازن جمعیتی در ایالت جورجیا با استفاده از خوشه‌یابی فضایی

علی ابوالحسنی\*<sup>id</sup> و سمیه تازی<sup>id</sup>

چکیده. تشخیص به هم خوردن نسبت جمعیتی در ساختار جمعیت، یکی از چالش‌هایی است که هر کشور با آن روبرو می‌شود. پیری جمعیت، نوعی ناهنجاری جمعیتی است که عدم توجه به آن سبب بروز مشکلات جمعیتی می‌شود. هشدار به موقع در مورد سالمند شدن جامعه، می‌تواند به برنامه‌ریزی در خصوص فرزندآوری از یک سو و فراهم کردن امکانات مناسب برای سالمندان از سوی دیگر مفید باشد. برای مثال، پایتخت ژاپن نمونه خوبی از محیط شهری مناسب برای سالمندان است. یکی از ابزارهای تشخیص ناهنجاری، خوشه‌یابی فضایی به کمک آماره اسکن است. در سه دهه گذشته، روش آماره اسکن حوزه‌ای بسیار مهم و فعال در تحقیقات آماری بوده است. تشخیص مناطقی بر روی نقشه‌های جغرافیایی که در آن‌ها تمرکز نقاط (سالمندان، بیماران، مجرمان، گونه حیوانی خاص و غیره) معنی‌دار باشد، در بسیاری از زمینه‌ها مانند اپیدمیولوژی، سیاست، جرم و جنایت، جانورشناسی و ... دارای اهمیت است. به کمک روش آماره اسکن می‌توان این نوع خوشه‌های فضایی را شناسایی کرد. در این مقاله، به معرفی آماره اسکن بر اساس توزیع پواسن می‌پردازیم. با استفاده از شبیه‌سازی، کارایی این روش در شناسایی خوشه‌های فضایی را مورد بررسی قرار می‌دهیم. بر اساس نتایج به دست آمده از شبیه‌سازی، روش آماره اسکن پواسن، روشی مناسب برای تشخیص ناهنجاری‌ها در داده‌های فضایی شمارشی است. به‌عنوان کاربردی از خوشه‌یابی فضایی، ساختار جمعیتی ایالت جورجیا را در نظر می‌گیریم و ناحیه‌هایی را مشخص می‌کنیم که در آنها جمعیت سالمند به‌طور معنی‌دار زیاد است. این ناحیه‌ها در اجرای برنامه‌های اصلاحی جمعیت بایستی در اولویت قرار گیرند.

### ۱. مقدمه

فرض کنیم بیماری خاصی مانند کوید-۱۹ در جامعه شیوع پیدا کرده است. برخی از شهرها درگیر این بیماری مسری هستند، با این حال در برخی از شهرها شیوع بیماری کمتر است. هدف، مشخص کردن ناحیه‌هایی در روی نقشه است که تراکم بیماران در آن‌ها زیاد است و این تراکم از نظر آماری معنی‌دار است. هشدار به موقع به دولت و مراکز بهداشتی، در خصوص وجود چنین مناطقی، می‌تواند در پیشگیری از مرگ و میر انسان‌ها و عدم تحمیل هزینه‌های اضافی درمان به دولت مؤثر باشد. با مشخص کردن این مکان‌ها، حتی ممکن است به نتایجی راجع به عامل ایجاد بیماری دست یابیم. به‌عنوان مثال، در زمان شیوع بیماری وبا، جان اسنو [۲۸]، پزشک انگلیسی، موقعیت جغرافیایی افرادی را که دچار مرگ (در اثر بیماری وبا) شده

عبارات و کلمات کلیدی: خوشه‌یابی فضایی، آماره اسکن، آزمون مونت کارلو، توزیع پواسن، نسبت درستیابی

نوع مقاله: پژوهشی

دبیرتخصصی رابط: افشین پرورده

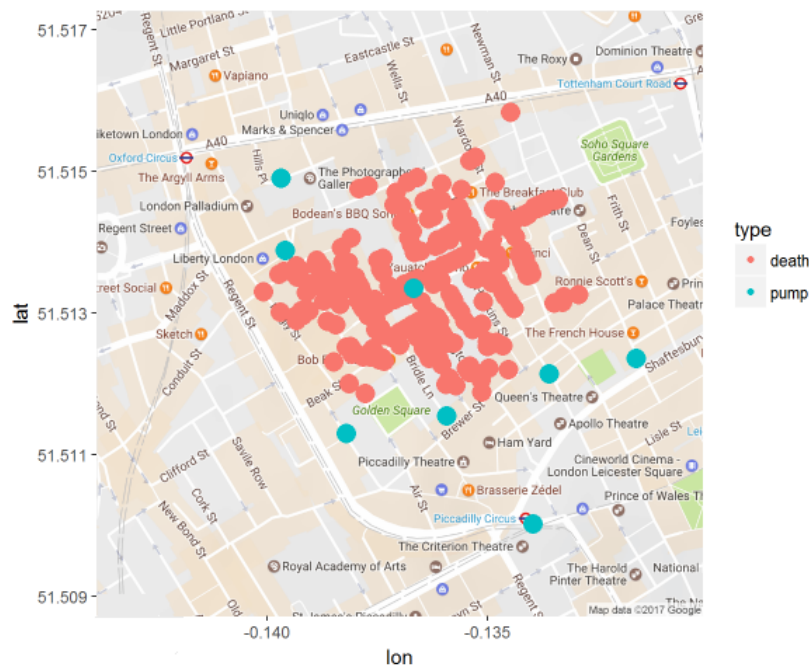
\* نویسنده مسئول

تاریخ دریافت: ۱۴۰۲/۰۳/۲۲ تاریخ پذیرش: ۱۴۰۳/۰۲/۰۸ تاریخ انتشار آنلاین: ۱۴۰۳/۰۹/۰۶

این مقاله مستخرج از نتایج طرح تحقیقاتی اجرا شده به شماره قرارداد ۹۰۶/۱۴۰۲ از محل اعتبار ویژه پژوهشی (Grant) دانشگاه شهید مدنی آذربایجان می‌باشد. ارجاع به مقاله: علی ابوالحسنی و سمیه تازی، بررسی توازن جمعیتی در ایالت جورجیا با استفاده از خوشه‌یابی فضایی، نشریه ریاضی و جامعه، ۹ شماره ۳ (۱۴۰۳) ۱۰۷-۱۲۴.

<http://dx.doi.org/10.22108/msci.2024.138014.1585>





شکل ۱. موقعیت جغرافیایی پمپ‌های آب (آبی) و مرگ و میر مبتلایان به وبا (قرمز) در نقشه لندن در سال ۱۸۵۶.

**Figure 1:** The geographic location of water pumps (blue) and cholera patient fatalities (red) on the map of London in 1856.

بودند، بر روی نقشه لندن نمایش داد و متوجه ارتباط مرگ و میر در اثر بیماری وبا و آلودگی منابع آب آشامیدنی شد، زیرا بیشتر افراد در نزدیکی منابع آب دچار بیماری وبا و مرگ شده بودند. برای مشاهده این نقشه می‌توان به [۳۴] مراجعه نمود. در این مقاله، نقشه مورد نظر با نرم افزار R تهیه و در شکل ۱ آورده شده است. هرچند میزان مرگ و میر در اطراف منابع آب بیشتر است، اما ممکن است این تراکم به صورت تصادفی اتفاق افتاده باشد. به عبارت دیگر، بررسی معنی‌داری این تراکم نیاز است. هرگاه تراکم جمعیت با یک ویژگی خاص (مانند ویژگی بیمار بودن، سالمند بودن، چاق بودن ...) در ناحیه‌ای معنی‌دار باشد، گوییم خوشه فضایی رخ داده است. یکی از ابزارهای آماری برای تشخیص این مناطق، روش آماره اسکن است که توسط [۱۷] معرفی شده است و در بخش ۲ معرفی می‌شود. یافتن خوشه‌های فضایی نه تنها در همه گیر شناسی، بلکه در سایر زمینه‌های تحقیقاتی کاربرد دارد. به عنوان مثال می‌توان به موارد زیر اشاره کرد:

- نجوم [۱۸]، [۴]، [۸]، [۱۵]
- تحلیل تصویر [۱۱]، [۳۲]، [۳]
- جرم‌شناسی [۲۰]، [۲۷]، [۱۳]، [۷]، [۹]
- بوم‌شناسی [۲۹]، [۲۱]
- جغرافیا [۲۲]، [۵]، [۱۹]، [۲۶]
- شناسایی الگو [۱۲]
- زیست‌شناسی [۱۰]، [۲۳]
- همه گیر شناسی [۲۵]، [۱۶]، [۶]، [۳۱]

• حمل و نقل عمومی [۳۰]، [۲]، [۱۴].

مرور کاملی بر روش آماره اسکن در [۱] آمده است، که در آن به مرور مطالب زیر پرداخته‌اند:

- مروری بر آماره اسکن برای توزیع‌های پیوسته و گسسته.
- روش‌های رگرسیونی برای خوشه‌یابی.
- روش‌های بیزی در خوشه‌یابی.
- یافتن خوشه‌های فضایی به شکل دایره‌ای و هم‌چنین خوشه‌های با شکل نامنظم.
- آماره اسکن برای خوشه‌یابی بر روی شبکه‌ها و گراف‌ها.
- روش‌های ناپارامتری در خوشه‌یابی فضایی.
- آماره اسکن برای داده‌های بقا.
- یافتن خوشه‌های زمانی-مکانی.
- نرم‌افزارها و بسته‌های نرم‌افزاری جهت یافتن خوشه‌های فضایی.

در ادامه و در بخش ۲، آماره فضایی اسکن پواسن را معرفی می‌کنیم. بخش ۳، به مطالعه شبیه‌سازی اختصاص داده شده است. در بخش ۴، مثالی از خوشه‌یابی فضایی روی داده‌های واقعی ارائه می‌کنیم و ناحیه‌های با جمعیت سالمند را در ایالت جورجیا مشخص می‌کنیم. نتیجه‌گیری در بخش ۵ آمده است.

## ۲. خوشه‌یابی فضایی و روش آماره اسکن

کالدورف و ناگاروالا [۱۷] برای شناسایی خوشه‌های فضایی، روش آزمون نسبت درستنمایی (LRT) را معرفی کردند. در این روش، یک پنجره دوبعدی نقشه مورد مطالعه را اسکن می‌کند. ضمن اسکن کردن نقشه، اندازه پنجره نیز تغییر می‌کند. گاهی اوقات، این روش در مهندسی «تحلیل پنجره متحرک» نامیده می‌شود. پس از اسکن کردن نقشه، بر اساس یک آزمون فرض آماری که در رابطه (۱) معرفی شده است، وجود خوشه‌هایی فضایی از «نقاط» یا «موارد» بررسی می‌گردد. لازم به توضیح است که هر عضو از جامعه را که دارای ویژگی خاصی باشد، «نقطه» یا «مورد» می‌نامند. در ادبیات خوشه‌یابی فضایی، واژه‌ی «کنترل» به اعضای از جامعه اطلاق می‌گردد که فاقد آن ویژگی خاص باشند.

همان‌طور که بیان شد، به کمک انجام یک آزمون فرض آماری، مسأله خوشه‌یابی حل می‌شود. در این آزمون فرض، عدم وجود خوشه در نقشه مورد مطالعه، به‌عنوان فرض صفر در نظر گرفته می‌شود. همچنین فرض یک عبارت است از وجود ناحیه‌ای در روی نقشه، به گونه‌ای که تراکم نقاط در آن ناحیه زیاد است.

فرض کنیم که نقشه مورد مطالعه دارای  $I$  شهر یا سلول است و برای شهر  $i$ ام،  $i = 1, \dots, I$ ، تعداد موارد و جمعیت به ترتیب،  $x_i$  و  $n_i$  است. هم‌چنین فرض کنیم تعداد موارد در آن شهر دارای توزیع پواسن است، یعنی  $X_i \sim Pois(n_i \theta_i)$ . هر زیرناحیه  $Z$  از نقشه می‌تواند به‌عنوان کاندیدایی برای خوشه در نظر گرفته شود. بنابراین کلاس کاندیداها به تعداد  $2^I - 1$  عضو خواهد داشت، که در آن  $|I|$  عدد اصلی مجموعه است. کلاس تمام کاندیداها ممکن را با  $Z$  نشان می‌دهند. تعداد تمام موارد در  $Z$  عبارت است از  $X_Z = \sum_{i \in Z} X_i$ . به‌طور مشابه  $X_{\bar{Z}} = \sum_{i \in \bar{Z}} X_i$ ، تعداد موارد در  $\bar{Z}$  است. ناحیه  $\bar{Z}$  متمم ناحیه  $Z$  است. تعداد کل جمعیت در ناحیه  $Z$  و  $\bar{Z}$  به‌ترتیب عبارت است از  $n_Z = \sum_{i \in Z} n_i$  و  $n_{\bar{Z}} = \sum_{i \in \bar{Z}} n_i$ . بنابراین جمعیت کل در نقشه  $n = n_Z + n_{\bar{Z}}$  است. فرض کنیم برای هر ناحیه  $i \in Z$  داریم  $\theta_i = \theta_Z$ . به‌طور مشابه برای هر ناحیه  $i \in \bar{Z}$  داریم  $\theta_i = \theta$ . برای یافتن مکان خوشه در روی نقشه، آزمون فرض

$$(1) \quad H_0 : \theta_Z = \theta, \forall Z \in \mathcal{Z} \quad v.s. \quad H_1 : \exists Z \in \mathcal{Z}, \theta_Z > \theta.$$

انجام می‌شود. هر عضو از  $Z$  می‌تواند به‌عنوان خوشه احتمالی در نظر گرفته شود. از آنجایی که مکان واقعی خوشه مجهول است، بنابراین  $Z$  نیز به‌عنوان یک پارامتر در نظر گرفته می‌شود. تابع درستنمایی عبارت است از

$$(۲) \quad L(Z) = L(Z, \theta_0, \theta_Z) = \left[ \prod_{i \in Z} \frac{e^{-n_i \theta_Z} (n_i \theta_Z)^{x_i}}{x_i!} \right] \left[ \prod_{j \notin Z} \frac{e^{-n_j \theta_0} (n_j \theta_0)^{x_j}}{x_j!} \right].$$

با در نظر گرفتن زیرناحیه  $Z$  به‌عنوان خوشه احتمالی، نسبت درستنمایی عبارت است از:

$$(۳) \quad \lambda(Z) = \frac{\sup_{\theta_Z > \theta_0} L(Z, \theta_Z, \theta_0)}{\sup_{\theta_Z = \theta_0} L(Z, \theta_Z, \theta_0)}.$$

آماره اسکن به‌صورت  $\lambda = \sup_{Z \in \mathcal{Z}} \lambda(Z)$  معرفی می‌شود.

به منظور یافتن خوشه فضایی،  $\lambda(Z)$  برای هر  $Z \in \mathcal{Z}$  محاسبه می‌گردد. قدم بعدی، مشخص کردن ناحیه  $Z$  است که منجر به بیشینه شدن  $\lambda(Z)$  می‌گردد. این ناحیه را محتمل‌ترین خوشه<sup>۱</sup> (MLC) می‌نامند. از آنجایی که توزیع آماره اسکن مشخص نیست، آزمون فرض بر اساس روش مونت کارلو انجام می‌شود. این روش به شرح زیر است:

برای انجام آزمون به کمک روش مونت کارلو، ابتدا تعداد موارد برای هر شهر تحت فرض صفر در (۱) تولید می‌شود. مقدار  $\lambda(Z)$  بر اساس داده‌های شبیه‌سازی شده، محاسبه و MLC مشخص می‌گردد. این فرایند ۴۹۹ بار انجام می‌شود و در هر تکرار مقدار آماره اسکن محاسبه می‌شود. فرض کنیم مقادیر محاسبه شده عبارت است از  $\lambda_i, i = 1, 2, \dots, 499$ . این ۴۹۹ مقدار، به‌ترتیب صعودی مرتب می‌شوند، یعنی  $\lambda_{(1)} \leq \lambda_{(2)} \leq \dots \leq \lambda_{(499)}$ . مقدار  $\lambda$  برای داده واقعی را که با نماد  $\lambda_R$  نشان می‌دهیم، در بین مقادیر مرتب شده قرار می‌دهیم.  $p$ -مقدار برابر است با  $\frac{1 + \sum_{i=1}^{499} I(\lambda_i > \lambda_R)}{500}$  به‌طوری که  $I()$  تابع نشانگر است. مراحل به‌دست آوردن  $p$ -مقدار در خوشه‌یابی فضایی در الگوریتم<sup>۱</sup> آورده شده است:

### الگوریتم ۱ محاسبه $p$ -مقدار در خوشه‌یابی فضایی

- (۱) تعداد موارد در هر شهر را بر اساس فرض صفر تولید کنید.
- (۲) آماره اسکن را بر اساس داده‌های تولید شده در مرحله قبل، به‌دست آورید.
- (۳) مراحل ۱ و ۲ را ۴۹۹ بار تکرار کنید.
- (۴) مقدار  $\lambda_R$  را بر اساس داده واقعی به‌دست آورید.
- (۵)  $p$ -مقدار عبارت است از  $\frac{1 + \sum_{i=1}^{499} I(\lambda_i > \lambda_R)}{500}$ ، که  $I()$  تابع نشانگر است.

در روش آمار اسکن دایره‌ای، اعضای  $Z$  به شکل دایره هستند. برای ساختن این اعضا، مرکز یک شهر به‌عنوان مرکز دایره در نظر گرفته می‌شود. سپس مرحله افزودن شهرها به شهر اول شروع می‌گردد. در هر مرحله، نزدیک‌ترین شهر به شهر اولیه اضافه می‌شود. روند افزودن شهرها تا زمانی که ناحیه به‌دست آمده ۵۰٪ از نقشه را پوشش دهد، ادامه می‌یابد. با تغییر مرکز شهر، به شهر دیگر روند اضافه شدن شهرها مانند قبل ادامه دارد. به این ترتیب، عناصر کلاس کاندیدا، یعنی عناصر  $Z$ ، تعیین می‌گردد. لازم به توضیح است که یافتن خوشه‌های با شکل دایره‌ای و غیردایره‌ای در [۳۳] بررسی شده است. در ادامه پنج معیار را معرفی می‌کنیم که میزان کارایی روش خوشه‌یابی فضایی را اندازه‌گیری می‌کنند. این معیارها عبارت‌اند از: حساسیت<sup>۲</sup>، دقت<sup>۳</sup>، میانگین هارمونیک<sup>۴</sup> یا مقدار  $F1$ <sup>۵</sup>، اریبی<sup>۶</sup> و توان<sup>۷</sup>.

<sup>1</sup>cluster likelihood most <sup>2</sup>recall <sup>3</sup>precision <sup>4</sup>mean harmonic <sup>5</sup>score F1 <sup>6</sup>biasness <sup>7</sup>power

همان‌طور که می‌دانیم، بعد از تشخیص مکان خوشه به کمک آماره اسکن، حالت‌های مختلفی ممکن است رخ دهد. ممکن است ناحیه‌ای که به‌عنوان خوشه تشخیص داده شده، کاملاً خوشه واقعی را در برگیرد و برخی از ناحیه‌های غیرخوشه را شامل شود. این امکان وجود دارد که خوشه شناسایی شده کاملاً درون خوشه واقعی قرار بگیرد. ممکن است ناحیه شناسایی شده، دقیقاً بر خوشه واقعی منطبق باشد و در نهایت ممکن است ناحیه تشخیص داده شده، بخشی از خوشه واقعی و بخشی از ناحیه غیرخوشه را دربرگیرد. بنابراین نسبت اشتراک خوشه واقعی و شناسایی شده به خوشه واقعی (یا شناسایی شده) می‌تواند معیاری برای مقایسه روش‌های خوشه‌یابی باشد. به عبارت دیگر

$$\text{حساسیت} = \frac{\#(\text{خوشه واقعی} \cap \text{خوشه شناسایی شده})}{\#(\text{خوشه واقعی})}$$

$$\text{دقت} = \frac{\#(\text{خوشه واقعی} \cap \text{خوشه شناسایی شده})}{\#(\text{خوشه شناسایی شده})}$$

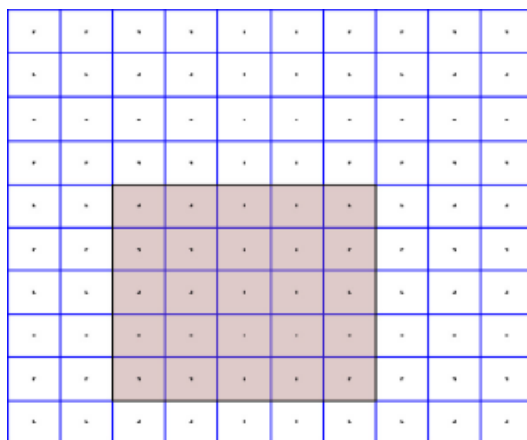
که در آن‌ها  $\#(A)$  به معنی تعداد اعضای مجموعه  $A$  است. لازم به توضیح است که هرچه این معیارها نزدیک به عدد یک باشند، نشان دهنده آن است که خوشه واقعی با خطای کمتری شناسایی شده است. بنابراین عدد یک، مقدار بهینه برای این معیارها است. این کمیت‌ها هر چقدر برای روشی به عدد یک نزدیک‌تر باشند، نشان دهنده برتری آن روش نسبت به روشی است که مقدار این کمیت‌ها در آن از یک فاصله بیشتری دارد. حساسیت و دقت به تنهایی گویای برتری یک روش نسبت به روش دیگر نیست و با ادغام این دو معیار، معیار  $F_1$  به دست می‌آید که میانگین هارمونیک دقت و حساسیت است و به صورت زیر تعریف می‌شود:

$$F_1 = \frac{2 \times (\text{دقت} \times \text{حساسیت})}{\text{دقت} + \text{حساسیت}}$$

علاوه بر دقت، حساسیت و میانگین هارمونیک، ارزیابی روش‌های خوشه‌یابی باشد. پریتمس و همکاران [۲۴] در مورد ارزیابی روش‌های خوشه‌یابی بحث کرده‌اند. نسبت پارامتر مدل در داخل خوشه به خارج آن را ریسک نسبی<sup>۸</sup> می‌نامند و با نماد  $RR$  نشان می‌دهند. بعد از انجام خوشه‌یابی، در مطالعات شبیه‌سازی، می‌توان این نسبت را برای خوشه شناسایی شده نیز محاسبه کرد و برآورد ریسک نسبی را به دست آورد. این برآورد با نماد  $\hat{RR}$  نشان داده می‌شود. نسبت واقعی ریسک نسبی به برآورد آن، ارزیابی نامیده می‌شود. هرچه ارزیابی به عدد یک نزدیک‌تر باشد، نشان دهنده آن است که ریسک نسبی واقعی به ریسک نسبی برآورد شده نزدیک‌تر است. این به معنی آن است که روشی که با آن خوشه‌یابی انجام شده است، ریسک نسبی را بهتر برآورد کرده است.

آخرین معیار برای مقایسه روش‌های خوشه‌یابی، توان آزمون است. روش متداول در خوشه‌یابی بر مبنای آزمون فرض است. بنابراین توان آزمون می‌تواند در مقایسه روش‌های متفاوت به کار رود. توان آزمون عبارت است از احتمال به درستی رد کردن فرض صفر. اگر روشی که در خوشه‌یابی استفاده شده است، توان کمتر داشته باشد به این معنی است که با احتمال بالا، در نقشه‌ای که شامل خوشه است، خوشه‌ای شناسایی نمی‌کند. در فصل بعد، با در نظر گرفتن سناریوهای مختلف، مطالعه شبیه‌سازی را انجام می‌دهیم و کارایی روش خوشه‌یابی فضایی را بررسی می‌کنیم.

relative risk<sup>۸</sup>



شکل ۲. نقشه‌ای با ۱۰۰ سلول و شامل یک خوشه فضایی (رنگ بنفش)

Figure 2: A map with 100 cells and a spatial cluster (purple area).

### ۳. شبیه‌سازی

در این بخش، با در نظر گرفتن سناریوهای مختلف، به مطالعه شبیه‌سازی می‌پردازیم. ابتدا نقشه‌ای مشابه شکل ۲ در نظر می‌گیریم. هر یک از سلول‌های نقشه شامل: تعداد بیماران، تعداد کل افراد در سلول، مختصات طول و عرض جغرافیایی مرکز سلول است. برای شبیه‌سازی تعداد بیماران در هر سلول از توزیع پواسن استفاده می‌کنیم. در سناریوی اول، تعداد بیماران در داخل ناحیه بنفش رنگ (محل وقوع خوشه)، از توزیع پواسن با پارامتر ۲۴ تولید می‌شود. در خارج این ناحیه، تعداد بیماران از توزیع پواسن با پارامتر ۲۰ تولید شده است. خوشه‌یابی را مطابق روش توضیح داده شده در بخش ۲ انجام می‌دهیم. به عبارت دقیق‌تر، ابتدا عضوهای کلاس  $Z$  را می‌سازیم. همان‌طور که قبلاً بیان کردیم این اعضا به صورت دایره‌ای ساخته می‌شوند. سپس، برای هر کدام از عضوهای  $Z \in \mathcal{Z}$  مقدار  $\lambda(Z)$  را از رابطه (۳) محاسبه می‌کنیم. ناحیه  $Z$  که  $\lambda(Z)$  را بیشینه می‌کند، به عنوان محتمل‌ترین خوشه در نظر می‌گیریم و مقدار آماره اسکن را برای آن ذخیره می‌کنیم و مقدار آن را با نماد  $\lambda_R$  نمایش می‌دهیم. سپس تحت فرض صفر، تعداد بیماران را برای هر سلول تولید می‌کنیم. مرحله بعد، یافتن آماره اسکن برای نقشه شبیه‌سازی شده است. مکان محتمل‌ترین خوشه را مشخص کرده و مقدار آماره اسکن را برای آن به دست آورده و با نماد  $\lambda_1$  ذخیره می‌کنیم. معیارهای آریبی، توان، دقت، حساسیت و  $F1$  را برای تعیین کارایی خوشه‌یابی محاسبه و ذخیره می‌کنیم. شبیه‌سازی نقشه‌ها تحت فرض صفر را به تعداد ۴۹۹ بار تکرار می‌کنیم. در هر بار، محل خوشه شناسایی شده، معیارهای معرفی شده و مقدار آماره اسکن را محاسبه می‌کنیم. بنابراین ۴۹۹ مقدار برای آماره اسکن داریم. این مقادیر را مرتب می‌کنیم و داریم:  $\lambda(499) < \dots < \lambda(2) < \lambda(1)$ . در نهایت مقدار  $\lambda_R$  را در دنباله مرتب شده به گونه‌ای قرار می‌دهیم تا دنباله همچنان صعودی باقی بماند. اگر مکان  $\lambda_R$  در ۵٪ بالایی باشد، فرض صفر را رد می‌کنیم و بیان می‌کنیم که شواهد آماری کافی را برای رد کردن فرض نداریم. فرآیند فوق را ۱۰۰۰ بار تکرار می‌کنیم و توان آزمون را نیز محاسبه می‌کنیم. نتایج مربوط به این سناریو در جدول ۱ و در اولین ستون از سمت چپ آمده است. همان‌گونه که مشاهده می‌شود، توان آزمون کم و حدود ۰.۱۹ است. این به معنی آن است که در این سناریو تنها در ۱۹٪ از تکرارها، وجود خوشه توسط آماره اسکن هشدار داده می‌شود. از طرف دیگر، مقدار آریبی، حساسیت، دقت و  $F1$  از مقدار بهینه ۱، فاصله دارند. دور بودن این معیارها از عدد بهینه ۱ به این معنی است که خوشه‌های شناسایی شده، خوشه واقعی را با خطای بیشتری ردیابی کرده‌اند.

## جدول ۱. نتایج شبیه‌سازی برای سه سناریوی متفاوت

Table 1: Simulation results for three different scenarios.

$Pois(24)$ در مقابل $Pois(20)$	$Pois(30)$ در مقابل $Pois(20)$	$Pois(60)$ در مقابل $Pois(20)$	
۱/۱۷۵۲	۱/۰۵	۱/۰۱	ارزیبی
۰/۸۱	۱	۱	حساسیت
۱/۹	۱	۱	دقت
۰/۱۹	۰/۹۳	۱	توان
۰/۸۳	۱	۱	F۱

در دومین سناریو، پارامتر داخل خوشه  $30$  و پارامتر خارج خوشه  $20$  است. همانند مرحله قبل خوشه‌یابی را انجام و پنج معیار معرفی شده را محاسبه می‌کنیم. انتظار داریم مقادیر این معیارها، در مقایسه با سناریوی قبل، به مقدار بهینه نزدیک‌تر باشند. نتایج در جدول ۱ و در ستون میانی آمده است. نه تنها توان از  $0.19$  به  $0.93$  افزایش یافته است، بلکه مقادیر سایر معیارها نیز به مقدار بهینه نزدیک‌تر شده است. به عبارت دیگر، وجود خوشه در  $93\%$  از تکرارها هشدار داده شده است. علاوه بر این، محل خوشه واقعی دقیق‌تر ردیابی شده است.

در سناریوی پایانی، مقدار پارامتر توزیع پواسن داخل خوشه را  $60$  و مقدار پارامتر در خارج از خوشه را  $20$  در نظر گرفته‌ایم. نتایج خوشه‌یابی در جدول ۱ در ستون سمت راست آمده است. در این سناریو، در تمام موارد، در خصوص وجود خوشه در نقشه هشدار داده شده است (توان آزمون برابر ۱ است) و خوشه شناسایی شده دقیقاً بر خوشه واقعی منطبق است.

## ۴. مثالی از خوشه‌یابی بر روی نقشه واقعی

در این بخش از داده‌های georgia استفاده می‌کنیم که در پکیج GISTools در نرم‌افزار R موجود است. این داده‌ها شامل ۱۴ متغیر از جمله: طول و عرض جغرافیایی هر منطقه، جمعیت هر منطقه، جمعیت سالمند در هر منطقه و غیره است. برای رسم نقشه ایالت جورجیا که شامل ۱۵۹ منطقه است، ابتدا داده georgia را فراخوانی می‌کنیم. فرمت داده‌ها به صورت sp ذخیره شده است. فرمت داده‌ها را به sf تغییر می‌دهیم. برای این منظور از دستور زیر استفاده می‌کنیم.

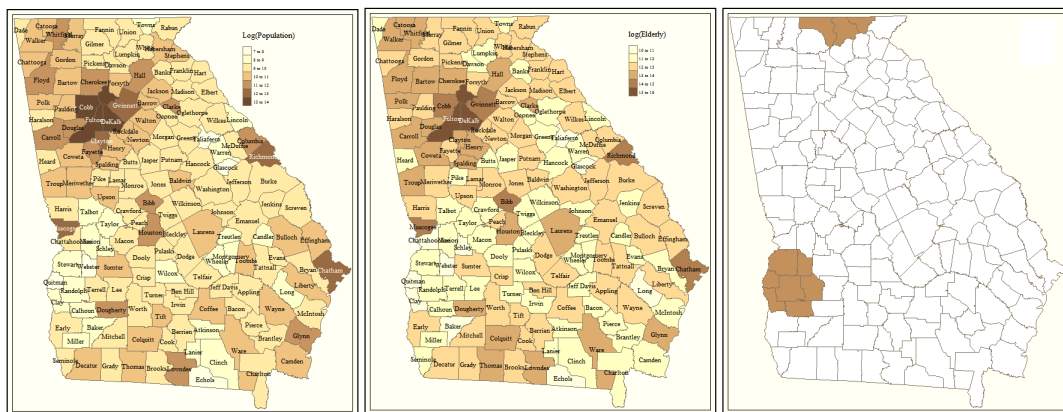
```
georgia_sf <- st_as_sf(georgia)
```

قصد داریم بررسی کنیم که در کدام مناطق تراکم جمعیت سالخورده بیشتر و این تراکم از نظر آماری معنی‌دار است. برای این منظور ابتدا نقشه ایالت جورجیا را بر اساس تعداد جمعیت و تعداد سالخوردگان در شکل ۳ رسم می‌کنیم. برای رسم نقشه‌ها می‌توان از تابع  $qtm()$  و به صورت زیر استفاده کرد:

```

georgia_sf$logTotPop90<-log(georgia_sf$TotPop90)
qtm(georgia_sf,fill="logTotPop90",text="Name",
    text.size=0.70,style="classic",
    fill.title="Log(Population)")+
tm_layout(
    legend.title.size=1,
    legend.text.size = 0.43,
    legend.position = c("right","top"),
    legend.bg.color = "white",
    legend.bg.alpha = 1)
georgia_sf$logEld<-log(georgia_sf$TotPop90*georgia_sf$PctEld)
qtm(georgia_sf,fill="logEld",text="Name",
    text.size=0.70,style="classic",
    fill.title="log(Elderly)")+
tm_layout(
    legend.title.size=1,
    legend.text.size = 0.43,
    legend.position = c("right","top"),
    legend.bg.color = "white",
    legend.bg.alpha = 1)

```



شکل ۳. نقشه ایالت جورجیا. شکل چپ: رنگ‌بندی بر اساس تعداد جمعیت، شکل میانی: رنگ بندی بر اساس نرخ سالمندان، شکل راست: مکان خوشه‌های شناسایی شده.

Figure 3: The map of the state of Georgia. Left figure: Color coded based on population count, middle figure: Color coded based on the elderly rate, right figure: Location of identified clusters.

پس از انجام خوشه‌یابی به کمک آماره اسکن، دو خوشه در غرب و شمال ایالت جورجیا شناسایی می‌شود. در مناطق مشخص شده، جمعیت سالمند بیش از سایر مناطق است. به عبارت دقیق‌تر، این مناطق نسبت به سایر مناطق مسن‌تر هستند.



بنابراین توصیه می‌شود که این مناطق در برنامه‌ریزی‌های جمعیتی در اولویت قرار بگیرند و تدابیر مناسب از نظر بسترسازی شهری مناسب برای سالمندان و فراهم کردن شرایط برای فرزندآوری برای جوانان مد نظر قرارگیرد.

## ۵. نتیجه‌گیری

پیری جمعیت، نوعی ناهنجاری جمعیتی است که عدم توجه به آن سبب بروز مشکلات جمعیتی می‌شود. هشدار به موقع در مورد سالمند شدن جامعه می‌تواند به برنامه‌ریزی درخصوص فرزندآوری از یک‌سو و فراهم کردن امکانات مناسب برای سالمندان از سوی دیگر مفید باشد. به‌عنوان مثال، پایتخت ژاپن نمونه خوبی از محیط شهری مناسب برای سالمندان است. در حال حاضر، کشور ایران نیز مرحله انتقال ساختار سنی جمعیت از جوانی به سالخوردگی را تجربه می‌کند و به نظر می‌رسد که در شروع قرن پانزدهم هجری شمسی با افزایش جمعیت سالمند در کشور مواجه هستیم. بنابراین نیاز است که هشدارهای به موقع جهت کنترل پیری جمعیت صورت پذیرد و با اندیشیدن تدابیر سنجیده، امکانات شهری و رفاهی مناسب برای این قشر از جامعه فراهم شود.

آماره اسکن فضایی یکی از ابزارهایی است که به تازگی برای تشخیص ناهنجاری‌ها معرفی شده است. در این مقاله ما از این ابزار برای ردیابی مناطقی با جمعیت سالمند استفاده کردیم. امروزه به دلیل استفاده از موبایل، تبلت، دوربین‌های دیجیتال و غیره، موقعیت جغرافیایی افراد جامعه به متغیر مورد مطالعه ضمیمه می‌شود و به کمک این اطلاعات ضمیمه شده می‌توان تحلیل‌های جامع‌تری را ارائه داد. هرچند در تحقیق حاضر، داده‌های جغرافیای سالمندان کشور در دسترس نویسندگان مقاله نیست، اما به دلیل اهمیت موضوع و معرفی آماره اسکن فضایی، از داده‌های مربوط به سالمندان در ایالت جورجیا استفاده شده است تا خوشه‌های فضایی سالمندان در این ایالت ردیابی شود. روش معرفی شده را می‌توان در مورد کشور ایران نیز به کار برد.

## مراجع

- [1] A. Abolhassani, M. O. Prates, An up-to-date review of scan statistics, *Stat. Surv.*, **15** (2021) 111–153.
- [2] A. Abolhassani, M. O. Prates, F. Castellares and S. Mahmoodi, Zero-inflated Bell scan: A more flexible spatial scan statistic, *Spat. Stat.*, **36** (2020) 18 pp.
- [3] A. Abolhassani, M. O. Prates and S. Mahmoodi, Irregular shaped small nodule detection using a robust scan statistic, *Statistics in Biosciences*, **15** (2023) 141–162.
- [4] K. L. Adelberger, C. C. Steidel, M. Pettini, A. E. Shapley, N. A. Reddy and D. K. Erb, The spatial clustering of star-forming galaxies at redshifts  $1.4 < z < 3.5$ , *The Astrophysical Journal*, **619** (2005) 619–697.
- [5] L. Anselin, Local indicators of spatial association—LISA, *Geographical Analysis*, **27** (1995) 93–115.
- [6] L. H. Duczmal, G. J. P. Moreira, D. Burgarelli, R. H. Takahashi, F. C. Magalhães and E. C. Bodevan, Voronoi distance based prospective space-time scans for point data sets: a dengue fever cluster analysis in a southeast Brazilian town, *International Journal of Health Geographics*, **10** (2011) 1–4.
- [7] J. Eck, S. Chainey, J. Cameron and R. Wilson, *Mapping crime: Understanding hotspots*, 2005.
- [8] M. D. Gladders and H. K. Yee, A new method for galaxy cluster detection. I. The algorithm, *The Astronomical Journal*, **120** (2000) 2148–2162.
- [9] T. H. Grubestic, On the application of fuzzy clustering for crime hot spot detection, *Journal of Quantitative Criminology*, **22** (2006) 77–105.



- [10] A. Gutteridge, G. J. Bartlett and J. M. Thornton, Using a neural network and spatial clustering to predict the location of active sites in enzymes, *Journal of Molecular Biology*, **330** (2003) 719–734.
- [11] R. Haralick and I. H. Dinstein, A spatial clustering procedure for multi-image data, *IEEE Trans. Circuits and Systems*, **22** (1975) 440–450.
- [12] R. M. Haralick and G. L. Kelly, Pattern recognition with measurement space and spatial clustering for multiple images, *Proceedings of the IEEE*, **57** (1969) 654–665.
- [13] K. D. Harries, *Mapping crime: Principle and practice*, US Department of Justice, Office of Justice Programs, National Institute of Justice, Crime Mapping Research Center, 1999.
- [14] L. Huang, D. G. Stinchcomb, L. W. Pickle, J. Dill and D. Berrigan, Identifying clusters of active transportation using spatial scan statistics, *Am. J. Prev. Med.*, **37** (2009) 157–166.
- [15] R. S. Kim, J. V. Kepner, M. Postman, M. A. Strauss, N. A. Bahcall, J. E. Gunn, R. H. Lupton, J. Annis, R. C. Nichol, F. J. Castander and J. Brinkmann, Detecting clusters of galaxies in the sloan digital sky survey. i. monte carlo comparison of cluster detection algorithms, *Astron. J.*, **123** (2002) 20–36.
- [16] M. Kulldorff, A spatial scan statistic, *Comm. Statist. Theory Methods*, **26** (1997) 1481–1496
- [17] M. Kulldorff and N. Nagarwalla, Spatial disease clusters: detection and inference, *Stat. Med.*, **14** (1995) 799–810.
- [18] H. J. Mo and S. D. White, An analytic model for the spatial clustering of dark matter haloes, *Mon. Not. R. Astron. Soc.*, **282** (1996) 347–361.
- [19] A. T. Murray and V. Estivill-Castro, Cluster discovery techniques for exploratory spatial data analysis, *Int. J. Geogr. Inf. Sci.*, **12** (1998) 431–443.
- [20] A. T. Murray, T. H. Grubestic and R. Wei, Spatially significant cluster detection, *Spat. Stat.*, **10** (2014) 103–116.
- [21] N. Myers, R. A. Mittermeier, C. G. Mittermeier, G. A. B. Da Fonseca and J. Kent, Biodiversity hotspots for conservation priorities, *Nature*, **403(6772)** (2000) 853–858.
- [22] J.K. Ord, A. Getis, Local spatial autocorrelation statistics: distributional issues and an application, *Geographical Analysis*, **27(4)** (1995) 286–306.
- [23] D. Pellin and C. Di Serio, A novel scan statistics approach for clustering identification and comparison in binary genomic data, *BMC Bioinformatics*, **17** (2016) 61–71.
- [24] M. O. Prates, M. Kulldorff and R. M. Assunção, Relative risk estimates from spatial and space-time scan statistics: are they biased?, *Stat. Med.*, **33** (2014) 2634–2644.
- [25] C. J. Ribeiro, A. D. Dos Santos, S. V. Lima, E. R. da Silva, B. V. Ribeiro, A. M. Duque, M. V. Peixoto, P. L. Dos Santos, I. M. de Oliveira, M. W. Lipscomb and K. C. de Araújo, Space-time risk cluster of visceral leishmaniasis in Brazilian endemic region with high social vulnerability: an ecological time series study, *PLoS Neglected Tropical Diseases*, **15** (2021) 1–20.
- [26] P. Rogerson and I. Yamada, *Statistical detection and surveillance of geographic clusters*, CRC Press, 2008.
- [27] L. W. Sherman and D. Weisburd, General deterrent effects of police patrol in crime hot spots: A randomized, controlled trial, *Justice Quarterly*, **12** (1995) 625–648.
- [28] J. Snow, *On the mode of communication of cholera*, John Churchill, 1849.

- [29] T. J. Stohlgren, D. Binkley, G. W. Chong, M. A. Kalkhan, L. D. Schell, K. A. Bull, Y. Otsuki, G. Newman, M. Bashkin and Y. Son, Exotic plant species invade hot spots of native plant diversity, *Ecological Monographs*, **69** (1999) 25-46.
- [30] Y. Tanoue, D. Yoneoka, T. Kawashima, S. Uryu, S. Nomura, A. Eguchi, K. Makiyama and K. Matsuura, Public transportation network scan for rapid surveillance, *Biostatistics and Epidemiology*, **7** (2022) 1-15.
- [31] S. C. Wieland, J. S. Brownstein, B. Berger and K. D. Mandl, Density-equalizing Euclidean minimum spanning trees for the detection of all disease cluster shapes, *Proceedings of the National Academy of Sciences*, **104** (2007) 9404-9409.
- [32] L. Zhang and Z. Zhu, Spatial multiresolution cluster detection method, (2012), *arXiv preprint arXiv*.  
 [۳۳] ع. ابوالحسنی، چند روش جدید در خوشه‌بندی فضایی، پایان‌نامه دکتری آمار، دانشگاه صنعتی اصفهان، ۱۳۹۹.  
 [۳۴] م. محمدزاده، آمار فضایی و کاربردهای آن، دانشگاه تربیت مدرس، ۱۳۹۸.

### علی ابوالحسنی

گروه ریاضی، دانشگاه شهید مدنی آذربایجان، آذربایجان، تبریز، ایران  
 ali.abolhassani@azaruniv.ac.ir

علی ابوالحسنی عضو هیات علمی دانشگاه شهید مدنی آذربایجان است. وی تحصیلات کارشناسی را در رشته آمار در دانشگاه تبریز به اتمام رسانده است. تحصیلات کارشناسی ارشد وی در رشته آمار و در دانشگاه صنعتی امیرکبیر بوده است. تحصیلات ایشان در مقطع دکتری نیز در رشته آمار و در دانشگاه صنعتی اصفهان می‌باشد. یک دوره فرصت مطالعاتی به مدت یک سال در کشور برزیل حضور داشته و در زمینه خوشه‌بندی فضایی به تحقیق پرداخته است. هم‌اکنون نیز علاقمند به توسعه روش‌های خوشه‌بندی فضایی است.



### سمیه تاری

گروه ریاضی، دانشگاه شهید مدنی آذربایجان، آذربایجان، تبریز، ایران  
 s\_tari@azaruniv.ac.ir

سمیه تاری، عضو هیات علمی دانشگاه شهید مدنی آذربایجان و دانش‌آموخته دکترای ریاضی محض، گرایش منطق ریاضی از دانشگاه تبریز است. علائق پژوهشی وی در منطق ریاضی و کاربردهای آن، نظریه مدل و ساختارهای ترتیب کمینه است.

