

УДК 311.17:311:213

**ЗАДАЧИ ИНФОРМАЦИОННОЙ ПОДДЕРЖКИ СОЦИАЛЬНОГО ИССЛЕДОВАНИЯ В РАМКАХ ПРОЕКТА ИЗУЧЕНИЯ СОЦИОКУЛЬТУРНЫХ ФАКТОРОВ НОВОЙ ИНДУСТРИАЛЬНОЙ МОДЕРНИЗАЦИИ В РЕГИОНАХ****Романчуков С.В.***ФГАОУ ВО «Национальный исследовательский Томский политехнический университет», Томск, e-mail: inoy@vtomske.ru*

В современном обществе велика потребность в проведении разного рода социальных и психологических исследований. Единоразово достигнутые результаты требуют непрерывного уточнения, подтверждения, корректировки в условиях быстрого изменения экономической, политической и социальной обстановки. В Российской Федерации как в силу известных событий, так и в связи с необходимостью разрешения накопившихся системных проблем, велика потребность в широком спектре социальных исследований для определения возможных областей роста, и напротив, наиболее угрожаемых направлений внутренней политики, состояния человеческого и экономического потенциала регионов. Объемы информации, накапливаемые при рассмотрении таких масштабных задач, крайне велики и требуют машинной обработки. Но при этом сложно требовать от специалистов в области гуманитарного знания навыков программирования и анализа данных в специализированных пакетах, данная работа требует привлечения специалистов в области прикладной математики, математического моделирования и анализа данных. Информационной поддержке одного из подобных исследований и посвящена данная работа. В работе показан процесс формирования системы, охватывающей все этапы исследования от подготовительных работ по формированию выборки до подготовки структуры данных к итоговой обработке. Работа выполнена в рамках гранта РГНФ №15-03-00366 Социокультурные факторы новой индустриальной модернизации в регионах (на материалах исследований в Томской области).

**Ключевые слова:** выборка, репрезентативность, структура населения, статистика, модернизация**OBJECTIVES OF SOCIAL STUDIES INFORMATION SUPPORT WITHIN THE RESEARCH OF SOCIOCULTURAL FACTORS OF THE NEW INDUSTRIAL MODERNIZATION IN REGIONS****Romanchukov S.V.***FSAEI HE «National Research Tomsk Polytechnic University», Tomsk, e-mail: inoy@vtomske.ru*

In modern society there is a high need for all sorts of social and psychological research. Once achieved results require continuous refinement, validation, adjustments to the rapid changes in the economic, political and social environment. In the Russian Federation as a result of certain events, and in connection with the need to resolve the accumulated systemic problems, there is a great need for a wide range of social research to identify possible points of growth, and on the other hand the most threatened areas of domestic policy, the state of the regions human and economic potential. Volumes of information collected when dealing with such large-scale problems, are great and require computer processing. But it is difficult to demand from professionals in the field of human knowledge programming knowledge and skills in specialized data analysis package, so this work requires the involvement of specialists in the field of applied mathematics, mathematical modeling and data analysis. This work is devoted to the information support of one of these studies. The paper shows the process of forming a system that covers all stages of the study from the preparatory work on the formation of the sample prior to the preparation of the data structure to the final processing. This work is performed under the grant RHF №15-03-00366 Socio-cultural factors of a new industrial modernization in the regions (based on research in the Tomsk region).

**Keywords:** sampling, representativity, the structure of the population, statistics, modernization

Развитие России в настоящее время требует успешных экономических преобразований, которые невозможны без модернизации, в первую очередь, региональной, как составляющей модернизации России. Между тем, такого рода преобразования возможны лишь при наличии полной и исчерпывающей информации о социальной обстановке, человеческих ресурсах и скрытом модернизационном потенциале регионов. Именно данную задачу на местах решают исследовательские группы, участвующие в проекте «Социокультурный портрет региона». В Томской области данный проект выполняется в рамках гранта РГНФ № 15-03-00366 Социокультурные

факторы новой индустриальной модернизации в регионах (на материалах исследований в Томской области).

Научная значимость проекта заключается в расширении методологии исследования, что будет способствовать совершенствованию инструментария социальных исследований российской модернизации.

Проект выполняется на базе факультета психологии, связей с общественностью и рекламы ТГПУ коллективом, состоящим из учащихся и сотрудников Томского государственного педагогического университета и Национального исследовательского Томского политехнического университета, и охватывает всю территорию Томской об-

ласти. Полученные результаты предоставляются вышестоящим проектам федерального масштаба и служат основанием для формирования рекомендаций на высшем уровне.

Очевидно, что объём подлежащих обработке данных и масштаб решаемых задач требуют мощной информационной поддержки и сопровождения. Задачи, которые необходимо решить, чтобы обеспечить успешное проведение исследования, достаточно массивны, даже если временно оставить вне рассмотрения вопрос статистической обработки полученных данных. Особого внимания требуют задачи построения выборки, разработки инструментария для обеспечения взаимодействия участников проекта и разработки структуры данных для обеспечения их дальнейшей обработки. Обратимся к данным вопросам более подробно.

### **Население и экономика Томской области**

Прежде чем говорить о построении репрезентативной выборки, необходимо коснуться структуры населения и экономики рассматриваемого региона. По данным Росстата [6], на территории Томской области по состоянию на начало 2015 г. проживало 1075,5 тысяч человек, из них порядка 586 тысяч – в городе Томске. Население Северска составляет порядка 115 тысяч человек, Стрежевого – 41,8 тысячи. Населённость различных районов Томской области варьируется от 6,4 тыс. человек в Тегульдетском районе до 34,5 в Асиновском. Плотность населения Томской области составляет 3,42 чел./км<sup>2</sup>, при этом 71,95% населения проживает в городах.

Томская область делится на 4 городских округа (Томск, Стрежевой, Колпашево, ЗАТО Северск), 16 муниципальных районов, 3 городских и 118 сельских поселений, 576 сельских населённых пунктов. Административным центром Томской области является город Томск.

Одной из наиболее важных и динамично развивающихся отраслей является высшее образование. В Томске 6 государственных вузов, по числу студентов на душу населения Томск занимает одно из первых мест в стране. Также имеется один вуз в Северске, городе-спутнике Томска. По оценке Минэкономразвития России, Томская область в 2004 году относилась к региону с уровнем развития выше среднего (18-е место по стране). При этом область занимала 14-е место по размеру среднемесячной заработной платы (9640 руб.), объёму платных услуг на душу населения (18,8 тыс. руб.), поступлению налогов в бюджет на душу на-

селения (60,4 тыс. руб.). Основные приоритеты экономического развития – топливно-энергетический, научно-образовательный комплексы и малый бизнес [2].

### **Построение выборки по Томской области**

Карта исследования составляется таким образом, чтобы при проведении опросов полученная выборка являлась репрезентативной, достаточно близко отображая реальные показатели расселения жителей региона, половозрастной состав и т.д., это означает необходимость выбора маршрутных точек таким образом, чтобы гарантировать эффективный охват жилых территорий основных населённых пунктов Томской области с частотой размещения точек пропорциональной их доле в населении региона.

С учётом опыта прошлых лет и в рамках типового методического инструментария проекта «Социокультурный портрет региона» предполагается формирование выборки из 1000 респондентов, отражающей географические особенности распределения населения в регионе. При этом оптимальным числом респондентов, опрошенных на каждом из маршрутов, считается число 10.

Простой и хорошо зарекомендовавшей себя процедурой отбора маршрутных точек является методика случайного выбора по принципу совмещения спирали и треугольника. Треугольник и спираль накладываются на план (схему) города, ПГТ или станицы. Точки их пересечения определяют улицу и начало маршрута, на котором необходимо опросить 10 респондентов. В тех случаях, когда в городе необходимо определить большое количество маршрутов, было рекомендовано использовать процедуру развернутой спирали [5].

Однако в силу неравномерности застройки, существования больших административных и производственных комплексов, размещённых в городской черте, лесопарковых зон, неровностей ландшафта и естественных водоёмов, имеющих значительную площадь, равномерно распределённые по спирали точки размещены недостаточно эффективно. Во-первых, часть точек попадает в незаселённые районы, офисную застройку, в лесополосы, реки и другие места, где проведение опросов жителей невозможно [3]. Во-вторых, равномерное распределение точек в пространстве приводит к тому, что район малоэтажной застройки может предоставить такое же или даже большее количество респондентов, чем улица, компактно застроенная многоэтажными домами (при очевидно неравной численности населения).

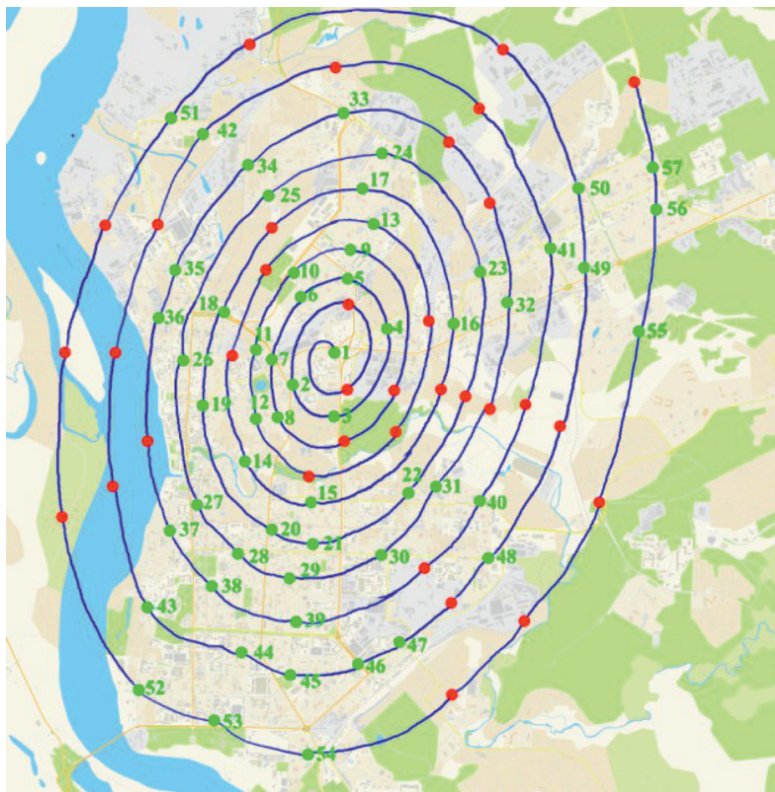


Рис. 1. Определение точек маршрута в Томске

Чтобы избежать подобных несовершенств системы формирования выборки, предложенный типовой алгоритм был расширен с использованием методов, близких к математическим методам оптимизации. Коротко говоря – некая случайная спираль с равномерно распределёнными точками была взята за основу, после чего каждый из узлов был оценён с использованием целевой функции, завязанной на плотность населения в его окрестности, недостаточно информативные узлы (с нулевым или крайне малым населением) были исключены. После выполнения данной операции из 57 выбранных изначально случайных маршрутов более 20 были отвергнуты. На втором этапе недостающие точки размещались в окрестности «удачных» узлов, в соответствии с плотностью населения районов, расположенных между указанными «удачными» метками, но и с некоторым элементом случайности. Операция добавления, переоценки и удаления точек повторялась до тех пор, пока все узлы не были размещены в населённых районах. После выполнения данной операции полученная карта была откорректирована вручную с целью переноса проблемных точек, отвергнутых в силу факторов, слабо поддающихся формализации. Результат выполнения всей совокупности данных процедур показан на рис. 1.

Аналогичные работы выполнялись для Северска (рис. 2).

В Стрежевом, Кожевниково и Асино поиск маршрутов осуществлялся по методу совмещения треугольника и произвольной спирали, вписанных в границы поселения, как это показано на рис. 3, а–в.

В меньших по размеру населённых пунктах для проведения опроса была выбрана главная жилая улица поселения. Таким образом был осуществлён поиск необходимого числа маршрутов, опросы населения по которым позволили создать репрезентативную выборку, отражающую структуру населения Томской области.

#### **Инструментарий для подготовки данных к обработке**

Анкета для проведения социологического опроса по месту жительства в рамках программы «Социокультурные факторы новой индустриальной модернизации в регионах» (в модификации 2015 года) содержит 62 вопроса. Все они сгруппированы в 5 тематических блоков: вопросы о месте проживания, о трудовой деятельности, о культурной жизни региона и степени вовлечённости респондента в оную, об участии в общественной жизни региона и блок вопросов биографического характера. Существует множество



способов организации компьютерных вычислений от использования языков программирования высокого уровня и до привлечения узкоспециальных утилит и пакетов прикладных программ. Учитывая специфику решаемой задачи, а именно: спектр необходимых для реализации методик анализа (статистическая обработка данных, методы кластерного и факторного анализа), потребность в визуализации результатов, необходимость обеспечения сопоставимости результатов исследования в разных регионах и результатов работы разных исследовательских групп – для данной работы был выбран программный продукт SPSS Statistics.

SPSS Statistics («Statistical Package for the Social Sciences») – программный пакет для статистического анализа, реализующий функции анализа данных, управления данными, добычи данных, визуализации данных с привлечением статистических методов, один из лидеров рынка в области коммерческих статистических продуктов, предназначенных для проведения прикладных исследований в социальных науках.

IBM SPSS Statistics и продукты IBM SPSS Amos, Sample Power, VizDesigner, Data Collection, Collaboration and Deployment Services образуют модульный, полностью интегрированный программный комплекс, охватывающий все этапы аналитического процесса:

- планирование исследования;
- сбор данных;
- доступ и управление данными;
- всесторонний анализ (от базовых процедур выведения итогов и классической статистики до моделирования с применением новейших алгоритмов);
- создание отчетов;
- хранение и распространение результатов.

Большой выбор процедур в базовом модуле IBM SPSS Statistics дает широкие возможности анализа данных различных типов. Встраиваемые дополнительные модули расширяют аналитические возможности настолько, насколько это необходимо. SPSS обладает широкими графическими возможностями, позволяет выводить информацию в виде различных типов графиков, все компоненты графиков настраиваются [4].

С помощью статистического пакета SPSS Statistics разработана структура данных, представляющая в электронной форме результаты исследований. При этом были, по сути, заново разработаны инструкции для заполнения базы данных и описания отдельных компонентов анкеты, т.к. аналогичные указания 2010 года, используемые другими проектными группами, устарели по отношению к анкете в модификации 2015 года. Предоставленные анкеты были удачно внесены в созданную структуру.



Рис. 2. Определение точек маршрута в Северске

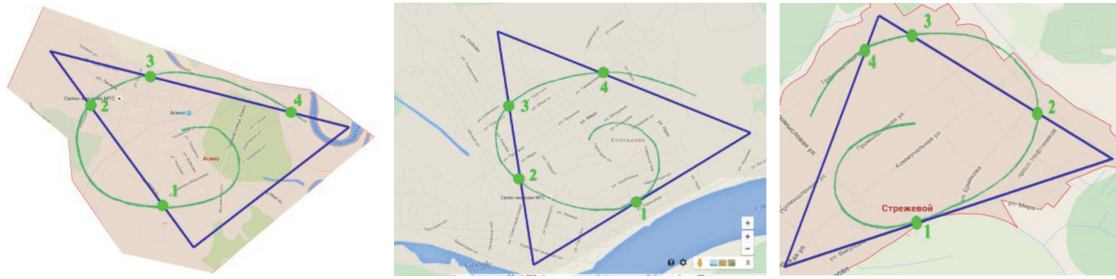


Рис. 3. Определение точек маршрута (а – Асино, б – Колпашево, в – Стржевой)

	N_ank	start	q1	q2	w_q2_1	w_q2_1dr	q3	q4	q5	q6	q6_1	q6_2
415	472	16:00	1	2	2		3	1	2		2	
416	475	9:35	3	1	2		1	1	1			2
417	476	22:05	3	2	1		1	1	1			2
418	477	15:45	4	1	1		2	2	4			
419	478	11:30	2	2	1		1	1	1	88		
420	479	9:40	1	3	1		2	2	2			
421	480	11:15	4	1	2		1	1	4			
422	481	13:30	4	2	2		2	2	3			2
423	482	20:15	3	2	1		1	2	2			2
424	483	12:00	2	2	2		1	1	2		2	
425	484	9:35	4	2	1		1	1	2			2
426	485	18:00	4	2	2		2	3	2			
427	486	9:00	4	1	1		1	77	1			
428	487	19:40	3	2	2		1	2	4			2
429	488	12:20	3	2	2		1	2	2	77		
430	489	22:35	2	3	1		1	1	1			2
431	490	18:50	4	1	2		1	2	1			2
432	491	19:00	3	2	2		1	2	1			2

Рис. 4. Информация о времени тестирования и списке анкет



Исследовательский проект  
Социокультурные факторы новой  
индустриальной модернизации в регионах  
на материалах исследований в Томской области

Главная О проекте **Документация** Карта исследований Контакты Построение маршрута для городов Томской области

Рис. 5. Закладки разделов на портале проекта

Налажен процесс цифровой обработки результатов опросов, проводимых на территории Томской области. Фрагмент одной из таблиц, содержащих результаты анкетирования, показан на рис. 4.

Был накоплен и подготовлен к дальнейшей обработке большой массив данных. Результаты их обработки будут представлены позднее, т.к. работы продолжаются и внесены в план на 2015–2016 год.

### Веб-интерфейс

Не менее важной задачей является обеспечение взаимодействия участников

проекта. Распространение нормативной документации, распределение заданий и маршрутных точек для отдельных интервьюеров, сбор анкетных данных и подготовку анкет к дальнейшей обработке можно значительно ускорить, используя ресурсы сети Интернет.

Существует некоторый набор общепринятых требований, которым должна соответствовать разрабатываемая система. Как минимум, она должна обеспечивать:

- возможность авторизации (с дифференцированным по признакам роли в проекте доступом к материалам);

