

С.Н. Мартышенко

S.N. Martyshenko

АНАЛИЗА ДАННЫХ МОНИТОРИНГА СОЦИАЛЬНО- ЭКОНОМИЧЕСКИХ ПРОЦЕССОВ В МУНИЦИПАЛЬНЫХ ОБРАЗОВАНИЯХ

DATA ANALYSIS OF MONITORING SOCIAL AND ECONOMIC PROCESSES IN MUNICIPAL UNIONS

В работе рассматривается компьютерная технология обработки данных анкетных опросов, производимых при мониторинге социально-экономических процессов в муниципальных образованиях.

Ключевые слова: анкетный опрос, мониторинг, компьютерная технология.

In the work the computer technology of data processing of questionnaires of social and economic processes of municipal unions made at monitoring is considered.

Keywords: questionnaire, monitoring, computer technology.

Владивостокский государственный университет экономики и сервиса

В настоящее время актуальной является проблема исследования взаимодействия участников социально-политических и трудовых процессов, протекающих на уровне муниципальных образований. Одним из наиболее эффективных средств их исследования является мониторинг, осуществляемый на основе независимых анкетных опросов жителей, проживающих на территории муниципальных образований.

Термин «мониторинг» английского происхождения. В России стал употребляться во второй половине XX века. Значение этого термина – постоянное наблюдение за каким-либо процессом, цель которого – выявление соответствия наблюдаемого процесса стандартам, желаемым результатам, первоначальным предположениям. Мониторинг это важнейший инструмент оценки эффективности принятия управленческих решений и оценки устойчивости состояния систем под воздействием внешних факторов. В зависимости от назначения мониторинга и уровня социально-экономической системы, в отношении которой он применяется, изменяется структура мониторинга, порядок его проведения и состав исследуемых показателей [1].

Сегодня актуальной проблемой является проблема анализа взаимоотношений участников социально-экономических и трудовых процессов, протекающих на уровне муниципальных образований. В предлагаемой работе сделан акцент на технологическом обеспечении методик анализа данных социологических опросов, проводимых на уровне муниципальных образований с целью стратегического управления развитием.

Проблема обработки данных мониторинга социально-экономических процессов

Одним из основных инструментов изучения социально-экономических процессов и явлений являются анкетные опросы населения. Профессиональные анкетные опросы, как правило, проводятся с некоторой периодичностью. Поэтому у исследователей возникает необходимость в выполнении расчетов не по всем данным, а по отдельным группам данных.

Анализ данных мониторинга с использованием различных анкет вызывает ряд трудностей технического характера. В настоящей работе мы рассмотрим некоторые дополнительные средства анализа данных мониторинга, позволяющие повысить эффективность работы менеджеров муниципальных образований. Предлагаемые средства предназначены для обработки данных, которые собираются с помощью анкетных форм, а затем вводятся в компьютер для их дальнейшей обработки в EXCEL.

Для обработки данных мониторинга было разработано два программных модуля, которые позволяют производить анализ данных признаков различной природы.

Программные модули включены в состав комплекса программных средств, являющегося надстройкой к EXCEL. При их разработке были учтены некоторые требования, предъявляемые ранее разработанным комплексом программ [2].

Прежде чем перейти к описанию функций разработанных программных модулей, необходимо проанализировать процесс сбора данных в мониторинговых исследованиях и определить некоторые понятия, связанные с ним.

Сначала определимся с формой представления данных. Будем считать, что данные, полученные в результате опроса, заносятся в некоторую таблицу, которую принято называть «объект-свойство». Главной особенностью сбора данных в мониторинговых исследованиях является то, что в каждой анкете должна быть указана дата сбора информации, которая при компьютерной обработке заносится в отдельный столбец таблицы данных.

Чаще всего анкетные данные собираются на бумажном носителе, а затем уже вводятся в компьютер для обработки. В реальных исследованиях ежегодно по одной анкете может опрашиваться значительное количество респондентов (от нескольких сотен до нескольких тысяч). При сборе данных могут быть задействованы множество технических работников, которые непосредственно контактируют с респондентами. К такой работе часто привлекаются студенты.

Сбор данных осуществляется в течении некоторого временного периода (иногда до нескольких месяцев). Считается, что за время сбора данных изменений состояния исследуемой социально-экономической системы не происходит. Социально-экономические системы вообще достаточно инертны.

Необходимо различать понятия «периоды сбора данных» и периоды, на который распространяются полученные оценки состояния системы – «периоды оценки состояния системы». В большинстве случаев, оценки распространяются на календарный год, но могут быть назначены и другие временные рамки.

Данные от технических, осуществляющих сбор анкетных данных, для ввода в компьютер, поступают неравномерно. Каждая анкета должна иметь свой уникальный номер. По этому номеру можно идентифицировать анкету.

На ввод данных в компьютер тоже требуется достаточно много времени. В этой работе может участвовать несколько операторов. Обычно для удобства ввода данных по конкретной анкете создается специализированная форма, облегчающая работу оператора. После ввода в компьютер все данные собираются в единую таблицу, в которой они упорядочиваются по номеру анкеты. Это удобно при сверке компьютерных данных с оригиналом на бумажном носителе с целью обнаружения ошибок. Данные на бумажном носителе сохраняются на весь период, пока существует перспектива сбора информации по данной анкете.

Будем считать, что в конечном итоге будет сформирована некоторая таблица EXCEL, данные которой и нужно обработать. Разработанный программный комплекс требует чтобы «Лист EXCEL», где располагается таблица «объект-свойство», имел название «Данные». Никаких прочих данных или расчетов на листе «Данные» быть не должно, кроме данных таблицы «объект-свойство».

Методика расчета и интерфейс программных модулей

Прежде чем приступить к анализу данных, полученных в процессе мониторинга, желательно проанализировать процесс сбора данных. На первом этапе желательно представить процесс сбора данных в виде диаграммы, например, по месяцам или кварталам. Для расчета графика был разработан специальный программный модуль в среде EXCEL - «Календарный график сбора данных». Интерфейс обращения к программе «Календарный график сбора данных» представлен на рис. 1.

Рассмотрим содержание параметров, которые задаются при обращении к программе:

- «столбец дат» - указывается ссылка на столбец таблицы данных, содержащий даты сбора анкет;
- «начальная дата» – дата, начиная с которой производится расчет графика сбора данных (выбирается из выпадающего списка);
- «конечная дата» – дата, до которой производится расчет графика сбора данных (выбирается из выпадающего списка);
- «шаг по дате» – размер интервалов, по которым рассчитывается календарный график сбора данных (год, квартал, месяц - выбирается из выпадающего списка);

– «результаты разместить» – указывается ячейка листа EXCEL, выбранная для размещения таблицы результатов.

Рис. 1. Интерфейс программы «Календарный график сбора данных»

Для демонстрации результатов работы мы выбрали анкету «Исследование времяпрепровождения отпусков жителями Приморского края». Пример таблицы результатов, рассчитанных с помощью программы «Календарный график сбора данных» по этой анкете представлен на рис. 2, а календарный график, построенный по данным этой таблицы на рис. 3. Из таблицы и графика на рис. 3 можно увидеть, что данные анкет в период наблюдений собрались неравномерно.

Анализ графика процесса сбора данных по анкетному опросу позволяет спланировать работу по решению задач мониторинга исследуемого процесса или явления. График дает информацию о том, каким статистическим материалом располагает исследователь к моменту решения задач мониторинга.

	R	S	T	U	V	W
	Номер интервала	Год	Начальная дата интервала	Конечная дата интервала	Количество анкет	Надпись на оси графика
	1	2008	01.07.2008	31.07.2008	89	июль. 08
	2	2008	01.08.2008	31.08.2008	78	авг. 08
	3	2008	01.09.2008	30.09.2008	39	сен. 08
	4	2008	01.10.2008	31.10.2008	281	окт. 08
	5	2008	01.11.2008	30.11.2008	258	нояб. 08
	40	2011	01.10.2011	31.12.2005	212	окт. 11
	41	2011	01.11.2011	31.03.2006	292	нояб. 11
	42	2011	01.12.2011	30.06.2006	354	дек. 11

Рис. 2. Результаты, полученные с помощью программы «Календарный график сбора данных»

Прежде чем приступить к обработке данных мониторинга, необходимо сопоставить даты сбора данных и периоды оценки состояния системы. Например, для другой анкеты - «Исследование времяпрепровождения отпусков и каникул», сбор данных обычно начинался в конце календарного года и продолжался в течение двух-трех месяцев, а иногда и более. Для выполнения расчетов необходимо определить временные интервалы сбора данных для каждого периода оценки состояния системы.

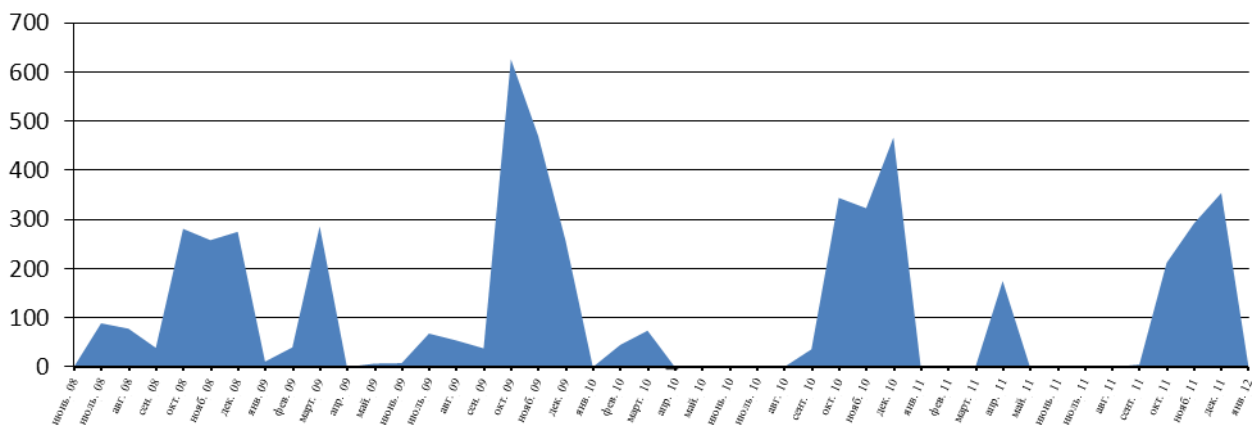


Рис. 3. Временная диаграмма сбора статистических данных

Для исследования времяпрепровождения отпусков было выбрано четыре периода сбора данных и четыре периода оценки состояния системы. Соответствие дат сбора анкетных данных и периодов оценки состояния системы представлено в табл. 1. Как видно из таблицы периоды сбора данных и периоды оценки системы не совпадают.

Таблица 1. Таблица соответствия периодов оценки состояния системы и периодов сбора анкетных данных

Номер периода	Периоды оценки состояния системы	Даты сбора данных		Количество анкет
		Начальная дата периода сбора данных	Конечная дата периода сбора данных	
1	2008 г.	01.07.2008	31.08.2009	1494
2	2009 г.	01.09.2009	31.08.2010	1510
3	2010 г.	01.09.2010	31.08.2011	1345
4	2011 г.	01.09.2011	31.08.2012	863
Итого		01.07.2008	31.08.2012	5212

Расчет частотных рядов какого-либо исследуемого признака (свойства) по периодам оценки состояния системы производится с помощью

специального модуля EXCEL «Мониторинг частотных рядов». Интерфейс программы представлен на рис. 4.

Рис. 4. Интерфейс программы «Мониторинг частотных рядов»

В программе необходимо определить столбец таблицы данных, в котором размещены даты сбора анкет. Затем определяется таблица соответствия «период-дата». В ней задаются даты начала и конца периодов сбора данных (даты сбора данных в табл. 1). Эти данные задаются на любом листе EXCEL, кроме листа «Данные». «Диапазон данных» – это ссылка на один из столбцов таблицы данных, в котором представлены данные исследуемого признака.

С помощью этой программы можно обрабатывать различные типы данных, которые могут встретиться при проведении анкетных опросов. В настоящее время в программе предусмотрена обработка двух типов данных: «числовые данные», «простые текстовые», «составные текстовые». Для различных типов данных используется своя схема расчета. Поэтому при обращении к программе необходимо задать тип данных. Сначала рассмотрим обработку числовых данных. Под этот тип подпадают и ранговые данные.

Далее в программе определяется диапазон ячеек, в которых размещаются «границы интервалов» для расчета частотного ряда исследуемого признака. Считается, что интервалы следуют один за другим. Поэтому задаются только верхние границы интервалов. Сама верхняя граница включается в соответствующий интервал и не входит в следующий интервал. Примером задания верхних границ может служить первый столбец таблицы ре-

зультатов (рис. 5). В этой таблице приведены результаты обработки данных признака по затратам времени на занятие спортом студентами, полученные с помощью программы «Мониторинг частотных рядов».

Параметр «результаты разместить» определяет ячейку листа EXCEL, выбранную для размещения таблицы (рис. 5).

При анализе данных мониторинга может возникнуть необходимо построить серию частотных рядов не по всем данным, а по данным, отвечающим условиям для некоторой группы респондентов, например, только для женщин или только для мужчин. Для определения условий, описывающих такую группу, служит кнопка «установить фильтр на таблицу данных». По этой команде вызывается стандартный «пользовательский автофильтр» EXCEL, настраиваемый пользователем.

При обработке текстовых данных границы интервалов не задаются. Список уникальных текстовых значений для построения частотного ряда определяется программой. Для того, чтобы избежать получения громоздких таблиц, на промежуточной стадии программа сообщает количество найденных уникальных текстовых значений. Пользователь должен определить нужно ли продолжать расчеты или необходимо произвести дополнительные преобразования по унификации данных.

	R	S	T	U	V	W
129						
130		Верхние границы интервалов	Периодов оценки состояния системы			Итого
131			1	2	3	
132			01.07.2004 — 31.12.2006	01.01.2007 — 31.12.2009	01.01.2010 — 31.12.2012	
133		0	205	87	29	321
134		1	72	38	18	128
135		3	240	130	41	411
136		5	95	51	16	162
137		7	67	21	16	104
138		9	15	15	7	37
139		11	26	15	2	43
140		13	12	5	1	18
141		15	11	8	2	21
142		17	5	0	0	5
143		56	18	6	3	27
144		Итого	766	376	135	1277

Рис. 5. Результаты работы программы «Мониторинг частотных рядов»

Кроме простых текстовых данных программа рассчитывает частотные ряды для составных текстовых данных. Составной признак возникает тогда, когда при ответе на вопрос респондент может указать сразу несколько ответов [2]. Например, при ответе на вопрос анкеты: «Какие крупные города вы посетили за последние три года», респондент может указать сразу несколько городов. Таким образом, составной признак состоит из нескольких простых ответов. Для идентификации составного признака в компью-

терном представлении вводится какой-либо единый знак разделителя. Простой ответ может состоять из нескольких слов или даже может быть сформулирован в форме целого предложения. При обработке более сложных случаев необходимо производить предварительную обработку составных признаков, производя переход от неструктурированных данных к структурированным. Методика обработки таких данных представлена в работе [3].

Заключение

Предлагаемая технология предназначена для массовых исследований и при большом количестве исследуемых признаков. В простых случаях отдельные задачи мониторинга можно решить и с помощью стандартного набора инструментальных средств EXCEL (например, с использованием «сводных таблиц»). В различных ситуациях обработки мониторинговых данных придется сделать серию дополнительных операций, формировать промежуточные данные и т.п..

Рассмотренные модули являются составной частью целого комплекса специализированных программных средств обработки анкетных данных. Специализация программных модулей на определенном классе задач позволяет повысить эффективность работы исследователя, существенно сокращая затраты времени на обработку данных. Рассмотренная технология прошла апробацию при обработке данных целой серии массовых социологических опросов, проводимых на территории муниципальных образований Приморского края силами научных подразделений Владивостокского государственного университета экономики и сервиса.

Литература

1. Малышев М. Л. Мониторинг социально-трудовой сферы: Учеб. пособие / М. Л. Малышев. – М.: Издательство РГСУ «Союз»; «Перспектива», 2007. – 276 с.
2. Мартышенко Н.С., Мартышенко С.Н., Кустов Д.А. Совершенствование математического и программного обеспечения обработки первичных данных в экономических и социологических исследований // Вестник Тихоокеанского государственного экономического университета. – 2006. – №2. С. 91–103.
3. Мартышенко Н.С. Исследование структуры потребления туристских продуктов в Приморском крае // Региональная экономика: теория и практика. – 2010. – №26(161). С. 60-68.
4. Мартышенко Н.С. Компьютерная технология обработки качественных данных опросов потребителей туристских услуг // Маркетинг и маркетинговые исследования. – 2011. – №3(93). С. 184–192.