

УДК 303.7:004.9

*Е.Е. Фомина, Н.К. Жиганов***МЕТОДИКА ОБРАБОТКИ РЕЗУЛЬТАТОВ АНКЕТИРОВАНИЯ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МНОГОМЕРНОЙ И ПАРАМЕТРИЧЕСКОЙ СТАТИСТИКИ**

Анкетирование как один из основных методов изучения общественного мнения широко применяется в разных областях, позволяя узнать мнение респондентов на интересующие темы. Результаты первичной обработки анкет представляют собой, как правило, базы данных большого объема, требующие дальнейшего анализа и интерпретации результатов, более глубокой проработки и поиска взаимосвязей между агрегированными переменными и социально-демографическими характеристиками респондента. Для упрощения этой проблемы на практике могут быть применены методы многомерной и параметрической статистики. Данная статья посвящена использованию факторного анализа и критерия хи-квадрат для обработки результатов анкетирования. Предложена методика анализа результатов, включающая в себя два этапа: поиск общих факторов, оказывающих влияние на ответы респондентов, и установление взаимосвязи между найденными факторами и дополнительными переменными (такими как пол, возраст и доход респондента).

Ключевые слова: обработка результатов, факторный анализ, критерий хи-квадрат, таблицы сопряженности, анкетирование.

Введение

К методу анкетирования как популярному инструменту изучения общественного мнения обращаются компании, желающие получить отзыв о своей продукции и результатах своей деятельности; а также научные организации, стремясь узнать оценку эффективности своей научной работы. Анкетирование широко применяется в социологии, психологии, педагогике и других науках. Основное его достоинство в том, что исследователь может одновременно опросить большое количество респондентов в разных регионах и получить сопоставимые данные, которые удобно анализировать с использованием методов математической статистики [1].

Из методов, наиболее часто используемых для обработки результатов анкетирования, отметим несколько: расчет показателей описательной статистики, ранжирование и шкалирование, выявление корреляционной зависимости между отдельными признаками, графическая обработка информации. При этом важную роль играют методы и алгоритмы интерпретации, придающие смысл результатам анкетирования.

Выделим один из них – метод факторного анализа, позволяющий вести поиск скрытой структуры взаимосвязей во множестве переменных, значительно сокращать их объем и выделять факторы, влияющие на ответы респондентов [2; 3]. Преимущество данного метода заключается также в том, что он реализован во многих пакетах обработки данных, в частности, в таких широко известных, как SPSS и STATISTICA [4].

Однако, как показывает анализ литературы, в большинстве случаев обработка результатов заканчивается этапом выделения и интерпретации факторов, тогда как в ряде задач логика исследования требует более глубокой проработки и поиска взаимосвязей между найденными факторами и социально-демографическими характеристиками респондента, такими как, например, пол, образование, уровень дохода.

В связи с этим актуальной становится задача разработки эффективных методик анализа результатов анкетирования. В статье предлагается такая методика. Она базируется на методах многомерной и параметрической статистики, в частности – на методе факторного анализа и критерии хи-квадрат.

Материалы и методы исследования

Комбинация метода факторного анализа и процедуры проверки гипотезы о независимости признаков с использованием критерия хи-квадрат была применена к изучению общественного мнения на тему «Отношение тверских студентов к институту выборов».

Сбор информации осуществлялся через распространение стандартизированных анкет со следующими вопросами:

1. Интересуетесь ли Вы политикой? 1) Да; 2) Нет.
2. Принимаете ли Вы участие в общественной жизни? 1) Да; 2) Нет.
3. Доверяете ли Вы результатам выборов различного уровня, проходящих в нашей стране?

	Уровень	Да	Скорее да	Скорее нет	Нет	Не знаю
3.1	Президент	1	2	3	4	5
3.2	Государственная Дума	1	2	3	4	5
3.3	Городская дума	1	2	3	4	5
3.4	Законодательное собрание	1	2	3	4	5
3.5	Мэр города	1	2	3	4	5

4. Какой смысл Вы вкладываете в понятие «выборы»?
 - 1) Это – почетное право гражданина, гарантированное конституцией.
 - 2) Это – главный признак демократического государства.
 - 3) Это – возможность граждан влиять на власть.
 - 4) Это – пустая формальность, создающая видимость демократии.
 - 5) Затрудняюсь ответить.
5. Как бы Вы оценили степень развитости в нашей стране института выборов?
 - 1) Институт выборов в нашей стране развит хорошо.
 - 2) Институт выборов в нашей стране еще не достаточно развит.
 - 3) Институт выборов в нашей стране развит слабо.
 - 4) Институт выборов в нашей стране не развит совсем.
 - 5) Мне сложно об этом судить.
6. Участвовали ли Вы в последних выборах в различные властные структуры?

	Уровень	Да	Нет	Не помню
6.1	Президент	1	2	3
6.2	Государственная Дума	1	2	3
6.3	Городская дума	1	2	3
6.4	Законодательное собрание	1	2	3
6.5	Мэр города	1	2	3

7. Укажите Ваш пол:
 - 1) Мужской; 2) Женский.
8. Укажите Вашу специальность _____.
9. К какой из перечисленных групп населения по уровню материального положения Вы могли бы себя отнести?
 - 1) Мы едва сводим концы с концами. Денег не хватает даже на питание.
 - 2) На продукты денег хватает, но покупка одежды вызывает затруднения.
 - 3) Денег хватает на продукты и одежду, но покупка вещей длительного пользования (телевизор, холодильник) вызывает затруднения.
 - 4) Мы можем приобретать вещи длительного пользования, но не дорогие покупки (машины).
 - 5) Без труда приобретаем вещи длительного пользования. Можем накопить деньги на недорогую машину, но купить квартиру или коттедж не можем.
 - 6) У нас нет материальных проблем. Можем приобрести недвижимость, дорогой автомобиль, дорогой заграничный тур.

Анкета была разработана на кафедре «Социологии и социальных технологий» Тверского государственного технического университета.

В анкетировании приняли участие 100 респондентов.

Предлагаемая методика обработки включает в себя два этапа.

На первом этапе к результатам анкетирования применяется метод факторного анализа, позволяющий выделить агрегатные переменные, влияющие на ответы респондентов. В рассматриваемом примере в качестве исходных данных для факторного анализа выступала таблица с данными, включающими в себя ответы на вопросы 1–6.

На втором этапе исследования анализируются взаимосвязи выделенных факторов с дополнительными социально-демографическими переменными, что позволяет получить представление о том, какие значения факторных переменных наиболее ярко выражены в той или иной социально-демографической группе, и проверить гипотезу о независимости признаков.

Автоматизированная обработка полученных данных осуществлялась в пакете STATISTICA.

Результаты

Для выявления основных факторов, обуславливающих отношение респондентов к институту выборов, применялся факторный анализ с использованием метода максимального правдоподобия.

Цель первого шага метода – решение вопроса об оптимальном количестве факторов. В итоге выделены два фактора (два собственных числа больше единицы [2]), влияющих на результаты анкетирования (табл. 1).

Таблица 1

Собственные значения

Значение	Собственное значение	Процент общей дисперсии	Кумулятивный процент
1	3,69	26,40	26,40
2	2,50	17,86	44,26

Как можно видеть из табл. 1, процент дисперсии, объясняемой первым генеральным фактором, составил 26,40; вторым фактором – 17,86. Общий процент дисперсии, объясняемый всеми факторами, равен 44,26.

Следующий шаг исследования – расчет и анализ факторных нагрузок. Для более наглядной интерпретации решения был применен метод вращения Варимакс исходных, позволивший проследить четкую факторную структуру и выделить переменные, отмеченные высокими значениями коэффициентов корреляции с тем или иным фактором. Корреляция считалась сильной, если ее коэффициент принимал значение более 0,70. Результаты расчетов представлены в табл. 2. Факторные нагрузки, величина которых мене 0,30, в табл. 2 не указаны.

Таблица 2

Факторные нагрузки, полученные с использование метода максимального правдоподобия и метода вращения Варимакс исходных

Переменная	Фактор 1	Фактор 2
Вопрос 1		
Вопрос 2		
Вопрос 3.1	0,77	
Вопрос 3.2	0,87	
Вопрос 3.3	0,92	
Вопрос 3.4	0,87	
Вопрос 3.5	0,79	
Вопрос 4		0,30
Вопрос 5	0,36	
Вопрос 6.1		0,75
Вопрос 6.2		0,87
Вопрос 6.3		0,71
Вопрос 6.4		0,70
Вопрос 6.5		0,70

Анализ табл. 2. показывает, что первый, то есть генеральный, фактор, включает в себя 5 переменных с положительными значениями коэффициента корреляции: вопрос 3.1–3.5. При ответе на эти вопросы респонденты выражали свое доверие к результатам выборов различного уровня, проходящих в нашей стране. Таким образом, первый фактор можно охарактеризовать как «Доверие». В целом объяснительная «сила» фактора, или факторный вес, составил 26,40 %. С содержательной точки зре-

ния, полученные результаты не противоречат здравому смыслу, то есть степень доверия респондентов к результатам выборов в различные властные структуры скоррелирована.

Второй фактор тесно связан с вопросами 6.1–6.5, в которых респонденты отмечали свою активность на выборах. Условно фактор можно обозначить как «Активность». Его факторный вес составил 17,86 %.

Помимо вопросов, раскрывающих отношение к выборам, в анкете содержалась дополнительная информация, касающаяся пола, образования и дохода респондента. Следующим шагом анализа было установление взаимосвязи между полученными факторами и дополнительными переменными.

Для решения этой задачи применялись таблицы сопряженности и критерий хи-квадрат, позволяющий проверить гипотезу о независимости признаков. Факторные значения 1-го и 2-го факторов были разбиты на 4 группы, в результате чего были получены две новые факторные переменные (обозначим их f_1 и f_2), содержащие значения от 1 до 4.

Для 1-го фактора значение 1 соответствует позиции «сильное доверие»; значение 2 – позиции «скорее доверяю»; значение 3 позиции «скорее не доверяю»; значение 4 позиции «не знаю».

Для 2-го фактора значение 1 соответствует позиции «да»; значение 2 – позиции «не во всех»; значение 3 – позиции «нет»; значение 4 – позиции «не помню».

С использованием инструмента таблицы сопряженности программы STATISTICA было получено совместное распределение таких переменных, как пол – f_1 , образование – f_1 , доход – f_1 , пол – f_2 , образование – f_2 , доход – f_2 , послужившее в качестве исходных данных для расчета критерия хи-квадрат и, кроме того, позволило проанализировать, какие значения факторных переменных f_1 и f_2 наиболее ярко выражены для каждой социально-демографической группы.

Применение критерия хи-квадрат дало возможность проверить гипотезу о независимости признаков пол – f_1 , образование – f_1 , доход – f_1 , пол – f_2 , образование – f_2 , доход – f_2 .

Для примера в статье приводятся 4 таблицы сопряженности, отражающие распределение переменных пол – f_1 и доход – f_1 , пол – f_2 и доход – f_2 (табл. 3–6). *Примечание:* в табл. 4, 6 присутствует только уровень дохода, соответствующий категориям 3, 4 и 5, так как респондентов с уровнем дохода из 1 и 6 категорий в анкетировании не было.

Таблица 3

Таблица сопряженности (пол – f_1)

Пол	F1 (Фактор 1) хи-квадрат Пирсона: 4,2; $p=0,23$, $\alpha=0,05$				Всего
	1 «сильное доверие»	2 «скорее доверяю»	3 «скорее не доверяю»	4 «не знаю»	
1	3	22	20	5	50
2	6	26	11	7	50
Всего	9	48	31	12	100

Таблица 4

Таблица сопряженности (доход – f_1)

Доход	F1 (Фактор 1) хи-квадрат Пирсона: 9,62; $p=0,14$, $\alpha=0,05$				Всего
	1 «сильное доверие»	2 «скорее доверяю»	3 «скорее не доверяю»	4 «не знаю»	
3	4	19	11	4	38
4	2	25	11	7	45
5	3	4	9	1	17
Всего	9	48	31	12	100

Анализ табл. 3 и 4 показывает, что в целой выборке наиболее ярко выражено 2-е значение («скорее доверяю» результатам выборов) факторной переменной f_1 (48 % респондентов).

Среди респондентов мужского пола наиболее часто присутствуют варианты «скорее доверяю» или «скорее не доверяю» (табл. 3), оба встречаются приблизительно с одинаковой частотой. Среди

респондентов женского пола значение «скорее доверие» выражено ярче (табл. 3). Тем не менее, тест по критерию хи-квадрат говорит о незначительном различии уровня доверия избирателей в зависимости от пола ($p=0,23 > \alpha=0,05$).

Данные табл. 4 показывают, что в группах респондентов с уровнем дохода из 3 и 4 категории наиболее часто встречается «скорее доверие» к результатам выборов. Среди респондентов с уровнем дохода из 5 категории наиболее часто присутствует вариант «скорее не доверяю». Тест по критерию хи-квадрат говорит о незначительном различии уровня доверия избирателей в зависимости от дохода ($p=0,14 > \alpha=0,05$).

Таблица 5

Таблица сопряженности (пол – f2)

Пол	F2 (Фактор 2) хи-квадрат Пирсона: 3,2; $p=0,36$, $\alpha=0,05$				Всего
	1 «да»	2 «не во всех»	3 «нет»	4 «не знаю»	
1	5	20	13	12	50
2	9	21	14	6	50
Всего	14	41	27	18	100

Таблица 6

Таблица сопряженности (доход – f2)

Доход	F2 (Фактор 2) хи-квадрат Пирсона: 4,83; $p=0,56$, $\alpha=0,05$				Всего
	1 «да»	2 «не во всех»	3 «нет»	4 «не знаю»	
3	8	17	7	6	38
4	5	18	14	8	45
5	1	6	6	4	17
Всего	14	41	27	18	100

Анализируя табл. 5 и 6, можно прийти к выводу, что в целой выборке наиболее ярко выражена позиция 2, т.е. большая часть респондентов участвует не во всех выборах в различные властные структуры (41 % респондентов).

Респонденты мужского пола в большинстве случаев либо принимали участие не во всех выборах, либо не принимали участие вовсе, либо не помнят о своем посещении избирательного участка. Среди респондентов женского пола частичное участие в выборах выражено ярче (табл. 5). В свою очередь тест по критерию хи-квадрат говорит о незначительном различии активности избирателей в зависимости от пола ($p=0,36 > \alpha=0,05$).

В группе респондентов с уровнем дохода из 3-й категории преобладает позиция 2 (участие не во всех выборах), а в группе респондентов с уровнем дохода из 4-й и 5-й категорий позиции 2 и 3 встречаются с почти одинаковой частотой (табл. 6). Тест по критерию хи-квадрат говорит о незначительном различии активности избирателей в зависимости от уровня дохода ($p=0,56 > \alpha=0,05$).

Выводы

В статье предложена и проиллюстрирована методика обработки результатов анкетирования, основанная на применении методов факторного анализа и критерия хи-квадрат, позволяющая выделять из всего объема вопросов общие агрегированные переменные и устанавливать взаимосвязи между этими переменными и дополнительными характеристиками респондента, такими, как пол, доход и др. Рассмотрен пример обработки и анализа анкет, разработанных с целью получения общественного мнения на тему «Отношение тверских студентов к институту выборов». Построена двухфакторная модель; на основе применения критерия хи-квадрат сделаны выводы о взаимосвязи между выделенными факторами и социально-демографическими характеристиками респондента.

СПИСОК ЛИТЕРАТУРЫ

1. URL: <http://cito-web.yspu.org/link1/metod/met93/node4.html>.
2. Буреева Н.Н. Многомерный статистический анализ с использованием ППП «STATISTICA». Учеб.-метод. материал по программе повышения квалификации «Применение программных средств в научных исследованиях и преподавании математики и механики». Нижний Новгород, 2007. 112 с.
3. Самсонова А.В. Факторный анализ. Направления применения и неиспользованные возможности // Вестник Балтийской Педагогической Академии. Вып. 62. 2005. С. 67-75.
4. Фомина Е.Е. Обзор программных продуктов, используемых при проведении социологических исследований // Социосфера. 2016. № 2. С. 99-102.

Поступила в редакцию 16.12.16

E.E. Fomina, N.K. Zhiganov

THE METHODOLOGY OF PROCESSING OF SURVEY RESULTS WITH THE USE OF MULTIVARIATE AND PARAMETRIC STATISTICS

The survey is one of the main methods of public opinion research. It is widely used in different areas and allows one to ascertain the opinion of the respondents on the topics of interest. The results of primary processing of questionnaires are usually a large database requiring further analysis and interpretation of the results, a deeper study and search for the relationship between aggregate variables and socio-demographic characteristics of the respondent. In practice, to solve this problem, the methods of multivariate and parametric statistics can be applied. This article is devoted to the application of factor analysis and chi-square test for processing of survey results. The article proposes a method of analysis of the results, which includes two phases: the search for common factors influencing the respondents' answers, and establishing the relationship between the factors and additional variables, such as gender, age and income of the respondent.

Keywords: processing of results, factor analysis, chi-square test, contingency tables, survey.

Фомина Елена Евгеньевна,
кандидат технических наук, доцент
E-mail: f-elena2008@yandex.ru

Жиганов Николай Константинович,
доктор физико-математических наук, профессор
E-mail: jjiganov2005@yandex.ru

Тверской государственный технический университет
170020, Россия, г. Тверь, наб. Афанасия Никитина, 22

Fomina E.E.,
Candidate of Technical Sciences, Associate Professor
E-mail: f-elena2008@yandex.ru

Zhiganov N.K.,
Doctor of Physics and Mathematics, Professor
E-mail: jjiganov2005@yandex.ru

Tver State Technical University
Nab. Afanasiya Nikitina, 22, Tver, Russia, 170020