

УДК 316:519.2

**ПОДГОТОВКА И АНАЛИЗ РЕЗУЛЬТАТОВ АНКЕТИРОВАНИЯ
С ПРИМЕНЕНИЕМ МАТЕМАТИЧЕСКИХ МЕТОДОВ****Е. Е. Фомина***Кандидат технических наук, доцент,
ORCID 0000-0002-1028-0750,
e-mail: f-elena2008@yandex.ru,
Тверской государственный
технический университет,
г. Тверь, Россия***PREPARATION AND ANALYSIS OF THE SURVEY RESULTS
WITH THE APPLICATION OF MATHEMATICAL METHODS****E. E. Fomina***Candidate of technical sciences, associate professor,
ORCID 0000-0002-1028-0750,
e-mail: f-elena2008@yandex.ru,
Tver State Technical University,
Tver, Russia*

Abstract. The survey is used by sociologists as the main tool for monitoring the state of public opinion. Development of reliable and valid questionnaires, the correct interpretation of the results on the basis of which can make important decisions is an actual problem. To solve this problem can be used the mathematical methods, which will optimize all stages of the survey. This article is devoted to the review of methods that can be used to check the validity and reliability of the measuring scale and methods for primary processing of questionnaires meaningful, methods for interpretation of the survey results and methods for analysis of contingency tables.

Keywords: survey; reliability of the scale; validity of the scale; processing of the survey results; loglinear analysis; factor analysis; CatPCA.

1. Введение

Метод анкетирования – основной инструмент социолога, предназначенный для получения статических и динамических статистических представлений о состоянии общественного мнения по тем или иным вопросам. Метод позволяет проводить массовые исследования за достаточно короткий промежуток времени с минимальными затратами. Неточности и ошибки, допущенные на этапе проектирования анкеты и на этапе анализа результатов, могут привести к некорректным результатам. Внедрение математического аппарата, позволяющего оптимизировать процедуры подготовки и интерпретации результатов анкетирования, является актуальной задачей. Цель настоящей статьи – обзор основных математических методов, используемых на всех этапах анкетирования.

2. Математические методы, используемые на этапе подготовки измерительной шкалы

Для того, чтобы измерительная шкала (список вопросов анкеты) была пригодна для использования она должна быть валидной и надежной.

Под валидностью шкалы понимается однозначность и правильность получаемых результатов относительно измеряемого свойства объектов, т. е. относительно предмета измерения [2]. Для оценки валидности шкалы (в частности критериальной и конструктивной) используется метод корреляционного анализа.

Под надежностью шкалы понимается воспроизводимость получаемых результатов измерения, их устойчивость к действию случайных факторов. Надежная шкала не должна содержать неясно или

неоднозначно сформулированных вопросов. Вопросы, связанных с колебаниями внимания респондентов [2].

Основные математические способы оценки надежности измерительных шкал делятся на следующие группы:

1. Проверка внутренней согласованности между утверждениями, образующими шкалу путем расчета статистики α Кронбаха [4]. Статистика α Кронбаха показывает, насколько правильно составлены вопросы анкеты с точки зрения их взаимосогласованности. Значение α может находиться в пределах от 0 до 1. Надежные измерительные шкалы, используемые при массовых социологических опросах должны иметь $\alpha \geq 0,8$ по зарубежным нормам и $\alpha \geq 0,7$ по отечественным нормам. Для разведочного исследования возможно отсечение $\alpha \geq 0,6$ [5, 6].

Преимущество статистики α Кронбаха – простота расчета и удобство интерпретации результата. Основной недостаток – прямо пропорциональное увеличение значения с ростом числа вопросов, что может привести к ошибочным положительным результатам для плохо структурированной шкалы [4].

2. Расщепление шкалы и анализ корреляции между ее частями. Согласно процедуре расщепления шкала разбивается на две половины и исследуется корреляция между ними. Если они коррелируют между собой, исходная шкала считается надежной.

Для анализа надежности в этом случае рассчитываются коэффициенты Спирмена-Брауна и лямбда Гуттмана. Нужно отметить, что на коэффициент надежности Спирмена-Брауна сильно влияют способы разделения пунктов на две половины, которое предпочтительно осуществлять случайным образом, чтобы получить равенство дисперсий двух частей опросника. В свою очередь коэффициент лямбда Гуттмана не требует равенства дисперсий. К тому же его значения не возрастают прямо пропорционально длине опросника,

что позволит более адекватно оценить результат надежности [4]. Значения коэффициентов Спирмена-Брауна и Гуттмана большие 0,90 говорят об отличной надежности; значения в промежутке от 0,8 до 0,89 – о хорошей надежности; значения от 0,7 до 0,79 – об удовлетворительной надежности [4, с. 14].

3. Проверка ретестовой надежности, которая базируется на сравнении и анализе результатов измерений с использованием одной и той же шкалы на одной и той же выборке респондентов в разные промежутки времени. Шкала считается надежной, если ее устойчивость достаточно высокая, т.е. имеет максимальное количество совпадения между первой и второй сериями измерений. Проверка устойчивости может быть произведена путем расчета коэффициента Спирмена-Брауна [4] и коэффициента устойчивости по каждому вопросу [3].

К существенным недостаткам методов можно отнести сложности, связанные с повторным опросом той же выборки респондентов, который не всегда возможен при проведении исследования.

3. Математические методы, используемые на этапе анализа результатов анкетирования

Первым этапом анализа является расчет показателей описательной статистики по вопросам анкеты; подсчет распределения ответов в зависимости от значений дополнительных переменных. Этот этап позволяет получить первоначальное представление об имеющихся тенденциях и предпочтениях респондентов [9].

Второй этап – анализ частотных таблиц или таблиц сопряженности между значениями переменных.

Первичным результатом обработки анкет является база данных, поля которой содержат варианты ответов на те, или иные вопросы. В этом случае столбцы с вариантами ответов выступают в качестве анализируемых переменных, измеренных, в большинстве случаев, в категориальной

шкале. Для исследования взаимосвязи между значениями двух переменных используется критерий хи-квадрат, который позволяет проверить гипотезу о независимости признаков для двух номинальных переменных [11]. В случае анализа большого числа переменных, расчет статистики хи-квадрат для нескольких десятков пар признаков – трудоемкая операция. Для её решения может использоваться метод логарифмического линейного анализа, позволяющий проверить статистическую значимость переменных и их совместных эффектов, присутствующих в таблице сопряженности.

Логлинейная модель – линейная модель множественной регрессии, в которой зависимая переменная – натуральный логарифм соответствующей частоты таблицы сопряженности, а независимые переменные – категориальные переменные (или факторы) и их совместные взаимодействия. Например, логлинейная модель, включающая три переменные А, В, С, имеет

$$\ln(n_{ijk}^*) = u_0 + u_i^A + u_j^B + u_k^C + u_{ij}^{AB} + u_{ik}^{AC} + u_{jk}^{BC} + u_{ijk}^{ABC}$$

[1, 8].

Если модель включает все факторы и их возможные взаимодействия, то она является насыщенной. Модель, содержащая только часть значимых факторов и их взаимодействий, называется ненасыщенной. На практике насыщенные модели не всегда бывают оптимальными. Основная задача метода заключается в выборе такой ненасыщенной модели, которая адекватно описывает исходные данные и содержит наименьшее возможное число значимых эффектов. Для определения значимости эффекта его исключают из модели и рассчитывают статистики хи-квадрат Пирсона и хи-квадрат метода максимального правдоподобия с уровнем значимости p . Если $p < 0,05$, то эффект считается значимым и должен присутствовать в модели. В противном случае, эффект считается незначимым и может быть исключен из модели.

Метод логарифмического линейного анализа включает в себя следующие шаги: построение модели, адекватно описывающей исходные данные: для этого исследуется насыщенная модель, из которой постепенно исключаются сложные составные взаимодействия до тех пор, пока не останутся только значимые эффекты и интерпретация построенной модели.

Третий этап – содержательная интерпретация результатов. На этом этапе могут использоваться факторный анализ и категориальный метод главных компонент (CatPCA), которые позволяют осуществлять поиск закономерностей во множестве анализируемых переменных и выявлять скрытые факторы, оказывающие влияние на результаты анкетирования.

Методы базируются на предположении, что исследуемое явление, описываемое системой признаков, может быть описано с помощью меньшего числа латентных (скрытых) переменных, называемых факторами, которые объясняют причины этих изменений [10, 12, 13]. Факторы – группы определенных переменных, коррелирующих между собой больше, чем с переменными, входящими в другой фактор.

Процедура факторного анализа состоит из следующих этапов: построение корреляционной матрицы системы переменных; извлечение факторов и расчет факторных нагрузок; вращение факторного решения, позволяющее провести более наглядную интерпретацию результата.

Основное ограничение метода связано со шкалой измерения исходных данных. Факторный анализ может применяться для исследования интервальных переменных или порядковых переменных с большим числом значений [7]. Однако в анкетах это требование может нарушаться. В этом случае альтернативой факторному анализу выступает CatPCA.

Метод CatPCA не накладывает никаких ограничений на исходные данные и обладает большими возможностями, в частности, при обработке результатов ан-

кетирования. Обработка данных методом CatPCA включает в себя два этапа. На первом этапе происходит процедура «оцифровки» переменных, которая опирается на принципы оптимального шкалирования; на втором этапе выполняется редукция размерности данных [12].

4. Выводы

В статье рассмотрены основные математические методы, которые могут применяться при проектировании измерительной шкалы и для анализа результатов анкетирования. Описана область применения каждого метода, его возможности и ограничения.

Библиографический список

1. Аптон Г. Анализ таблиц сопряженности. – М. : Финансы и статистика. – 1982. – 143 с.
2. Девятко И. Ф. Методы социологического исследования. – Екатеринбург : Издательство Уральского университета. – 1998. – 208 с.
3. Мазорчук М. С., Добряк В. С., Базилевич К. А., Хазай М. Ю. Оценка эффективности одномерных шкал при анализе данных в социологических исследованиях. // Радиотехника и компьютерные системы. – 2013. – № 1 (60). – С. 99–104.
4. Мазорчук М. С., Соколова Е. О., Добряк В. С., Сухобрус А. А. Обоснование выбора методов измерения надежности педагогических тестов // Радиотехника и компьютерные системы. – 2014. – № 4 (68). – С. 131–137.
5. Митина О. В. Альфа Кронбаха: когда и зачем ее считать // Современная психодиагностика России. Преодоление кризиса: сборник материалов III Всероссийской конференции: в 2 т. / редколлегия: Н. А. Батурин (отв. ред.) и др. – Челябинск: Издательский центр ЮУрГУ. – 2015. – Т. 1. – С. 232–240.
6. Наследов А. IBM SPSS Statistics 20 и AMOS: профессиональный статистический анализ данных. – СПб. : Питер, 2013. – 416 с.
7. Толстова Ю. Н. Измерение в социологии. – М. : КДУ, 2007. – 288 с.
8. Трофимов Д. А. Логлинейный анализ таблиц мобильности: обзор основных моделей // Социология. – 2008. – № 26. – С. 119–138.
9. Фомина Е. Е. Обзор программных продуктов, используемых при проведении социологических исследований // Социосфера. – 2016. – № 2. – С. 99–102.
10. Фомина Е. Е. Применение факторного анализа для обработки результатов анкетирования // Социосфера. – 2016. – № 3. – С. 122–127.
11. Фомина Е. Е., Жиганов Н. К. Математические методы анализа данных в социологии с использованием пакетов MS Excel и Statistica. Тверь. – 2017. – 68 с.
12. Фомина Е. Е. Факторный анализ и категориальный метод главных компонент: сравнительный анализ и практическое применение для обработки результатов анкетирования // Гуманитарный вестник. – 2017. – № 10 (60). – С. 3.
13. Фомина Е. Е., Жиганов Н. К. Методика обработки результатов анкетирования с использованием методов многомерной и параметрической статистики // Вестник Пермского национального исследовательского политехнического университета. Социально-экономические науки. – 2017. – № 1. – С. 106–115.
14. Халафян А. А., Темердашев З. А., Якуба Ю. Ф., Киселева Н. В., Гугучкина Т. И., Антоненко Т. И. Позиционный анализ как метод оценки согласованности экспертных оценок // Заводская лаборатория. Диагностика материалов. – 2015. – № 12. – Том 81. – С. 69–78.

Bibliograficheskij spisok

1. Apton G. Analiz tablicz sopryazhennosti. – M. : Finansy` i stati-stika. – 1982. – 143 s.
2. Devyatko I. F. Metody` sociologicheskogo issledovaniya. – Ekate-rinburg : Izdatel'stvo Ural'skogo universiteta. – 1998. – 208 s.
3. Mazorchuk M. S., Dobryak V. S., Bazilevich K. A., Hazaj M. Yu. Ocenka e`ffektivnosti odnomerny`x shkal pri analize danny`x v sociologicheskix issledovaniyax. // Radioe`lektronika i komp`yuterny`e sistemy`. – 2013. – № 1 (60). – S. 99–104.
4. Mazorchuk M. S., Sokolova E. O., Dobryak V. S., Suxobrus A. A. Obosnovanie vy`bora metodov izmereniya nadezhnosti pedagogicheskix testov // Radioe`lektronika i komp`yuterny`e sistemy`. – 2014. – № 4 (68). – S. 131–137.
5. Mitina O. V. Al`fa Kronbaxa: kogda i zachem ee schitat` // Sovremennaya psixodiagnostika Rossii. Preodolenie krizisa: sbornik materialov III Vserossijskoj konferencii: v 2 t. / redkollegiya: N. A. Baturin (otv. red.) i dr. – Chelyabinsk: Izdatel'skij centr YuUrGU. – 2015. – T. 1. – S. 232–240.
6. Nasledov A. IBM SPSS Statistics 20 i AMOS: professional`ny`j statisticheskij analiz danny`x. – SPb. : Piter, 2013. – 416 s.
7. Tolstova Yu. N. Izmerenie v sociologii. – M. : KDU, 2007. – 288 s.

8. Trofimov D. A. Loglinejny`j analiz tablicz mobil`nosti: obzor osnovny`x modelej // Sociologiya. – 2008. – № 26. – S. 119–138.
9. Fomina E. E. Obzor programmny`x produktov, ispol`zuemy`x pri provedenii sociologicheskix issledovanij // Sociosfera. – 2016. – № 2. – S. 99–102.
10. Fomina E. E. Primenenie faktornogo analiza dlya obrabotki rezul`tatov anketirovaniya // Sociosfera. – 2016. – № 3. – S. 122–127.
11. Fomina E. E., Zhiganov N. K. Matematicheskie metody` analiza danny`x v sociologii s ispol`zovaniem paketov MS Excel i Statistica. Tver`. – 2017. – 68 s.
12. Fomina E. E. Faktorny`j analiz i katedral`ny`j metod glavny`x komponent: sravnitel`ny`j analiz i prakticheskoe primenenie dlya obrabotki rezul`tatov anketirovaniya // Gumanitarny`j vestnik. – 2017. – № 10 (60). – S. 3.
13. Fomina E. E., Zhiganov N. K. Metodika obrabotki rezul`tatov anketirovaniya s ispol`zovaniem metodov mnogomernoj i parametricheskoj statistiki // Vestnik Permskogo nacional`nogo issledovatel`skogo politexnicheskogo universiteta. Social`no-e`konomicheskie nauki. – 2017. – № 1. – S. 106–115.
14. Xalafyan A. A., Temerdashev Z. A., Yakuba Yu. F., Kiseleva N. V., Guguchkina T. I., Antonenko T. I. Pozicionny`j analiz kak metod ocenki soglasovannosti e`kspertny`x ocenok // Zavodckaya labopatopiya. Diagnostika materialov. – 2015. – № 12. – Tom 81. – S. 69–78.

© Фомина Е. Е., 2018.