DOI 10.15593/2224-9354/2018.3.16
УДК 303.621.3-047.44

Е.Е. Фомина

ВОЗМОЖНОСТИ ЛОГЛИНЕЙНОГО АНАЛИЗА ПРИ ОБРАБОТКЕ РЕЗУЛЬТАТОВ АНКЕТИРОВАНИЯ

Анкетирование представляет собой основной инструмент, предназначенный для изучения общественного мнения в работе социолога. Обработка базы данных с результатами анкетирования осуществляется с применением математических методов, среди которых можно отметить факторный анализ, категориальный метод главных компонентов, расчет показателей описательной статистики. Одним из наиболее часто используемых методов является анализ частотных таблиц. Для исследования зависимости в двумерных частотных таблицах, т.е. зависимости между двумя переменными, рассчитывается значение критерия хи-квадрат. Однако особый интерес представляют многомерные таблицы сопряженности с числом категорий переменных, большим двух. Для изучения сложных внутренних взаимосвязей в таких таблицах может использоваться логарифмический линейный анализ. Задача настоящей работы – продемонстрировать возможности метода логлинейного анализа при обработке анкет. Метод применен для обработки результатов анкетирования по вопросу притока мигрантов в РФ. Исследовано влияние пола, возраста, уровня дохода и типа населенного пункта на отношение к притоку иностранных мигрантов. Построена оптимальная ненасыщенная модель, включающая в себя пять значимых взаимодействий. Адекватность модели установлена проверкой критериев хи-квадрат Пирсона и хи-квадрат метода максимального правдоподобия.

Ключевые слова: *логлинейный анализ, частотные таблицы, анкетирование.*

Введение. Метод анкетирования представляет собой основной инструмент изучения состояния общественного мнения. С помощью анкетирования можно проанализировать отношение различных групп респондентов к определенным событиям и процессам, выявить их интересы и потребности.

Нужно отметить, что изучение общественного мнения имеет не только практическое, но и теоретическое значение, способствует развитию наук об обществе [1].

Первичный результат обработки анкет представляет собой базу данных, поля которой содержат варианты ответов на вопросы. Дальнейшая обработка результатов осуществляется с использованием математических методов и моделей [2–4]. В этом случае столбцы с вариантами ответов выступают в качестве анализируемых переменных, измеренных, как правило, в категориальной шкале. Среди наиболее часто используемых методов можно отметить расчет показателей описательной статистики; составление таблиц сопряженности (частотных таблиц) между значениями двух переменных и вычисление

© Фомина Е.Е., 2018

Фомина Елена Евгеньевна – канд. техн. наук, доцент кафедры информатики и прикладной математики, ФГБОУ ВО «Тверской государственный технический университет», e-mail: f-elena2008@yandex.ru.

критерия хи-квадрат для изучения зависимости между ними [5, 6]. В случае анализа большого набора переменных-столбцов базы данных с результатами анкетирования расчет статистики хи-квадрат для нескольких десятков пар признаков – трудоемкая операция, поэтому актуальная задача – комплексная оценка взаимосвязей в многомерных таблицах частот, т.е. изучение зависимости между несколькими переменными одновременно.

Для решения этой задачи может использоваться метод логарифмического линейного анализа (логлинейного анализа – ЛЛА), который позволяет проверить статистическую значимость факторов и их совместных эффектов, присутствующих в таблице сопряженности.

Основным условием применения метода является независимость наблюдений и категориальная шкала анализируемых переменных. Таким образом, применение ЛЛА является оправданным для обработки анкет. К сожалению, метод ЛЛА не часто используется в социологических исследованиях [7].

Задача данного исследования – продемонстрировать основные возможности метода ЛЛА при обработке результатов анкетирования.

Метод исследования. Математическая модель логлинейного анализа. Логлинейная модель (ЛЛМ) – линейная модель множественной регрессии, где зависимая переменная – это натуральный логарифм соответствующей частоты таблицы сопряженности, а независимые переменные (предикторы) – категориальные переменные (или факторы) и их совместные взаимодействия [8, 9].

Рассмотрим ЛЛМ, которая включает в себя три переменные (три фактора) A, B и C. Переменная A может принимать I значений, переменная B – J значений, переменная C – K значений.

Информация о совместном распределении переменных представлена в трехмерной таблице сопряженности, ячейки которой содержат наблюдаемые частоты n_{ijk}, где $i = 1,2, \ldots, I$, $j = 1,2, \ldots, J$, $k = 1,2, \ldots K$.

В этом случае логлинейная модель имеет вид [10, 11]

$$\ln(n^*_{ijk}) = u_0 + u_i^{\mathrm{A}} + u_j^{\mathrm{B}} + u_k^{\mathrm{C}} + u_{ij}^{\mathrm{AB}} + u_{ik}^{\mathrm{AC}} + u_{jk}^{\mathrm{BC}} + u_{ijk}^{\mathrm{ABC}}, \qquad (1)$$

где n^*_{ijk} – теоретические частоты;

u_0 – общее среднее значение;

u_i^{A} – эффект i-й категории переменной A, $i = 1,2, \ldots, I$;

u_j^{B} – эффект j-й категории переменной B, $j = 1,2, \ldots, J$;

u_k^{C} – эффект k-й категории переменной C, $k = 1,2, \ldots, K$;

u_{ij}^{AB} – эффект совместного взаимодействия i и j категории двух переменных A и B соответственно;

u_{ik}^{AC} – эффект совместного взаимодействия i и k категории двух переменных A и C соответственно;

u_{jk}^{BC} – эффект совместного взаимодействия j и k категории двух переменных B и C соответственно;

u_{ij}^{AB} – эффект совместного взаимодействия i, j и k категории трех переменных A, B и C соответственно.

Если какая-либо ячейка таблицы сопряженности окажется равной нулю, то ее значение заменяется на 0,5, что не влияет на конечный результат.

Чтобы число параметров не превышало числа ячеек таблицы сопряженности, должны выполняться следующие ограничения [10, 12]:

$$\sum_i u_i^{\mathrm{A}} = \sum_j u_j^{\mathrm{B}} = \sum_k u_k^{\mathrm{C}} = 0,$$

$$\sum_i u_{ij}^{\mathrm{AB}} = \sum_j u_{ij}^{\mathrm{AB}} = \sum_i u_{ik}^{\mathrm{AC}} = \sum_k u_{ik}^{\mathrm{AC}} = \sum_j u_{jk}^{\mathrm{BC}} = \sum_k u_{jk}^{\mathrm{BC}} = 0,$$

$$\sum_i u_{ijk}^{\mathrm{ABC}} = \sum_j u_{ijk}^{\mathrm{ABC}} = \sum_k u_{ijk}^{\mathrm{ABC}} = 0.$$

Модель (1) называется насыщенной, если она включает все факторы и их возможные взаимодействия, т.е. все вклады, вносимые различными эффектами, отличны от нуля.

Модель, которая содержит только часть статистически значимых факторов и их взаимодействий, называется ненасыщенной.

При решении практических задач насыщенные модели не всегда бывают оптимальными, так как очень редко все факторы и их взаимодействия значимы. Основная задача метода заключается в выборе такой ненасыщенной модели, которая адекватно описывает анализируемые данные и содержит наименьшее возможное число значимых эффектов [13].

Для определения значимости эффекта его исключают из модели и рассчитывают статистики хи-квадрат Пирсона (хи-квадрат МП, χ^2) с уровнем значимости p и хи-квадрат метода максимального правдоподобия (хи-квадрат ММП, G^2 с уровнем значимости p [14]:

$$\chi^2 = \sum_{ijk} \frac{(n_{ijk} - n_{ijk}^*)}{n_{ijk}^*}; \tag{2}$$

$$G^2 = 2\sum_{ijk} n_{ijk} \ln\left(\frac{n_{ijk}}{n_{ijk}^*}\right), \tag{3}$$

где n_{ijk} – наблюдаемые частоты; n^*_{ijk} – теоретические частоты.

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.

Другими словами, критерии (2) и (3) позволяют проверить утверждение: «Отличается ли статистически значимо от нуля тот член насыщенной модели, который не включен в текущую модель».

Проверка равенства отдельных эффектов нулю равносильна проверке отсутствия связи между переменными. Так, например, равенство нулю эффекта u_{ij}^{AB} означает отсутствие связи между переменными А и В.

Важным отличием статистики G^2 от статистики χ^2 является то, что отношение правдоподобия G^2 обладает свойством аддитивности для частных связанных моделей, которое позволяет проверять значимость отдельных ее эффектов. Однако ряд исследователей полагают, что статистика χ^2 дает более точный результат оценки качества модели. На практике рекомендуется рассчитывать значения двух статистик для большей уверенности в выводах [13].

Таким образом, метод ЛЛА включает в себя следующие этапы:

1) построение модели, адекватно описывающей исходные данные. Для решения этой задачи исследуется насыщенная модель, из которой постепенно исключаются сложные составные взаимодействия до тех пор, пока не останутся только значимые эффекты;

2) интерпретация построенной модели.

Материалы исследования. Применение метода логлинейного анализа для обработки результатов анкетирования. Рассмотрим практическое применение метода при обработке результатов анкетирования на тему «Отношение жителей РФ к притоку мигрантов».

Проблема миграции в последние годы очень важна для России. Даже несмотря на кризис, наша страна входит в пятерку государств, которые более всего привлекают мигрантов [15].

Мигранты оказывают существенное влияние на все сферы жизни общества: экономику, культуру политику, криминальную обстановку. Разное восприятие мира и религиозные убеждения могу стать причиной социальных конфликтов. Поэтому тема исследования является актуальной.

Обрабатываемая анкета содержала вопросы, касающиеся отношения респондентов к притоку мигрантов, оценку их влияния на экономику и культуру.

В качестве примера рассмотрим вопрос анкеты, при ответе на который респонденты выражали свое мнение относительно ограничения потока мигрантов, отличающихся по национальности, религиозным убеждениям и культурным традициям от большинства жителей нашей страны. Исследуем, какое влияние на мнение респондентов оказывают его пол, возраст, уровень дохода и тип населенного пункта.

Вопрос 1. Следует ли позволить мигрантам, которые по национальности и расовой принадлежности отличаются от большинства населения страны переезжать жить и работать в нашу страну?

Варианты ответа: 1 – следует позволить многим таким людям переезжать; 2 – поток мигрантов должен быть ограничен; 3 – следует позволить переезжать лишь немногим из них; 4 – никому не разрешать.

Пол: 1 – мужской; 2 – женский.

Возраст: 1 – от 18 до 25; 2 – от 26 до 45; 3 – от 46 до 65; 4 – более 66.

Уровень дохода: 1 – денег хватает только на питание, покупка одежды – проблема; 2 – на питание и недорогую одежду денег хватает, но покупка вещей длительного пользования – проблема; 3 – на питание, одежду (в том числе и дорогую) и покупку вещей длительного пользования денег хватает, но машина, дача – не по карману; 4 – нет материальных проблем.

Тип населенного пункта: 1 – большой город; 2 – пригород или окраина большого города; 3 – небольшой город или поселок городского типа; 4 – деревня/село.

Анкета была разработана на кафедре социологии и социальных технологий Тверского государственного технического университета. Объем выборки составил 1654 чел. Обработка данных проводилась в пакете STATISTICA, в котором реализован модуль ЛЛА, содержащий полный набор процедур для исследования многомерных частотных таблиц с числом измерений от 2 до 7 [16–18].

На рис. 1 представлен фрагмент базы данных, содержащей результаты первичной обработки анкет.

	1 Мигранты другой	2 Пол	3 Возраст	4 Доход	5 Тип населенно
7	3	2	1	2	1
8	3	1	1	4	1
9	1	2	3	4	1
10	3	2	3	3	1
11	3	2	1	4	1
12	3	1	2	4	1
13	2	1	1	2	1
14	2	1	1	4	1
15	4	1	1	4	1
16	2	1	1	4	1
17	4	1	3	4	1
18	3	1	3	3	1
19	2	2	1	4	1

Рис. 1. Фрагмент базы данных с результатами анкетирования

Для лучшего понимания структуры данных рассчитаем частоты встречаемости по каждой переменной.

Анализ таблицы частот с результатами ответа на *Вопрос 1* показывает, что преобладает вариант 3 – «Следует позволить лишь немногим мигрантам другой национальности и расовой принадлежности приезжать жить и рабо-

тать в нашу страну» (37,8 %); 31,7 % считает, что «Поток мигрантов должен быть ограничен»; 23,3 % высказались за то, что «Нужно запретить приезд мигрантов»; и лишь 7,2 % респондентов считают, что «Можно позволить многим таким людям переезжать жить и работать в нашу страну» (рис. 2, *а*).

Таблица частот: Вопрос 1

Группа	Частота	Кумул. Частота	Процент	Кумул. Процент
1	119	119	7,19468	7,1947
2	525	644	31,74123	38,9359
3	625	1269	37,78718	76,7231
4	385	1654	23,27690	100,0000
Пропущ.	0	1654	0,00000	100,0000

а

Таблица частот: Пол

Группа	Частота	Кумул. Частота	Процент	Кумул. Процент
1	636	636	38,45224	38,4522
2	1018	1654	61,54776	100,0000
Пропущ.	0	1654	0,00000	100,0000

б

Таблица частот: Возраст

Группа	Частота	Кумул. Частота	Процент	Кумул. Процент
1	217	217	13,11971	13,1197
2	581	798	35,12696	48,2467
3	676	1474	40,87062	89,1173
4	180	1654	10,88271	100,0000
Пропущ.	0	1654	0,00000	100,0000

в

Таблица частот: Доход

Группа	Частота	Кумул. Частота	Процент	Кумул. Процент
1	282	282	17,04958	17,0496
2	415	697	25,09069	42,1403
3	580	1277	35,06651	77,2068
4	377	1654	22,79323	100,0000
Пропущ.	0	1654	0,00000	100,0000

г

Таблица частот: Тип населенного пункта

Группа	Частота	Кумул. Частота	Процент	Кумул. Процент
1	696	696	42,31003	42,3100
2	65	761	3,95137	46,2614
3	541	1302	32,88754	79,1489
4	343	1645	20,85106	100,0000
Пропущ.	0	1645	0,00000	100,0000

д

Рис. 2. Таблицы частот: *а* – вопрос 1; *б* – пол; *в* – возраст; *г* – уровень дохода; *д* – тип населенного пункта

На первом этапе можно сделать вывод о настороженном отношении респондентов к мигрантам другой национальности, религиозных убеждений и традиций.

Большинство опрошенных респондентов – женщины (61,5 %) (рис. 2, *б*).

Преобладают возрастные категории от 46 до 65 лет (40,8 %) и от 25 до 45 лет (35,1 %) (рис. 2, *в*).

По уровню дохода преобладает 3-я категория (35,1 %), далее следуют 2-я и 4-я категории – 25,1 и 22,8 % соответственно, наименьшая по численности – 1-я категория – 17,0 % (рис. 2, *г*).

По типу населенного пункта респонденты распределены неоднородно: преобладают респонденты 1-й категории (42,3 %), далее следуют респонден-

ты 3-й категории – 32,9 %, затем респонденты 4-й категории – 20,8 % и наименьшее число респондентов 2-й категории – 4,0 % (рис. 2, *д*).

Для исследования взаимосвязей между вопросами анкеты применим метод логлинейного анализа. Введем следующие обозначения: *Вопрос 1* – переменная 1; *Пол* – переменная 2; *Возраст* – переменная 3; *Уровень дохода* – переменная 4; *Тип населенного пункта* – переменная 5. Все переменные измерены в категориальных шкалах. Наблюдения являются независимыми. Следовательно, применение метода является правомерным.

Построим и проанализируем таблицы одновременных критериев хи-квадрат Пирсона и хи-квадрат метода максимального правдоподобия для всех возможных *k*-факторных взаимодействий (рис. 3).

	Результаты подгонки *k*-факторн. взаимодействий (Анкета 1) Это одновременная проверка того, что все *k*-факторные взаимодействия равны нулю				
k-фактор	**Число ст.своб.**	МП хи-квад.	Вероятн. p	Пирсона хи-квад.	Вероятн. p
1	14	1897,263	0,000000	3210,306	0,000000
2	76	557,165	0,000000	682,604	0,000000
3	198	134,345	0,999835	138,756	0,999526
4	243	105,150	1,000000	103,076	1,000000
5	108	62,277	0,999873	62,389	0,999868

Рис. 3. *k*-Факторные взаимодействия

При $k = 1$ и $k = 2$ вероятности $p < 0{,}05$, следовательно, влияние факторов и их попарных взаимодействий статистически значимо, а при включении в модель всех трех, четырех или пятифакторных взаимодействий она слабо согласуется с исходными данными. Таким образом, наименее сложная ненасыщенная модель содержит только двухфакторные взаимодействия. Однако, исходя из содержательной постановки решаемой задачи, можно проанализировать модели, содержащие отдельные двух, трех, четырех и пятифакторные эффекты.

Используя критерий маргинальных и частных связей, выявим значимые взаимодействия. Из таблицы рис. 4 видно, что статистически значимыми являются следующие эффекты ($p < 0{,}05$):

14 – между переменными *Вопрос 1* и *Уровень дохода*;

15 – между переменными *Вопрос 1* и *Тип населенного пункта*;

23 – между переменными *Пол* и *Возраст*;

24 – между переменными *Пол* и *Уровень дохода*;

34 – между переменными *Возраст* и *Уровень дохода*;

45 – между переменными *Уровень дохода* и *Тип населенного пункта*.

Таким образом, оптимальной оказалась ненасыщенная ЛЛМ, включающая в себя факторные взаимодействия 14, 15, 23, 24, 34, 45. Адекватность модели была установлена по критериям хи-квадрат Пирсона и хи-квадрат ме-

тода максимального правдоподобия. Основная проверяемая гипотеза заключалась в равенстве наблюдаемых и ожидаемых частот. Уровень значимости для обоих критериев больше 0,05 ($p > 0{,}05$), следовательно, модель является адекватной исходным данным:

- значение хи-квадрат метода максимального правдоподобия равно 308,94, уровень значимости: $p = 0{,}99$;
- значение хи-квадрат Пирсона равно 316,43, уровень значимости: $p = 0{,}99$.

	Критерии маргинальных и частных связей (Анкета				
Эффект	Число ст.своб.	Част.св. хи-квад.	Част.св. p	Марг.св. хи-квад.	Марг.св. p
1	3	332,600	0,000000	332,600	0,000000
2	1	74,391	0,000000	74,391	0,000000
3	3	396,361	0,000000	396,361	0,000000
4	3	92,167	0,000000	92,167	0,000000
5	4	1001,744	0,000000	1001,744	0,000000
12	3	5,349	0,147933	4,961	0,174644
13	9	11,834	0,222812	16,879	0,050652
14	9	16,264	0,041562	24,509	0,003565
15	12	28,349	0,004917	37,359	0,000195
23	3	16,727	0,000804	28,963	0,000002
24	3	37,866	0,000000	43,900	0,000000
25	4	6,741	0,150248	2,998	0,558234
34	9	159,764	0,000000	184,474	0,000000
35	12	12,077	0,439518	23,930	0,020794
45	12	205,401	0,000000	216,439	0,000000
123	9	2,694	0,975209	2,641	0,976845
124	9	4,021	0,909996	5,195	0,816961
125	12	4,826	0,963543	6,272	0,901763
134	27	18,486	0,887973	23,133	0,677824
135	36	19,605	0,988072	21,702	0,971206
145	36	27,075	0,858488	29,562	0,767121
234	9	7,012	0,635914	8,187	0,515452
235	12	4,947	0,959735	4,452	0,973829
245	12	8,346	0,757518	7,781	0,802011
345	36	29,059	0,787402	33,680	0,579410
1234	27	12,954	0,989537	16,432	0,944203
1235	36	12,464	0,999909	13,494	0,999764
1245	36	10,165	0,999993	12,975	0,999852
1345	108	44,053	1,000000	46,775	1,000000
2345	36	19,419	0,989072	20,629	0,981247

Рис. 4. Критерий маргинальных и частных связей

Дополнительным критерием качества моделей служит график наблюдаемых и подогнанных частот, точки на котором расположены вдоль прямой линии (рис. 5).

Особый интерес представляет вопрос влияние *Пола* респондента, его *Возраста*, *Уровня дохода* и *Типа населенного пункта*, в котором он проживает, на отношение к ограничению потока мигрантов. В контексте сформулированной задачи переменные *Пол*, *Возраст*, *Уровень дохода* и *Тип населенного пункта* будем считать независимыми (объясняющими переменными или предикторами), а переменную *Вопрос 1* – зависимой.

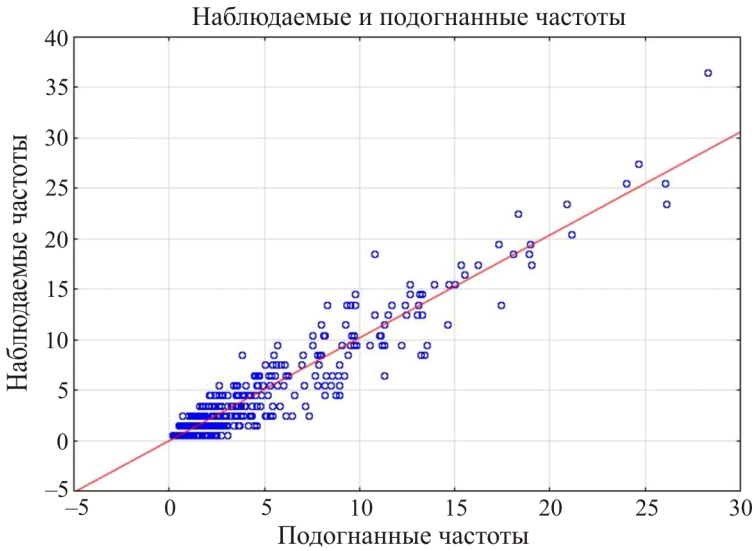

Рис. 5. График наблюдаемых и подогнанных частот

Исходя из анализа значимых взаимодействий (см. рис. 3), влияние на мнение респондентов по *Вопросу 1* (переменная 1) оказывают *Уровень дохода* (взаимодействие 14) и *Тип населенного пункта* (взаимодействие 15), поскольку величина уровня значимости для них менее 0,05.

Таблица рис. 4 позволяет также определить силу влияния каждого объясняющего фактора (4 и 5) на зависимую переменную (1) [19, 20]. Для этого рассчитывается величина *K*:

$$K = \frac{\chi_m^2}{\sum \chi_m^2} \cdot 100\ \%,$$

где χ_m^2 – значение статистики хи-квадрат Пирсона, соответствующее *m*-му эффекту; $\sum \chi_m^2$ – сумма значений статистик хи-квадрат Пирсона всех эффектов.

Отношение *K* было рассчитано для эффектов 1, 14, 15 (таблица).

Расчет отношения *K*

Эффект	Хи-квадрат Пирсона	*K*
1	332,6	88 %
14	16,3	4 %
15	28,3	8 %
Сумма	377,2	100 %

Значения величины *K* позволяет сделать вывод, что отношение к потоку мигрантов на 4 % зависит от уровня дохода и на 8 % – от типа населенного

пункта. На 88 % отношение к притоку мигрантов другой национальности определяется неучтенными в модели факторами.

Таким образом, было установлено, что из всех независимых переменных только *Уровень дохода* и *Тип населенного пункта* оказывают влияние на мнение респондентов относительно притока мигрантов, однако степень этого влияния незначительна.

Используя возможности модуля ЛЛА пакета STATISTICA, можно провести более содержательный анализ частот с помощью маргинальных двумерных таблиц для попарных уровней факторов.

Рассмотрим признаки *Вопрос 1 – Тип населенного пункта* (рис. 6).

Тип населенного пункта	**Мигранты другой национальности 1**	Мигранты другой национальности 2	Мигранты другой национальности 3	Мигранты другой национальности 4	Сумма
1	71,0000	256,0000	287,0000	146,0000	760,000
2	19,0000	41,0000	41,0000	28,0000	129,000
3	48,0000	191,0000	209,0000	157,0000	605,000
4	45,0000	100,0000	146,0000	116,0000	407,000
Сумма	183,0000	588,0000	683,0000	447,0000	1901,000

Рис. 6. Маргинальная таблица факторов *Вопрос 1 – Тип населенного пункта*

По таблице рис. 6 можно установить, что для больших городов отношение к притоку мигрантов другой национальности распределяется примерно как 1:4:4:2. Иначе говоря, на одного респондента, который считает, что «Следует позволить многим таким людям переезжать», приходится по четыре респондента, которые полагают, что «Поток мигрантов должен быть ограничен» или «Следует позволить лишь немногим из них приезжать жить и работать в нашу страну», и два респондента, которые высказались против приезда мигрантов (сама многочисленная категория респондентов) (см. рис. 2, *д*).

Для пригородов и окраин больших городов отношение равно примерно 1:2:2:1; для небольших городов и поселков городского типа 1:4:4:3; для деревень и сел – 1:2:3:2.

В крупных городах и поселках городского типа к притоку мигрантов относятся более негативно, чем в пригородах и окраинах больших городов и селах.

Аналогично рассмотрим признаки *Вопрос 1 – Уровень дохода* (рис. 7).

По таблице рис. 7 можно установить, что для респондентов с низким уровнем дохода отношение к притоку мигрантов другой национальности распределяется примерно как 1:3:3:3. Следовательно, на одного респондента, который считает, что «Следует позволить многим таким людям переезжать», приходится по три респондента, которые высказались за ограничение потока или за запрет на приезд мигрантов.

Доход	Мигранты другой национальности 1	Мигранты другой национальности 2	Мигранты другой национальности 3	Мигранты другой национальности 4	Сумма
1	37,0000	112,0000	97,0000	98,0000	344,000
2	39,0000	137,0000	185,0000	116,0000	477,000
3	55,0000	191,0000	237,0000	156,0000	639,000
4	52,0000	148,0000	164,0000	77,0000	441,000
Сумма	183,0000	588,0000	683,0000	447,0000	1901,000

Рис. 7. Маргинальная таблица факторов *Вопрос 1 – Уровень дохода*

Для респондентов с уровнем дохода, относящимся ко 2-й категории, отношение равно примерно 1:4:5:3, т.е. на одного респондента, который считает, что «Следует позволить многим таким людям переезжать», приходится четыре респондента, считающих, что поток мигрантов нужно ограничит, пять респондентов, высказавшихся за то, что следует пускать лишь немногих и три человека, которые считают, что приезд мигрантов нужно запретить.

Для респондентов с уровнем дохода 3-й категории отношение составляет 1:3:4:3. Для респондентов с уровнем дохода, относящимся к 4-й категории (не имею материальных проблем), отношение равно 1:2:3:1; 2, 3, 4-я категории являются самыми многочисленными (см. рис. 2, *г*).

Таким образом, можно сделать вывод, что наиболее лояльно относятся к притоку мигрантов респонденты с высоким уровнем дохода и самое негативное отношение у респондентов с уровнем дохода из 2-й категории («На питание и недорогую одежду денег хватает, но покупка вещей длительного пользования – проблема»).

Выводы. Целью проведенного исследования являлась демонстрация возможностей метода логарифмического линейного анализа при обработке результатов анкетирования. ЛЛА позволяет исследовать взаимосвязи между набором категориальных переменных (результатами ответов на вопросы анкеты), определять значимые взаимодействия, строить оптимальные модели, адекватно описывающие исходные данные.

Метод ЛЛА был применен для изучения общественного мнения по вопросу притока мигрантов в нашу страну. В частности, проведено исследование влияния пола, возраста, уровня дохода и типа населенного пункта, в котором проживает респондент на мнение по анализируемому вопросу. Построена оптимальная ненасыщенная модель, включающая пять значимых взаимодействий. Адекватность модели была подтверждена проверкой критериев хи-квадрат Пирсона и хи-квадрат метода максимального правдоподобия.

Установлено, что на отношение к мигрантам, которые по национальности, религиозным убеждениям и культурным традициям отличаются от большинства жителей нашей страны, в большей степени влияет тип населенного пункта и в меньшей степени уровень дохода. Сила этого влияния незначительна.

С помощью маргинальных двумерных таблиц проведен более детальный анализ частот и установлено, что в крупных городах и поселках городского типа к притоку мигрантов относятся более негативно, чем в пригородах и окраинах больших городов и селах. Наиболее лояльно относятся к приезду мигрантов респонденты с высоким уровнем дохода и наиболее негативно респонденты с невысоким и средним уровнем дохода.

Список литературы

1. Общественное мнение [Электронный ресурс]. – URL: https://ru.wikipedia.org/wiki/%D0%9E%D0%B1%D1%89%D0%B5%D1%81%D1%82%D0%B2%D0%B5%D0%BD%D0%BD%D0%BE%D0%B5_%D0%BC%D0%BD%D0%B5%D0%BD%D0%B8%D0%B5 (дата обращения: 26.01.2018).

2. 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.

3. 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.

4. Бессокирная Г.П. Факторный анализ: традиции использования и новые возможности // Социология: методология, методы, математическое моделирование. – 2000. – № 12. – С. 142–153.

5. Фомина Е.Е. Обзор методов и пакетов прикладных программ для обработки результатов анкетирования // Современные технологии в мировом научном пространстве: сб. ст. Междунар. науч.-практ. конф.: в 6 ч. – Ч. 6. – Самара, 2017. – С. 190–192.

6. Толстова Ю.Н. Анализ социологических данных: методология, дескриптивная статистика, изучение связей между номинальными признаками. – М.: Научный мир, 2000. – 352 с.

7. Толстова Ю.Н., Рыжова А.В. Анализ таблиц сопряженности: использование отношения преобладаний и логлинейных моделей // Социология 4М. – 2003. – № 16. – С. 150–164.

8. Логлинейный анализ [Электронный ресурс]. – URL: http://www.statmethods.ru/konsalting/statistics-metody/114-loglinejnyj-analiz.html (дата обращения: 21.01.2018).

9. Буре В.М., Гливинская О.А., Сотников А.В. Логлинейный анализ базы данных по инфаркту миокарда у больных молодого и среднего возраста // Вестник Санкт-Петербургского университета. – Сер. 10. – 2010. – Вып. 1. – С. 35–41.

10. Аптон Г. Анализ таблиц сопряженности. – М.: Финансы и статистика, 1982. – 143 с.

11. Трофимов Д.А. Логлинейный анализ таблиц мобильности: обзор основных моделей // Социология 4М. – 2008. – № 26. – С. 119–138.

12. Трошин Л.И., Балаш В.А., Балаш О.С. Статистический анализ нечисловой информации / Моск. гос. ун-т экономики, статистики и информатики. – М., 2001. – 67 с.

13. Макарова Н.В. Статистический анализ медико-биологических данных с использованием пакетов статистических программ Statistica, SPSS, NCSS, SYSTAT: метод. пособие / Всерос. центр экстрен. и радиац. медицины им. А.М. Никифорова МЧС России. – СПб.: Политехника-сервис, 2012 – 178 с.

14. Репина Е.Г. Статистический анализ нечисловой информации – непараметрический подход. – Самара: Изд-во Самар. гос. экон. ун-та, 2009. – 96 с.

15. Миграция – 2017 [Электронный ресурс]. – URL: http://center-yf.ru/data/stat/migraciya-2017.php (дата обращения: 21.01.2018).

16. Наследов А. SPSS 19: профессиональный статистический анализ данных. – СПб.: Питер, 2011. – 400 с.

17. Программа статистического анализа Statistica [Электронный ресурс]. – URL: http://bourabai.ru/tpoi/statistica/gl11.html (дата обращения: 26.01.2018).

18. Калинина В.Н., Соловьев В.И. Компьютерный практикум по прикладной статистике и основам эконометрики: учеб. пособие. – М.: Вега-Инфо, 2010. – 140 с.

19. Крымзин Д.Н. Применение логлинейного анализа для исследования зависимости оценки кадрового потенциала вуза от характеристик преподавателей // Вестник НГУЭУ. – 2014. – № 2. – С. 134–141.

20. Анализ качественных признаков на основе логлинейной модели [Электронный ресурс]. – URL: https://studopedia.ru/6_28922_analiz-kachestvennih-priznakov-na-osnove-loglineynoy-modeli.html (дата обращения: 26.01.2018).

References

1. Obshchestvennoe mnenie [Public opinion], available at: https://ru.wikipedia.org/wiki/% (accessed 26 January 2018).

2. Fomina E.E., Zhiganov N.K. Metodika obrabotki rezul'tatov anketirovaniia s ispol'zovaniem metodov mnogomernoi i parametricheskoi statistiki [Methodology of survey results processing with the use of multivariate and parametric statistics]. *PNRPU Sociology and Economics Bulletin*, 2017, no. 1, pp. 106–115.

3. Fomina E.E. Faktornyi analiz i kategorial'nyi metod glavnykh komponent: sravnitel'nyi analiz i prakticheskoe primenenie dlia obrabotki rezul'tatov anketirovaniia [Factor analysis and categorial principal component analysis: comparative analysis and practical application for processing of questionnaire survey results]. *Humanities Bulletin of BMSTU*, 2017, no. 10 (60), p. 3.

4. Bessokirnaia G.P. Faktornyi analiz: traditsii ispol'zovaniia i novye vozmozhnosti [Factor analysis: usage traditions and new opportunities]. *Sociology: Methodology, Methods, Mathematical Modeling*, 2000, no. 12, pp. 142–153.

5. Fomina E.E. Obzor metodov i paketov prikladnykh programm dlia obrabotki rezul'tatov anketirovaniia [The overview of methods and software packages for survey results processing]. *Sovremennye tekhnologii v mirovom nauchnom prostranstve. Sbornik statei Mezhdunarodnoi nauchno-prakticheskoi konferentsii*, 2017, part 6, pp. 190–192.

6. Tolstova Iu.N. Analiz sotsiologicheskikh dannykh: metodologiia, deskriptivnaia statistika, izuchenie sviazei mezhdu nominal'nymi priznakami [Analysis of sociological data: methodology, descriptive statistics, study of relationship between nominal characteristics]. Moscow, Scientific world, 2000, 352 p.

7. Tolstova Iu.N., Ryzhova A.V. Analiz tablits sopriazhennosti: ispol'zovanie otnosheniia preobladanii i loglineinykh modelei [The analysis of contingency tables: using the relations of predominance and log-linear models]. *Sociology: Methodology, Methods, Mathematical Modeling*, 2003, no. 16, pp. 150–164.

8. Loglineinyi analiz [Log-linear analysis], available at: http://www.statmethods.ru/konsalting/statistics-metody/114-loglinejnyj-analiz.html (accessed 21 January 2018).

9. Bure V.M., Glivinskaia O.A., Sotnikov A.V. Loglineinyi analiz bazy dannykh po infarktu miokarda u bol'nykh molodogo i srednego vozrasta [Log-linear database analysis by cardiac infarction data among young and middle age patients]. *Vestnik of the Saint Petersburg University*, 2010, ser. 10, iss. 1, pp. 35–41.

10. Apton G. Analiz tablits sopriazhennosti [Analysis of conjugacy tables]. Moscow, *Finansy i statistika*, 1982, 143 p.

11. Trofimov D.A. Loglineinyi analiz tablits mobil'nosti: obzor osnovnykh modelei [Loglinear analysis of mobility tables: Basic models review]. *Sociology: Methodology, Methods, Mathematical Modeling*, 2008, no. 26, pp. 119–138.

12. Troshin L.I., Balash V.A., Balash O.S. Statisticheskii analiz nechislovoi informatsii [Statistical analysis of non-numeric information]. Moscow State University of Economics, Statistics and Informatics, 2001, 67 p.

13. Makarova N.V. Statisticheskii analiz mediko-biologicheskikh dannykh s ispol'zovaniem paketov statisticheskikh programm Statistica, SPSS, NCSS, SYSTAT [Statistical analysis of biomedical data using statistical software packages Statistica, SPSS, NCSS, SYSTAT]. Saint Petersburg, Politekhnika-SERVIS, 2012, 178 p.

14. Repina E.G. Statisticheskii analiz nechislovoi informatsii – neparametricheskii podkhod [Statistical analysis of non-numeric information – nonparametric approach]. Samara State University of Economics, 2009, 96 p.

15. Migratsiia–2017 [Migration–2017], available at: http://center-yf.ru/data/stat/migraciya-2017.php (accessed 21 January 2018).

16. Nasledov A. SPSS 19: professional'nyi statisticheskii analiz dannykh [SPSS 19: professional statistical data analysis]. Saint Petersburg, Peter, 2011, 400 p.

17. Programma statisticheskogo analiza Statistica [Statistical analysis program Statistica], available at: http://bourabai.ru/tpoi/statistica/gl11.html (accessed 26 January 2018).

18. Kalinina V.N., Solov'ev V.I. Komp'iuternyi praktikum po prikladnoi statistike i osnovam ekonometriki [Computer workshop on applied statistics and basics of econometrics]. Moscow, VEGA-Info, 2010, 140 p.

19. Krymzin D.N. Primenenie loglineinogo analiza dlia issledovaniia zavisimosti otsenki kadrovogo potentsiala vuza ot kharakteristik prepodavatelei [Application of log-linear analysis for researches of dependence of the assessment of the personnel capacity of higher education institution from characteristics of teachers]. *Vestnik NSUEM*, 2014, pp. 134–141.

20. Analiz kachestvennykh priznakov na osnove loglineinoi modeli [Analysis of qualitative characteristics based on log-linear models], available at: https://studopedia.ru/6_28922_analiz-kachestvennih-priznakov-na-osnove-logliney-noy-modeli.html (accessed 26 January 2018).

Оригинальность 89 %

Получено 12.03.018 Принято 13.04.2018 Опубликовано 28.09.2018

E.E. Fomina

THE POSSIBILITY OF LOG-LINEAR ANALYSIS FOR SURVEY RESULTS PROCESSING

The questionnaire survey is the main tool for studying public opinion in the work of a sociologist. To process databases with the survey results mathematical methods are used, including factor analysis, categorical principal component method, calculation of indicators of descriptive statistics. One of the most commonly used methods is frequency table analysis. To study the dependence in two-dimensional frequency tables, i.e. dependencies between two variables, the Chi-square test is applied. However, multidimensional frequency tables with more than two variable categories are of special interest. To study the complex internal relationships in these tables linear logarithm analysis can be used. The purpose of this article is to demonstrate possibilities of log-linear analysis method in processing survey results. The method is used to process survey results on influx of migrants in Russia. The influence of the sex, age, income level and type of settlement on attitude to foreign migrants influx is investigated. The optimal unsaturated model including five significant interactions is constructed. The adequacy of the model is established by checking the Pearson Chi-square and Chi-square criteria of maximum likelihood method.

Keywords: *log-linear analysis, frequency table, questionnaire survey*.

Elena E. Fomina – Candidate of Technical Sciences, Associate Professor, Department of Informatics and Applied Mathematics, Tver State Technical University, e-mail: f-elena2008@yandex.ru.

Received 12.03.2018 Accepted 13.04.2018 Published 28.09.2018