Ю.Н. Толстова, А.В. Рыжова
(*Москва*)

АНАЛИЗ ТАБЛИЦ СОПРЯЖЕННОСТИ: ИСПОЛЬЗОВАНИЕ ОТНОШЕНИЯ ПРЕОБЛАДАНИЙ И ЛОГЛИНЕЙНЫХ МОДЕЛЕЙ[1]

В статье представлены два метода анализа таблиц сопряженности: вычисление отношения преобладаний и построение логарифмически линейных моделей. На примерах демонстрируется, какое новое знание о связях между изучаемыми признаками может получить социолог с помощью этих методов. Кратко описывается положение дел с использованием соответствующих подходов в отечественной социологии.

Ключевые слова: номинальная переменная, категория номинальной переменной, связь между признаками, частота, многомерная таблица сопряженности, отношение преобладаний, модель частоты, логарифмически линейная модель, логлинейный анализ.

Частотные таблицы и их роль в эмпирической социологии. Место рассматриваемых методов в анализе частотных таблиц

Частотные таблицы являются основным способом представления социологических данных. Они используются (рассчитыва-

Юлиана Николаевна Толстова – доктор социологических наук, профессор ГУ-ВШЭ, главный научный сотрудник Института социологии РАН.
Анастасия Валентиновна Рыжова – студентка V курса механико-математического факультета МГУ.

[1] Статья написана при поддержке Российского фонда фундаментальных исследований, проект № 02-06-80403.

ются и анализируются) практически в любом эмпирическом исследовании. Это не случайно. Средства изучения статистических закономерностей (а именно такие закономерности и пытается найти социолог на основе изучения частотных таблиц), предлагаемые современной наукой, так или иначе базируются на положениях математической статистики, объектом изучения которой служат случайные величины, предметом – параметры их распределений. Все те закономерности, которые мы можем найти с помощью математической статистики, по существу фигурируют в виде наборов параметров распределений случайных величин. Частотные же таблицы являются, строго говоря, выборочными представлениями одномерных и многомерных случайных величин (в выборочных исследованиях вместо термина «случайная величина» используется термин «признак» или его синонимы – «величина», «характеристика», «переменная»). Таким образом, опора на анализ частотных таблиц, к чему побуждает социолога здравый смысл, в действительности оправдывается тем серьезным изучением понятия статистической закономерности, которое осуществлялось в течение нескольких столетий в рамках математической статистики[1].

Важность анализа частотных таблиц давно осознана учеными. В настоящее время имеется большое количество методов соответствующего плана. К сожалению, далеко не все из них активно задействованы в нашей социологии (парадоксальным выглядит то обстоятельство, что среди весьма слабо используемых в широкой отечественной практике методов есть и очень интересные алгоритмы, разработанные советскими и российскими исследователями).

Частотные таблицы размерности два и выше обычно называют таблицами сопряженности, поскольку содержащаяся в них информация говорит о сопряжении значений двух и более признаков.

---

[1] Более подробно о соотношении понятий «анализ данных» и «математическая статистика» можно прочесть в [1].

В настоящей статье речь идет о двух методах, очень широко применяющихся на Западе и, к сожалению, почти забытых нашими социологами (хотя в свое время делались попытки внедрить их в отечественную социологическую практику, но об этом будет сказано ниже). Рассматриваемые методы не являются равноценными. Отношение преобладания – простое, но очень важное понятие, которое активно используется, а в некоторых случаях является базовым при построении сложных моделей для анализа таблиц сопряженности. Одной из таких сложных моделей является логарифмически линейная модель. Это хорошо разработанный метод, позволяющий исследователю с равным успехом находить и простые, и сложные связи между двумя и большим числом переменных.

Отношение преобладаний

Об истории использования отношений преобладания в социальных науках можно прочесть в работе венгерского ученого Т. Рудаша [2]. Эти отношения и раньше были активно задействованы в процессе применения разного рода логлинейных моделей. Однако именно Т. Рудаш выделил их в явном виде и показал, как можно использовать эти отношения сами по себе, без привязки к каким-либо более сложным моделям.

Для наглядности описания понятия отношения преобладаний обратимся к реальным данным[1]. Рассмотрим таблицу сопряженности, которая классифицирует респондентов в зависимости от их возраста и участия в выборах (см. табл. 1).

Рассмотрим основной вопрос, который встает перед социологом, занимающимся анализом таблицы сопряженности – о характере связи между лежащими в основе таблицы переменными X (возраст) и Y (участие в выборах). Мы видим, что представи-

[1] Данные, собранные в процессе изучения электорального поведения москвичей, были любезно предоставлены авторам М.И. Тарарухиной.

тели обеих возрастных категорий не склонны к участию в выборах. Найдем отношение количества респондентов, которые не участвуют в выборах, к количеству участвующих. Для возрастных категорий «18–45 лет» и «старше 45 лет» подобные отношения будут иметь, соответственно, вид:

$$\frac{228}{121} = 1{,}88 \tag{1}$$

и

$$\frac{295}{58} = 5{,}09 \tag{2}$$

Таблица 1

ТАБЛИЦА СОПРЯЖЕННОСТИ ДЛЯ ДВУХ ПРИЗНАКОВ «ВОЗРАСТ» И «УЧАСТИЕ В ВЫБОРАХ»

Возраст	Участие в выборах	
	Не участвуют*	Всегда участвуют
18–45 лет	295	58
Старше 45 лет	228	121

* Градация «не участвуют» отвечает тем респондентам, которые дали один из двух ответов: «не всегда участвую» и «вообще не участвую».

Каждую из этих величин называют преобладанием (речь идет о преобладании количества не участвующих в выборах над количеством участвующих). В данном случае ее можно связать со склонностью к неучастию в выборах. Таким образом, в рамках возрастной категории «18–45 лет» количество респондентов, которые не участвуют в выборах, более чем в 5 раз превышает количество голосующих. На каждых 100 голосующих приходится 509 человек, которые игнорируют выборы. В возрастной же категории «старше 45 лет» на каждых 100 голосующих приходится 188 воздержавшихся.

Введем в оборот еще одно отношение:

$$\frac{5{,}09}{1{,}88} = 2{,}7 \tag{3}$$

Оно говорит о том, что старшее поколение в 2,7 раза более склонно участвовать в выборах, чем молодое. Другими словами, имеет место довольно сильная зависимость участия в выборах от возраста.

Введем формулу, отражающую отношения типа рассмотренных. В общем случае отношение преобладания θ имеет вид:

$$\theta = \frac{n_{11} / n_{12}}{n_{21} / n_{22}} = \frac{n_{11} n_{22}}{n_{12} n_{21}},$$

где $n_{11}, n_{12}, n_{21}, n_{22}$ – наблюдаемые частоты в ячейках таблицы (индексы обозначают номера строк и столбцов таблицы).

Подчеркнем, что для получения содержательных выводов не имеет значения, вычислять ли отношение величины (1) к величине (2), или наоборот. Нам просто важно понять, насколько сильно различаются эти величины. Поэтому при работе с таблицами сопряженности наряду с θ можно использовать 1/θ.

Проведем анализ еще одной таблицы сопряженности (см. табл. 2).

Таблица 2

ТАБЛИЦА СОПРЯЖЕННОСТИ ДЛЯ ДВУХ ПРИЗНАКОВ: «УЧАСТИЕ В ВЫБОРАХ» и «ОБРАЗОВАНИЕ»

Образование	Участие в выборах	
	Не участвуют	Всегда участвуют
Высшее	157	49
Другое	364	129

Преобладание для респондентов с высшим образованием равно:

$$\frac{157}{49} = 3{,}2,$$

а для респондентов без высшего образования:

$$\frac{364}{129} = 2{,}82.$$

Отношение полученных преобладаний равно:

$$\theta = \frac{3,20}{2,82} = 1,13.$$

Таким образом, респонденты, имеющие высшее образование, всего в 1,13 раза более склонны к участию в выборах, чем респонденты без высшего образования. Следовательно, участие в выборах очень слабо зависит от образования.

Вообще, можно сказать, что, чем ближе значение отношения преобладания к единице, тем слабее связь между признаками. Равенство этого отношения единице свидетельствует о полной независимости признаков. Однако здесь следует сделать следующее замечание. Допустим, для двух таблиц сопряженности получены значения отношений преобладания, равные величинам $\theta_1 = \frac{1}{4}$ и $\theta_2 = 2$ соответственно. Ясно, что, хотя значение $\theta_1 = \frac{1}{4}$ численно ближе к единице, оно указывает на более слабую связь между переменными, чем значение $\theta_2 = 2$. Чтобы не возникало подобной путаницы, нужно следить за направлением приближения отношения преобладания к единице. В нашем примере целесообразно сравнивать не $\theta_1 = \frac{1}{4}$ и $\theta_2 = 2$, а либо $\frac{1}{\theta_1} = 4$ и $\theta_2 = 2$, либо $\theta_1 = \frac{1}{4}$ и $\frac{1}{\theta_2} = \frac{1}{2}$.

Заметим, что отношения типа (3) могут обобщаться на таблицы сопряженности большего размера – трехмерные, четырехмерные и т.д. Такие обобщенные отношения преобладания называются, соответственно, трехмерными, четырехмерными и т.д.

Логарифмически линейные модели для таблиц сопряженности

Логарифмически линейные модели позволяют определить, как наблюдаемая частота в ячейке таблицы сопряженности зависит от категорий номинальных переменных, соответствующих этой ячейке. Поясним это на гипотетическом примере.

Допустим, что изучается зависимость занятости человека от его пола, и полученные данные сгруппированы в таблицу сопряженности размерности 2×2. Основой для построения таблицы служат признаки X (пол) (с категориями X_1 (женщина) и X_2 (мужчина)), и Y (занятость) (с категориями Y_1 (работает) и Y_2 (не работает)). Предположим, что опрошено 1000 человек, причем наблюдаемая частота в клетке с индексом (11) («работающая женщина») получилась равной 50 (см. табл. 3). Столь малая (по отношению к объему выборки) величина наблюдаемой частоты, на первый взгляд, может привести нас к выводу, что женщины не склонны работать.

Таблица 3

ФРАГМЕНТ ГИПОТЕТИЧЕСКОЙ ТАБЛИЦЫ СОПРЯЖЕННОСТИ, ПОЗВОЛЯЮЩЕЙ ОЦЕНИТЬ СВЯЗЬ МЕЖДУ ПРИЗНАКАМИ «ЗАНЯТОСТЬ» И «ПОЛ»

Пол	Занятость	
	Работает	Не работает
Женщина	50	
Мужчина		

Однако не все так очевидно. В действительности рассматриваемая наблюдаемая частота может зависеть, например, от трех причин (эти причины отвечают той модели частоты, которая соответствует данному подходу; известны модели, в которых частота раскладывается на другие компоненты, как это делается, например, в каноническом анализе):

а) может быть, в выборку вообще попало мало женщин, т.е. категория X_1 (женщина) встречается реже, чем категория X_2 (мужчина);

б) может быть, мало работающих людей попало в выборку, т.е. категория Y_1 (работает) встречается реже, чем категория Y_2 (не работает), вне зависимости от пола;

в) наконец, может быть, действительно мало работающих женщин, т.е. сочетание X_1Y_1 встречается реже, чем можно было бы ожидать, если бы переменные X и Y были независимы (в таких случаях говорят о наличии взаимодействия между первыми градациями рассматриваемых признаков).

Итак, у нас есть три фактора, которые воздействуют на наблюдаемые частоты в ячейках таблицы сопряженности, и нам нужна математическая модель, позволяющая количественно сравнить силу этих воздействий.

Такую модель предложил Гудмен [3]. Она имеет вид:

$$v_{ij} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \tag{4}$$

где $v_{ij} = \ln p_{ij}$, μ – «средний» член, он равен той вероятности, которая отвечает равномерному распределению; чтобы получить его выборочную оценку, надо общее количество объектов в выборке разделить на число клеток изучаемой частотной таблицы; $\lambda_i^X, \lambda_j^Y, \lambda_{ij}^{XY}$ соответствуют влияниям трех названных выше факторов на наблюдаемые частоты (эти «лямбды» обычно называют вкладами упомянутых факторов в частоту). Параметры λ имеют надстрочные индексы, показывающие, к каким переменным они относятся, и подстрочные индексы, говорящие о том, к каким категориям они прилагаются.

Поясним смысл выражения «логарифмически линейная модель». Дело в том, что, как показал Гудмен, адекватными реальности являются так называемые мультипликативные модели изучаемых частот p_{ij}, т.е. такие модели, в которых частота выражается как произведение вкладов отдельных градаций и их взаимодействий (ясно, что эти модели нелинейны). Однако оценивать такие вклады, т.е.

рассчитывать их конкретные величины, очень трудно. При решении уравнений легче работать с суммами определяемых величин, чем с их произведениями. Но эта трудность легко преодолевается. Как известно, логарифм произведения каких бы то ни было величин равен сумме логарифмов каждой из них. Поэтому, прологарифмировав мультипликативную модель, мы получим выражение (4), т.е. аддитивную, линейную модель. Таким образом, речь идет о таких моделях частот, которые являются линейными только после их логарифмирования. И в результате расчета параметров такой модели находится конкретный вид (разложение на отдельные фрагменты) не самой частоты, а ее логарифма. Для практических целей это особой роли не играет. В любом случае найденные параметры модели оцениваются лишь на качественном уровне. Для получения выводов о причинно-следственных отношениях между переменными обычно бывает достаточно лишь оценки того, какие коэффициенты больше, какие – меньше, а такие выводы инвариантны относительно замены частоты на ее логарифм.

Чтобы в модели не получился избыток, т.е. чтобы число параметров не превысило числа ячеек, на значения λ налагаются следующие ограничения (в противном случае параметры модели в принципе не смогут быть определены):

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = 0 \qquad (5)$$

Особое внимание следует обратить на λ_{ij}^{XY}. Этот параметр характеризует взаимодействие между переменными, и чем дальше его значение отстоит от нуля, тем сильнее связь между категорией i переменной X и категорией j переменной Y.

Рассмотрим случай трехмерной таблицы. Пусть у нас есть таблица с переменными X, Y и Z, имеющими I, J и K категорий соответственно. Обозначим за p_{ijk} вероятность того, что случайно выбранное наблюдение попадает в ячейку ($i\ j\ k$) и положим $v_{ijk} = \ln(p_{ijk})$.

Для такой таблицы модель имеет вид:

$$v_{ijk} = \mu + \lambda_i^X + \lambda_j^Y + \lambda_k^Z + \lambda_{ij}^{XY} + \lambda_{ik}^{XZ} + \lambda_{jk}^{YZ} + \lambda_{ijk}^{XYZ}. \qquad (6)$$

Чтобы в модели не получился избыток, т.е. чтобы число параметров не превысило числа ячеек ($I \times J \times K$), на значения λ в (6) накладываются следующие ограничения (аналогичные ограничениям (5)):

$$\sum_i \lambda_i^X = \sum_j \lambda_j^Y = \sum_k \lambda_k^Z = \ldots \sum_i \lambda_{ij}^{XY} = \sum_j \lambda_{ij}^{XY} = \ldots \sum_k \lambda_{ijk}^{XYZ} = 0. \quad (7)$$

Существует простой алгоритм, обеспечивающий определение всех значений λ. Коротко опишем его.

Введем обозначение:

$$v_{\ldots} = \sum_i \sum_j \sum_k \frac{v_{ijk}}{IJK}. \quad (8)$$

Ясно, что $v_{\ldots}$ будет общим средним для логарифмов вероятностей. Пусть

$$v_{i\cdot\cdot} = \sum_j \sum_k \frac{v_{ijk}}{JK}. \quad (9)$$

Тогда $v_{i\cdot\cdot}$ будет средним по всем тем логарифмам вероятностей, для которых фактор X находится на уровне i. Подставляя выражение (6) в правые части уравнений (8) и (9), мы получим:

$$\lambda_i^X = v_{i\cdot\cdot} - v_{\ldots}. \quad (10)$$

Таким образом, λ_i^X оказывается мерой того, насколько более (или менее) вероятна категория X_i по сравнению со средним по всем категориям этого фактора.

Сходным образом можно получить формулы и для других λ. Так, например, если ввести обозначения:

$$v_{ij\cdot} = \sum_k \frac{v_{ijk}}{K}, \; v_{i\cdot k} = \sum_j \frac{v_{ijk}}{J} \text{ и т.д.,}$$

то будут справедливы равенства

$$\lambda_{ij}^{XY} = v_{ij\cdot} - v_{i\cdot\cdot} - v_{\cdot j\cdot} + v_{\ldots}, \quad (11)$$

$$\lambda_{ijk}^{XYZ} = v_{ijk} - v_{ij\cdot} - v_{i\cdot k} - v_{\cdot jk} + v_{i\cdot\cdot} + v_{\cdot j\cdot} + v_{\cdot\cdot k} - v_{\ldots}. \quad (12)$$

Соотношение (11) показывает, что λ_{ij}^{XY} – это мера того, насколько совместное появление категорий X_i и Y_j более (или менее) вероятно, чем можно было бы ожидать, если бы они были независимы. Аналогично λ_{ijk}^{XYZ} – это мера того, насколько независимость факторов *X* и *Y* сама зависит от категории фактора *Z*.

Когда все три переменные *X*, *Y*, *Z* являются дихотомическими (т.е. когда $I = J = K = 2$), соотношения типа (8) и (9) значительно упрощаются, обеспечивая тем самым возможность более ясно описать смысл величин λ. Мы обнаруживаем, например, что справедливо соотношение

$$\lambda_{111}^{XYZ} = \frac{1}{8}\ln\left(\frac{p_{111}p_{221} / p_{121}p_{211}}{p_{112}p_{222} / p_{122}p_{212}}\right).$$

Другими словами, значение λ_{111}^{XYZ} пропорционально логарифму отношения отношений преобладаний для двух таблиц 2×2, соответствующих двум категориям переменной *Z*.

Если в этих двух таблицах переменные *X* и *Y* связаны друг с другом примерно в одинаковой степени, то и значения соответствующих отношений преобладаний θ_1 и θ_2 будут примерно одинаковыми. Тогда $\frac{\theta_1}{\theta_2}$ будет близко к единице, а $\lambda_{111}^{XYZ} = \frac{1}{8}\ln(\frac{\theta_1}{\theta_2})$ близко к нулю.

Для трехмерной таблицы произвольной размерности $I \times J \times K$ данные представляют собой совокупность наблюдаемых частот $\{n_{ijk}\}$, каждой из которых отвечает одноименная ячейка с индексом $(i\,j\,k)$.

Обозначим $f_{ijk} = \ln(n_{ijk})$ и найдем оценки параметров, заменяя значения v в (11) и (12) на соответствующие им значения *f*. Например, оценкой для λ_{ij}^{XY} из (11) будет

$$\overline{\overline{\lambda}}_{ij}^{XY} = f_{ij.} - f_{i..} - f_{.j.} + f_{...},$$

где $f_{ij.} = \sum_k \frac{v_{ijk}}{K}$, $f_{i..} = \sum_j \sum_k \frac{f_{ijk}}{JK}$ и т.д.

С помощью изложенного выше метода было проведено исследование трехмерной частотной таблицы (см. табл. 4).

Основанием табл. 4 служат три переменные: *X* (возраст), *Y* (участие в выборах), *Z* (образование).

Таблица 4

ТАБЛИЦА СОПРЯЖЕННОСТИ ДЛЯ ТРЕХ ПРИЗНАКОВ: «ВОЗРАСТ», «УЧАСТИЕ В ВЫБОРАХ» И «ОБРАЗОВАНИЕ»

Образование	Возраст	Участие в выборах	
		Всегда участвуют	Не всегда участвуют
Другое	18–45 лет	33	197
	Старше 45 лет	96	167
Высшее	18–45 лет	25	96
	Старше 45 лет	24	61

Для примера была выбрана ячейка с индексом (111) и найдены следующие значения оценок параметров λ:

$$\bar{\lambda}_1^X = -0{,}05;\ \bar{\lambda}_1^Y = -0{,}58;\ \bar{\lambda}_1^Z = 0{,}42;$$

$$\bar{\lambda}_{11}^{XY} = -0{,}20;\ \bar{\lambda}_{11}^{YZ} = -0{,}01;\ \bar{\lambda}_{11}^{XZ} = -0{,}17;\ \bar{\lambda}_{111}^{XYZ} = -0{,}44.$$

Проанализируем полученные результаты. Наибольший интерес для нас представляют значения тех $\bar{\lambda}$, которые характеризуют взаимодействие между переменными. Во-первых, имеет место довольно сильный эффект взаимодействия $\bar{\lambda}_{11}^{XY} = -0{,}20$. Знак «минус» указывает на то, что вследствие этого взаимодействия наблюдаемая частота n_{111} имеет меньшее значение, чем если бы X и Y были независимы. Интересным является тот факт, что значение $\bar{\lambda}_{11}^{YZ}$ оказалось близким к нулю. Это указывает на почти полное отсутствие связи между участием в выборах и образованием. Аналогичный вывод был получен на основе анализа табл. 2 методом отношения преобладаний. Значение $\bar{\lambda}_{11}^{XZ}$ в данном случае особого интереса не представляет, так как мы не рассматриваем вопрос о связи между переменными X (возраст) и Z (образование).

Вычислим значения отношений преобладаний для двух двумерных таблиц, составляющих табл. 4: $\theta_1 = 3{,}43$ для респондентов без высшего образования, $\theta_2 = 1{,}51$ для респондентов с высшим образованием. Таким образом, для первой группы респондентов существует довольно сильная зависимость участия в выборах от возраста. Для второй же группы эта связь слаба. Столь заметную разницу в силе взаимодействия между X и Y для выделенных групп респондентов отражает также довольно большое по абсолютной величине значение $\bar{\lambda}_{111}^{XYZ} = -0{,}44$.

Полагаем, что мы сумели убедить читателя в эффективности использования методов отношения преобладаний и построения логарифмически линейных моделей в процессе решения социологических задач изучения связей между переменными.

В западной социологии рассмотренные методы широко применяются для исследования многомерных таблиц сопряженности. Они давно включаются практически во все учебники по анализу данных (например, [4]; см. также хорошо известную на Западе и переведенную на русский язык работу [5]). В настоящее время соответствующие подходы в методической литературе обычно преподносятся как частный случай более общего класса методов – алгоритмов построения так называемых обобщенных линейных моделей [4] (коротко о них см. в [1]). Не столь благополучно обстоит дело в отечественной социологии.

Описание логлинейного анализа в отечественной социологической литературе

Как мы уже отмечали, в российской социологической практике логлинейный анализ, несмотря на свою перспективность, используется крайне редко (что касается отношений преобладания, то нам известны только два факта их упоминания [5, с. 112; 1]). Тем не менее работы, посвященные логлинейному анализу, в отечественной литературе имеются.

Первым в нашей стране применил логлинейный анализ в социально-демографических исследованиях Мирзоев [7]. В работах [8–9] им описана математическая сторона логлинейного анализа. В коллективной монографии [10, с. 68–72, 273–284] тот же автор представил свое видение возможных стратегий использования рассматриваемого метода в социологическом исследовании. В публикациях [6, с. 108–115; 11, с. 166–187; 12] логлинейный анализ сравнивается с другим известным статистическим методом, основанным на моделировании частот многомерной таблицы сопряженности, – каноническим анализом. Мы считаем, что для читателя-социолога было бы небесполезно сравнить соответствующие подходы, вдуматься в то обстоятельство, что модели частот могут быть разными, что выбор той или иной модели зависит от исследовательских предположений о тех обстоятельствах, которые обусловливают различие частот в таблице сопряженности.

В [1] осуществляется сравнение логлинейного анализа с другими методами поиска взаимодействий. Принятые в математической статистике и анализе данных понимания термина «взаимодействие» при этом обобщаются (напомним, что в анализе данных этот термин трактуется как сочетание значений признаков, детерминирующих интересующее социолога явление).

Строгое математическое описание логлинейного анализа представлено в [13].

ЛИТЕРАТУРА

1. *Толстова Ю.Н.* Анализ социологических данных: методология, дескриптивная статистика, изучение связей между номинальными признаками. М.: Научный мир, 2000.

2. *Rudas T.* Odds Ratios in the Analysis of Contingency Tables // Sage University Paper Series on Quantitative Applications in the Social Sciences, Series No. 07-119. Thousand oaks, CA: Sage, 1998.

3. *Goodman L.A.* The Multivariate Analysis of Qualitative Data: Interactions among Multiple Classification. J. Amer. Statist. Assoc., 1970. V. 65.

4. *Agresti A.* An Introduction to Categorical Data Analysis. N.Y.: John Wiley&Sons Inc., 1996.

5. *Аптон Г.* Анализ таблиц сопряженности. М.: Финансы и статистика, 1982.

6. *Елисеева И.И.* Статистические методы измерения связей. Л.: Изд-во ЛГУ, 1982.

7. *Мирзоев А.А.* Применение логлинейного анализа в социально-демографических исследованиях: Автореф. дис. ... канд. наук. М., 1980.

8. *Мирзоев А.А.* Логлинейный анализ социологической информации // Многомерный анализ социологических данных (методические рекомендации, алгоритмы, описание программ). М.: ИСИ АН СССР, 1981.

9. *Мирзоев А.А.* Применение логлинейного анализа для обработки данных социологических исследований // Математико-статистические методы анализа данных в социологических исследованиях. М.: ИС АН СССР, 1980.

10. Типология и классификация / Под ред. В.Г. Андреенкова, Ю.Н. Толстовой. М.: Наука, 1982.

11. *Елисеева И.И., Рукавишников В.О.* Логика прикладного статистического анализа. М.: Финансы и статистика, 1982.

12. Интерпретация и анализ данных в социологических исследованиях / Под ред. В.Г. Андреенкова, Ю.Н. Толстовой. М.: Наука, 1987.

13. *Миркин Б.Г.* Анализ качественных признаков и структур. М., 1980.