

---

---

А.И. Коченков, Ю.Н. Толстова  
(Москва)

## ИДЕИ ЛАТЕНТНО-СТРУКТУРНОГО АНАЛИЗА ЛАЗАРСФЕЛЬДА В СОВРЕМЕННОЙ СОЦИОЛОГИИ<sup>1</sup>

В статье коротко формулируются основные положения латентно-структурного анализа П.Ф. Лазарсфельда. Показывается, как они отражены в отечественной социологической литературе. Коротко описываются основные направления развития латентно-структурного анализа, существующие в западной литературе.

*Ключевые слова:* латентно-структурный анализ, тест, латентно-классовый анализ, латентно-классовая модель, модель латентных переменных (модель основной переменной), модель Раша, аксиома локальной независимости, локальная зависимость, критерий «Хи-квадрат».

### *О творчестве Лазарсфельда и его отражении в советской литературе*

Известный американский (австрийский, в 1933 г. эмигрировал в США) социолог-эмпирик, крупнейший специалист по методологии социологических исследований П.Ф. Лазарсфельд (1901–1976) значительную часть своего творчества посвятил проблеме операционализации понятий, разработке таких способов опроса респондентов, которые дали бы возможность измерить их латент-

---

**Андрей Игорьевич Коченков** – аспирант социологического факультета Московского государственного университета им. М.В. Ломоносова.

**Юлиана Николаевна Толстова** – доктор социологических наук, профессор Государственного университета – Высшая школа экономики.

<sup>1</sup>Статья подготовлена при финансовой поддержке РФФИ. Проект № 02-06-80403.

ные характеристики (мнения, установки и т.д.; латентной мы называем переменную, не поддающуюся измерению с помощью прямого обращенного к респонденту вопроса). Эта сторона деятельности Лазарсфельда коротко описана, например, в [1].

Вершиной разработок американского ученого, осуществленных в указанном направлении, теорией, в которой воплотились многие содержательные соображения (а Лазарсфельд подчеркивал, что содержательная социологическая теория и методика исследования – это не две разные области социологии, но очень часто – одно и то же) явился латентно-структурный анализ (ЛСА, в англоязычной литературе – LSA) [2; 3]. См. также переведенные на русский язык работы [4; 5].

Суть модели, предложенной Лазарсфельдом, сводилась к следующему. Мы полагаем, что существует некоторая латентная глубинная переменная, которая объясняет внешнее поведение респондентов. Поведение же это выражается в виде ответов каждого человека на определенные дихотомические вопросы анкеты. Латентная переменная номинальна, число ее значений заранее известно исследователю. Объясняющая способность латентной переменной обуславливается тем, что именно она служит причиной наличия связи между наблюдаемыми переменными<sup>1</sup>. Фундаментальной для лазарсфельдовского варианта ЛСА явилась *аксиома локальной независимости*: при фиксации значения латентной переменной связи между наблюдаемыми переменными исчезают (заметим, что эта аксиома лежит в основе очень многих ме-

---

<sup>1</sup> Современные варианты ЛСА предусматривают возможность использования любых шкал и для наблюдаемых переменных, и для латентной. Воспользуемся этим и поясним сказанное с помощью порядковых признаков. Для них можно привести простой и «прозрачный» пример. Латентная переменная объясняет связи между наблюдаемыми, если одновременный рост значений всех переменных при переходе от одного индивида к другому связан с тем, что первый индивид имеет меньшее значение латентной переменной, чем второй (этот рост понимается статистически, и для отдельных переменных могут быть незначительные «сбои»).

тодов шкалирования, активно использующихся в социологических исследованиях; Лазарсфельд первым сформулировал ее в явном виде). Более подробное описание лазарсфельдовской модели мы приведем в следующем разделе.

Далее предложенная Лазарсфельдом относительно простая модель ЛСА глубоко изучалась западными авторами и постепенно усложнялась. Основные моменты этого процесса будут кратко описаны нами ниже. Некоторые итоги были подведены самим Лазарсфельдом в ставшей классической работе [3].

С простейшим вариантом ЛСА можно познакомиться также по посвященным ЛСА фрагментам подготовленных Институтом социологии АН СССР книг [6, с. 140–151; 7; 8; 9]. Там же указаны некоторые известные на момент написания каждой работы направления развития первоначальных лазарсфельдовских идей.

Можно назвать и такие работы советских ученых, в которых ЛСА (в его простейшем варианте) был использован для решения конкретных социологических задач. Например, в работе [10] с помощью ЛСА были измерены такие синтетические латентные переменные как «обеспеченность материальным имуществом», «удовлетворенность жизнью», «склонность к общественной деятельности», «удовлетворенность рабочей обстановкой». В статье [11] были выявлены латентные факторы, определяющие отношение респондентов к отдельным характеристикам городской среды. В работе [12] построена типология общественно-политической активности молодых рабочих (использовался гудменовский вариант ЛСА, о котором будет сказано ниже). Еще один пример практического использования ЛСА приведен в публикации [13].

На связь идеи ЛСА с задачей поиска однородных совокупностей объектов указано в работе [14, с. 137]. Автор показывает, что при разумном понимании однородности, однородными имеет смысл считать такие группы объектов, которые попадают в один класс при использовании ЛСА, т.е. такие группы, в которых наблюдаемые переменные независимы (автор принимает аксиому локальной неза-

висимости). Другими словами, социальная группа считается однородной при слабой статистической связи основных показателей ее жизнедеятельности. «Действительно, сильная связь между показателями означала бы, что люди, отличающиеся, например, по образованию, отличаются и по другим характеристикам: по-разному ведут трудовую деятельность... используют свое время и т.п. В этом случае уровень образования будет существенно дифференцировать людей по характеру жизнедеятельности. Напротив, при отсутствии связи между признаками выделение лиц с тем или иным уровнем образования никак не дифференцирует людей по другим характеристикам их жизнедеятельности: группы разного уровня образования имеют одинаковую структуру трудовой деятельности, досуга, быта и т.п.». Автор говорит о том, что представления, уточняющие этот подход к пониманию однородности, сформулированы в работе [15].

К сожалению, в последние годы интерес российских исследователей к рассматриваемому подходу, практически, не проявлялся. Почти никакие изменения и усовершенствования первоначальной лазарсфельдовской модели, насколько нам известно, не использовались в отечественной социологической практике (исключение составляет одна из моделей, предложенных Гудменом в 70-х гг., о чем пойдет речь ниже).

### *Основные положения латентно-структурного анализа Лазарсфельда*

Напомним основные идеи, заложенные в первоначальном ЛСА. Без этого трудно говорить о том, в каком направлении развивались идеи Лазарсфельда.

Итак, полагаем, что латентная переменная номинальна и нам известно число ее градаций (т.е. количество искомых латентных классов). Для простоты предположим, что это число равно двум. Наблюдаемые переменные – дихотомические, принимают значения 0 и 1. Будем условно называть первый ответ отрицательным, а второй –

положительным. Введем обозначения:  $v^1$  и  $v^2$  – доли объектов, попавших в первый и второй латентные классы соответственно;  $P_i$  – доля всех респондентов, положительно ответивших на  $i$ -й вопрос анкеты;  $p_i^1$  и  $p_i^2$  – аналогичные доли для первого и второго латентного класса соответственно,  $P_{ij}$  – доля людей, положительно ответивших на  $i$ -й и  $j$ -й вопросы одновременно;  $p_{ij}^1$  и  $p_{ij}^2$  – аналогичные доли для первого и второго класса соответственно. Аналогично определяются величины  $P_{ijk}$ ,  $p_{ijk}^1$ ,  $p_{ijk}^2$  и т.д. Заметим, что здесь и ниже мы говорим о частотах, о частотных распределениях признаков, т.е. только о выборочных данных, что не вполне корректно. Следовало бы использовать и терминологию, имеющую смысл для генеральной совокупности – говорить о распределении вероятностей встречаемости значений соответствующих случайных величин. При этом, конечно, нужно осознавать необходимость переноса результатов с выборки на генеральную совокупность хотя бы в виде построения доверительных интервалов для всех искомых величин.

По *формуле полной вероятности* (предполагаем, что наши классы являются взаимоисключающими и исчерпывающими, т.е. каждый объект попадает в один и только один латентный класс),

$$\begin{aligned} P_i &= v^1 \cdot p_i^1 + v^2 \cdot p_i^2 \\ P_{ij} &= v^1 \cdot p_{ij}^1 + v^2 \cdot p_{ij}^2 \end{aligned} \tag{1}$$

.....

Добавляя сюда очевидное соотношение

$$v^1 + v^2 = 1 \tag{2}$$

получим систему уравнений, в которой известными являются величины, стоящие в левых частях уравнений (1), а неизвестными – величины, стоящие в правых частях тех же уравнений.

Эту систему, вообще говоря, нельзя решить, поскольку число уравнений в ней гораздо меньше числа неизвестных. На помощь приходит *аксиома локальной независимости*. В соответствии с ней, имеет место соотношение:

$$p_{ij}^1 = p_i^1 \cdot p_j^1, p_{ij}^2 = p_i^2 \cdot p_j^2 \quad (3)$$

Благодаря этому соотношению, количество неизвестных резко снижается. В результате система поддается решению.

На практике использование описанной модели ЛСА осуществляется примерно таким образом. Выбирается ориентировочная латентная переменная (культурный уровень респондента, уровень урбанизации региона и т.д.). Подбираются анкетные вопросы, наиболее связанные, по мнению исследователя, с этой переменной. Число градаций латентной переменной выбирается из содержательных соображений. На основе ответов на вопросы анкеты респондентов *якобы* строится латентная объясняющая переменная, т.е. вся совокупность индивидов разбивается на классы так, чтобы внутри каждого класса ответы на вопросы анкеты стали статистически независимыми. Подчеркнем, что слово «якобы» мы не напрасно выделили. В действительности происходит следующее. Решается система уравнений (1) и (2) с учетом условий (3). Найденные неизвестные  $v^1, v^2, p_i^1, p_{ij}^1, p_i^2, p_{ij}^2$  и т.д. интерпретируются описанным выше образом. Другими словами, мы полагаем, что нашли распределения значений наших признаков в латентных классах и соотношение этих классов по их объему. Преобразуя указанные величины с помощью формул Байеса, можно найти вероятность принадлежности к каждому рассматриваемому классу любого респондента, давшего произвольный набор ответов на вопросы данной анкеты. Таким образом, поиск нашей латентной переменной сводится к построению однородных в статистическом смысле классов, причем для каждого объекта определена лишь вероятность его отнесения к тому или иному классу, нахождение объемов этих классов и вероятностей встречаемости любых наборов ответов на вопросы анкеты для респондентов, принадлежащих к каждому классу.

Подчеркнем, что доказательство существования латентной переменной при этом остается как бы на совести исследователя. Решив формально систему уравнений (1), (2), мы будем считать,

что полученные величины отражают именно одномерные распределения, отвечающие входящим в нашу анкету признакам. Однако логика, на которой базируется это утверждение, весьма уязвима. Поясним это.

Прежде всего, необходимо отдавать себе отчет в том, что с гносеологической точки зрения довольно сложным и «темным» является само понятие «латентная переменная». Нелегко ответить на вопрос о том, что значит существование этой переменной.

Вспомним, что понятие признака почти всегда является результатом абстрагирования весьма высокого уровня. Мы уходим от уникальности каждого изучаемого объекта (у нас – респондента), выделяем в каждом из них нечто похожее на другого. И не так-то просто понять, правы ли мы в своих предположениях. Вероятно, доказательством адекватности наших посылок может служить только наблюдение реального отличия какого-то интересующего нас (связанного с латентной переменной) поведения респондентов, принадлежащих к разным латентным классам. Однако мы чаще всего не задаемся целью тщательной проверки справедливости сформулированных гипотез. Решили систему уравнений (1), (2) в предположении справедливости (3) и полагаем, что нашли распределения анкетных признаков. А ведь логика формирования соотношений (1), (2), (3) была другая.

Да, если искомые латентные классы действительно существуют, то эти соотношения должны быть верны. А если не существуют? И в таком случае мы вполне можем решить систему (1), (2) при условии справедливости (3). Разве это решение будет служить доказательством справедливости предположения о существовании латентной переменной? Нет, конечно. Тем не менее, на практике мы обычно полагаем, что имеем дело с латентной переменной, даже не задумываясь над поставленными выше вопросами.

В литературе имеются некоторые результаты, говорящие о том, что описанные сомнения возникали в сознании ведущих ученых. Одним из существенных обстоятельств, приводящих к со-

ответствующим размышлениям, послужило то, что решением нашей системы иногда являются числа, большие единицы. Ясно, что в таком случае их нельзя интерпретировать как какие бы то ни было вероятности. Что же делать? Сомневаться в существовании латентной переменной? Может быть, это и было бы справедливо, но тогда в разработке формализма двигаться дальше некуда. Необходимо в корне пересматривать наши содержательные послышки. Однако некоторые ученые пошли по другому пути – в качестве причины описанного противоречия стали рассматривать нарушение аксиомы локальной независимости. Об этом – ниже.

### *Основные направления развития идей латентно-структурного анализа в современной западной социологии*

В настоящее время идеи Лазарсфельда достаточно популярны среди западных социологов. Представления о ЛСА включаются в современные образовательные программы для студентов-социологов (об этом косвенно свидетельствует, в частности, то, что данной проблематике посвящены два тома из известной серии книг по количественным методам в социальных исследованиях [16; 17]).

За последние 50 лет идеи, предложенные П.Ф. Лазарсфельдом, активно развивались. Назовем некоторые направления. При этом не будем претендовать на полноту приводимых сведений. Остановимся лишь на некоторых моментах, представляющихся нам наиболее существенными.

### *Сравнение моделей ЛСА и факторного анализа*

В работе [18] было показано, что факторный анализ является частным случаем общей модели латентно-структурного анализа с несколькими латентными переменными. Здесь, однако, имеет смысл отметить, что довольно часто в литературе полагают, что ЛСА –



метод, решающий для порядковых и номинальных данных те же задачи, что делает для интервальных данных факторный анализ; см., например, [17]. Более глубокий анализ соотношения между факторным и латентно-структурным анализом осуществлен, например, в [19]. В этой работе после краткого описания развития факторного анализа и латентно-структурных моделей показывается, как последние могут быть обобщены для исследования соотношений между количественными измерениями таким способом, чтобы избежать некоторых проблем факторного анализа. Возникающая в результате модель латентного профиля применяется затем к эмпирическим данным с целью демонстрации эффекта ее использования.

Ниже речь пойдет о соображениях, предложенных в работах [20; 21; 22; 23] (коротко о соответствующих моделях говорится в [12]). Можно сказать, что именно в этих публикациях был заложен целый ряд направлений модификации лазарсфельдовских представлений. Заметим, что варианты ЛСА, о которых ниже пойдет речь, в наше время называются также латентно-классовым анализом (ЛКА, или LCA). Это не касается случая, когда латентная переменная непрерывна.

#### *Использование критерия «Хи-квадрат» при поиске латентных параметров модели*

В названных работах предложен способ поиска описанных выше латентных параметров (объемов классов, распределений наблюдаемых признаков в каждом классе, вероятностей отнесения респондента, давшего определенный ответ, к тому или иному классу), отличный от лазарсфельдовского.

Отличие прежде всего состоит в переходе к логике математической статистики – логике статистического оценивания параметров генеральной совокупности по выборочным данным; см. также [24]. Предполагаем, что вовсе не обязательно стремиться к точному соблюдению равенства (1). Будем считать, что это равенство отвечает теоретическому (ожидаемому) распределе-

нию наблюдаемых признаков – тому распределению, которое справедливо для генеральной совокупности. Для эмпирического распределения, наблюдаемого по выборке, равенство (1) может нарушаться.

Система уравнений для расчета этих параметров опирается на использование только таких частот, которые отвечают наборам ответов на все вопросы анкеты сразу. Решение находится с помощью итеративного процесса. Цель итерации состоит в том, чтобы теоретические частоты, которые определяют с помощью соотношений (1) и (2), как можно лучше соответствовали бы наблюдаемым данным. Поясним, что означает такое соответствие. При этом для простоты предположим, что в анкету включены три вопроса.

На очередном шаге посредством найденных латентных параметров рассчитываются так называемые теоретические (ожидаемые) частоты  $\hat{P}_{ijk}$  (речь идет о вычислении с помощью модельных расчетов количества людей, давших заданный набор ответов на наши три вопроса), и эти частоты сравниваются с реальными, наблюдаемыми частотами  $P_{ijk}$  (т.е. теми частотами, которые фактически служили нам исходными данными). Сравнение осуществляется посредством известного критерия «Хи-квадрат». Если окажется, что для очередного шага нашей итерации величина этого критерия (естественно, при заданном уровне значимости) превышает табличное значение, то переходим к следующему уровню итерации. Если не превышает – считаем, что найденные латентные параметры достаточно хорошо отражают реальность.

#### *Обеспечение возможности выбора количества латентных классов (числа значений латентной переменной)*

Каждый шаг описанной выше итерации определяется числом латентных классов. Результаты будут тем более эффективными, чем большее количество латентных классов (т.е. значений латентной переменной) мы рассматриваем. Более того, скажем,

при двух латентных классах вообще можно не найти удовлетворительного решения, а при трех – найти и т.д.

Реализацию итеративного процесса начинают с ситуации, когда количество латентных классов равно двум. Далее возьмем ситуацию, когда латентная переменная предположительно имеет три значения, и повторим все сначала. Вероятность нахождения адекватного решения (и качество такового) повышается. Однако необходимо иметь в виду, что слишком большое количество латентных классов может привести к невозможности удачно их проинтерпретировать.

О том, насколько увеличение числа латентных классов от  $n$  до  $n+1$  позволяет улучшить различие между наблюдаемыми и ожидаемыми частотами, можно судить по разнице соответствующих значений критерия «Хи-квадрат» [25; 12]. Обычно процесс останавливают, когда переход от  $n$ -го к  $(n+1)$ -му шагу итерации дает малое изменение этого критерия.

Кроме критерия «Хи-квадрат», существуют и другие способы оценки соответствия теоретических и эмпирических частот: информационные критерии и индексы модельного соответствия. С этим можно подробно ознакомиться в [17, с. 13].

### *Обеспечение возможности рассмотрения непрерывной латентной переменной*

Попытки обобщения лазарсфельдовских идей на ситуацию, когда латентная переменная является непрерывной (и отвечающей числовой шкале), осуществлялись неоднократно. Поначалу такие обобщения носили теоретический характер. Это отмечали авторы работы [3]. Затем соответствующие соображения были переведены во вполне конструктивное русло. Примером могут служить описанные выше предложения Гудмена и Хабермана по поиску оптимального числа значений латентной переменной. Рассмотрим ситуацию, когда латентная переменная непрерывна (является числовой, с бесконечным числом значений), а вопросы в анкете – ди-

хотомические. Установим соответствие между основными понятиями и соотношениями «дискретного» и «непрерывного» вариантов латентно-структурного анализа. Введем новые обозначения:  $x$  – действительное число, значение нашей латентной переменной (в литературе чаще всего в качестве такой переменной рассматривают установку респондентов);  $\Phi(x)$  – плотность распределения латентной переменной, являющейся случайной величиной (напомним, что случайная величина по определению характеризуется заданием функции распределения или плотности распределения);  $f_i(x)$  – так называемый график  $i$ -го вопроса, т.е. вероятность положительного ответа на этот вопрос для респондента, имеющего значение  $x$  латентной переменной.

Нетрудно понять, что между этими и введенными ранее обозначениями можно установить следующее соответствие.

Таблица 1

СВЯЗЬ МЕЖДУ ДИСКРЕТНЫМ И НЕПРЕРЫВНЫМ  
ВАРИАНТАМИ ЛСА

Обозначения для дискретного случая	Обозначения для непрерывного случая
$v^1, v^2$	$\int_{d_k}^{d_{k+1}} \Phi(x) dx$ , где $(d_k, d_{k+1})$ – произвольный отрезок числовой оси
$v^1 + v^2$	$\int_{-\infty}^{+\infty} \Phi(x) dx$
$p_i^1, p_i^2$	$f_i(x)$
$v^1 \cdot p_i^1 + v^2 \cdot p_i^2$	$\int_{-\infty}^{+\infty} f_i(x) \cdot \Phi(x) dx$

Описание моделей с непрерывной латентной переменной можно найти в русскоязычных работах [8, с. 109–110; 26, с. 56–62].

Большое внимание в литературе уделяется виду графиков вопросов (функциям  $f_i(x)$ ). Рассматриваются разные случаи. При-

меры различных графиков вопросов приведены в [8, с. 111]. Выделим наиболее часто использующуюся в тестовых опросах функцию – функцию Раша:

$$f_i(x) = \frac{c_i x}{1 + c_i x}.$$

Специфика этой функции определяет специфику поиска значения латентной переменной для каждого респондента. Соответствующая техника становится не очень похожей на традиционную технику ЛСА и ЛКА. При ее применении говорят об использовании модели Раша. Описание модели Раша можно найти в работах [27; 8, с. 284–290].

#### Связь ЛСА с теорией тестов

Рассмотрение латентной переменной как непрерывной характеристики сближает ЛСА с теорией тестов. Собственно, о сравнении соответствующих подходов писал и сам Лазарсфельд [5]. Мы коснемся результатов, связанных с использованием упомянутой выше модели Раша. Они лежат в рамках не только ЛСА, но и определенного направления теории тестов – работе с моделями с латентной основной переменной (latent trait models, LTM). Подобных публикаций довольно много; назовем, например, переведенную на русский язык статью [28], которая в подлиннике называется «The relation of test score to the trait underlying the test». Значимость этого подхода станет более объяснимой, если вспомнить, что упомянутая выше функция Раша обычно изменяется следующим образом:

$$P_{ij} = \frac{1}{1 + \exp(T_i - B_j)},$$

где  $P_{ij}$  – вероятность того, что испытуемый с  $T_i$ -м уровнем знаний правильно выполнит задание, трудность которого равна  $B_j$ .

В последние годы российские исследователи активно занимаются внедрением и модификацией модели Раша, созданием соответствующих пакетов компьютерных программ в рамках

разработки системы эффективной оценки знаний учащихся в системе российского образования [29; 30; 31]. Одна из них – диалоговая система «Rasch Measurement» – описана в [32].

Сравнение подходов, лежащих в русле тестовой традиции, с подходами ЛСА, осуществлялось неоднократно. Следует отметить, например, названную выше работу самого Лазарсфельда [5]. Западными исследователями в последние годы активнейшим образом проводится сравнение модели Раша с более каноническими приемами ЛКА (ЛСА сравнивается с LTM). Об этом красноречиво свидетельствует появление сборника [33], с которым можно ознакомиться в Интернете по адресу: <http://www.ipn.uni-kiel.de/aktuell/buecher/rostbuch/inhalt.htm>. Этот сборник состоит из 40 статей, каждая из которых описывает использование рассматриваемых подходов в приложении к образовательной сфере.

В качестве дополнительного примера сравнительного использования ЛКМ и других моделей подобного рода можно назвать работу [34]. Применение ЛКМ в кластерном, факторном и регрессионном анализе в сочетании с примерами реализации этих моделей с помощью компьютерных программ можно найти в [35].

### *Обеспечение возможности рассмотрения недихотомических вопросов*

Предполагается, что на включенные в анкету вопросы может быть предложено не два, а любое конечное количество ответов. Пример (с тремя вариантами ответов на каждый вопрос) будет приведен в следующем пункте.

**Связь ЛСА со шкалограммным анализом Гуттмана.** Предположим, что мы на базе дихотомических вопросов в анкете (будем отождествлять вопрос в анкете с соответствующим ему признаком, характеризующим респондента) пытаемся измерить некоторую установку респондента с помощью техники шкалограммного анализа Гуттмана [36]; см. также, например, [6; 9]. Напомним, что эта техника опирается на предположение о том, что,

если измеряемая латентная переменная действительно существует и наблюдаемые признаки хорошо ее отражают, то эти признаки могут быть упорядочены в соответствии со следующим принципом: один признак предшествует другому, если положительный ответ на первый вопрос влечет положительный ответ на второй (это утверждение обычно воспринимается как нечто приближительное: полагаем, что оно верно в том случае, если признаки удалось проранжировать лишь с некоторой погрешностью; о том, как строится соответствующий критерий, можно прочитать в названной выше литературе).

Так же, как и ЛКА, идеи Гуттмана были обобщены на случай, когда наблюдаемые переменные имеют более двух значений. При этом предполагается, что эти переменные получены, по крайней мере, по порядковой шкале.

Поясним суть описанного ранжирования признаков, считая, например, что у нас имеется три переменных – *A*, *B*, *B*, каждая из которых может принимать три значения – 1, 2, 3 ( $1 < 2 < 3$ ).

Допустим, что нам удалось упорядочить признаки таким образом: *B*, *B*, *A*. Это означает, что у нас встречаются только следующие сочетания этих признаков (с точностью до небольшой погрешности).

*Таблица 2*  
ДОПУСТИМЫЕ СОЧЕТАНИЯ ЗНАЧЕНИЙ  
УПОРЯДОЧЕННЫХ ПРИЗНАКОВ

Номера сочетаний значений признаков	Признаки		
	Б	В	А
1	3	3	3
2	3	3	2
3	3	3	1
4	3	2	2
5	3	2	1
6	3	1	1
7	2	2	2
8	2	2	1
9	2	1	1
10	1	1	1

Итак, если у нас встречаются только упомянутые сочетания значений, т.е. если мы имеем основания полагать, что предположения о существовании латентной переменной и адекватности наблюдаемых не лишены основания (как и выше, возможность упорядочения признаков не доказывает нам справедливости этих предположений, а только говорит о том, что наблюдаемые данные не противоречат им), то это обстоятельство можно считать и косвенным подтверждением того, что решение уравнений типа (1), (2) тоже отражает некоторую структуру латентных классов.

В [22] введены понятия шкалируемых и существенно нешкалируемых респондентов.

К первым относятся респонденты, дающие ответы типа наших наборов 1-10. Подобные наборы называются разрешенными. Помимо латентных классов, отвечающих значениям искомой латентной переменной, рассматривается еще один латентный класс – класс нешкалируемых респондентов, которые дают неразрешенные ответы. Таких классов может быть и несколько.

Вспомним, что техника ЛСА позволяет сказать, к какому именно классу с наибольшей вероятностью будет отнесен тот или иной набор из числа пронумерованных нами в табл. 2. Наверное, мы можем считать, что шкалограммный анализ подтверждает модель ЛСА, если с примерно одинаковой вероятностью окажутся попадающими в один и тот же класс соседние наборы: скажем, к первому классу с примерно одинаковой и достаточно большой вероятностью окажутся отнесенными наборы 1, 2, 3. К второму – 4, 5, 6, 7, к третьему – 8, 9, 10.

Аналогичным образом применение ЛКА может помочь исследователю определить, как оптимальным образом проранжировать признаки. Более того, можно выйти и на такую ситуацию, когда окажется, что ранжировку можно осуществить более, чем одним способом.

Совместное использование ЛКА и шкалограммного анализа можно найти, например, в [17].



*Замена локальной независимости на квазинезависимость*

Вспомним упомянутую выше ситуацию, когда решение системы уравнений Лазарсфельда дает значения вероятностей, больших единицы. Выходом из нее может служить допущение того, что внутри одного класса (при одном и том же значении латентной переменной) признаки могут зависеть друг от друга. Сомнения в необходимости принятия аксиомы локальной независимости могут возникнуть и из других соображений. Эта тема требует специального обсуждения, к которому мы перейдем ниже. Здесь отметим, что в гудменовских моделях предполагается, что некоторая зависимость между наблюдаемыми признаками в латентных классах имеется, но не сильная (другими словами, речь идет о том, что наша латентная переменная лишь частично объясняет связи между наблюдаемыми). Гудмен назвал ее квазинезависимостью и предложил способ построения соответствующих моделей.

*Отказ от аксиомы локальной независимости (АЛН)*

Споры по поводу правомочности принятия АЛН начались почти сразу после ее четкого введения в научный обиход Лазарсфельдом. Выше мы уже говорили об одной из причин таких споров: именно нарушением этой аксиомы объясняли то, что решением систем (1), (2) иногда служили числа, большие единицы. Приведем некоторые другие соображения, позволяющие прийти к выводу о целесообразности отказа от обязательного соблюдения АЛН. Эти соображения очень близки к тем, которые лежат в основе причинного анализа [37]. Покажем это на примере.

Пусть латентная переменная – логические способности учащегося. В качестве наблюдаемых переменных выступают некие тестовые логические вопросы, на каждый из которых учащийся должен дать ответ «да-нет». Предполагаем, что связь между ответами объясняется именно тем, что у учащихся разный уровень логических способностей: человек с высокими способностями в основном дает правильные ответы, со средними – «серединок-

наполовинку», с низкими – отвечает в основном неправильно. Казалось бы, вполне допустимо предположение о том, что для каждой из указанных групп учащихся связь между их ответами на разные вопросы пропадает. Однако вполне может быть так, что, скажем, связь в классе слабых учащихся остается, поскольку здесь начинает «работать» другая латентная переменная – например, усидчивость. Среди слабых учащихся усидчивые лучше решают задачи, неусидчивые – хуже. В целом оценки низкие, но все же одни лучше, другие – хуже, и связь между ними имеется!

Лазарсфельдовская техника латентно-структурного анализа не дает нам возможности находить классы, в которых наблюдаемые переменные были бы связанными друг с другом. АЛН – основное предположение, на котором базируется вся лазарсфельдовская техника. Выше мы уже отмечали, что в гудменовских моделях предусмотрена возможность некоторого ослабления рассматриваемой аксиомы (мы имеем в виду квазинеzависимость переменных внутри латентных классов). Однако этим предложения западных исследователей не ограничиваются. Были разработаны и способы выявления ситуаций, когда имеет смысл отказаться от АЛН, и такие приемы использования техники ЛКА, при котором зависимость определенных признаков в латентных классах «вылезает», если она отвечает реальности.

В качестве способа выявления того, стоит ли отказываться от АЛН и, если стоит, то для каких именно переменных, назовем диагностическую процедуру, описанную в работе [38] (другой подход можно найти, например, в [39]). Эти авторы предложили диагностировать каждую пару дихотомических переменных. Строятся две четырехклеточные частотные таблицы – одна состоит из наблюдаемых частот, другая – из ожидаемых, вычисленных на основе модели в соответствии с описанным выше гудменовским подходом. Авторы разработали критерий, основанный на вычислении отношений преобладания, нормально распределенный при условии локальной независимости рассматриваемых двух

признаков во всех латентных классах. Рассуждаем традиционным образом: если значение этого критерия превышает табличное (естественно, предусматривается задание определенного уровня значимости), то считаем, что признаки в каких-то латентных классах должны быть зависимы. И для них применяется один из описанных ниже способов учета такой зависимости (методы ЛКА).

Одним из способов допущения появления связанных переменных в латентных классах является использование *метода объединения признаков*. Предположим, например, что диагностическая процедура дала нам основания думать, что из заданных четырех дихотомических признаков  $A, B, C, D$  признаки  $B$  и  $C$  могут быть локально зависимыми. Тогда с помощью обычной техники ЛКА рассмотрим последние два признака не отдельно, а введя новую переменную  $BC$ , принимающую четыре значения, каждое из которых отвечает одной из четырех возможных комбинаций исходных значений признаков  $B$  и  $C$ . Алгоритм строит латентные классы, в каждом из которых признаки  $A, BC, D$  независимы. А признаки  $B$  и  $C$  вполне могут оказаться зависимыми.

Другим методом допущения локальной зависимости может служить *метод множественных индикаторов* [20; 39]. На том же примере поясним его сущность. Мы предполагаем, что, наряду с нашей основной переменной  $X$ , обуславливающей взаимосвязь всех четырех переменных, действует еще одна переменная  $Y$  только для зависимости  $B$  и  $C$ , а  $X$  действует не непосредственно на  $B$  и  $C$ , а опосредованно, через  $Y$ . Модель подстраивается под эту ситуацию. Количество искомым латентных параметров возрастает.

Еще одним способом поиска латентной переменной, при котором возможна зависимость наблюдаемых переменных в отдельных классах, является использование *логлинейного анализа* (ЛЛА) [23]; о самом ЛЛА можно узнать, например, в [40]. Изучается частотная таблица, размерность которой на единицу больше количества наблюдаемых переменных, поскольку к последним

добавляется еще и латентная переменная  $X$ . Рассматриваются двумерные взаимодействия каждой наблюдаемой переменной с латентной. И, если мы хотим допустить локальную зависимость, скажем, наблюдаемых переменных  $B$  и  $C$ , то в подлежащую идентификации логлинейную модель, наряду с взаимодействиями  $BX$  и  $CX$ , мы «закладываем» также взаимодействие  $BC$ . В соответствии с правилами ЛЛА мы получаем вклад каждого из указанных двумерных взаимодействий в частоты рассматриваемой многомерной частотной таблицы.

Отметим, что применение ЛЛА с указанной целью в западной литературе противопоставляется использованию моделей Лазарсфельда и Гудмена, называемых классическими [41]. О представлении латентно-классовой модели в виде логлинейной говорилось в [23].

Перечисленные выше ситуации описаны в [42]. Там же можно найти рекомендации по применению компьютерных программ, позволяющих выявлять локально зависимые признаки и реализовывать ЛСА любым из трех описанных способов.

### *Программное обеспечение*

Для указанных выше моделей уже разработаны и давно используются различные программные пакеты, в основном западные. Описания их применения и соответствующие интернет-ссылки можно найти в [17]. Полный список компьютерных пакетов приведен на сайте: <http://members.aol.com/KMarkus/lca.html>.

Среди отечественных продуктов можно назвать Универсальный психодиагностический комплекс «ЭКСПЕРТ-ПСИконтроль», который применяют для психологических и кадровых служб. С ним можно ознакомиться на: <http://www.effinf.ru/7Ceni/Expert.shtml>.

Упоминание о другом программно-методическом комплексе «Эксперт» также можно найти в Интернете на сайте: [http://www.expert-plus.ru/expert\\_benefits.htm](http://www.expert-plus.ru/expert_benefits.htm). По заявлению разработ-

чиков данный продукт создает с помощью встроенного латентно-структурного анализа эталонные профили по выборке и экспертные системы на основе эмпирического и теоретического анализа и коррекции этих профилей.

#### ЛИТЕРАТУРА

1. *Лазарсфельд П.* Измерение в социологии // Американская социология: перспективы, проблемы, методы. М.: Прогресс, 1972.
2. *Lazarsfeld P.* A Conceptual Introduction to Latent Structure Analysis // Mathematical Thinking in the Social Sciences. N.Y.: Free Press, 1954.
3. *Lazarsfeld P.F., Henry N.W.* Latent Structure Analysis. Boston: Houghton Mifflin Co., 1968.
4. *Лазарсфельд П.* Логические и математические основания латентно-структурного анализа // Математические методы в современной буржуазной социологии. М.: Прогресс, 1966.
5. *Лазарсфельд П.Ф.* Латентно-структурный анализ и теория тестов // Математические методы в социальных науках. М.: Прогресс, 1973.
6. *Осинов Г.В., Андреев Э.П.* Методы измерения в социологии. М.: Наука, 1977.
7. Статистические методы анализа социологической информации. М.: Наука, 1979.
8. Типология и классификация в социологических исследованиях. М.: Наука, 1982.
9. *Толстова Ю.Н.* Измерение в социологии. М.: Инфра-М, 1998.
10. *Мамедов А.Г.* Методы анализа и прогнозирования социально-экономического поведения и задачи социального управления (на примере рабочей молодежи г. Баку): Автореф. дис. ... канд. эконом. наук. М.: ЦЭМИ АН СССР, 1984.
11. *Барабаш Н.Б.* Выявление отношений населения к элементам городской среды // Вестн. Моск. ун-та. Сер. 5, География. 1975. № 6.
12. *Дегтярев Г.П.* Построение типологии с помощью модели латентных классов // Математические методы в социологическом исследовании. М.: Наука, 1981.
13. *Комаровский В.С.* Некоторые проблемы измерения социальных установок // Вопросы философии. 1970. № 7.
14. *Миркин Б.Г.* Группировки в социально-экономических исследованиях. М.: Финансы и статистика, 1985.
15. *Гаврилец Ю.Н.* Социально-экономическое планирование: системы и модели. М.: Экономика, 1974.
16. *McCutcheon A.L.* Latent Class Analysis: Sage University Papers Series on Quantitative Applications in the Social Sciences. 07-64. Thousand oaks, CA: SAGE, 1987.

17. *Dayton C.M.* Latent Class Scaling Analysis: Sage University Papers Series on Quantitative Applications in the Social Sciences. 07-126. Thousand oaks, CA: SAGE, 1998.

18. *Anderson T.W.* Some Scaling Methods and Estimation Procedures in the Latent Class Model in Probability and Statistics. N.Y., 1959.

19. *Гибсон В.* Факторный, латентно-структурный и латентно-профильный анализ // Математические методы в социальных науках. М.: Прогресс, 1973.

20. *Goodman L.A.* The Analysis of System of Qualitative Variables when Some of the Variables Are Unobservable. P. 1: Modified Latent Structure Approach // *Amer. J. Sociol.* 1974a. Vol. 79. № 5.

21. *Goodman L.A.* Explanatory Latent Structure Analysis Using Both Identifiable and Unidentifiable Models // *Biometrika.* 1974b. Vol. 61. № 2.

22. *Goodman L.A.* A New Model for Scaling Response Patterns: An Applications of the Quasi-independent Concepts // *Journal of the American Statistical Association.* 1975. Vol. 70.

23. *Haberman S.J.* Qualitative Data Analysis. V. 2: New Developments. N.Y.: Academic Press, 1979.

24. *Anderson T.W.* On Estimation Parameters in Latent Structure Analysis // *Psychometrika.* 1954. Vol. 19.

25. *Миркин Б.Г.* Анализ качественных признаков и структур. М.: Статистика, 1980.

26. Моделирование социальных процессов. М.: Изд-во РЭА им. Г.В. Плеханова, 1993.

27. *Рэск Дж.* Индивидуальный подход к анализу вопросов // Математические методы в социальных науках. М.: Прогресс, 1973.

28. *Лорд Ф.М.* Отношение между тестовым баллом и исследуемой способностью // Математические методы в социальных науках. М.: Прогресс, 1973.

29. *Анисимова Т.С., Маслак А.А., Осипов С.А.* Сравнительный анализ модели Rasch и классической модели по точности оценивания // Материалы конференции «Анализ качества образования и тестирование». М.: Издательство МЭСИ, 2001.

30. *Маслак А.А.* Исследование эффективности системы тестирования на основе модели Раша / Третья Всероссийская научно-практическая конференция «Оценка эффективности образовательных инноваций и технологий». Славянск-на-Кубани: Изд-во СФ АГПИ, 2001.

31. *Maslak A.A.* Efficiency Estimation of Testing on Rasch Model // International Conference on Objective Measurement. Chicago: University of Illinois, 2001.

32. *Маслак А.А., Осипов С.А., Бянов В.В.* Программно-алгоритмическое обеспечение анализа результатов тестирования на основе модели Раша // XII Международная конференция-выставка «Информационные технологии в образовании». Web page, 2002 (<http://www.bitpro.ru/ITO/2002/VI/VI-0-906.html>).

33. Applications of Latent Trait and Latent Class Models in the Social Sciences. Münster, New York, München, Berlin: Waxmann, 1997.
34. *Greene W.H., Hensher D.A.* A Latent Class Model for Discrete Choice Analysis: Contrast with Mixed Logit. The University of Sydney and Monash University, 2002 ([http://www.its.usyd.edu.au/publications/2002\\_working\\_papers.asp](http://www.its.usyd.edu.au/publications/2002_working_papers.asp)).
35. *Magidson J., Vermunt J.K.* Latent Class Models. Statistical Innovations Inc. Tilburg University, 2002 (<http://www.statisticalinnovations.com/lg/sage11.pdf>).
36. *Гуттман Л.* Основные компоненты шкального анализа // Математические методы в современной буржуазной социологии. М.: Прогресс, 1966.
37. *Хейс Д.* Причинный анализ в статистических исследованиях. М.: Финансы и статистика, 1981.
38. *Garrett E.S., Zeger S.L.* Latent Class Model Diagnosis // *Biometrics*. 2003.
39. *Hagenaars J.A.* Latent Structure Models with Direct Effects between Indicators: Local Dependence Models // *Sociological Methods and Research*. 1988. № 16.
40. *Антон Дж.* Анализ таблиц сопряженности. М.: Финансы и статистика, 1982.
41. *Clogg C.C.* Latent Class Models // *Handbook of Statistical Modeling for the Social and Behavioral Sciences*. Ch. 6. N.Y.: Plenum, 1995.
42. *Uebersax J.* A Practical Guide to Local Dependence in Latent Class Models. Web page, 2000 (<http://ourworld.compuserve.com/homepages/jsuebersax/condep.htm>).