

## МАТЕМАТИКА В СОЦИОЛОГИИ: ФЕНОМЕНОЛОГИЧЕСКИЕ ЗАМЕТКИ

### От автора

Это эссе продолжает тему матриц данных в социологии, начатую в предыдущем номере [1]. Именно через матрицы данных осуществляется контакт эмпирической социологии с математикой. Прежде, чем перейти к специфике этого контакта, остановлюсь на некоторых особенностях матриц данных в социологии.

С.В. ЧЕСНОКОВ

### МАТРИЦА ДАННЫХ СОЦИОЛОГИЧЕСКОГО ОПРОСА

Любой опрос это серия диалогов социолога с респондентами. Диалогов, в которых реплики социолога (вопросы) фиксированы, а реплики респондентов (ответы) могут быть любыми.

Напомню, что привычная запись такой серии широко известна как *матрица данных*, чаще всего в форме таблицы «объект-признак»:

$g$	$x_1$	$x_2$	$\dots$	$x_m$	$u$
$g_1$	$x_{11}$	$x_{21}$	$\dots$	$x_{m1}$	$U$
$g_2$	$x_{12}$	$x_{22}$	$\dots$	$x_{m2}$	$U$
$\vdots$	$\vdots$	$\vdots$	$\dots$	$\vdots$	$\vdots$
$g_n$	$x_{1n}$	$x_{2n}$	$\dots$	$x_{mn}$	$U$

(1)

Символы в клетках таблицы (1) обозначают реплики социолога и респондентов, общее количество реплик равно  $(m+2)(n+1)$ . Символы в клетках верхней строки обозначают вопросы (реплики) социолога. Их ровно  $m+2$ , столько, сколько столбцов. Символы в клетках остальных строк обозначают ответы (реплики) респондентов. Их  $(m+2)n$ . В любой матрице данных есть два особых столбца. В таблице (1) это первый столбец ( $g$ ) и последний ( $u$ ). В столбце  $g$  все ответы  $g_1, g_2, \dots, g_n$  попарно различные, это уникальные имена опрошенных. В столбце  $u$  все ответы  $U$  одинаковы, это имя опроса (имя выборки респондентов).

---

**Чесников Сергей Валерианович** — старший научный сотрудник Института социологии РАН, научный директор компании «Контекст Медиа». Автор детерминационного анализа и детерминационной логики.  
Телефон: +7(499) 132-36-11. Электронная почта: sergeych@context.ru

**Первичные факты.** Матрица данных (1) замечательна тем, что содержит *все элементарные эмпирические факты, установленные в ходе данного опроса*. Все до одного. Дальше я буду называть такие факты *первичными*, в отличие от *вторичных*, которые получены путем преобразования первичных. Каждый первичный факт это ответ определенного респондента на определенный вопрос в условиях конкретного опроса. Первичных фактов столько, сколько ответов респондентов, то есть  $(m + 2)n$ .

Каждый первичный факт описывается высказыванием типа «в опросе  $U$  на вопрос  $x_1$  респондент  $g_2$  дал ответ  $x_{12}$ ». В любом таком высказывании ровно четыре реплики из матрицы данных (1): имя опроса  $U^1$ , вопрос  $x_i$ , уникальное имя  $g_k$  респондента, и ответ  $x_{ik}$  респондента на вопрос  $x_i$ .

**Реплики в матрице данных.** По М.М. Бахтину [3], репликой в диалоге может быть любая *последовательность* слов, образов, действий, *начало* и *конец* которой определяется отражением хода конкретного диалога в сознании любого из участников. Отражение, может быть разным для разных участников. В одном и том же диалоге реплики с точки зрения одного из участников могут отличаться от реплики с точки зрения другого участника (эффект Расёмон). Это принципиально.

Вот вопрос: чью точку зрения на диалоги представляют реплики в социологических матрицах данных? Ответ: точку зрения социолога. Матрицы данных в социологии содержат феномены сознания социолога.

Именно привязка к конкретному диалогу и конкретному сознанию делает реплику (а не слово, как принято думать) первоэлементом языка. Бахтин настаивал на этом [3]. Его точка зрения находит все большее понимание и развитие в современной лингвистике [4].

В социологии диалоги активно используются при создании эмпирического базиса социальных теорий. Занимаясь включенным наблюдением или интерпретацией вербальных текстов, социолог еще может позволить себе забыть, что всякое наблюдение, всякое чтение вербального (тем более — невербального) текста это диалог с наблюдаемым объектом (в частности — с текстом), обмен репликами с ним. Но, занимаясь опросами, социолог непосредственно имеет дело с техникой планирования и проведения опросов как техникой планирования и проведения диалогов.

<sup>1</sup> По имени опроса можно установить характеристики опроса, которые необходимы для оценки его качества (метод отбора респондентов, контекст, в котором происходит диалог с респондентом, способ фиксации реплик в ходе опроса и способ переноса этих реплик в матрицу данных). В профессиональных опросах эти характеристики указаны. Если не указаны, это знак, что опрос, возможно, непрофессиональный или недобросовестный.

Матрица данных (1) результат длинной цепи действий социолога.

Надо поставить задачу, создать вопросник и рабочие документы, спроектировать выборку, отобрать респондентов, опросить их, зафиксировать реплики и перенести их в матрицу данных. Вариативность этих действий чрезвычайно велика. Но результат один: матрица данных, независимо от задачи, вопросника, количества и природы реплик (вербальные – невербальные), содержания реплик (осмысленные – бессмысленные с чьей-либо точки зрения), от вида потерь и искажения реплик при их записи и переносе в базу данных.

Полная совокупность первичных эмпирических фактов, полученных в ходе опроса и подлежащих дальнейшему анализу, всегда имеет вид матрицы данных. В этой связи хочу заметить следующее.

Через сто тысяч лет (если человечество выживет) радикально изменятся национальные языки, грамматика и лексика, понятия, социальные теории, формы социальной организации. Но серии диалогов с фиксированными вопросниками, а с ними и матрицы данных, останутся в принципе такими же, какими их видим мы. Осваивая новые части мира, социального или физического, люди будут иметь дело с матрицами данных точно так же, как они имели дело с ними 1000 и 100000 лет назад.

**Разные представления матриц данных.** Изучая свойства матриц данных как природных объектов полезно иметь в виду, что таблица (1) лишь *один из многих* способов представления матрицы данных.

Есть другие способы. Выпишем  $(m + 2)n$  высказываний вида «В опросе  $U$  респондент  $g_k$  на вопрос  $x_i$  дал ответ  $x_{ik}$ » при всех допустимых  $i, k$ . Получим представление матрицы данных (1) в виде текста. Способ представления изменился, а объект тот же: матрица данных. Еще одно широко известное представление матрицы данных — пачка заполненных анкет. Можно представлять матрицу данных в виде так называемого «веера отображений» [5]. Матрица данных представима также в виде графа — схемы со стрелками, соединяющими реплики.

Таблица (1) удобнее других способов, поэтому используется чаще. Форма таблицы дала жизнь слову «матрица» в словосочетании «матрица данных». Но матрица данных остается матрицей данных независимо от представления.

Любое высказывание, справедливое для матрицы данных в форме (1), справедливо для любого другого представления той же матрицы.

**Что такое анализ данных.** Проведя опрос, социолог не может ограничиться первичными фактами. В матрице данных их  $(m + 2)n$ . При  $m \geq 50, n \geq 1000$  таких фактов очень много — десятки, а то и сотни тысяч. Эти факты поневоле необходимо преобразовать в другие, более обобщенные, более удобные для восприятия и размышлений. Это функция анализа. Здесь и включается в дело математика. Любой мате-

математический метод анализа данных есть метод преобразования первичных фактов в другие, вторичные, факты, которые преподносятся социологом тоже как «эмпирические», «полученные путем опроса».

Отсюда следует наиболее общая точка зрения на анализ данных в социологии. Анализ это конструирование вторичных фактов из первичных, содержащихся непосредственно в матрице данных. Очевидно, что такое конструирование не всегда социально нейтрально. Анализ, как и сам опрос, есть социальное действие.

**Социальный смысл математических методов.** Бывает, что социальный смысл математических операций, применяемых для анализа первичных фактов в матрицах данных, не вызывает сомнений. Например, подсчет числа респондентов, давших одинаковый ответ на определенный вопрос. Социальный смысл операции прост: вместо объекта «респондент» возникает объект «группа респондентов». Операция натурального счета дает объем этой группы, натуральное число, смысл которого ясен и легко поддается социальному контролю.

Но есть другие случаи, когда социальный смысл операций над первичными фактами неясен, и социальному контролю не поддается. Таковы операции, используемые в кластерном анализе, факторном анализе, ряде других популярных методов. Математика действует здесь как «бог из машины», рождая результаты, которые *похожи* на эмпирические факты, но таковыми не являются. Вот почему проблема социальной осмысленности математических операций в социологии относится к важнейшим проблемам методологии получения эмпирических социальных знаний. Это важная часть темы «математика в социологии», я обязательно вернусь к ней позже.

**Матрица данных как математический объект.** Матрицы данных связывают математику и социологию. Пользуясь для анализа данных математическими понятиями, описывающими реплики в матрице данных, социолог уверен, что эти понятия пришли в социологию извне, из естественных наук. Это правдоподобно, но неверно. Почему? Потому что матрица данных не только природный объект, просто-напросто серия диалогов, но также и объект математический. Подробности в следующем эссе.

#### ЛИТЕРАТУРА

1. Чесноков С.В. Матрица данных как природный объект // Социологический журнал. 2008. № 3.
2. Чесноков С.В. Феноменология диалогов в гештальт-теории, математике, логике. М.: URSS, 2009.
3. Бахтин М.М. Проблема речевых жанров // Бахтин М.М. Собр. соч. Т. 5. М.: Русские словари, 1996.
4. Крейдлин Г.С. Невербальная семиотика (монография). М.: Новое литературное обозрение, 2002.
5. Чесноков С.В. Детерминационный анализ социально-экономических данных. 2-е изд. М.: УРСС, 2009.