

138
делает динамику численности населения. Его влияние оценивается регрессионным анализом действия независимых факторов на по-казатель численности населения. Получаемые в результате анализа коэффициенты регрессии формируются под влиянием кластеров хо-лодных показателей, формирующихся в свою очередь, соответствую-щий им фактор. В силу ограниченной возможности предлагается приби-нуть прямой метод «расщепления» базового коэффициента регрессии на женный способ «расщепления» отражающие их влияние на численность населения. Все исследования проводятся путем сравнения полученных результатов по 2002 и 2005 годам. Для оценки динамики изменения уровня жизни используется **показатель-индекс УРЖ**, вычисляемый в процентах от достигнутого уровня.

Литература

1. Материалы круглого стола «Качество жизни россиян в современных условиях» // Вестн. Моск. ун-та. Сер. 18. Социология и политология, 2009.
2. Статистика. Учебник / Под ред. И.И. Елисеевой. — М.: КНОРИС, 2006.
3. Шенюк К. Работы по теории информации и кибернетике. — М. ИЛ, 1963.
4. Кендалл М., Стюарт А. Теория распределений. — М.: «Наука», 1965.
5. Айвазян С.А. и др. Прикладная статистика. — М.: «Финансы и статисти-ка», 1989.
6. Вушков И. и др. Прикладной линейный регрессионный анализ. — М.: «Финансы и статистика», 1987.

ВЕРОЯТНОСТНЫЕ И НЕВЕРОЯТНОСТНЫЕ МОДЕЛИ ПОРОЖДЕНИЯ ДАННЫХ В СОЦИОЛОГИИ

В настоящей статье внимание социологов обращается на важность выявления того, какую модель порождения данных — вероятность выбора или невероятностную — он выбирает, показывается, что только при правильном определении модели возможен корректный выбор методов анализа данных. Констатируется то, что, хотя речь идет, по большому счету, об известных вещах, но на практике исследователи, как правило, не обращают на них внимания, поскольку (а) указанное обстоятельство мало известно социологам; (б) при описании в литературе вероятностных свойств исходных данных отсутствует некото-рая терминология; более того, отсутствуют четкие определения некото-рых терминов; (в) нигде не анализируется то, какая реальная, «содер-жательная» ситуация лежит в основе явления или неадекватной вероят-ностную модель порождения данных. Цель статьи — в определенной мере ликвидировать пробелы, отраженные в перечисленных пунктах. Ниже для краткости иногда будем говорить просто о вероятностной или невероятностной модели.

Значение выбора социологом модели порождения данных

Известно, что проблема интерпретации получаемых социологом данных при решении практически каждой конкретной задачи. Выбор (невы-ческой задачи стоит интересуют нас интересует лишь один момент: принять или не-В данной статье вероятностного порождения данных. Выбор (невы-бор) указанной модели много детерминирует в определении и спосо-ба сбора данных, и методов их анализа. Укажем только два примера.
Первый пример состоит в том, что принятие модели самое идею ис-пользования выборочной совокупности для изучения интересующих,1 пользования выборочной совокупности для изучения интересующих,1 пользования выборочной совокупности для изучения интересующих,1

¹ Сама идея выборки, вошедшая в науку в конце XIX — начале XX века (и в неявном виде присутствовавшая уже в законе больших пред-д. Вернулли, опубликованном в 1713 году), по существу и охватывает, из которой положение о существовании некой генеральной совокупности, из которой

дает исследователю возможность использования главного достижения математической статистики — переноса результатов статистического анализа на генеральную совокупность, т.е. осуществления статистического оценивания параметров и проверки статистических гипотез. Без такого переноса смысл выборки теряется. Ясно, что это обстоятельство касается и сбора, и анализа данных. Так, скажем, мы можем, желая сделать вероятностную модель адекватной, пытаться строить случайную выборку из некоторой совокупности, которую априори считаем генеральной. Это сбор данных. А что касается анализа, то, как известно, упомянутые положения математической статистики заложены в регрессионных методах, активно используемых в социологии: в регрессионном, факторном, дисперсионном анализе и т.д.

Второй пример состоит в следующем. Как известно, в любом исследовании ищутся причины тех или иных явлений. Понятие причинности — вещь «смутная». Много всего между учеными происходит споры о том, как это надо трактовать. И, в зависимости от этой трактовки, предлагаются абсолютно разные способы выявления причин причинности. Основные линии практической трактовки причинности к выбору вероятностной или вероятностной статистической модели часто называют, соответственно, статистической и нестатистической (их часто называют, опираясь на первую трактовку, обычно ской). Линия с Лапласом, на вторую — с Юмом и Миллем. (Коротко об этом говорится в работе [1]).²

Итак, выбор (невыбор) вероятностной модели порождения данных в значительной мере определяет результаты исследования. Учитывает ли это социолог в процессе работы с эмпирическими данными? Да и всегда ли он об этом знает? Рассмотрим эти вопросы.

с определенными вероятностями «черпаются» объекты с определенными свойствами. Другими словами, считалось, что исследователю, используя понятие выборки, тем самым делает предположение о вероятностном порождении данных. Однако сейчас эти два понятия «разошлись», скажем, иногда говорят о правилах построения выборки в т.н. качественных исследованиях, тогда не идет речи о генеральной совокупности.

² Естественно, ни один серьезный исследователь, даже считающий статистический подход адекватным его задаче, никогда не отождествляет наличие причинных связей между переменными с существованием между ними причинных отношений.

Отношение широких кругов социологов к проблеме выбора модели порождения исходных данных

В настоящее время пришло время для анализа социологических данных пакета программ содержит алгоритмы, как описанные на вероятностные модели, так и не опирающиеся. Социолог обычно нажимает кнопки компьютера, не задумываясь об адекватности используемых моделей решаемой содержательной задаче.³ Да и узнать о том, что такое вероятностный или вероятностный характер предположений брошено по широкому кругу публикаций, отнюдь не ориентированных на социолога. Точного определения вероятностной модели нет. Терминология не унифицирована (так, упомянутая модель называют еще и статистической). И, самое главное, нигде не обсуждается вопрос о том, когда, в какой реальной ситуации соответствующее предположение приемлемо, в какой — нет. В результате работает т.н. «кнопочная» психология: компьютер все решит за исследователя, знай, нажимай кнопки (как говорил известный классик, следовало бы знать не обязательно, извозчики доведут).

То же подтверждается состоянием учебных программ по анализу данных, т.е. наиболее концентрированного отражения представлений научного сообщества о том, что должен знать социолог о соответствующих вопросах. Все усилия преподавателей по анализу данных, как правило, направлены на пропаланду вероятностной точки зрения: студенты обучаются математической статистике, осваивают способы оценки параметров распределений одномерных и многомерных случайных величин (напомним, что очень многие методы анализа данных, начиная с методов регрессионного и факторного анализа, по существу являются методами оценки параметрической модели обычно не заправляется. Чаще всего вообще не говорится о том, что пользуются т.д. При этом сама суть фактически задействованной модели обычно не затрагивается. Чаще всего вообще не говорится о том, что пользуются (или неадекватной) вероятностную модель исходных данных (хотя в литературе, прямо не ориентированной на социолога, об этом говорится (см., например, [2]). Тем более, не встает вопрос об адекватности этой модели конкретной социологической ситуации (обсуждение соответствующих вопросов, как мы уже упоминали, практически отсутствует в литературе).

³ К сожалению, это касается не только моделей, рассматриваемых в настоящей статье. Социолог, как правило, вообще не задумывается о моделях, заложенных в используемых им методах.

ски наблюдаемой конкретике. И здесь важно отметить, что частота, служащие оценками генеральных вероятностей, сами-то колмогоровской аксиоматике, вообще говоря, НЕ удовлетворяют! И мы должны приложить усилия к тому, чтобы доказать, что наша содержательная модель (носителем которой служат конкретные значения наших признаков, т.е., типотетически, — выборочных реализации случайных величин, а часто — совокупность частот) действительно отвечает модели первого этапа. Собственно, с подобной ситуацией мы сталкиваемся всегда, когда пользуемся каким бы то ни было формальным методом для анализа реальных данных: в любом методе заложена та или иная модель, и, применяя его, мы предполагаем эту модель адекватной нашей содержательной ситуации. Вопрос в том, чтобы определить, в чем эта модель состоит.

Представляется, что на втором этапе мы фактически переходим к тому содержательному представлению о вероятности, которое господствовало в науке до 1933 года, мешая своей нестрогостью восприятию теории вероятностей как части математики, обуславливая отнесение её к естественным наукам. Появляется интересная связь между дедуктивным построением теории вероятностей и индуктивным историческим генезисом её основных понятий. Будем считать, что именно бытовавшая в науке в течение веков подуодержательная трактовка вероятности и может послужить основой для предложения адекватности модели, отвечающей нашему первому этапу (возможно, существует и другой путь трактовки, интерпретации исходных данных, приводящий к тому же, нам это неизвестно). Тогда разговор о «комплекс» условий будет весьма кстати.⁹

Итак, будем полагать, что для того, чтобы полученные на втором этапе относительные частоты действительно можно было использовать в качестве оценок вероятностей из нашей триады (полученной на первом этапе), и применять к ним известные положения теории вероятностей и математической статистики, нам надо быть уверенными в том, что выполнены предположения, носящие не математический, а содержательный характер: наблюдаемые частоты действительно отчаются как-то интуитивно понимаемым вероятностям (что невозможно без предположения о том, что элементарные события совершаются

отся при реализации одного и того же комплекса условий) понятие генеральной совокупности осмыслено, из неё возможно «черпать» бесконечное количество случайных выборок и наша выборка — одна из них (напомним, что на базе предположений о возможности такой «черпаний» доказываются известные положения математической статистики). По существу об этом и должно говорить принятое вероятностной модели порождения исходных данных. А всегда ли такие предположения оправданы? Рассмотрим, какие практические проблемы возникают в связи с желанием ответить на этот вопрос. При этом будем давать себе отчет в том, что, решая встающие здесь серьезные проблемы, мы решаем вопрос об адекватности интересующей нас модели.

Содержательные условия, обеспечивающие адекватность модели вероятностного порождения данных

Резюмируя сказанное выше, будем полагать, что в основе интересующей нас модели вероятностного порождения данных лежат следующие утверждения, выполнение (невыполнение) которых опирается на содержательную трактовку исследователем эмпирических данных.

1. Исходные данные суть случайная выборка (одна из бесконечного числа мыслимых выборок заданного объема) из генеральной совокупности, отождествляемой со случайной величиной, вероятностным распределением, примером вероятностного пространства (элементарными событиями в котором служат события, состоящие в том, метрическая величина принимает те или иные значения).

2. На выборке случайным величинам (отдельным составляющим многомерной случайной величины) отвечают некоторые наблюдаемые (измеряемые) признаки; вероятность признаков (трактуется как реализовались тот или иной набор значений признаков) пося как предел частоты его реализации (имеется в виду сходимость вероятности) при неограниченном числе повторений одного и того же комплекса условий¹⁰ (заметьте, что такое повторение здесь заведомо считается возможным; понятие «комплекс условий» — ясно).

⁹ Если выражаться более точно, то следует сказать, что построение комплекса условий требуется при многократных реализациях испытаний, в результате которых появляется то событие, вероятность которого мы опираемся. В качестве испытаний для социолога чаще всего выступают опрос одного респондента, многократная реализация испытания — опрос разных респондентов.

¹⁰ Так, прибегая к самому распространённому в учебниках по теории вероятностей примеру, можно сказать, что, изменив положение центра тяжести брошенной монеты, мы перейдем к другому комплексу условий реализации наших событий — выпадений орла или решки (точнее, следовало бы говорить не о реализации событий, а о проведении опыта, испытании, эксперимента, коим здесь является акта бросания монеты).

Эта модель действительно зачастую далека от того, с чем имеет дело социолог. Отказ от нее (модели) может быть обусловлен разными причинами: (1) у исследователя могут быть сомнения в справедливости каких-либо из сформулированных положений; (2) исследователь может принципиально отрицать адекватность модели вероятностного порождения данных для решаемой им задачи. Рассмотрим оба аспекта.

Причины, заставляющие исследователя сомневаться в справедливости содержательных аспектов адекватности модели вероятностного порождения данных

Перечисленные условия нередко нарушаются в социологическом исследовании. Приведем примеры нарушений.

1. В рассматриваемой генеральной совокупности нет некоторого единого распределения рассматриваемой случайной величины, а, к примеру, имеется т.н. смесь нескольких распределений. Тогда все наши оценки получившейся доли встречаемости рассматриваемого события становятся бессмысленными. А мы ведь заранее, как правило, не можем знать, имеем ли мы дело с подобной ситуацией.

Скажем, мы можем измерять зарплату респондентов, податая, что в генеральной совокупности ей отвечает какое-то единое (наприм. нормальное) распределение. Пользуемся наблюдаемыми значениями, строим доверительные интервалы для математического ожидания и других параметров этого распределения. А в действительности у нас имеется два четко выраженных нормальных распределения: для одного математическое ожидание равно 200 000 рублей, а для другого — 20 000. Какова цена наших формальных математико-статистических построений?

Или другой пример. Социолог анализирует данные, собранные кем-то другим, с неизвестными целями, генеральной совокупностью (при том, конечно, неизвестно, с какой случайной величиной эта совокупность отождествляется) и способом построения выборки (может быть, она не случайная, а, скажем, квотная). Можно ли тут говорить об адекватной модели вероятностного порождения данных решаемой задаче? Отметим, что именно желание анализировать подобные массивы данных привело к рождению Data mining.

Можно привести еще один распространянный пример. Исследование состоит в том, что социолог изучает какую-то конкретную группу респондентов, не зная заранее, какова генеральная совокупность. Он «прошупывает» ситуацию, пытается «подолить» гипотезы,

вызывать проблему. Такого рода работа может помочь исследователю сформировать представление о генеральной совокупности. Но возможность использования результатов математической статистики (что мы по существу отождествляем с адекватностью вероятностной модели порождения данных) возникнет только в результате перехода именно к этой неизвестной генеральной совокупности, осуществление случайного отбора из нее, что может быть неосуществимо практически.

2. Понятие комплекса условий, соблюдение которого лежит в основе самого определения вероятности, часто бывает невозможно четко обрисовать. Может быть неясно, изменилась ли доля реализации или рассматриваемого события в серии экспериментов (повторений) или рассматриваемого комплекса условий) за счет перехода к другому рассматриваемому комплексу условий) или же за счет перехода к другому действию закона больших чисел), или же за счет перехода к другому комплексу условий. Другими словами, неясно, имеет ли место вариабельность частоты в рамках выборки из одной и той же генеральной совокупности, или же неявно осуществляется переход к другой.

Принципиальное отрицание исследователем адекватности модели вероятностного порождения данных для решаемой им задачи

1. Некоторые методы анализа данных заведомо не опираются на модель вероятностного порождения данных. К таким методам относят, например, многие методы классификации, многомерного шкалирования, теории измерений.

2. Наиболее яркий пример именно для социолога — т.н. качественный подход к сбору и анализу данных. В отечественной социологической литературе много говорится о неадекватности вероятностных моделей в ситуации, когда речь идет о сборе данных с помощью неформализованных методов (например, свободного интервью). Но качественный метод может быть и подход к анализу данных анкетных опросов [9]. И в таком случае вероятностная модель порождения данных тоже отрицается исследователями.

3. Отметим еще одну ситуацию, когда исследователь просто не думает об адекватности модели вероятностного порождения данных, зная о том, что правила переноса результатов с выборки на генеральную совокупность для используемого им метода не разработаны (при этом модель вероятностного порождения данных может выглядеть вполне приемлемой). Скажем, в случае реализации какого-либо алгоритма классификации объектов на выборке трудно сформулировать

«вооружение» статистический подход только в середине XIX века, когда под воздействием развития теории газов родилась статистическая физика.

Русские ученые (тог же А.И. Чупров и, в большей мере, — А.А. Чупров) уже в конце XIX — начале XX века активно анализировали гносеологическую роль статистического подхода, отвечая, что это — общенаучный подход, и доказывая своим читателям, обществоведам, что в естественных науках статистический подход может эффективно использоваться не менее эффективно, чем в обществознании). Специфика этого подхода основана не на том, что он применяется в определенном классе наук. По мнению русских ученых, типологический подход должен применяться там, где причины, обуславливающие изучаемое явление, постоянны, а статистический — когда эти причины постоянно меняются и если «в одном, например, случае изменяющиеся причины отклоняют явление от постоянного причин, в сторону плюс, а в другом в сторону минус, а при массовом наблюдении явление представляется исследователю в том виде, какой оно имело бы, будучи подвергнутым действию одних постоянных причин» [11, с. 16].

В наше время становится ясно что «постоянных» причин не бывает. Поэтому статистический подход очень популярен. Однако, как было сказано в предыдущем параграфе, он не универсален.¹³

И еще один исторический момент, не потерявший своего значения и в наше время, хотелось бы отметить. С XVII века социологи пользовались статистическим подходом, но до конца XIX века не считали, что он связан с использованием теории вероятностей. К использованию этой теории при изучении общества призывали Кондорсе, Кетле. Но об отношении статистиков к теории вероятностей яркое свидетельство считает нужным доказывать необходимость опоры статистики на теорию вероятностей. Его дипломная работа (он начал математическое отделение физико-математического факультета МГУ) называлась «Теория вероятностей как основа теоретической статистики» (1896). И в своих дальнейших публикациях он был активным сторонником внедрения идей рождающейся математической статистики в статистические исследования русских ученых. При этом он

¹³ Теория вероятностей, будучи до 1933 года неформализованной наукой, базировавшаяся на весьма неоднозначно трактуемом, но важном (в разных смыслах) понятии вероятности, нередко оказывалась в центре философских споров. Например, иногда её называли наукой о беспримечательных явлениях.

активно опирался на творчество петербургских ученых А.А. Маркова и П.Л. Чебышева и, в частности, на теорию средних. Другими словами, теория вероятностей и проблема изучения закономерностей «в среднем» в его творчестве не отделялись друг от друга.

Литература

1. Давыдова Ю.Н. Математико-статистические методы в социологии. М.: ГУ-ВШЭ, 2010. Глава 12.
2. Давыдова Ю.Н., Мухоморов В.С. Теория вероятностей и прикладная статистика. М.: ЮНИТИ, 2001.
3. Чупров А.А. Вопросы статистики. М.: Госстатиздат ЦСУ СССР, 1960.
4. Колмогоров А.Н. Основные понятия теории вероятностей. Изд. 2-е. М.-Л., 1974.
5. Крамер Г. Полюска с теорией вероятности // Эта статья Крамера имеется в интернете. Однако мы не можем дать ее адрес. См. также: Крамер Х. Полюска с теорией вероятностей: наброски воспоминаний // Современные проблемы математики. М.: Знание, 1979.
6. Давыдова Ю.Н. Сущность математики в преодолении к потребностям социологии: уроки истории // Математическое моделирование социальных процессов. Вып. 10. М.: КДУ, 2009. С. 376–423.
7. Орлов А.И. Математика случая. Вероятность и статистика — основные факты. Учебное пособие. М.: МЗ-Пресс, 2004.
8. Давыдова Ю.Н. Обеспечение однородности исходных данных в процессе применения математических методов // Социс, 1986, №3. М.: Наука. С. 149–154.
9. Фини В.К. Интеллектуальные системы и общество. М.: УРСС, 2006 (например, С. 267).
10. Пидуха М. Очерки по истории статистики XVII-XVIII веков. М.: ОИИЗ. Государственное издательство политической литературы, 1945.
11. Чупров А.И. Статистика. Лекции А.И. Чупрова. С.-Пб политехнический институт, 1907.