

Ротмистров Алексей Николаевич

НИУ ВШЭ

Доцент

Кандидат социологических наук

Rotmistrov Alexei Nikolaevich

NRU HSE

Docent

E-Mail: alexey.n.rotmistrov@gmail.com

Шулус Алексей Апполинариевич

ФГБОУ ВПО ГУУ

Доктор экономических наук, профессор

Shulus Alexei Appolinariievich

The state university of management

Professor

E-Mail: shulus@bk.ru

Проблема взаимодействия предикторов в регрессионном моделировании на примере исследования посещения московскими семьями развлекательных центров

A problem of interaction of predictors in regression models illustrated by a study of
Moscow families' visits to entertainment centers

Аннотация: В статье кратко описана суть регрессионного моделирования, рассмотрен феномен взаимодействия и его проявления в регрессионном моделировании, сформулирована проблема, являющаяся следствием этих проявлений, она проиллюстрирована на конкретном примере и для её решения предложены рекомендации.

Abstract: The article briefly describes essence of the regression model, focuses on phenomenon of interaction and its appearance in regression models, formulates the problem which is a consequence of these manifestations; it is illustrated by a specific example, and its solutions are proposed recommendations.

Ключевые слова: Регрессия, моделирование, предиктор, взаимодействие, качество модели, статистическая значимость.

Keywords: Regression modeling, predictor, interaction, quality of model, statistical significance.

Социология – сравнительно молодая и довольно неточная наука. Она в меру своих сил старается выполнять три главные общенаучные функции: описание, объяснение и предсказание (прогнозирование). И если с первыми двумя из них она более или менее справляется, то с последней – обычно нет. На ум сразу приходит свежий пример неверного прогноза явки москвичей на выборах мэра. Отклонение эмпирических значений от прогнозируемых было столь удручающим, что глава Фонда «Общественное мнение» Александр Ослон выступил с заявлением: «Да, я это признаю – фиаско социологов произошло. Мы не умеем прогнозировать явку <...> На будущее я откажусь от прогнозирования выборов. Наше дело проводить опросы» [5].

Сложность прогнозирования в социальных науках имеет объективные причины. «Сложность соответствующих явлений влечет сложность формализации наших представлений о них. Модели реальности, которые мы фактически строим, используя тот или иной метод анализа данных, оказываются чересчур приблизительными, соответствующие прогнозы не сбываются и т.д. Эти модели настолько субъективны, что исследователь все время рискует получить результаты, плохо отражающие реальность». [3, гл.1.3]. Примером исключительно сложного не то чтобы для прогнозирования – для описания и объяснения явления служит протестное поведение. Регистрация протестного поведения и составление перечня гипотетических его детерминант – большой труд [2].

Однако и в словах А. Ослона есть правда: социологи неумело пользуются математическим аппаратом обработки собранных данных. Другой вопрос: стоит ли научиться пользоваться этим аппаратом более умело или заниматься только опросами? Мы предпочитаем первый вариант и постараемся внести своей статьёй свой вклад в его реализацию.

1. Краткое описание регрессионного моделирования

Регрессионное моделирование – активно развивающийся класс методов. Они находятся на стыке анализа данных и моделирования явлений. Корень регрессионного моделирования – уравнение регрессии. В классическом виде оно выглядит так (1):

$$Y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (1)$$

где a_0 – константа, x_i – независимые переменные, или предикторы, a_i – коэффициенты при них, а Y – зависимая переменная.

В анализе социологических данных предикторы и зависимая переменная – некоторые признаки респондентов или иных изучаемых эмпирических объектов, а связь между правой и левой сторонами уравнения обычно не динамическая (функциональная), а статистическая; поэтому справа обычно стоит не просто зависимая переменная, а её выборочная оценка математического ожидания (т.е. среднее арифметическое значение). Например, зная возраст автовладельца, его доход и цену его нынешнего автомобиля, можно попробовать оценить примерную цену нового автомобиля (в случае наличия у автовладельца желания сменить автомобиль).

Регрессионная модель считается качественной, во-первых, если предикторы «объясняют» большую долю вариации зависимой переменной. Другими словами, насколько знание возраста автовладельца, его дохода и цены его нынешнего автомобиля позволяет повысить точность прогноза цены нового автомобиля по сравнению с ситуацией незнания этих параметров, настолько регрессионная модель качественна. Оценке качества модели служит параметр, называемый R^2 или его аналоги. Во-вторых, качественная регрессионная модель такая – в которой все предикторы статистически значимы. Это означает, что коэффициенты a_i не равны нулю ни в выборочной совокупности, ни в генеральной совокупности. Для проверки статистической значимости существует специальный математико-статистический инструментарий.

Здесь мы не будем углубляться в детали регрессионного моделирования: каков математико-статистический инструментарий проверки статистической значимости, каковы показатели объясняющей способности модели, каковы требования, предъявляемые ею к данным. Эти детали хорошо изложены в [8, ch.5].

Значимость регрессионного моделирования для социологов становятся совершенно очевидной, когда речь идёт о такой его ветви, как логистическое регрессионное моделирование. Дело в том, что слева в этой разновидности регрессионного моделирования

находится не метрическая или хотя бы интервальная переменная, не слишком характерная для социологии, а вероятность интересующего исследователя события (например, того, что респондент купит ту или иную модель автомобиля). Если быть точным, то слева стоит отношение вероятности того, что интересующее событие произойдёт к вероятности, что оно не произойдёт; причём от этого отношения с целью нормировки взят логарифм. Таким образом, в логистическом виде уравнение регрессии (1) превращается в:

$$\log(p/q) = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (2)$$

где p – вероятность того, что интересующее событие произойдёт, q – вероятность того, что интересующее событие не произойдёт, \log – логарифм по любому основанию, обычно в качестве основания выступает экспонента.

Логистическое регрессионное моделирование по большинству оснований сходно с классическим. Его особенности изложены, напр., в [7, ch.9.3]. В контексте настоящей статьи эти особенности не принципиальны.

Казалось бы, методам регрессионного моделирования не один десяток лет, он нашёл своё воплощение в различных статистических пакетах – какие здесь могут быть проблемы? По крайней мере, методологические? Мы представим пример такой проблемы, после чего дадим её строгую формулировку и теоретическое обобщение.

2. Иллюстрация проблемы

Недавно в Москве было проведено исследование на тему: «Потребление московскими семьями с детьми услуг развивающих и развлекательных детских центров». Цель: составить детальные описания того, как проводят своё время малолетние дети из московских семей, классифицировать семьи по этому признаку, выяснить какие характеристики семьи в целом и родителей в отдельности предопределяют попадание семьи в тот или иной класс. Соответственно цель распадается на 3 блока задач, вторая и третья из которых «напрашиваются» на применение логистической регрессии. Каждый из блоков задач включает разные опции времяпрепровождения детей: начиная с центров дошкольного образования через различные секции и кружки к развлекательным центрам. Генеральная совокупность – московские семьи (в пределах «старой» Москвы). Эмпирическая база – представители московских семей (мамы, папы, бабушки, дедушки). Выборочная совокупность – 500 респондентов (как раз средняя выборка).

Довольно глубокое исследование – на самом деле, несколько исследований в одном. В массиве данных около 700 переменных (до создания фиктивных) с очень неравномерным распределением пропусков.

В качестве иллюстрации к нашей статье мы взяли только один блок задач – посвящённый развлекательным центрам; и только 328 респондентов – имеющих мало пропусков по переменным, относящимся к интересующему нас блоку. Эти переменные таковы. Зависимая: «Часто ли Вы с ребенком (детьми) посещаете развлекательные детские мероприятия»: «Редко (раз в месяц и даже реже)» (0) и «Часто (чаще раза в месяц)» (1). Потенциальными предикторами выступают 6 категориальных переменных: «Ваше семейное положение», «Материальное положение Вашей семьи», «Ваше образование», «Ваша должность», «В каком административном округе Москвы Вы проживаете?», «Насколько далеко от центра проживаете?»¹.

¹ Категориальные переменные как таковые нельзя использовать в регрессионном моделировании в качестве предикторов. Чтобы обойти этот запрет, диктуемый математической статистикой, в анализе данных разработана процедура дихотомизации: каждая номинальная или порядковая переменная «рассыпается» на

Попробуем построить модель логистической регрессии общепринятым в социологии способом². Получаем:

- оценка качества модели (аналог R^2) равна 62%. Очень хороший результат, особенно для социологических исследований;
- среди всех значений (их 27) всех предикторов на 5-процентном уровне значимости значимо только значение «Менеджер» предиктора «Должность респондента». Коэффициент при этом значении равен 0,8.

Можем ли мы считать модель состоявшейся и приступить к её интерпретации? Большинство пользователей метода отвечают «да». Действительно практика «выдёргивания» значимой переменной из окружения статистически не значимых предикторов общепринята в социологических исследованиях. Чтобы не быть голословными, приведём фрагмент другого исследования компетентного и авторитетного российского учёного, подготовленного для международного круга читателей.

В новейшем и масштабном исследовании [9] автор отбирает для интерпретации статистически значимые предикторы (например, доход) [ibid, p.24]. Но если судить по приложениям к статье, эти предикторы находятся в окружении ряда статистически не значимых предикторов (например, «беспокойство по поводу дохода») [ibid, p.18]. Очень вероятно, что после применения более тонкого алгоритма отбора предикторов для итоговой модели выводы исследования претерпели бы некоторые изменения. Но это тема отдельной – скорее содержательной, а не методологической статьи.

Мы же вернёмся к нашему примеру и посмотрим, как «поведёт себя» найденный нами статистически значимый предиктор вне окружения статистически не значимых предикторов. Т.е. банально построим однофакторную регрессионную модель, предиктором³ которой выступит значение «Менеджер» предиктора «Должность респондента». Получаем:

- оценка качества модели (аналог R^2) равна 0,2%;
- предиктор на 5-процентном уровне значимости не значим. Коэффициент при этом значении равен -0,27.

10 минут назад мы получили хорошую модель со статистически значимым предиктором, который имел прямое влияние на вероятность того, что респондент часто водит своего ребёнка в развлекательные центры. Теперь та же модель никуда не годится – ни по оценке качества модели, ни по статистической значимости предиктора. Более того, теперь влияние предиктора на вероятность того, что респондент часто водит своего ребёнка в развлекательные центры, обратное. Что произошло с моделью за 10 минут? Мы исключили из неё все предикторы, кроме одного. Оказалось, что это был набор из 27 взаимодействующих

новые дихотомические переменные. Каждая новая дихотомическая переменная соответствует одной категории исходной категориальной переменной и кодируется «да/нет» (1/0). Новые дихотомические переменные выступают в роли предикторов регрессионного уравнения. Содержательно это оправданно в большинстве случаев номиналистичностью социологических шкал, о чём говорили выше. Оправданно это и в части применения МНК, поскольку он как раз предполагает расчёт средних арифметических значений предикторов и зависимой переменной. Чтобы полностью эмпирически оправдать применение техники дихотомизации категориальных предикторов в бинарной логистической регрессии, ниже на примере покажем, что интерпретация её результатов не меняется при изменении кодировки дихотомических шкал в рамках допустимых для данной шкалы преобразований (см. [4]).

² Для построения модели бинарной логистической регрессии мы пользуемся статпакетом SPSS и будем ссылаться на его алгоритмы и результаты там, где без этого не обойтись. Наш выбор имеет ту причину, что реализовать алгоритмы регрессионного моделирования вручную крайне затруднительно и большинство социологов прибегают для этого к статистическим пакетам – причём большинство к SPSS.

³ В данном контексте предиктор и фактор – синонимы.

предикторов. Никакие параметры регрессии не фиксировали это взаимодействие. Взаимодействие стало заметным только после исключения 26 предикторов из набора. Т.е. взаимодействие обнаружило себя только после того, как было разрушено. Проверяют ли социологи свои регрессионные модели на наличие взаимодействия? Обычно нет. Это практическая проблема. Располагают ли социологи хорошо разработанными инструментами для проверки своих регрессионных моделей на наличие взаимодействия? Нет. Есть только разрозненные приёмы и рекомендации. Это методологическая проблема. Мы собрали ряд рекомендаций из самых свежих источников и обобщили их. Но прежде чем изложить их, разберёмся в сути проблемы через рассмотрение феномена взаимодействия.

3. Описание феномена взаимодействия и теоретическое обобщение проблемы

Начнём с предостережений, которые даёт в своём учебнике К. Доугерти исследователям, исповедующим механический подход к регрессионному моделированию: «Свойства оценок коэффициентов регрессии в значительной мере зависят от правильности спецификации модели. Результаты неправильной спецификации переменных в уравнении могут быть в обобщенном виде выражены следующим образом. 1. Если опущена переменная, которая должна быть включена, то оценки коэффициентов регрессии, вообще говоря, хотя и не всегда, оказываются смещенными. Стандартные ошибки коэффициентов и соответствующие тесты в целом становятся некорректными. 2. Если включена переменная, которая не должна присутствовать в уравнении, то оценки коэффициентов регрессии будут несмещенными, однако, вообще говоря (хотя и не всегда), – неэффективными. Стандартные ошибки будут в целом корректны, но из-за неэффективности регрессионных оценок они будут излишне большими» [1, с.166]. Далее автор подробно, с примерами и упражнениями раскрывает эти два пункта. Несомненно, его предупреждения актуальны для исследователей не только экономических явлений. Но проблема смущённости и неэффективности регрессионной модели проистекает не только из неучтённой корреляции включённых и не включённых в модель предикторов – на чём акцентирует своё внимание К. Доугерти. Эта проблема шире. Оба описанных К. Доугерти изменения набора предикторов даже при отсутствии внутри набора какой-либо корреляции могут повлиять на поведение всех предикторов этого набора. Может незначимый коэффициент стать значимым? Да. Может, наоборот, значимый коэффициент стать незначимым? Да. Может величина коэффициента измениться? В разы. Может коэффициент поменять свой знак? Запросто. Это его величество взаимодействие. К. Доугерти и многие другие авторы не рассматривают проблему формирования набора предикторов в контексте этого явления.

Воспользуемся развёрнутым определением из [3, гл.2.2.1]: взаимодействие – феномен, проявляющий себя в любом методе изучения более чем 2-мерной связи как статистической, так и не статистической природы. Взаимодействие предполагает, что одна из переменных является (или назначается) зависимой, остальные – независимыми. И по сочетаниям значений независимых переменных прогнозируется (с более или менее высокой точностью) то или иное значение зависимой переменной. «Из того, что 3-й признак принимает 2-е значение одновременно с тем, что 4-й принимает 5-е значение, как правило, следует, что 6-й признак принимает либо 2-е, либо 3-е». Взаимодействие может «плавать» по значениям изучаемых независимых переменных, т.е. для прогноза каждого из значений зависимой переменной в сочетании не обязательно должны вступать значения всех изучаемых независимых переменных; для прогноза какого-то из значений зависимой переменной сочетаний может не оказаться вовсе. Взаимодействием может быть не только конъюнкция значений независимых переменных, а любые логические функции от них. Наконец, результатом взаимодействия может выступать не только значения зависимой переменной, но и любая логическая функцию значений независимых переменных.

Из этого определения следует, что для учёта в регрессионной модели взаимодействия некоторых предикторов необходимо сочетать их значения посредством некоторой функции и включить эти сочетания в уравнение. Обычно гипотетически взаимодействующие предикторы попарно перемножают и это перемножение включают в уравнение. Однако «уловить» взаимодействие перемножением возможно не всегда.

D.W. Hosmer и S. Lemeshow, внёсшие заметный вклад в развитие логистического регрессионного моделирования, заметили что взаимодействие «ловится» по-разному и предложили различать два типа взаимодействия, а также – в рамках логистической модели – формальный аппарат для их различения. «Взаимодействие принимает самые разные формы, поэтому начнём с описания ситуации, в которой взаимодействия нет. Представим ситуацию: у нас один дихотомический признак и один континуальный признак. <...> Если связь между континуальным признаком и результирующей переменной одинакова для каждого уровня дихотомического признака, то взаимодействия между дихотомическим континуальным признаками нет» [6, р.63]. Hosmer и Lemeshow различают 2 типа взаимодействия (interaction): «вмешательство» (confounding) и «модификация эффекта» (effect modifying). Формальным основанием для констатации наличия взаимодействия (любого из двух типов) является изменение регрессионного коэффициента одного предиктора при включении в модель другого предиктора (что мы и наблюдали на нашем примере). Формальным основанием для различения confounder и effect modifier является статистически не значимое или значимое (соответственно) повышение качества модели при включении в неё члена более высокого порядка (например, перемножения)⁴. [Ibid, р. 65-66].

Поскольку обозначенные два типа взаимодействия проявляются отнюдь не только в регрессионном моделировании, предлагаем альтернативную терминологию: взаимодействие первого типа, effect modifier – явное взаимодействие, взаимодействие второго типа, confounder – латентное взаимодействие.

4. Рекомендации по решению проблемы

Итак, в новейших западных работах содержатся предложения по обнаружению взаимодействия и его двух типов, по крайней мере для регрессионного моделирования. Если перевести их на язык рекомендаций, то они будут выглядеть так:

Шаг 1. Проверьте, меняются ли значения и/или знаки коэффициентов при всех предикторах, составляющих набор, при производимом Вами изменении этого набора (исключении из него каких-то предикторов или включении в него новых предикторов).

Шаг 2.А. Если значения и/или знаки коэффициентов при каких-либо предикторах не менялись, взаимодействия нет. Шаг 2.Б. Если же поменялись, выясните тип взаимодействия и переходите к шагу 3.

Шаг 3. Перемножьте предикторы, чьи коэффициенты поменяли значения и/или знаки. Если качество модели после включения в неё нового члена (представляющего собой описанное перемножение) статистически значимо выросло, то новый член должен остаться в уравнении и Вы обнаружили явный тип взаимодействия. Если же качество модели статистически значимо не выросло, то новый член должен быть удалён из уравнения и Вы обнаружили латентный тип взаимодействия.

Помимо содержательной стороны рекомендаций есть и техническая: когда у Вас 27 предикторов, проверять каждый из них на участие во взаимодействии – очень долго и утомительно. Есть различные процедуры, применение которых позволяет автоматизировать

⁴ Т.е. в регрессионном уравнении появится новый член вида $a_{ij}x_i x_j$.

процесс проверки предикторов на участие во взаимодействии. Мы говорим о довольно новых⁵ процедурах пошаговой регрессии. Они в разном сочетании реализованы в популярных стапакетах.

Разнообразие процедур пошаговой регрессии как по заложенным в них принципам, так и по решаемым ими задачам довольно велико, отношение к ним западного научного сообщества полярно, они до сих пор не имеют достаточного теоретического осмысления и обобщения. Поэтому, конечно, без их систематизации наши рекомендации не могут считаться законченными. Но такая систематизация – цель отдельной работы. В текущей же работе мы рекомендуем следовать при применении процедур пошаговой регрессии прилагаемыми к соответствующим статпакетам руководствам пользователей.

5. Иллюстрация применения рекомендаций

Применив наши рекомендации по учёту взаимодействия, мы смогли построить две качественные модели, которые при традиционном способе регрессионного моделирования совершенно не просматривались.

Первая модель:

- оценка качества модели (аналог R^2) равна 37,5%. Для социологических моделей вполне приемлемо;
- 5 предикторов, значимых на 5-процентном уровне значимости.

Таблица 1

Предикторы второй модели

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 5	PQR_3 ⁶	-2,92	0,34	73,92	1	0,00	0,05
	PR1_5 ⁷	-2,17	0,20	124,09	1	0,00	0,12
	PR2_6 ⁸	-3,63	0,48	56,53	1	0,00	0,03
	R1_7 ⁹	-2,21	0,28	61,71	1	0,00	0,11
	R1_11 ¹⁰	-3,17	0,24	172,67	1	0,00	0,04
	Constant	5,18	0,24	465,68	1	0,00	176,81

Интерпретация:

- если семья имеет доход ниже или выше, чем «Зарботков хватает на все, кроме покупки недвижимости», живёт в любом округе Москвы, кроме Западного, родители не разведены, не имеют учёной степени, то вероятность, что они будут водить детей на развлекательные детские мероприятия чаще, чем раз в месяц равна P . И эта вероятность P равна 0,99. Поскольку соотношение шансов $P/(1-P) = 176,81$ (строка, соответствующая константе таблицы 1). Назовём это состояние «status-quo». Теперь рассмотрим учтённые моделью потенциальные изменения (каждое изменение предполагает, что остальные признаки соответствуют status-quo):

⁵ Самой «старой» из них не более 20 лет.

⁶ «Разведен(а)».

⁷ «Зарботков хватает на все, кроме покупки недвижимости».

⁸ «Учёная степень».

⁹ «Западный административный округ».

¹⁰ «Московская область».

- если материальное состояние семьи меняется, то вероятность снижается до 0,95.
- если семья переезжает в Западный административный округ, то вероятность снижается до 0,88.
- если родители разводятся, то вероятность снижается до 0,91.
- если кто-то из родителей получает учёную степень, то вероятность снижается до 0,83.

Как видим, все предикторы в случае их «активации» (т.е. когда респонденты обладают этими свойствами) снижают вероятность того, что родители будут водить детей на развлекательные детские мероприятия чаще, чем раз в месяц.

Разумеется, модель доступна для более глубокой интерпретации. Например, интерпретации сочетаний предикторов (без перемножений).

Вторая модель:

- оценка качества модели (аналог R^2) равна 37,5%;
- 5 предикторов, значимых на 5-процентном уровне значимости.

Таблица 2

Предикторы второй модели

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 20	PQR_3	-3,16	0,36	75,24	1	0,00	0,04
	PR1_5	-2,45	0,22	128,59	1	0,00	0,09
	PR2_5 ¹¹	1,07	0,23	22,40	1	0,00	2,92
	PR4_1 ¹²	-2,12	0,29	51,75	1	0,00	0,12
	PR4_2 ¹³	-0,85	0,29	8,75	1	0,00	0,43
	PR4_3 ¹⁴	-1,02	0,27	14,21	1	0,00	0,36
	R1_7	-2,63	0,32	67,10	1	0,00	0,07
	R1_9 ¹⁵	-1,73	0,41	17,44	1	0,00	0,18
	R2_2 ¹⁶	3,20	0,44	52,94	1	0,00	24,41
	R2_3 ¹⁷	2,81	0,27	107,94	1	0,00	16,66
	Constant	3,12	0,26	142,19	1	0,00	22,65

Интерпретация:

Вторая оптимальная модель.

- если семья имеет доход ниже или выше, чем «Зарботков хватает на все, кроме покупки недвижимости», живёт в любом округе Москвы, кроме Восточного и Западного, в пределах Садового кольца или, наоборот, за МКАД, родители не разведены, не имеют высшего образования, не руководители, не менеджеры и

¹¹ «Высшее образование».

¹² «Руководитель».

¹³ «Менеджер».

¹⁴ «Служащий офиса».

¹⁵ «Восточный административный округ».

¹⁶ «В пределах Третьего кольца».

¹⁷ «В пределах МКАД».

не служащие офиса, то вероятность, что родители будут водить детей на развлекательные детские мероприятия чаще, чем раз в месяц равна P . И эта вероятность P равна 0,96. Поскольку соотношение шансов $P/(1-P) = 22,65$ (строка, соответствующая константе таблицы 2). Теперь рассмотрим учтённые моделью потенциальные изменения (каждое изменение предполагает, что остальные признаки соответствуют status-quo):

- если материальное состояние семьи меняется, то вероятность снижается до 0,66.
- если семья переезжает в Западный административный округ, то вероятность снижается до 0,62.
- если семья переезжает в Восточный административный округ, то вероятность снижается до 0,8.
- если семья переезжает на новое место жительства между Садовым и Третьим кольцами, то вероятность вырастает до 0,99.
- если семья переезжает на новое место жительства между Третьим кольцом и МКАД, то вероятность вырастает чуть меньше.
- если родители разводятся, то вероятность снижается до 0,5.
- если кто-то из родителей получает высшее образование, то вероятность вырастает до 0,99.
- если кто-то из родителей становится руководителем, то вероятность снижается до 0,73.
- если кто-то из родителей становится менеджером, то вероятность снижается до 0,91.
- если кто-то из родителей становится служащим офиса, то вероятность снижается до 0,89.

Как мы видим, смена семьёй места жительства относительно центра города и получение высшего образования кем-то из родителей в среднем способствуют частому посещению их детьми развлекательных детских мероприятий. Остальные признаки семьи и родителей, учтённые моделью, действуют в обратном направлении.

Следует ясно понимать, что это скорее риторический приём – рассматривать изменения в динамике. Наше исследование динамику не измеряло. Поэтому его выводы должны быть сформулированы более сухо: в терминах отнесения детей, чьи родители имеют такие-то признаки и чьи семьи имеют такие-то признаки, в категорию часто посещающих развлекательные детские мероприятия или в категорию редко посещающих. Именно поэтому логистическое регрессионное моделирование наряду с прочими задачами решает и задачу классификации объектов наблюдения. Фактически, это конечная задача логистического регрессионного моделирования – в отличие от более базового линейного регрессионного моделирования, конечной задачей которого выступает построение регрессионного уравнения.

6. Заключение

Мы продемонстрировали на конкретном примере, как при неумелом применении регрессии (п.2) прогноз получается неверным и существенно меняется в лучшую сторону при умелом её применении (п.5). Возможно, кому-то из читателей пример покажется недостаточно развёрнутым. Это ощущение оправданно хотя бы важностью самой области, из

которой взят пример: воспитание подрастающего поколения. Эта область действительно ключевая для воспроизводства нашего общества. Тем не менее, мы надеемся, нас извиняет тот факт, что мы планировали статью прежде всего как методную, нежели как содержательную. Кроме того, именно умелое применение метода позволило выявить интересные и не до конца понятные стороны взаимодействия родителей и детей. Тем самым открыта дорога к выдвижению новых гипотез и их проверке.

В начале статьи мы упоминали выдающийся своей сложностью для изучения феномен протестного поведения. Мы планируем применить наши наработки для его изучения.

ЛИТЕРАТУРА

1. Доугерти, К. Введение в эконометрику. М. ИНФРА-М, 1999. – 402 с.
2. Ротмистров А.Н. Сравнительный анализ факторов студенческого протестного движения в России на рубеже XIX-XX и в начале XXI века // Высшее образование сегодня. 2009. № 1. С. 36-41 // URL: http://www.hetoday.org/arxiv/2009/arxiv_0109.html
3. Толстова Ю.Н. Анализ социологических данных. Научный мир, 2000. – 352 с. // URL: <http://socioline.ru/pages/tolstova-yun-analiz-sotsiologicheskikh-dannyh>
4. Толстова Ю. Н. Измерение в социологии. КДУ, 2009. – 291 с., гл.13.3.1, URL: <http://socioline.ru/pages/yun-tolstova-izmerenie-v-sotsiologii>
5. Фонд "Общественное мнение" не будет больше прогнозировать результаты выборов // Электронный ресурс ИТАР-ТАСС. 2013. 12 сентября // URL: <http://www.itar-tass.com/c1/874380.html> (Дата обращения: 26.09.2013)
6. Hosmer, D. W., and S. Lemeshow. 2000. Applied Logistic Regression, 2nd ed. New York: John Wiley and Sons.
7. Agresti, A. An introduction to categorical data analysis. John Wiley & Sons, 1996. – 296 с.
8. Agresti, A. Statistical methods for the social sciences. Upper Saddle River Pearson Education International, 2009. – 609 с., ch.5 // URL: <http://bookfi.org/book/833357>
9. Roshchina Y. To drink or not to drink: the microeconomic analysis of alcohol consumption in Russia in 2006-2010 / Working papers by NRU Higher School of Economics. Series SOC "Sociology". 2013. No. 20 // URL: <http://publications.hse.ru/preprints/82519809>

Рецензент: Кирсанов Константин Александрович, д.э.н., профессор ФГБОУ ВПО ГУУ.