

УДК 519.25

ПРИМЕР «ВЫРАЩИВАНИЯ» РЕГРЕССИОННОЙ МОДЕЛИ СОЦИАЛЬНОГО ЯВЛЕНИЯ НА БАЗЕ КРИТЕРИЯ ПРАВДОПОДОБНОСТИ ИНТЕРПРЕТАЦИИ

Н.Б. Осипенко, А.Н. Осипенко, К.А. Осипенко

Гомельский государственный университет им. Ф. Скорины, Гомель

EXAMPLE OF THE «CULTIVATION» OF THE SOCIAL PHENOMENON'S REGRESSION MODEL ON THE BASIS OF CRITERION OF PLAUSIBLE INTERPRETATION

N.B. Osipenko, A.N. Osipenko, K.A. Osipenko

F. Scorina Gomel State University, Gomel

В статье описывается способ поэтапной корректировки регрессионной модели в задаче объяснения продолжительности жизни человека с помощью факторов его генетического потенциала. Для раскодирования этих факторов используется популярный среди практикующих психологов метод квадрата Пифагора. «Выращивание» модели ведется по двум направлениям. Отмечаются основные причины некорректных регрессионных построений при работе с качественными признаками. Интерпретация результирующих моделей позволяет по-новому взглянуть на реальные приоритеты в общественном устройстве и выработать рекомендации по их изменению с целью увеличения продолжительности жизни населения Беларуси.

Ключевые слова: регрессионная модель, социальное явление, продолжительность жизни, алгоритм Пифагора, критерий правдоподобности интерпретации, этапы «выращивания» модели.

In the article the way of step-by-step correction of regression model in the problem of explanation of a person's life expectancy with the help of factors of his genetic potential is described. For decoding these factors the popular among practicing psychologists method of a square of Pythagoras is used. «Cultivation» of the model is conducted in two directions. The main reasons for incorrect regression constructions are noted during the work with qualitative signs. Interpretation of resultant models allows to look at real priorities of a social system in a new way and develop recommendations about their change for the purpose of increase of life expectancy of the population of Belarus.

Keywords: regression model, social phenomenon, life expectancy, algorithm of Pythagoras, criterion of interpretation plausibility, stages of model «cultivation».

Введение

Статистическое исследование социального явления существенно отличается от анализа данных о работе технической системы. В частности, нередки случаи некорректного использования в уравнениях регрессии качественных признаков. В особенности это относится к номинальным или порядковым характеристикам объектов в виде цифровых значений. В предлагаемой работе описывается методологическая схема корректной обработки порядковых признаков на примере выявления обуславливающих продолжительность жизни человека факторов по дате рождения, преобразованной в последовательность из десяти порядковых признаков по алгоритму Пифагора [1]. Существуют гипотезы, что дата рождения каким-то образом отражает генетический потенциал человека. Многие практикующие психологи с успехом используют этот алгоритм для диагностики характера человека и его склонностей.

1 Исходный статистический материал

В качестве материала для исследования послужила выборка из 1376 жителей города Гомеля,

умерших с ноября 2006 г. по октябрь 2007 г. По каждому человеку брались данные в виде строки: имя, день, месяц, год рождения, день, месяц и год смерти. Для обработки данных использовался пакет «Statistica», в частности, его средства подготовки новых переменных путём того или иного преобразования исходных признаков, а также программы анализа вариантов, построения гистограмм, матриц корреляций признаков и уравнений множественной линейной регрессии.

Как известно, Пифагор, его ученики и последователи сократили все числа до цифр от 1 до 9 включительно, поскольку они являются исходными числами, из которых могут быть получены все другие. Алгоритм построения квадрата Пифагора опишем на примере. Пусть человек родился 11.07.1953. 1. Складываем все цифры даты рождения (получаем 27 – первое рабочее число). 2. Складываем цифры этого числа ($2 + 7 = 9$ – второе рабочее число). 3. Из первого числа отнимаем удвоенную первую цифру даты рождения ($27 - 2 = 25$ – третье рабочее число). 4. Сложим цифры третьего числа ($2 + 5 = 7$ – четвертое рабочее число). 5. Первый ряд цифр для квадрата

Пифагора формируется из цифр даты рождения (11, 7, 1953), второй – образуется из полученных выше четырех рабочих чисел (27, 9, 25, 7).
6. Строим таблицу в виде квадрата из трех столбцов и трех строк, в ячейках которых зарезервированы места для количеств одинаковых цифр 1, 2, ..., 9. 7. Подсчитываем во множестве цифр первого и второго ряда пункта 5 количество одинаковых цифр 1, 2, ..., 9 и заполняем соответствующие ячейки квадрата $KП1, KП2, \dots, KП9$.
8. Дополнительно определяется признак $KП0$ – встречаемость цифры 0 в рядах цифр из пункта 5.

Цифры в квадрате Пифагора характеризуют способности: 1 – волевой потенциал, желание управлять ситуацией; 2 – энергетический потенциал, заряженность на самореализацию; 3 – нацеленность на порядок и аккуратность, рациональность, склонность к научному объяснению событий; 4 – потенциал здоровья; 5 – наличие интуиции, способность предвидеть последствия решений; 6 – заземленность, тяга к физическому труду, 7 – талант, тяга к творчеству и новизне; 8 – желание быть полезным, готовность слышать других и помогать им, обязательность и ответственность; 9 – способность к запоминанию информации и к сложным умственным операциям.

2 Описание процесса «выращивания» регрессионной модели путем поэтапного выявления причин неправдоподобности ее интерпретации

Методологическая схема такого параллельного «выращивания» результата исследования вместе с инструментом исследования в рамках функционально-эволюционного подхода описана в [2].

Непосредственное использование множественной линейной регрессии признаков $KП0, KП1, KП2, \dots, KП9$ на целевой показатель продолжительности жизни дает следующее уравнение (объем выборки $N = 1376$, коэффициент множественной корреляции $R = 0,48$):

$$\text{ПродЖ} = 129,6 - 2,3 \cdot KП1 - 1,9 \cdot KП2 - 4,9 \cdot KП3 - 7,6 \cdot KП4 - 8,9 \cdot KП5 - 0,4 \cdot KП6 - 6,6 \cdot KП7 - 6,8 \cdot KП8 - 5,6 \cdot KП9 - 6,0 \cdot KП0.$$

Как видим, это уравнение не поддается разумной интерпретации.

Первая причина некорректности любых корреляционно-регрессионных моделей – это неоднородность исходной выборки. Для разрешения этой проблемы исходная выборка первоначально была разбита на 4 подвыборки: 1) мужчины с $KП0 = 0$; 2) мужчины с $KП0 > 0$; 3) женщины с $KП0 = 0$; 4) женщины с $KП0 > 0$.

При этом использовалась гипотеза, что наличие нулей в рядах цифр алгоритма Пифагора связано с раскрытием дополнительного генетического потенциала.

В результате, после отбраковки малозначимых признаков, были получены следующие уравнения:

1) мужчины с $KП0 = 0$ ($N = 380, R = 0,31$):

$$\text{ПродЖ} = 67 + 1,9 \cdot KП2 - 0,5 \cdot KП3 - 2,8 \cdot KП5 - 3,0 \cdot KП6 - 1,8 \cdot KП7;$$

2) мужчины с $KП0 > 0$ ($N = 470, R = 0,36$):

$$\text{ПродЖ} = 67,2 + 0,8 \cdot KП1 + 1,5 \cdot KП2 - 3,7 \cdot KП5 - 3,4 \cdot KП6 - 2,9 \cdot KП7 - 3,9 \cdot KП8;$$

3) женщины с $KП0 = 0$ ($N = 241, R = 0,48$):

$$\text{ПродЖ} = 82,8 + 0,4 \cdot KП1 + 1,3 \cdot KП2 - 4,6 \cdot KП3 - 1,7 \cdot KП4 - 3,7 \cdot KП5 - 4,0 \cdot KП6;$$

4) женщины с $KП0 > 0$ ($N = 284, R = 0,48$):

$$\text{ПродЖ} = 76,2 + 2 \cdot KП1 + 2,4 \cdot KП2 - 3,7 \cdot KП4 - 4,8 \cdot KП5 - 5,2 \cdot KП6 - 2,4 \cdot KП8 - 3,0 \cdot KП9.$$

Интерпретация этих уравнений в том смысле, что почти весь генетический потенциал способностей (за исключением волевого и энергетического) способствует снижению продолжительности жизни, также вызывает подозрения в некорректности построенных моделей.

Второй причиной некорректности анализа данных явилась необеспеченность монотонности изменения целевого количественного показателя (продолжительности жизни) при увеличении значений объясняющих порядковых признаков.

Для проверки этого факта было осуществлено:

1. Градация каждого признака $KП0, KП1, KП2, \dots, KП9$ согласно порядковым смыслам их психологической интерпретации. Так, для $KП1$ было выделено четыре градации: 1) $KП1 = 1$ или $KП1(1)$; 2) $KП1 = 2$ или $KП1(2)$; 3) $KП1 = 3$ или $KП1(3)$; 4) $KП1 \geq 4$ или $KП1(\geq 4)$.

2. Построение для каждого признака $KП0, KП1, KП2, \dots, KП9$ групп подпризнаков, соответствующих выделенным выше градациям. Добавление этих подпризнаков в матрицу обработки.

В итоге выяснилось, что для ряда признаков их подпризнаки в последовательности от наименьшего до наибольшего могут несколько раз поменять знак коэффициента корреляции с целевым показателем.

Третьей причиной некорректности предыдущих регрессионных моделей оказался тот факт, что в разные периоды жизни (до 50–55 лет и после 65–70 лет) работают разные механизмы «участия» факторов генетического потенциала человека в формировании события его смерти.

Наконец, четвертой причиной некорректности модели регрессионного прогноза является то, что признаки даты рождения человека не исчерпывают весь набор факторов, определяющих продолжительность его жизни (около 20–30% вклада). В связи с этим применительно к корреляционно-регрессионным построениям речь можно вести не столько о задаче прогноза

продолжительности жизни, сколько о характере влияния того или иного генетического потенциала человека на формирование события его смерти в определенном социальном окружении. Так, например, в таких странах, как Япония, Беларусь и Сомали, средние продолжительности жизни отличаются друг от друга на десятки лет. Естественно ожидать, что для них также будут отличаться и анализируемые нами статистические связи. Если в стране созданы условия для реализации, в частности, интуитивного и творческого потенциала, то соответствующие признаки $KП5$ и $KП7$ будут иметь положительную корреляцию с продолжительностью жизни. И, наоборот, без таких условий, попытки раскрытия этого потенциала будут встречать сопротивление со стороны социума, что, в свою очередь, будет сказываться на формировании события смерти.

Особенность статистических моделей, построенных по данным социальной природы, состоит в том, что они с большой долей условности удовлетворяют классическим требованиям однородности и представительности используемых выборок. Однородность выборки предполагает сохранение единого механизма генерации случайных событий. В нашем примере косвенным показателем стабильности этого механизма может служить относительная устойчивость знаков и величин коэффициентов корреляции объясняющих подпризнаков и целевого показателя при ограниченном (в 1,5–2 раза) сужении выборки. Как видим, сам процесс статистического исследования социальных объектов носит итеративный характер и обязывает специалиста на каждом этапе обнаружения очередной некорректности и неинтерпретируемости моделей выдвигать новые гипотезы о возможном различии механизмов генерации событий.

Итогом моделирования в настоящей работе стали следующие уравнения:

1) женщины с $ПродЖ \leq 55$, $N = 81$, $R = 0,63$ (анализ факторов уменьшения $ПродЖ$):

$$ПродЖ = 52,8 - 2,0 \cdot KП1(\geq 3) - 2,7 \cdot KП2(1) - 4,2 \cdot KП3(2) - 12,0 \cdot KП5(0) - 5,0 \cdot KП6(\geq 2) - 2,0 \cdot KП7(\geq 1) - 2,1 \cdot KП8(\geq 1) - 2,0 \cdot KП9(1);$$

2) женщины с $ПродЖ \geq 70$, $N = 328$, $R = 0,43$ (анализ факторов увеличения $ПродЖ$):

$$ПродЖ = 76,1 + 2,4 \cdot KП1(= 4) + 0,5 \cdot KП2(2) + 0,6 \cdot KП2(= 4) + 5,9 \cdot KП3(0) + 2,0 \cdot KП6(0) + 0,6 \cdot KП7(0) + 1,5 \cdot KП8(2) + 1,6 \cdot KП9(2) + 1,2 \cdot KП0(0) + 0,9 \cdot KП0(\geq 2);$$

3) мужчины с $ПродЖ \leq 50$, $N = 170$, $R = 0,41$ (анализ факторов уменьшения $ПродЖ$):

$$ПродЖ = 52 - 3,3 \cdot KП1(2-3) - 1,5 \cdot KП2(0-1) - 1,3 \cdot KП3(1) - 4,4 \cdot KП5(0) -$$

$$-4,5 \cdot KП6(0-1) - 2,4 \cdot KП7(\geq 2) -$$

$$-5,2 \cdot KП8(\geq 1) - 2,8 \cdot KП9(\geq 1);$$

4) мужчины с $ПродЖ \geq 65$, $KП0=0$, $N=203$, $R=0,43$ (анализ факторов увеличения $ПродЖ$):

$$ПродЖ = 70 + 4,5 \cdot KП1(\geq 4) + 3,2 \cdot KП2(\geq 3) + 2,3 \cdot KП3(0-1) + 1,6 \cdot KП4(0) + 0,8 \cdot KП5(1) + 1,8 \cdot KП6(\geq 2) + 1,1 \cdot KП7(\geq 2) + 3,7 \cdot KП8(2) + 1,0 \cdot KП9(1);$$

5) мужчины с $ПродЖ \geq 65$, $KП0>0$, $N=244$, $R=0,39$ (анализ факторов увеличения $ПродЖ$):

$$ПродЖ = 66,6 + 1,1 \cdot KП1(3) + 2,2 \cdot KП2(2) + 4,7 \cdot KП2(\geq 4) + 1,6 \cdot KП3(0) + 4,0 \cdot KП4(0) + 1,8 \cdot KП6(0-1) + 1,0 \cdot KП7(\geq 1) + 1,4 \cdot KП8(0) + 0,8 \cdot KП9(1) + 1,5 \cdot KП0(1).$$

Заметим, что коэффициенты этих уравнений отражают условный вклад подпризнаков (в годах) в изменение продолжительности жизни.

3 Интерпретация итоговых уравнений регрессионной модели

С целью обеспечения полноты обзора общей картины статистических связей на рисунке 3.1 показаны: направление связи (прямая и обратная) и сила (слабая, умеренная, сильная) вклада признаков $KП0, KП1, KП2, \dots, KП9$ в целевой показатель продолжительности жизни для пяти выделенных случаев.

Анализируя две модели для женщин с ранними смертями (1) и со смертями в зрелом возрасте (2), видим, что одни и те же подпризнаки $\{KП1(\geq 3)$ – высокий волевой потенциал, $KП5(0)$ – отсутствие потребности в инструкции и предвидении, $KП8(\geq 1)$ – внушаемость и зависимость от других людей} в первом случае вносят весомый вклад в раннюю смерть, во втором же случае – положительно сказываются на увеличении продолжительности жизни. Ранним смертям женщин способствуют также: слабый энергетический потенциал $KП2(1)$, рациональность и педантичность $KП3(2)$, существенная тяга к физическому труду $KП6(\geq 2)$, творческие задатки $KП7(\geq 1)$ и плохая память $KП9(1)$.

На увеличение продолжительности жизни женщин в зрелом возрасте влияют также: умеренные и очень высокие волевые показатели $KП2(2)$ и $KП2(\geq 4)$, отсутствие потребности в рационализации и прагматизации всех явлений жизни $KП3(0)$, низкая значимость физического труда $KП6(0)$ и творчества $KП7(0)$. Кроме того, благоприятными оказываются: отсутствие нулей $KП0(0)$ или значительное их присутствие $KП0(\geq 2)$.

| Признаки: | КП1 Воля | КП2 Энергия | КП3 Рацио- нальность | КП4 Здо- ровье | КП5 Инту- иция | КП6 Физ. труд | КП7 Твор- чество | КП8 Помощь другим | КП9 Память | КП10 Доп. потен- циал |
|---------------------------------|-------------|------------------|----------------------------|----------------------|----------------------|---------------------|------------------------|-------------------------|---------------|--------------------------------|
| Модели: | | | | | | | | | | |
| женщины ПродЖ ≤ 55 | (≥ 3) ↓ | (1) ↓ | (2) ↓ | | (0) ↓ | (≥ 2) ↓ | (≥ 1) ↓ | (≥ 1) ↓ | (1) ↓ | |
| женщины ПродЖ ≥ 70 | (4) ↑ | (2) (4) ↑ ↑ | (0) ↑ | | | (0) ↑ | (0) ↑ | (2) ↑ | (2) ↑ | (0) (≥ 2) ↑ ↑ |
| мужчины ПродЖ ≤ 50 | (2–3) ↓ | (0–1) ↓ | (1) ↓ | | (0) ↓ | (0–1) ↓ | (≥ 2) ↓ | (≥ 1) ↓ | (≥ 1) ↓ | |
| мужчины ПродЖ ≥ 65, КП0=0 | (≥ 4) ↑ | (≥ 3) ↑ | (0–1) ↑ | (0) ↑ | (1) ↑ | (≥ 2) ↑ | (≥ 2) ↑ | (2) ↑ | (1) ↑ | |
| мужчины ПродЖ ≥ 65, КП0>0 | (3) ↑ | (2) (≥ 4) ↑ ↑ | (0) ↑ | (0) ↑ | | (0–1) ↑ | (≥ 1) ↑ | (0) ↑ | (1) ↑ | (1) ↑ |

Обозначения связей в ячейках таблицы рисунка

прямая обратная

↑

↓

↑

↓

↑

↓

Слабая

Умеренная (значимая)

Сильная (уверенная)

Рисунок 3.1 – Направление и сила вклада признаков в изменение продолжительности жизни для пяти выделенных случаев, приведенных в вертикальной шапке рисунка таблицы

У мужчин с ранними смертями картина факторов уменьшения продолжительности жизни (3) аналогична соответствующей модели у женщин (1).

В свою очередь, модели роста продолжительности жизни в зрелом возрасте у мужчин ((4) и (5)) имеют некоторые отличия от соответствующей модели у женщин (2). Так, в них: появляется фактор КП4(0) – отсутствие потенциала физического здоровья (возможно, это способствует формированию интереса к своему здоровью из-за болезней ещё с детства); предполагается наличие некоторых творческих потребностей КП7(≥1); необязательно отсутствие интуиции и предвидения КП5(1) в модели (4). Во всех моделях роста продолжительности жизни (2), (4) и (5) наблюдаются высокие показатели волевого КП1 и энергетического КП2 потенциала, а также низкая рациональность КП3. Модели (4) и (5) у мужчин отличаются разной тягой к физическому труду КП6 и степенью зависимости от других людей КП8. В модели (4) наличие этих факторов увеличивает сроки жизни. А в модели (5) наоборот, их отсутствие КП8(0) или слабая выраженность КП6(0–1) повышают продолжительность жизни. Возможно, наличие нулей в рядах Пифагора открывает дополнительный потенциал самостоятельности и ответственности человека за свою жизнь.

Заключение

Обращаясь к практическим выводам не только для человека, но и для государства, отметим, что в нашей республике имеются огромные

возможности для увеличения средней продолжительности жизни. Прежде всего, это касается создания лучших условий для самореализации людей с творческими художественными или интеллектуальными способностями (модели (1), (3), (4) и (5)). Без такой самореализации люди (особенно со средним и высоким волевым потенциалом, плохим предвидением, прагматичные и с высокой внушаемостью со стороны) склонны отходить от своих основных жизненных задач, попадать в «опасные» компании и в итоге «формировать» событие смерти.

Резюмируя всё выше изложенное, отметим, что корректное статистическое исследование социальной системы предполагает не только хорошее владение инструментом анализа данных, но и одновременно глубокое знание природы этой системы.

ЛИТЕРАТУРА

1. Осипенко, К.А. Метод регрессионного моделирования продолжительности жизни по дате рождения / К.А. Осипенко, Н.Б. Осипенко // Творчество молодых 2012 : сборник научных работ студентов и аспирантов УО «ГГУ им. Ф. Скорины» : в 2 ч. / Гомельский гос. ун-т им. Ф. Скорины ; отв. ред. О.М. Демиденко. – Гомель, 2012. – Ч. 1. – С. 194–197.

2. Осипенко, А.Н. Метод и средства автоматизации моделирования активных систем: автореф. дис. ... канд. техн. наук: 05.13.18 / А.Н. Осипенко ; ГГУ им. Ф. Скорины. – Гомель, 1997. – 16 с.

Поступила в редакцию 16.10.13.