

по этому предмету? Что еще хотелось бы вам узнать о применении физики в быту?

4. Подведение итогов. (Оценивание работы и выставление отметок.)

5. Домашнее задание: *для всех* – повторить весь материал, изученный по темам «Тепловые явления», «Агрегатные состояния вещества», подготовиться к самостоятельной работе; *по желанию* – подготовить сообщение по одной из тем: «Принцип работы пароварки»; «Применение электромагнитных волн в быту»; «Рациональное применение физики в ванной комнате».

*Учитель:* Благодарю всех за активную работу на уроке и желаю успешной подготовки к самостоятельной работе. Всего вам доброго.

### Литература

- 1 Каменецкий, С. Е. Методика решения задач по физике в средней школе / С. Е. Каменецкий, В. П. Орехов. Книга для учителя. – М.: Просвещение, 1987. – 336 с.
- 2 Тульчинский, М. Е. Качественные задачи по физике в средней школе : пособие для учителей / М. Е. Тульчинский. – М.: Просвещение, 1972. – 240 с.
- 3 Балаш, В.А. Задачи по физике и методы их решения / В. А. Балаш. – М.: Просвещение, 1974. – 430 с.
- 4 Беликов, В. С. Решение задач по физике. Общие методы. – М.: Высшая школа, 1986. – 256 с.
- 5 Энгельгардт В. А. Познание явлений жизни / В. А. Энгельгардт. – М. : Наука, 1984. – 304 с.
- 6 Учебная программа дисциплины «Молекулярная физика и биология» [Электронный ресурс] / Направление 010787 Прикладные математика и физика – Режим доступа: - <http://rudocs.exdat.com/docs/index-67380.html>.

УДК 519.25

*К. А. Осипенко, А. Н. Осипенко*

### О КОРРЕКТНОСТИ ИСПОЛЬЗОВАНИЯ КАЧЕСТВЕННЫХ ПРИЗНАКОВ В РЕГРЕССИОННОЙ МОДЕЛИ ПРОГНОЗА КОЛИЧЕСТВЕННОГО ОКАЗАТЕЛЯ

*В статье описываются особенности статистической проверки гипотезы о связи даты рождения человека с его генетическим потенциалом на примере множественной регрессии продолжительности жизни на показатели квадрата Пифагора. Отмечаются основные причины некорректных регрессионных построений при работе с качественными признаками. Результаты проведенного анализа данных позволяют по-новому взглянуть на реальные приоритеты в общественном устройстве и выработать рекомендации по их изменению с целью увеличения продолжительности жизни населения Беларуси.*

В практике статистических исследований нередки случаи некорректного использования качественных признаков в регрессионных уравнениях. 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 прогноза продолжительности жизни человека по дате рождения, преобразованной в последовательность из десяти порядковых признаков по алгоритму Пифагора [1]. Многие практикующие психологи с успехом используют этот алгоритм для диагностики характера человека и его склонностей. Существуют гипотезы, что дата рождения каким-то образом отражает генетический потенциал человека.

В качестве материала для исследования послужила выборка из 1376-ти 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 КП1, КП2, ..., КП9. 8. Дополнительно определяется признак КП0 – 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.

Непосредственное использование множественной линейной регрессии признаков КП0, КП1, ..., КП9 на целевой показатель продолжительности жизни дает следующее уравнение (объем выборки  $N=1376$ , коэффициент множественной корреляции  $R=0,48$ ):

$$\text{ПродЖ} = 129,6 - 2,3 * \text{КП1} - 1,9 * \text{КП2} - 4,9 * \text{КП3} - 7,6 * \text{КП4} - 8,9 * \text{КП5} - 4 * \text{КП6} - 6,6 * \text{КП7} - 6,8 * \text{КП8} - 5,6 * \text{КП9} - 6,0 * \text{КП0}.$$

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  $\text{КП0} = 0$ ; 2) мужчины с  $\text{КП0} > 0$ ; 3) женщины с  $\text{КП0} = 0$ ; 4) женщины с  $\text{КП0} > 0$ .

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  $\text{КП0} = 0$  ( $N = 380$ ,  $R = 0,31$ ):

$$\text{ПродЖ} = 67,0 + 1,9 * \text{КП2} - 0,5 * \text{КП3} - 2,8 * \text{КП5} - 3,0 * \text{КП6} - 1,8 * \text{КП7};$$

2) мужчины с  $\text{КП0} > 0$  ( $N = 470$ ,  $R = 0,36$ ):

$$\text{ПродЖ} = 67,2 + 0,8 * \text{КП1} + 1,5 * \text{КП2} - 3,7 * \text{КП5} - 3,4 * \text{КП6} - 2,9 * \text{КП7} - 3,9 * \text{КП8};$$

3) женщины с  $\text{КП0} = 0$  ( $N = 241$ ,  $R = 0,48$ ):

$$\text{ПродЖ} = 82,8 + 0,4 * \text{КП1} + 1,3 * \text{КП2} - 4,6 * \text{КП3} - 1,7 * \text{КП4} - 3,7 * \text{КП5} - 4,0 * \text{КП6};$$

4) женщины с  $\text{КП0} > 0$  ( $N = 284$ ,  $R = 0,48$ ):

$$\text{ПродЖ} = 76,2 + 2 * \text{КП1} + 2,4 * \text{КП2} - 3,7 * \text{КП4} - 4,8 * \text{КП5} - 5,2 * \text{КП6} - 2,4 * \text{КП8} - 3,9 * \text{КП9}.$$

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.

Вторая причина некорректности анализа данных явилась необеспеченность монотонности изменения целевого количественного показателя (продолжительности жизни) с ростом значений объясняющих порядковых признаков.

Для проверки этого факта было осуществлено:

1. Градация каждого признака КР0, КР1, КР2,..., КР9 согласно порядковым смыслам их психологической интерпретации. Так, для КР1 было выделено четыре градации: 1) КР1 = 1 или КР1(1); 2) КР1 = 2 или КР1(2); 3) КР1 = 3 или КР1(3); 4) КР1 >= 4 или КР1(>= 4).

2. Построение для каждого признака КР0, КР1, КР2,..., КР9 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 КР5 и КР7 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 расширении выборки или ее сужении. 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 случайных событий.

Итогом моделирования в настоящей работе стали следующие уравнения:

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

$$\text{ПродЖ} = 52,8 - 2,0 * \text{КР1}(>= 3) - 2,7 * \text{КР2}(1) - 4,2 * \text{КР3}(2) - 12,0 * \text{КР5}(0) - 5,0 * \text{КР6}(>= 2) - 2,0 * \text{КР7}(>= 1) - 2,1 * \text{КР8}(>= 1) - 2,0 * \text{КР9}(1);$$

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

$$\text{ПродЖ} = 76,1 + 2,4 * \text{КР1}(= 4) + 0,5 * \text{КР2}(2) + 0,6 * \text{КР2}(= 4) + 5,9 * \text{КР3}(0) +$$

+ 2,0\*КП6(0) + 0,6\*КП7(0) + 1,5\*КП8(2) + 1,6\*КП9(2) + 1,2\*КП0(0) +  
0,9\*КП0(>= 2);

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

ПродЖ = 52 - 3,3\*КП1(2 - 3) - 1,5\*КП2(0 - 1) - 1,3\*КП3(1) - 4,4\*КП5(0) -  
- 4,5\*КП6(0 - 1) - 2,4\*КП7(>= 2) - 5,2\*КП8(>= 1) - 2,8\*КП9(>= 1);

4) мужчины с ПродЖ >= 65, КП0 = 0 (N = 203, R = 0,43, анализ факторов увеличения ПродЖ):

ПродЖ = 70 + 4,5\*КП1(>= 4) + 3,2\*КП2(>= 3) + 2,3\*КП3(0 - 1) + 1,6\*КП4(0) +  
0,8\*КП5(1) + 1,8\*КП6(>= 2) + 1,1\*КП7(>= 2) + 3,7\*КП8(2) + 1,0\*КП9(1);

5) мужчины с ПродЖ >= 65, КП0 > 0 (N = 244, R = 0,39, анализ факторов увеличения ПродЖ):

ПродЖ = 66,6 + 1,1\*КП1(3) + 2,2\*КП2(2) + 4,7\*КП2(>= 4) + 1,6\*КП3(0) +  
4,0\*КП4(0) + 1,8\*КП6(0 - 1) + 1,0\*КП7(>= 1) + 1,4\*КП8(0) + 0,8\*КП9(1) + 1,5\*КП0(1).

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

Анализируя две модели для женщин с ранними смертями (1) и со смертями в зрелом возрасте (2), видим, что одни и те же подпризнаки {КП1(>= 3) – высокий волевой потенциал, КП5(0) – отсутствие потребности в инструкции и предвидении, КП8(>= 1) – внушаемость и зависимость от других людей} в первом случае вносят весомый вклад в раннюю смерть, во втором же случае – положительно сказываются на росте продолжительности жизни. Ранним смертям женщин способствуют также: слабый энергетический потенциал (КП2(1)), рациональность и педантичность (КП3(2)), существенная тяга к физическому труду (КП6(>= 2)), творческие задатки (КП7(>= 1)) и плохая память (КП9(1)).

На увеличение продолжительности жизни женщин в зрелом возрасте влияют также: умеренные и очень высокие волевые показатели (КП2(2) и КП2(>= 4)), отсутствие потребности в рационализации и прагматизации всех явлений жизни (КП3(0)), низкая значимость физического труда (КП6(0)) и творчества (КП7(0)). Кроме того, благоприятными оказываются: отсутствие нулей КП0(0) или значительное их присутствие КП0(>= 2).

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 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.

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 социальных объектов предполагает не только хорошее владение инструментом анализа данных, но и одновременно глубокое знание природы самих объектов.

### 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.

УДК 519.25

*К. А. Осипенко, А. Н. Осипенко*

### **О ВОЗМОЖНОСТИ ИСПОЛЬЗОВАНИЯ ПАСПОРТНЫХ ДАННЫХ В СТАТИСТИЧЕСКОМ АНАЛИЗЕ ПРИЧИН СМЕРТНОСТИ НАСЕЛЕНИЯ**

*В статье описывается один из вариантов трехэтапного скрининга здоровья населения. Предлагается способ использования паспортных данных для предварительной экспресс-диагностики групп риска обследуемого человека. Даются алгоритмы перевода даты рождения, имени и фамилии в множество бинарных признаков для распознавания групп риска. Приводятся схема распознавания и формула дискриминантной функции. На примере пробного исследования показывается практическая целесообразность применения этих данных на первом этапе скрининга.*

#### **Обоснование целесообразности использования паспортных данных в статистическом анализе причин смертности населения**

Задача выявления причин смертности и, в частности, особенностей основных групп риска по статистическим данным является одной из ведущих в сфере санитарно-гигиенических исследований. Особенность современного момента состоит в том, что появляются технические возможности для массового скрининга здоровья населения. Разрабатываются и внедряются различные концепции такого скрининга по поводу ранней диагностики и профилактики онкозаболеваний, туберкулеза, сердечнососудистых и других заболеваний. В связи с этим возникает проблема согласования и синхронизации всех этих обследований в рамках единой концепции общереспубликанского скрининга здоровья населения.

В настоящей работе предлагается для этих целей разработать методологию и программно-технологическое обеспечение предварительной экспресс-диагностики на основе паспортных сведений, данных о группе крови, антропометрии, анамнезе и иной стандартной информации, имеющейся в поликлинических базах данных. В перспективе к этой информации могут быть добавлены данные биометрии (отпечатки пальцев, фотографии сетчатки глаза), фрагменты почерка. В еще более далекой перспективе – данные недорогого экспресс-анализа ДНК.

В результате такой диагностики для каждого человека будут сформированы оценки вероятностей принадлежности к основным группам риска.

На втором этапе скрининга должно осуществляться подробное анкетирование граждан по поводу соответствующих целевых проблем в наиболее вероятных для них группах риска.

Наконец, только на третьем этапе предполагается перейти к более дорогостоящему детальному клиническому и амбулаторному обследованию (если в этом появляется