

УДК 004:614.2(575.2)

**ОБУЧЕНИЕ МЕТОДАМ ОБРАБОТКИ СТАТИСТИЧЕСКИХ ДАННЫХ
ПО ЗДРАВООХРАНЕНИЮ КЫРГЫЗСКОЙ РЕСПУБЛИКИ
В СИТУАЦИОННЫХ ЗАДАЧАХ**

С.Ц. Манжикова, Ч.С. Исмаилова

Показано проведение анализа статистических данных с целью получения научно достоверной информации на примерах решения ситуационных задач, сформулированных применительно к условиям Кыргызской Республики. Использован метод обработки динамических (временных) рядов данных, ретроспективный и оперативный анализ данных, представленных в статистических ежегодниках КР с использованием инструментария табличного процессора MS Excel 2016. Показана эффективность влияния инвестиций в охрану окружающей среды на заболеваемость активным туберкулезом органов дыхания в республике, для чего построена регрессионная модель. Данные исследований рекомендованы для работников общественного здравоохранения, а также для преподавателей информатики, медицинской информатики, информационных технологий в медицине и здравоохранении и др. точных наук.

Ключевые слова: здравоохранение; туберкулез; инвестиции; динамический ряд; коэффициент детерминации; корреляция.

**КЫРГЫЗ РЕСПУБЛИКАСЫНЫН САЛАМАТТЫК САКТОО СИСТЕМАСЫ
БОЮНЧА КЫРДААЛДЫК МАСЕЛЕЛЕРДЕ СТАТИСТИКАЛЫК
МААЛЫМАТТАРДЫ ИШТЕТҮҮ ЫКМАЛАРЫНА ОКУТУУ**

С.Ц. Манжикова, Ч.С. Исмаилова

Бул макалада Кыргыз Республикасынын шарттарында колдонуу үчүн түзүлгөн кырдаалдык маселелерди чечүүнүн мисалында илимий так маалыматтарды алуу максатында статистикалык маалыматтарга талдоо жүргүзүлдү. Маалыматтардын динамикалык (убактылуу) катарларын иштеп чыгуу, MS Excel 2016 таблица түрүндөгү процессордун каражаттарын пайдалануу менен Кыргыз Республикасынын статистикалык күндөлүктөрүндө берилген маалыматтарга ретроспективдүү жана оперативдүү талдоо жүргүзүү ыкмалары колдонулду. Республикадагы дем алуу органдарынын ачык кургак учук дартына айлана-чөйрөнү коргоого инвестицияларды салуунун натыйжалуулугу көрсөтүлдү, бул үчүн регрессивдүү модель түзүлдү. Аталган изилдөөлөр коомдук саламаттык сактоо системасынын кызматкерлери, ошондой эле информатика, медициналык информатика, медицинадагы маалыматтык технологиялар, саламаттык сактоо жана башка так илимдер боюнча окутуучулар үчүн сунушталат.

Түйүндүү сөздөр: саламаттык сактоо; кургак учук; инвестициялар; динамикалык катар; детерминация; коэффициент; корреляция.

**STUDY IN METHODS OF PROCESSING STATISTICAL DATA ON HEALTHCARE
OF THE KYRGYZ REPUBLIC VIA SITUATIONAL TASKS**

S.Ts. Manzhikova, Ch.S. Ismailova

The analysis of statistical data is carried out in order to obtain scientifically reliable information on examples of solving situational tasks formulated in relation to the conditions of the Kyrgyz Republic. The method of processing dynamic (time)

data series, retrospective and operational analysis of data presented in the statistical yearbooks of the KR using the tools of the MS Excel 2016 spreadsheet was applied. The effectiveness of the impact of investments in environmental protection on the incidence of active respiratory tuberculosis in the republic is shown, for which a regression model is built. Research data are recommended for public health workers, as well as for teachers of informatics, medical informatics, information technology in medicine and healthcare, and other exact sciences.

Keywords: health care; tuberculosis; investments; time series; coefficient of determination; correlation.

Актуальность данного исследования обуславливают два основных фактора: состояние общественного здравоохранения и образование в медицинских вузах.

Общественное здравоохранение. Известно, что для нашего региона, к сожалению, характерен тревожный уровень заболеваемости туберкулезом и активным туберкулезом дыхательных путей [1]. По данным ВОЗ, туберкулез входит в десятку ведущих причин смертности в мире [2]. В глобальном масштабе ВОЗ констатирует снижение заболеваемости туберкулезом примерно на 2 % в год. Но для достижения целей, предусмотренных Стратегией по ликвидации туберкулеза, эти темпы снижения необходимо ускорить до 4–5 % в год [2]. Поэтому в мире проводятся многочисленные исследования, направленные на решение этой задачи. Наряду с успешными изысканиями, проведенными Британскими учеными по разработке оболочки для вакцины от БЦЖ, позволяющей долгое время хранить ее в тепле [3], проводились исследования по изучению влияния экологических факторов на распространение туберкулеза [4]. В работе приведены результаты анализа влияния инвестиционных вложений в охрану окружающей среды на распространение этого заболевания в КР, что, по мнению авторов, может послужить определенным вкладом в решение этой важной проблемы.

Образование в медицинском вузе. Опыт преподавания таких дисциплин как информатика, медицинская информатика, информационные технологии в медицине и здравоохранении и др. точных наук студентам медицинских специальностей показывает, что обойти рассмотрение вопросов обработки статистических данных невозможно по нескольким причинам. Во-первых, в статистических данных скрыта ценная информация, которую специалисты должны уметь «добывать», а выводы, сделанные на основе такой информации, могут стать основой для принятия важных профессиональных и/или управленческих решений. Во-вторых, анализ так называемых «больших данных» является одним из самых современных направлений развития и практического применения информационных и компьютерных технологий. Преподаватель вуза обязан дать студентам представление об этой области знаний и технологий уже в самом начале их образовательного процесса. В-третьих, уметь практически применять научные методы обработки статистических данных и анализировать получаемые результаты, означает работать в соответствии с принципами доказательной медицины, что также отвечает требованиям сегодняшнего дня.

Таким образом, чтобы было понятно, каким образом построена модель зависимости численности заболеваемости туберкулезом в Кыргызстане от размера инвестиций в охрану окружающей среды, в работе приводятся примеры еще двух ситуационных задач и показано, как и какую информацию можно получить из статистических данных. Построение модели позволило сделать очень интересные выводы, а также повысить привлекательность математической статистики для студентов медицинского направления. Необходимо отметить и повышение профессионального интереса к анализу официальных статистических данных по здравоохранению КР и её экономики, что отражено в публикациях [5–7].

Материал и методы исследования

Предмет и методы исследования. В статье использованы официальные статистические данные, опубликованные в Статистических ежегодниках Кыргызской Республики с 2006 по 2018 г. [8–10], издаваемых Национальным статистическим комитетом и размещаемых на официальном сайте. Использован метод обработки динамических (временных) рядов данных, ретроспективный и оперативный анализ данных с использованием инструментария табличного процессора MS Excel.

Результаты и обсуждение. Результаты исследования представлены на рисунках 1–6. Для оценки взаимозависимости показателей авторами были разработаны три ситуационные задачи, в которых эти

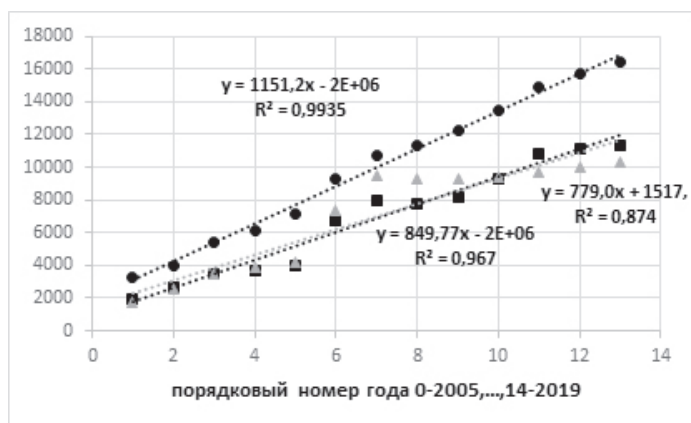


Рисунок 1 – Изменение численности врачей в КР в период 2006–2018 гг.:

• – исходные данные; ••• – линейная аппроксимация; ---- – полиномиальная аппроксимация

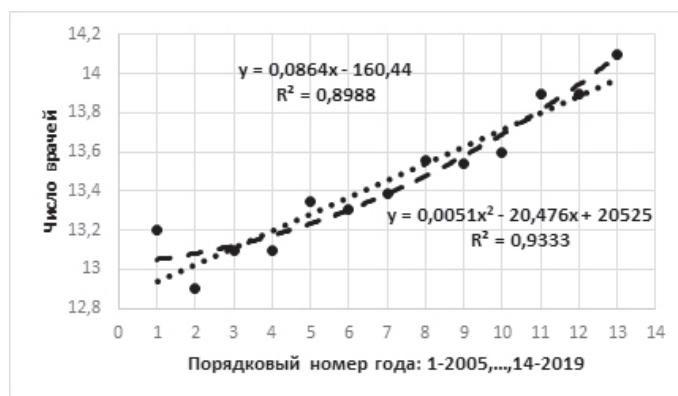


Рисунок 2 – Изменение номинальной з/п в КР в период с 2006 по 2018 г.:

■ – среднемесячная номинальная з/п работающих в экономике вообще;
 ▲ – з/п работающих в образовании; ■ – з/п работников здравоохранения и социальных услуг

показатели выступают в качестве исходных данных. Рисунки 1, 2 относятся к первой ситуационной задаче, рисунки 3, 4, 5 – ко второй задаче, рисунок 6 – к третьей ситуационной задаче.

Данные, представленные в работах [8–10], позволяют сгруппировать их в две группы: 1) факторы, или независимые, т. е. влияющие данные; 2) параметры, зависимые от факторов, которые могут перемещаться из одной группы в другую, в зависимости от поставленной задачи.

Ситуационная задача 1. Принимая во внимание официальные статистические данные Нацстаткома определить, является ли среднемесячная зарплата в КР решающим фактором при выборе профессии врача.

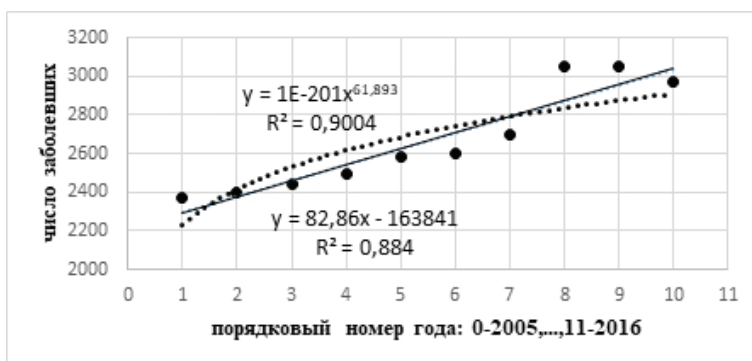


Рисунок 3 – Динамика заболеваемости в КР в период 2006–2016 гг.:

■ – исходные данные; ●●● – степенная аппроксимация; прямая – линейная аппроксимация

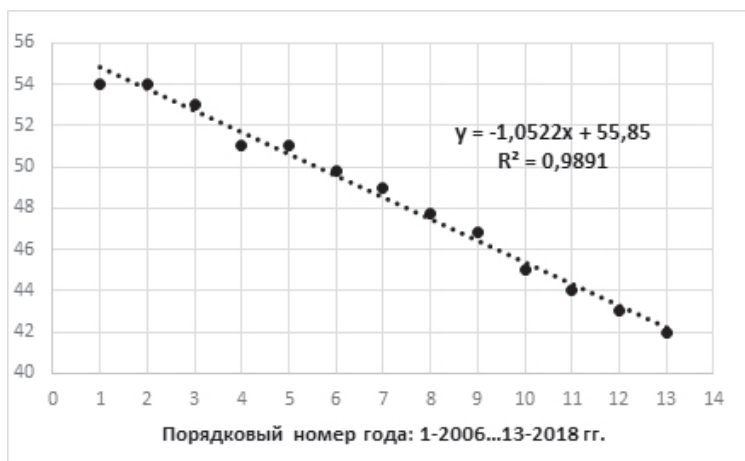


Рисунок 4 – Динамика изменения числа больничных коек на 10 тыс. населения в КР в период 2006–2018 гг.

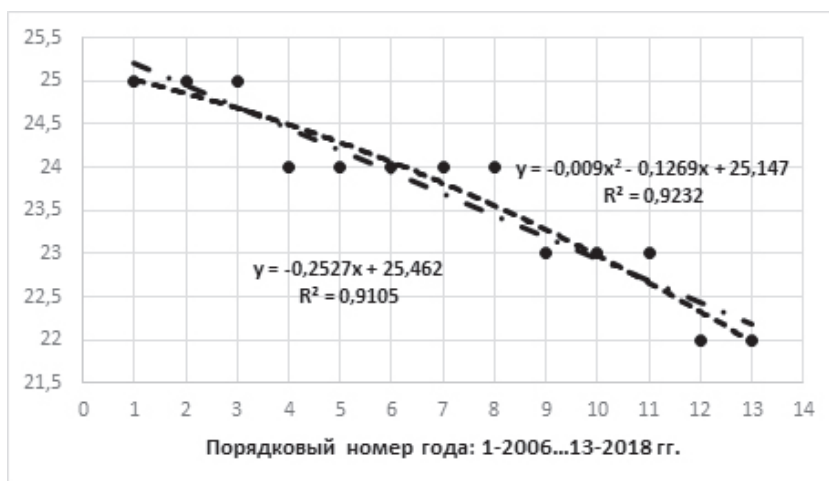


Рисунок 5 – Изменение численности врачей на 10 тыс. населения в КР в период 2006–2018 гг.:

■ – исх. данные; —●—●— — линейная аппроксимация; — — — — полиномиальная аппроксимация;

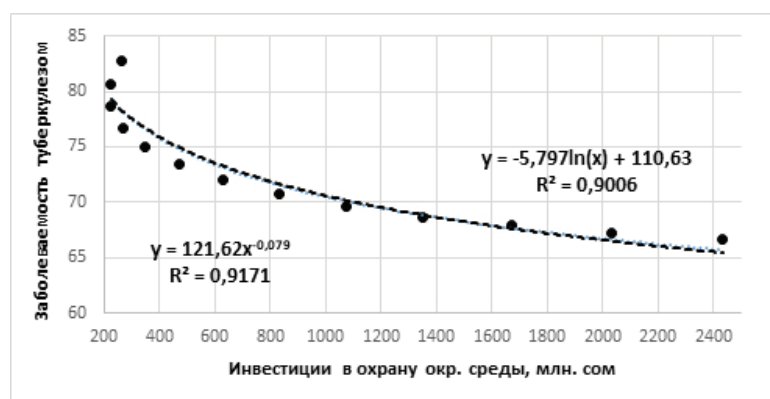


Рисунок 6 – Влияние инвестиций в охрану окружающей среды на заболеваемость активным туберкулезом органов дыхания на 100 000 человек: ● – исходные данные; ···· – степенная аппроксимация; ---- – логарифмическая аппроксимация

На рисунке 1 показано, как меняется этот показатель в КР с 2006 по 2018 г. Крупные точки соответствуют наблюдениям, приведенным в публикациях [8–10]. Данные хорошо аппроксимируются линейной зависимостью [11], в которой y – это число врачей, а x – это год. Коэффициент детерминации $R^2 = 0,8988$ свидетельствует о достаточно высоком качестве линейной аппроксимации, т. к. в этом случае ею описываются 89,88 %, почти 90 % наблюдений. Применение инструментария табличного процессора MS Excel показало, что аппроксимация исходного динамического ряда полиномиальной зависимостью $y = 0,0051x^2 - 20,47x + 20525$ лучше, чем линейная, описывает исходный временной ряд. Об этом говорит значение $R^2 = 0,9333$, т. е. вторая аппроксимирующая функция охватывает уже 93,33 % наблюдений. Если учесть связь коэффициентов детерминации R^2 и корреляции r , коэффициент корреляции числа врачей со временем (годом) составит в случае линейной аппроксимации:

$$r_{\text{лин}} = \sqrt{R^2} = \sqrt{0,8988} = 0,948,$$

а в случае полиномиальной зависимости:

$$r_{\text{поли}} = \sqrt{R^2} = \sqrt{0,9333} = 0,966,$$

т. е. обе аппроксимирующие функции указывают на очень тесную положительную и устойчивую связь численности врачей в КР, со временем оба коэффициента корреляции – $r_{\text{лин}}$ и $r_{\text{поли}}$ окажутся близкими к единице.

Чем же может быть вызван наблюдаемый устойчивый рост числа врачей в нашей стране? Рассмотрим такой важный фактор, влияющий на выбор профессии, как заработная плата. Данные Нацстаткома КР можно иллюстрировать графиком, представленным на рисунке 2. Каждый динамический ряд хорошо аппроксимируется линейной зависимостью – значения коэффициентов детерминации R^2 очень близки к единице. Это, в свою очередь, указывает на то, что динамика среднемесячной з/п имеет тесную положительную корреляционную связь со временем [8], и носит устойчивый характер. Однако средние темпы роста з/п в экономике КР вообще существенно превышают рост з/п работников здравоохранения и социальных услуг. В первом случае темп составляет 1151,2 сом. в год, а во втором случае – только 779,05 сом. в год. Даже в сфере образования годовой прирост з/п выше, чем в здравоохранении, и примерно равен 849,77 сом. Это свидетельствует о том, что заработная плата не является решающим фактором при выборе профессии врача.

Здесь необходимо дать некоторые пояснения относительно выбора показателя. Он напрямую зависит от условий поставленной задачи. Показатель численности врачей может рассматриваться в двух вариантах: абсолютное число врачей и число врачей на 10 000 человек (нормированный показатель). В нашем случае заработная плата рассмотренных категорий работников не зависит от численности населения республики, поэтому правильным в качестве показателя является выбор абсолютного числа врачей.

Неправильный выбор показателя может привести к абсолютно противоположному выводу. Например, если вместо абсолютного числа врачей взять нормированный показатель – число врачей на 10 000 населения, то мы увидим устойчивое его снижение в период с 2006 по 2018 г. (рисунок 5). И последующий анализ приведет к выводу о том, что заработная плата влияет на выбор профессии врача.

Очевидно, что студентов стоит учить видеть через статистику и такие связи. Далее, констатируя устойчивый рост числа врачей, можно задаться вопросом проверки предположения, что уровень заболеваемости зависит от числа врачей. Поиск ответа на этот вопрос возможен при рассмотрении следующей ситуационной задачи.

Ситуационная задача 2. Анализируя официальные статистические данные Нацстаткома КР, выявить, существует ли корреляционная связь между уровнем заболеваемости в нашей стране с одной стороны – и такими показателями состояния здравоохранения как коечный фонд и число врачей, с другой стороны.

В этой задаче однозначно следует выбрать нормированный показатель числа врачей на 10000 населения. Графики на рисунках 3–5 построены по официальным статистическим данным [8–10]. Все три группы данных представляют собой динамические ряды. Поэтому мы не можем определять коэффициенты корреляции непосредственно для рядов исходных данных. Здесь необходимо выполнить следующие шаги.

1. Определить, насколько тесно зависит от времени (года) каждый из исследуемых параметров. Для этого в среде MS Excel строим аппроксимирующие прямые и/или кривые и выводим соответствующие им уравнения и значения коэффициентов детерминации R^2 . На графиках видно, что лучшей аппроксимацией зависимости заболеваемости от времени является полиномиальная, обладающая более высоким значением коэффициента детерминации $R^2 = 0,5976$ и, следовательно, показывающим более тесную корреляцию $r \approx 0,773$, чем это делает линейная функция.

Число больничных коек на 10 000 населения хорошо представляется линейной функцией, что подтверждает и значение $R^2 = 0,9891$, соответствующее $r \approx 0,995$.

Для численности врачей на 10 000 населения предпочтительной является полиномиальная аппроксимация временной зависимости с $R^2 = 0,9232$, и дающая коэффициент корреляции $r \approx 0,9608$.

2. Аппроксимирующие функции выполняют сглаживание временных/динамических рядов [11] и при значениях R^2 , близких к единице, их можно использовать для получения расчетных рядов данных, для которых уже можно вычислить коэффициенты корреляции между исследуемыми параметрами. Результаты расчетов представлены в таблице 1.

Заболеваемость рассчитана по формуле: $y = -5,7887x^2 - 100,93x + 1193,4$ (рисунок 3), число больничных коек – по формуле: $x_3 = -1,0522x + 55,85$ (рисунок 4), численность врачей – по формуле: $x_5 = -0,009x^2 - 0,1269x + 25,147$ (рисунок 5). Парные коэффициенты корреляции определяются инструментами MS Excel и имеют следующие значения: $r_{yx_3} = -0,71946$ между заболеваемостью y и числом больничных коек x_3 ; $r_{yx_5} = -0,71946$ – между заболеваемостью y и численностью врачей x_5 .

И, наконец, третья задача, которая является довольно интересным исследованием.

Ситуационная задача 3. Проанализировать, каким образом и насколько эффективно влияют на заболеваемость туберкулезом инвестиции в охрану окружающей среды в нашей республике. Существует ли зависимость между этим параметром и деятельностью государства?

Алгоритм действий и ход рассуждений для решения этой задачи аналогичен предыдущей задаче.

Таблица 1 – Расчетные сглаженные ряды данных исследуемых параметров y, x_3, x_5

Год	№ п/п x	Заболееваемость, тыс. чел., y	Число больничных коек на 10 000 населения, x_3	Численность врачей на 10 000 тыс. население, x_5
2006	1	1288,5413	54,7978	25,2093
2007	2	1372,1052	53,7456	24,9566
2008	3	1444,0917	52,6934	24,7039
2009	4	1504,5008	51,6412	24,4512
2010	5	1553,3325	50,589	24,1985
2011	6	1590,5868	49,5368	23,9458
2012	7	1616,2637	48,4846	23,6931
2013	8	1630,3632	47,4324	23,4404
2014	9	1632,8853	46,3802	23,1877
2015	10	1623,83	45,328	22,935
2016	11	1603,1973	44,2758	22,6823
2017	12	1570,9872	43,2236	22,4296
2018	13	1527,1997	42,1714	22,1769

1. Исходные динамические ряды [8–10] сглаживаются [11] следующими аппроксимирующими функциями:

- для заболеваемости y активным туберкулезом органов дыхания на 100 000 населения в период с 2006 по 2018г.: $y = 0,0746x^2 - 2,388x + 85,099$;
- для инвестиций x_6 в охрану окружающей среды: $x_6 = 19,871x^2 - 97,412x + 340,14$, где x – порядковый номер года.

2. Сглаженные динамические ряды получены в результате расчета параметров по этим формулам и приводятся в таблице 2.

3. Между сглаженными динамическими рядами существует тесная отрицательная корреляция, которая характеризуется коэффициентом: $r_{yx_6} = -0,862$.

4. При такой тесноте корреляции возможно построение регрессионной зависимости для описания связи между исследуемыми параметрами. В среде MS Excel такая модель построена (рисунок 6). На поле графика приведены два математических выражения функций, пригодных для описания/моделирования регрессионной зависимости между параметрами. На рисунке y – это заболеваемость, а x – это инвестиции в охрану окружающей среды. Предпочтение для прогнозов следует отдать показательной модели:

$$y = 121,62x^{-0,079} \text{ при } R^2 = 0,9171.$$

5. Воспользуемся этой моделью, чтобы определить, при каких объемах инвестиций в охрану окружающей среды в КР может быть снижена заболеваемость активным туберкулезом органов дыхания до 50 на 100 000 населения. В этом случае $y = 50$, а x нужно определить. Запишем выражение модели в следующем виде: $50 = 121,62x^{-0,079}$ и прологарифмируем его:

$$\ln 50 = \ln 121,62 + \ln(x^{-0,079}) = \ln 121,62 - 0,079 \ln x,$$

отсюда $\ln x = (\ln 50 - \ln 121,62) / (-0,079)$ или $\ln x = (3,912 - 4,800) / (-0,079) = 11,25$, следовательно, $x = e^{11,25} = \exp(11,25) = 77005$ млн сом.

Использование такого метода обработки динамических рядов для решения поставленных задач обусловлено ограниченным размером выборки. Построение модели на основе многофакторного регрессионного анализа, конечно, предпочтительней, но требует гораздо большего объема материала,

Таблица 2 – Сглаженные динамические ряды данных по инвестициям на охрану окружающей среды x_6 и заболеваемости у активным туберкулезом органов дыхания на 100 тыс. населения

Год	№ п/п, x	Инвестиции на охрану окр. среды, млн сом x_6	Заболеваемость активным туберкулезом органов дыхания на 100 тыс. населения y
2006	1	262,599	82,785
2007	2	224,8	80,6202
2008	3	226,743	78,6046
2009	4	268,428	76,7382
2010	5	349,855	75,021
2011	6	471,024	73,453
2012	7	631,935	72,0342
2013	8	832,588	70,7646
2014	9	1072,983	69,6442
2015	10	1353,12	68,673
2016	11	1672,999	67,851
2017	12	2032,62	67,1782
2018	13	2431,983	66,6546

что не представляется возможным ввиду отсутствия официальных статистических данных ранее 2006 года. Кроме того, многофакторному регрессионному анализу должен предшествовать корреляционный анализ для выбора показателей, которые в последующем могут быть включены в многофакторный анализ, что не входило в задачу данного исследования.

Заключение. Приведенный анализ статистических данных позволяет сделать следующие выводы:

1. Сравнение динамики таких важных социально-экономических показателей как численность врачей (рисунок 1) и среднемесячная номинальная з/п (рисунок 2), позволяет заключить, что з/п не является в нашей стране стимулирующим фактором для молодых людей, выбирающих профессию врача.

2. Значения коэффициентов корреляции говорят о наличии отрицательной достаточно тесной и устойчивой связи между заболеваемостью населения и такими социально-экономическими факторами, как коечный фонд и численность врачей.

3. Анализ официальных статистических данных показал, что инвестиции в охрану окружающей среды эффективно влияют на заболеваемость активным туберкулезом органов дыхания. Между этими параметрами существует взаимная зависимость, которая может быть аппроксимирована показательной регрессионной моделью с высоким значением коэффициента детерминации $R^2 = 0,917$, построенная математическая модель позволяет рассчитать размер инвестиций, необходимых для снижения показателя заболеваемости туберкулезом до желаемого уровня.

4. Рассмотренные ситуационные задачи, в основе которых лежит научный анализ реальных/официальных статистических данных, позволяют усилить познавательный интерес студентов-медиков к обработке больших данных и практическому применению методов доказательной медицины/здравоохранения.

Конфликт интересов. Авторы статьи подтверждают отсутствие финансовой или какой-либо иной поддержки исследования, конфликта интересов, о которых необходимо сообщить.

Литература

1. Рыскулбекова А.Б. Эпидемиологическая ситуация по туберкулезу в г. Бишкек Кыргызской Республики / А.Б. Рыскулбекова, М.Б. Ырысова, В.С. Тойгомбаева, А.И. Буюклянов // Актуальные проблемы гуманитарных и естественных наук. Бишкек: Научное изд-во «Институт стратегических исследований». 2015. № 11-7. С. 84–87.
2. Информационный бюллетень ВОЗ. Туберкулез, 2018. URL: <https://www.who.int/ru/news-room/factsheets/detail/tuberculosis> (дата обращения: 15.10.2020).

3. *Wahid A.A.* Ensilication Improves the Thermal Stability of the Tuberculosis Antigen Ag85b and an Sbi-Ag85b Vaccine Conjugate / A.A. Wahid, A. Doekhie, A. Sartbayeva, J.M. H-van den Elsen // Scientific reports. 2019. Vol. 9. Article number 11409.
4. *Миронова С.А.* Загрязнение среды как один из факторов развития туберкулеза органов дыхания / С.А. Миронова // Известия Самарского научного центра РАН. 2011. Т. 13. № 5. С. 259–264.
5. *Ибраимова Д.Д.* Заболеваемость и смертность от болезней системы кровообращения в КР / Д.Д. Ибраимова // Вестник КРСУ. 2014. Т. 14. № 10.
6. *Баймуратов Т.Т.* Нерешенные проблемы медико-социальной экспертизы в КР / Т.Т. Баймуратов, М.У. Султаналиева, З.А. Айдаров // Вестник КРСУ. 2019. Т. 19. № 5.
7. *Гусева В.И.* Экономические и неэкономические факторы роста на современном этапе социально-экономического развития КР / В.И. Гусева, Ю.В. Гусева // Вестник КРСУ. 2019. Т. 19. № 5.
8. 01.06.2015 // Статистический ежегодник Кыргызской Республики 2006–2010. URL: <http://www.stat.kg/ru/publications/statisticheskij-ezhegodnik-kyrgyzskoj-respubliki/> (дата обращения: 15.02.2021).
9. 19.05.2017 // Статистический ежегодник Кыргызской Республики 2011–2015. URL: <http://www.stat.kg/ru/publications/statisticheskij-ezhegodnik-kyrgyzskoj-respubliki/> (дата обращения: 15.02.2021).
10. 24.02.2020 // Статистический ежегодник Кыргызской Республики 2014–2018. URL: <http://www.stat.kg/ru/publications/statisticheskij-ezhegodnik-kyrgyzskoj-respubliki/> (дата обращения: 15.02.2021).
11. *Воскобойников Ю.Е.* Эконометрика в Excel: учеб. пособие. Ч. 2. Анализ временных рядов / Ю.Е. Воскобойников. Новосибирск: НГАСУ (Сибстрин), 2008. 152 с.