

УДК 004.9

Т.В. КОРЧАК

*Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", Украина*

## ИНФОРМАЦИОННАЯ ТЕХНОЛОГИЯ АНАЛИЗА ВРЕМЕННЫХ РЯДОВ БИМЕДИЦИНСКИХ ДАННЫХ

*Работа посвящена созданию автоматизированного рабочего места для врача-эпидемиолога. Разработанная информационная технология решает три задачи: оценка степени напряженности территории по заболеваемости; вычисление эпидемических порогов; прогнозирование развития эпидемического процесса. Рассмотрены методы математической статистики для анализа временных рядов биомедицинских данных. Изучены проблемы обработки временных рядов в теории биомедицинских данных в условиях неопределенности. Описаны преимущества автоматизации рабочего места эпидемиолога. Представлено описание информационного обеспечения, осуществляющего накопление, переработку и анализ данных, обеспечивающее работников санитарно-эпидемиологической службы эффективным и точным сбором поступающей информации.*

**Ключевые слова:** прогнозирование, биомедицинские данные, информационная технология, эпидемиологический надзор, эпидемический процесс, временной ряд.

### Введение

Эпидемиологический надзор – это инструмент, с помощью которого учреждения общественного здравоохранения следят за состоянием здоровья населения. Целью эпидемиологического надзора является получение фактов, на основе которых эти учреждения могут определить приоритетные направления своей деятельности, планировать профилактические программы, проводить мероприятия, направленные на улучшение и охрану здоровья населения [1].

В последние годы очевидной стала важность информации о тенденции количественных и качественных изменений заболеваний для более результативного анализа и обработки биомедицинских данных. Однако с увеличением количества поступающих на исследование данных, а также с ростом нагрузки на персонал, становятся неизбежными ошибки при формировании сведений о результатах анализа и отчетной документации, дублирование исследований из-за потери полученных данных и другие недочеты в деятельности санитарно-эпидемиологической службы (СЭС). Всё это приводит к неоправданному расходу финансовых и трудовых ресурсов и к некачественному анализу и прогнозированию эпидемиологической ситуации. Поэтому отмечается растущий интерес к внедрению автоматизированных систем для качественного проведения эпидемиологической диагностики и принятию решения для выбора профилактических и противоэпидемических мероприятий.

**Целью данного исследования** является создание информационной технологии эпидемиологического надзора для анализа временных рядов биомедицинских данных.

### 1. Особенности временных рядов биомедицинских данных

Сбор эпидемических данных является чрезвычайно важным этапом работы эпидемиологической службы [2]. Качество данных является залогом успеха любого эпидемиологического исследования. Сбор данных требует много времени и усилий. Данные могут быть получены из литературных источников; уже существующих специальных баз данных; путем опроса (анкетирования); медицинского осмотра; лабораторных исследований материала от людей, животных и источников в окружающей среде; инструментальных и лабораторных исследований факторов окружающей среды и т. п.

Исходя из цели данной работы, основными данными, которые будут рассматриваться, являются биомедицинские данные – ежемесячные, еженедельные абсолютные показатели заболеваемости, которые представлены в виде временных рядов.

Основными причинами возникновения неопределенностей в биомедицинских данных являются:

- 1) неопределенности, проявляющиеся из-за изменчивых свойств окружающей среды;
- 2) неточность при сборе и формировании данных: неполная регистрация инфекционных больных; сознательное занижение показателей заболеваемо-

сти, за счет того, что за пределами официальной регистрации и учета инфекционных больных нередко оказываются неработающие люди; занижение регистрации и учета инфекционных больных, особенно страдающих легкими формами заболеваний;

3) неопределенность возникающая в результате допущений, сделанных во время анализа или из-за использования упрощенных моделей.

Создание информационного обеспечения поможет свести к минимуму перечисленные выше неопределенности. Эффективный и точный сбор поступающей информации обеспечит качественный анализ и обработку биомедицинских данных, с последующим прогнозированием эпидемической ситуации.

## 2. Анализ биомедицинских данных на основе методов математической статистики

Сравнительная оценка состояния здоровья населения и окружающей среды на различных территориях позволяет выделить неблагополучные территории. Для оценки одного и того же явления, определяемого несколькими элементами, необходимы комплексные показатели, которые позволят выделить наиболее неблагополучные территории, оценить тенденции изменения здоровья населения и состояния окружающей среды [3].

### 2.1. Определения комплексного показателя заболеваемости

Для определения комплексного показателя заболеваемости (КПз) вычисляются несколько интегральных коэффициентов заболеваемости (ИКЗ), характеризующих различные особенности оцениваемого явления.

Получив ИКЗо для каждой территории, можно оценить ее положение в сравнении с другими территориями области. Коэффициент, характеризующий степень групповой заболеваемости (ИКЗ<sub>гр</sub>), вычисляется по абсолютным значениям заболеваний.

Интегральный коэффициент, характеризующий степень превышения среднегодовых уровней (ИКЗ<sub>см</sub>), вычисляется аналогично первому методу, только за основание берется не средний областной показатель, а средний многолетний показатель заболеваемости на данной территории за тот же период.

И далее считается комплексный показатель заболеваемости [4]:

$$КП_3 = \frac{ИКЗ_о + ИКЗ_{см} + ИКЗ_{гр}}{3}.$$

По комплексному показателю заболеваемости может быть рассчитана оценка степени напряженно-

сти ситуации по городу или области. Данная оценка может быть представлена как в числовой, так и в лингвистической форме, с лингвистическими терминами: удовлетворительная – [0;1,5], неудовлетворительная или относительно напряженная – [1,5;2], напряженная или существенно напряженная – [2;3,6], угрожаемая или чрезвычайная – [3,6; 10].

### 2.2. Метод вычисления эпидемических порогов

Методика расчета эпидемических порогов заболеваемости базируется на алгоритме, описанном в [4]. Исходными данными для вычисления эпидемических порогов являются статистические сведения о суммарной заболеваемости в интенсивных показателях за календарные недели каждого года в течение последних 5 лет. Согласно методике, до проведения расчета необходимо исключить из исходных данных недели, которые приходились на период эпидемий в городе.

Пусть  $X'_1, X'_2, \dots, X'_n$  – заболеваемость за  $i$ -тую неделю неэпидемического периода за  $N_i$  – количество лет из рассматриваемого промежутка времени, для которых данные за неделю номер  $i$  имеются и неделя была неэпидемической. Вычисления проводятся одним из двух методов, в зависимости от количества наблюдений  $N_i$ .

Первый способ вычисления эпидемических порогов: при достаточном числе наблюдений  $N_i \geq 5$  вычисляем [4]:

– среднее значение заболеваемости:

$$\bar{X}_i = \frac{1}{N_i} \sum_{n=1}^{N_i} X'_n; \quad (2.1)$$

– среднее квадратичное отклонение:

$$S_i = \sqrt{\frac{1}{N_i - 1} \sum_{n=1}^{N_i} (X'_n - \bar{X}_i)^2}; \quad (2.2)$$

– верхний толерантный предел – порог:

$$X'_n = \bar{X}_i + Q_{N_i - 2} \cdot \sqrt{\frac{N_i - 1}{N_i - 2 + Q_{N_i - 2}^2}} \cdot S_i, \quad (2.3)$$

где  $Q_{N_i - 2}$  – значение критерия Стьюдента для доверительной вероятности 95% с  $N_i - 2$ .

Второй способ вычисления эпидемических порогов: при малом числе наблюдений  $N_i < 5$  используем приближенный метод. Для каждой недели с числом наблюдений  $N_i > 1$ , за которые в рассматриваемый интервал лет было два и более года, когда для недели с номером  $i$  данные есть, и она была неэпидемической, производим вычисления по формулам (2.1) – (2.3).

Затем для всех этих недель определяем коэффициент вариации по формуле (2.4):

$$V_i = \frac{S_i}{X_i} \cdot 100\%, \quad (2.4)$$

где  $i$  – порядковый номер недели в году,  $1 \leq i \leq 52$ .

По выборке  $V_i$ , число наблюдений в которой 52 или менее, вычисляют средний коэффициент вариации по формуле (2.5):

$$\bar{V} = \frac{\sum V_i}{k}, \quad (2.5)$$

где  $k$  – количество недель, по которым вычислены коэффициенты вариации  $V_i$ .

Для каждой недели с малым числом наблюдений вычисляем оценку среднего квадратичного отклонения  $\hat{S}_i$  по формуле (2.6):

$$\hat{S}_i = \frac{\bar{V} + \bar{X}_i}{100}, \quad (2.6)$$

где  $\bar{V}$  – средний коэффициент вариации, вычисленный по формуле (2.5),  $\bar{X}_i$  – среднее значение для данной недели, вычисленной по формуле (2.1).

Далее значение эпидемического порога определяем по формуле (2.7):

$$X'_n = \bar{X}_i + 1,67 \cdot \hat{S}_i. \quad (2.7)$$

### 2.3. Вычисление прогнозного значения

Пусть  $X = [x_1, x_2, \dots, x_n]^T$  – исследуемый временной ряд, хранящий биомедицинскую информацию по выбранной заболеваемости, упорядоченный по некоторому временному параметру  $\Delta t$ ;  $x_1, x_2, \dots, x_n$  – набор дискретных значений. Временная последовательность  $X$  была реконструирована по следующему правилу:

$$X = \begin{pmatrix} X_i^1 & X_{i+L}^1 & X_{i+2L}^1 & X_{i+nL}^1 \\ X_i^2 & X_{i+L}^2 & X_{i+2L}^2 & X_{i+nL}^2 \\ \dots & \dots & \dots & \dots \\ X_i^M & X_{i+L}^M & X_{i+2L}^M & X_{i+nL}^M \end{pmatrix} \rightarrow \begin{pmatrix} X_{i+(n+1)L}^1 \\ X_{i+(n+1)L}^2 \\ \dots \\ X_{i+(n+1)L}^M \end{pmatrix}.$$

где  $L$  – лаг или реконструируемая сдвигка ( $L=12$ ),  $X^i$  – ряд значений заболеваемости для  $i$ -го месяца,  $X_{i+(n+1)L}^1, X_{i+(n+1)L}^2, \dots$  – спрогнозированное значение,  $n$  – размерность вложения.

Прогноз строится помесячно на основании данных за конкретный месяц в разные годы с использованием адаптивной модели Брауна [5].

### 3. Автоматизация рабочего места эпидемиолога

Выполнение требований, предъявляемых к обработке информации, связано с определенными трудностями, преодоление которых возможно лишь с помощью электронно-вычислительной техники (ЭВТ). Средства ЭВТ освобождают человека от выполнения огромного объема механической работы и становятся одним из основных факторов роста производительности работы эпидемиолога.

В данной работе был разработан алгоритм, реализация которого представляет собой программный продукт, обрабатывающий исходную эпидемиологическую информацию, представленную в виде временного ряда за определенный период и дающий лингвистическую оценку развития эпидемического процесса. С помощью разработанной системы решаются три задачи:

- 1) оценка степени напряженности территории по заболеваемости (рис.1);
- 2) вычисление эпидемических порогов (рис.2);
- 3) прогнозирование развития эпидемического процесса (рис.3,4).

Выходные данные реализованы в числовом и графическом виде, что позволяет наглядно отобразить результаты анализа.

На рисунке 1 показана экранная форма оценки степени напряженности территории по заболеваемости, где в качестве исходных данных были взяты показатели заболеваемости острыми кишечными инфекциями (ОКИ) – дизентерия, ГЭК, сальмонеллез. По результатам расчетов можно сделать вывод, что как по районам г. Харькова, так и по области степень напряженности ситуации по всей группе ОКИ удовлетворительная.

На рисунке 2 изображена экранная форма вычисления эпидемических порогов, которая представляет собой два блока: первый – численное решение, второй – графическое представление. В блоке с численным решением столбец с нумерацией (27-37) – это порядковые номера недель. График можно построить как для отдельной недели, так и для всего периода, для этого необходимо выбрать номер недели и нажать кнопку «Рисовать». В численном решении представлены строки зеленого, синего и красного цвета.

При расчете эпидемических порогов красным цветом отмечаются недели, которые не рассматриваются в вычислении (напр. неделя под номером 34). Синим – недели, у которых число наблюдений меньше пяти (напр. недели под номерами 27, 28, 30, 36, 38); зеленым – недели, у которых достаточное количество данных (напр. недели под номерами 29, 31-33, 35, 37).

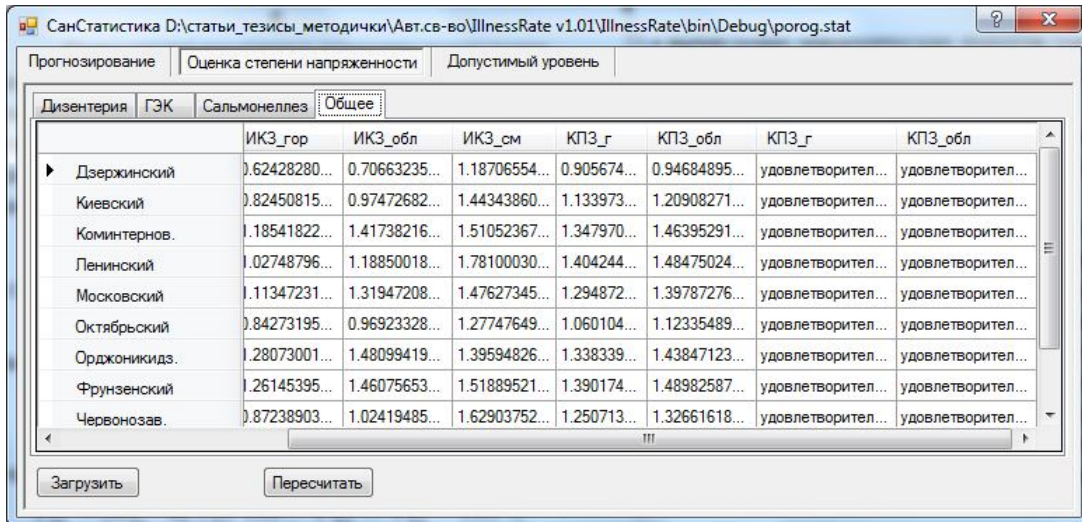


Рис. 1. Результат оценки степени напряженности ситуации по заболеваемости на территории Харькова и области

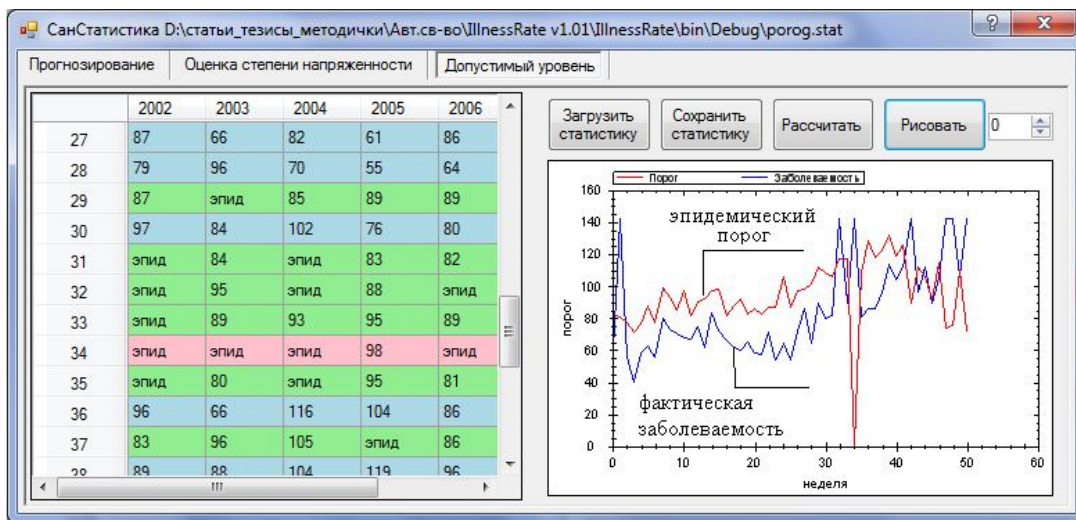


Рис. 2. Экранная форма вычисления эпидемических порогов

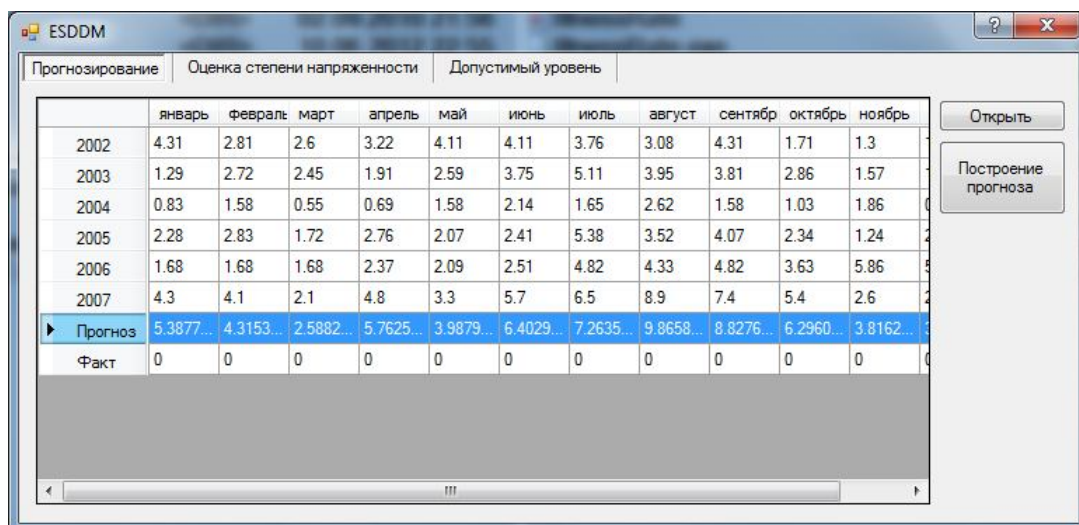


Рис. 3. Численное отображение результатов прогнозирования

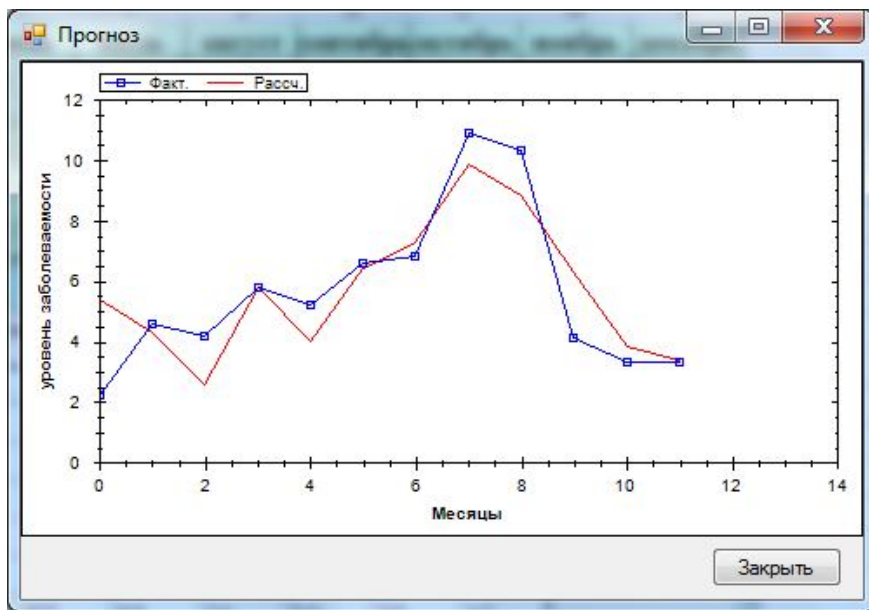


Рис. 4. Графическое отображение результатов прогнозирования

Вычисление эпидемических порогов позволит врачу-эпидемиологу:

- обнаружить превышения многолетнего уровня инфекционных заболеваний;
- определить количественную оценку интенсивности течения эпидемического процесса при вспышках, эпидподъемах и превышениях многолетнего уровня;
- формировать заключения о «возможном начале», «продолжении», «возможном окончании» и «окончании» вспышек, эпидподъемов и превышений многолетних уровней.

На рисунках 3 и 4 приведена экранная форма результатов прогнозирования заболеваемости сальмонеллезом на 2008 год. Прогноз строился по модели Брауна на основании периода 2002-2007 г. На рисунке 3 есть строка, в которую записываются прогнозные значения, и строка фактических данных, в которую по мере поступления данных врач-эпидемиолог может вносить реальные значения по заболеваемости.

При получении прогнозного значения, может возникнуть две ситуации. Первая – когда полученное значение превышает допустимый уровень; вторая – когда полученное значение не превышает допустимого уровня, но в середине месяца количество заболеваний стало близким прогнозируемому значению. В обоих случаях работа врача-эпидемиолога заключается в проведении мероприятий по предотвращению возникновения эпидемической ситуации и необходимо начать работу по проведению программ по профилактике инфекционных болезней и обеспечению санитарно-эпидемиологического благополучия населения.

## Заключение

В данной работе рассмотрены основные задачи, встречаемые в работе СЭС, и проблемы, возникающие при их решении. Описаны методы математической статистики для анализа временных рядов биомедицинских данных. Изучены особенности временных рядов биомедицинских данных, описаны свойственные им неопределенности. Представлена информационная технология прогнозирования биомедицинских данных. Приведены примеры работы информационной технологии в виде графических отображений результатов прогнозирования и задачи вычисления эпидемических порогов для анализируемого периода.

Автоматизация рабочего места эпидемиолога позволяет решить следующие основные задачи:

- сведение к минимуму количество ошибок при проведении исследований и обработке данных;
- ускорение и контроль качества исследований;
- обеспечение оперативного доступа к результатам исследований;
- уменьшение затрат рабочего времени персонала при подготовке текущей и отчетной документации.

Информационная технология позволяет эффективно проанализировать оперативную информацию, а также дать лингвистическую оценку состояния эпидемического процесса. А это позволяет значительно сократить время обработки статистических данных, дает возможность использовать большие объемы информации, что в свою очередь, влечет за собой возможность получить информативный материал для разработки мер профилактики.

## Литература

1. *Principles of Epidemiology: An Introduction to Applied Epidemiology and Biostatistics [Text]. – 2<sup>nd</sup> ed. – Atlanta, GA, 1992. – 458 p.*
2. Гюрджян, К.Т. *Современный эпидемиологический надзор [Текст]: учеб. пособие / К.Т. Гюрджян, В.А. Давидянц. – Ереван: Авторское издание, 2007. – 67 с.*
3. Зайцев, В.М. *Прикладная медицинская статистика [Текст] / В.М. Зайцев, В.Г. Лифляндский, В.И. Маринкин. – СПб.: Фолиант, 2006. – 432 с.*
4. *Методические рекомендации по оперативному анализу и прогнозированию эпидемиологической ситуации по гриппу и острым респираторным инфекциям (ОРВИ) [Текст]. – М., СПб.: ГУ НИИ гриппа Российской АМН, 2006. – 72 с.*
5. Радивоненко, О.С. *Оброблення невизначеності в медичних часових рядах для підвищення ефективності прогнозування [Текст] / О.С. Радивоненко, Т.В. Корчак, А.Н. Скаковська. – Радіоелектронні і комп'ютерні системи. – 2010. – № 3(44). – С. 50 – 55.*

Поступила в редакцію 15.06.2012

**Рецензент:** канд. техн. наук, доцент, доцент кафедри інформатики М.С. Мазорчук, Национальный аэрокосмический университет им. Н.Е. Жуковского “ХАИ”, Харьков.

## ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ АНАЛІЗУ ЧАСОВИХ РЯДІВ БІОМЕДИЧНИХ ДАНИХ

*Т.В. Корчак*

Робота присвячена створенню автоматизованого робочого місця для лікаря-епідеміолога. Розроблена інформаційна технологія вирішує три проблеми: оцінка ступеня напруженості території по захворюваності; обчислення епідемічних порогів; прогнозування розвитку епідемічного процесу. Розглянуто методи математичної статистики для аналізу часових рядів біомедичних даних. Вивчено проблеми обробки часових рядів в теорії біомедичних даних в умовах невизначеності. Створена інформаційна технологія епідеміологічного нагляду для аналізу часових рядів біомедичних даних. Описано переваги автоматизації робочого місця епідеміолога. Подано опис інформаційного забезпечення, що здійснює накопичення, переробку та аналіз даних, що допоможе працівникам СЕС ефективніше і точніше проводити збір інформації.

**Ключові слова:** прогнозування, біомедичні дані, інформаційна технологія, епідеміологічний нагляд, епідемічний процес, часовий ряд.

## INFORMATION TECHNOLOGY OF TIME SERIES ANALYSIS OF BIOMEDICAL DATA

*T.V. Korchak*

The work is devoted to creating a workstation for the physician-epidemiologist. The developed information technology solves three problems: evaluation of disease tension degree on the territory, the calculation of epidemic thresholds, epidemic process forecasting. The mathematical statistics methods for time series analysis of biomedical data are considered. The problem of time series processing in the theory of biomedical data is studied. The advantages of workplace automatization of epidemiologist are given. The information management description that carries out the storage, processing and analyzing data is presented.

**Keywords:** forecasting, biomedical data, information technology, epidemiological surveillance, epidemic process, time series.

**Корчак Татьяна Викторовна** – ассистент каф. информатики, Национальный аэрокосмический университет им. Н.Е. Жуковского “ХАИ”, Харьков, Украина, e-mail: kotavi@i.ua.