

СРАВНИТЕЛЬНЫЙ АНАЛИЗ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ МЕДИКО-ДЕМОГРАФИЧЕСКИХ ПОКАЗАТЕЛЕЙ

В. А. Голуб, Н. А. Гладских, С. С. Пронин

Воронежский государственный университет

Воронежская государственная медицинская академия им. Н. Н. Бурденко

В статье представлен сравнительный анализ результатов работы программ прогнозирования медико-демографических показателей, реализующие регрессионный анализ, корреляционный анализ, экспоненциальное сглаживание, а также модель множественной регрессии.

1. ВВЕДЕНИЕ

Прогнозирование динамики медико-демографических данных является важной актуальной задачей, так как представляет собой основу планирования и управления в здравоохранении.

В настоящее время разработаны программные продукты, позволяющие осуществлять прогнозирование статистических показателей. Однако большинство из этих программ сложны в использовании, не ориентированы на недостаточно подготовленного пользователя и требуют значительных затрат для обучения работы с ними.

В настоящей работе рассматриваются два программных продукта, предназначенных для прогнозирования медико-демографических показателей, разработанных в расчете на пользователя, не имеющего специальной подготовки.

2. ПРОВЕДЕНИЕ СРАВНИТЕЛЬНОГО АНАЛИЗА ПРОГРАММ

Первая из анализируемых программ — программа прогнозирования медико-демографических данных «ПРОГНОЗ 1» отличается простым дружественным пользователю интерфейсом, совместимостью с электронными таблицами Excel, что обеспечивает дополнительные удобства при вводе данных.

Алгоритм этой программы базируется на оригинальной методике, представляющей комбинацию методов, основанных на множественной регрессии, экспоненциальном сглаживании и корреляционном анализе [1, 2]. Алгоритм предполагает выполнение следующей последовательности действий:

1. Задаются n значений показателей, необходимо получить прогнозные значения для n -го показателя на заданный временной интервал.

2. Производится расчет попарных коэффициентов корреляции $n-1$ значений с n -м.

3. Выбираются пара параметров, имеющих наибольшую по модулю корреляцию.

4. По выбранным параметрам рассчитываются коэффициенты множественной регрессии по следующим формулам:

$$reg_B1 = \frac{(k_{xy}(x_1, y) - k_{xy}(x_2, y) \cdot k_{xy}(x_1, x_2)) \cdot S_y^2}{S_{x_1}^2 (1 - k_{xy}^2(x_1, x_2))}$$

$$reg_B2 = \frac{(k_{xy}(x_2, y) - k_{xy}(x_1, y) \cdot k_{xy}(x_1, x_2)) \cdot S_y^2}{S_{x_2}^2 (1 - k_{xy}^2(x_1, x_2))} \quad (1)$$

$$reg_A = \bar{y} - reg_B1 \cdot \bar{x}_1 - reg_B2 \cdot \bar{x}_2$$

$$Y = reg_A + reg_B1 \cdot X_1[i] + reg_B2 \cdot X_2[i] \quad (2)$$

где $k_{xy}(x_2, y)$ - корреляционный момент.

5. Строится прогноз коррелируемых параметров на заданный временной интервал с помощью метода экспоненциального сглаживания.

6. После расчетов коэффициентов уравнения множественной регрессии, строится прогноз на заданное число временных интервалов для последнего n показателя.

Программа предоставляет возможность задания периода, на который необходимо получить прогноз, обеспечивается графическое представление прогнозируемого показателя и коррелирующих факторов, имеются возможности работы, как в автоматическом режиме, так и в режиме, позволяющем пользователю осуществлять выбор независимых показателей.

Алгоритм второй программы «ПРОГНОЗ 2» базируется на методике построения оптимальной регрессионной модели. Регрессионное мо-

делирование проводится с помощью метода аналитического выравнивания с использованием наиболее распространенных аппроксимирующих функций. Оптимальная модель подбирается с помощью метода наименьших квадратов так, что сумма квадратов отклонений реальных и выровненных данных должна быть минимальной [1,3,4,5,7,8]. В ходе работы программы рассчитываются коэффициенты уравнения регрессии и выбирается оптимальная из нескольких предложенных регрессионная модель по критерию минимальности коэффициентов расхождения и вариации [6,7,8]:

1. Линейная модель $y = a_1 + a_2 \cdot x$.

$$\begin{cases} a_1 \cdot n + a_2 \cdot \sum x = \sum y, \\ a_1 \cdot \sum x + a_2 \cdot \sum x^2 = \sum y \cdot x, \end{cases}$$

2. Квадратичная модель $y = a_1 + a_2 \cdot x + a_3 \cdot x^2$.

$$\begin{cases} a \sum x^4 + b \sum x^3 + c \sum x^2 = \sum x^2 y \\ a \sum x^3 + b \sum x^2 + c \sum x = \sum x y \\ a \sum x^2 + b \sum x + nc = \sum y \end{cases}$$

3. Логарифмическая модель $y = a_1 + a_2 \cdot \ln x$.

$$\begin{cases} a_1 \cdot n + a_2 \cdot \sum \ln x = \sum y \\ a_1 \cdot \sum \ln x + a_2 \cdot \sum (\ln x)^2 = \sum y \cdot \ln x \end{cases}$$

4. Гиперболическая модель $y = a_1 + \frac{a_2}{x}$.

$$\begin{cases} a_1 \cdot n + a_2 \cdot \sum (1/x) = \sum y \\ a_1 \cdot \sum (1/x) + a_2 \cdot \sum (1/x)^2 = \sum (y/x) \end{cases}$$

5. Экспоненциальная модель $y = \exp(a_1 + a_2 \cdot x)$.

$$\begin{cases} a_1 \cdot n + a_2 \cdot \sum x = \sum \ln y \\ a_1 \cdot \sum x + a_2 \cdot \sum x^2 = \sum \ln y \cdot x \end{cases} \quad [7, 8].$$

Следует отметить, что метод, основанный на регрессионном анализе, допуская сравнительно несложную реализацию, является достаточно эффективным при работе с медико-демографическими данными [1].

В качестве объектов прогнозирования при сравнительном анализе работы программ использовались базовые медико-демографические показатели: общая и первичная заболеваемость взрослого и детского населения, рождаемость, смертность, численность населения, обеспеченность врачами, общий уровень госпитализации, число амбулаторных посещений, число койко-дней и др.

В данной статье в качестве примера был использован показатель обеспеченности врачами по Воронежской области в целом. Исходные данные за 1994—2006 гг., полученные из справки Облздора, приведены в таблице 1.

Результаты прогнозирования показателя обеспеченности врачами на период с 2007 года по 2012 год, полученные с помощью программы «ПРОГНОЗ 1» представлены на рис. 1 и в табл. 2. Аналогичные результаты, полученные с помощью программы «ПРОГНОЗ 2», приведены на рис. 2 и в табл. 3 соответственно.

На следующих этапах данного исследования был проведен сравнительный анализ эффективности представленных программ. В основу методики сравнительного анализа заложен следующий алгоритм.

Таблица 1

Данные по обеспеченности врачами по Воронежской области

Год	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006
Показатель	39,6	40,5	41,5	42,2	42,8	42,8	42,7	43,8	43,9	44,3	44,6	44,9

Таблица 2

Результаты прогнозирования показателя обеспеченности врачами, полученные по программе «ПРОГНОЗ 1»

Год	2006	2007	2008	2009	2010	2011	2012
Область Врач	53,56	54,01	54,47	54,92	55,38	55,83	56,29

Таблица 3

Результаты прогнозирования показателя обеспеченности врачами, полученные по программе «ПРОГНОЗ 2»

Год	2006	2007	2008	2009	2010	2011	2012
Область Врач	44,63	44,73	44,79	44,78	44,73	44,62	44,46

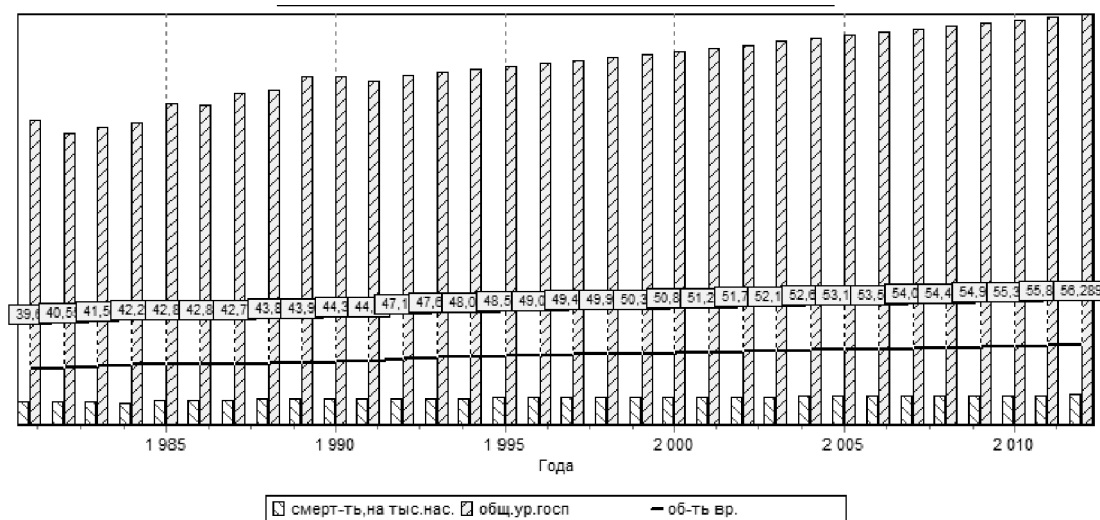


Рис. 1. Графическое представление результатов прогнозирования показателя обеспеченности врачами, полученные по программе «ПРОГНОЗ 1»

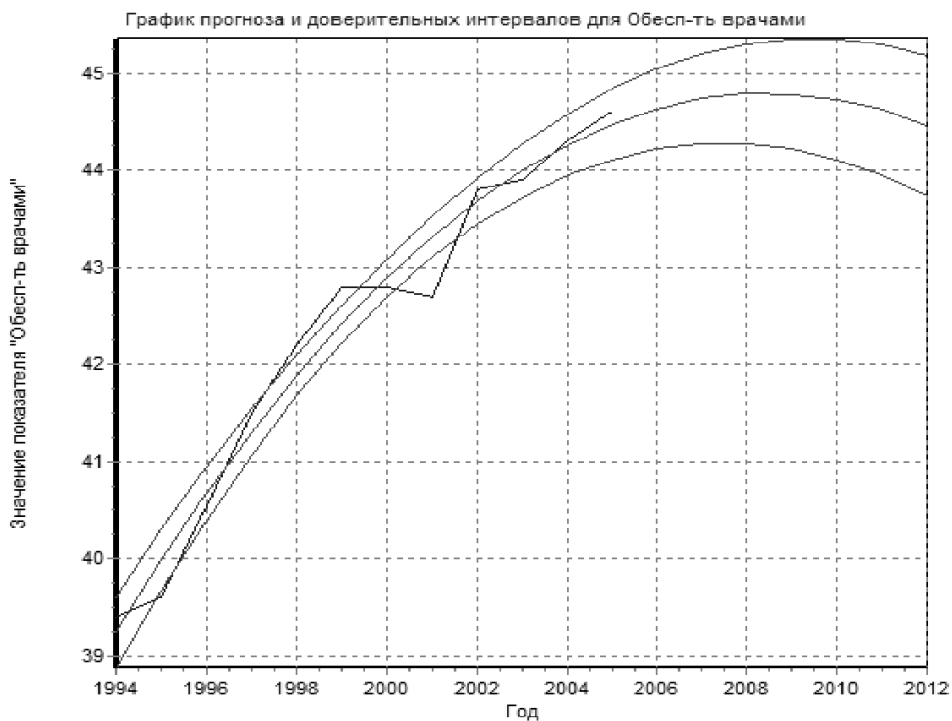


Рис. 2. Графическое представление результатов прогнозирования показателя обеспеченности врачами, полученные по программе «ПРОГНОЗ 2»

1. В качестве исходной информации на первом этапе используются данные по обрабатываемому показателю за 1994—1999 гг. Длина интервала равна 5 годам.

2. Строится прогноз на 3 года, т.е. на 2000—2002 гг.

3. Длина интервала увеличивается на один год, и этапы 1—2 повторяются, т.е. в качестве исходной информации используются данные за 1994—2000 гг. Прогноз строится уже на 2001—2003 гг.

4. Используются данные за 1994—2003 гг. Прогноз строится на 2004—2006 гг.

5. Рассчитывается ошибка прогнозирования на 1 год, на 2 года и на 3 года по следующей формуле:

$$\text{ОШИБКА_ПРОГНОЗА} = \frac{|\text{ФАКТИЧ.ЗНАЧЕНИЕ} - \text{ПРОГНОЗ.ЗНАЧЕНИЕ}|}{\text{ФАКТИЧ.ЗНАЧЕНИЕ}} \cdot 100 \%$$

Результаты рассчитанных ошибок прогнозирования по результатам работы программ «ПРОГНОЗ 1» и «ПРОГНОЗ 2» приводятся в таблице 4.

Сравнительный анализ точности прогнозирования показателя обеспеченности врачами по результатам работы программ «ПРОГНОЗ 1» и «ПРОГНОЗ 2» представлен на рисунке 3.

3. ЗАКЛЮЧЕНИЕ

Сравнивая ошибки прогнозирования, приведенные в таблице 4, а также графики зависимостей ошибки прогноза от длины интервала, представленные на рисунке 3, следует сделать вывод о том, что программа прогнозирования медико-демографических показателей «ПРО-

ГНОЗ 2», базирующаяся на регрессионном моделировании с использованием аналитического выравнивания на основе наиболее распространенных аппроксимирующих функций, дает намного более точные результаты, чем программа, в алгоритм которой заложен метод множественной регрессии. Объяснением этого факта может служить то, что множественная регрессия показывает общую тенденцию, являясь линейной функцией, тогда как программа «ПРОГНОЗ 2» выбирает из заданных функций ту, которая наиболее точно описывает эмпирические данные.

ЛИТЕРАТУРА

1. Руководство по статистике в медицине и биологии. В 2-х томах/под ред. проф. Ю. М. Комарова. Том 2. Прикладная статистика здоровья / В. А. Медик, Б. Б. Фишман, М.С. Токмачев — М.: Медицина, 2001. — 352 с.

2. Юнкеров В.И. Математико-статистическая обработка данных медицинских исследований/ В. И. Юнкеров, С. Г. Григорьев. — СПб.: ВМедА, 2002. — 266 с.

3. Шиган Е.Н. Методы прогнозирования и моделирования в социально-гигиенических исследованиях/ Е. Н. Шиган. — М.: Медицина, 1986. — 208 с.

4. Вальвачев Н.И. Статистический метод в медицинской практике с применением микроЭВМ и персональных компьютеров/ Н. И. Вальвачев, М. И. Римжа. — Мн.: Беларусь, 1989. — 112с.

5. Тюрин Ю.Н. Статистический анализ данных на компьютере/ Ю. Н. Тюрин, А. А. Макаров; под ред. В. Э. Фигурнова. — М.: ИНФРА-М, 1998. — 528 с.

6. Боровиков В.П. Прогнозирование в системе STATISTICA в среде Windows. Основы теории и интенсивная практика на компьютере/ В. П. Боровиков, Г. И. Ивченко — М.: Финансы и статистика, 1999. — 384 с.

Таблица 4

Ошибки прогноза по результатам программ «ПРОГНОЗ 1» и «ПРОГНОЗ 2» в зависимости от длины интервала.

	Длина интервала	5	6	7	8	9
ПРОГНОЗ 1	ош. прогноза на 1 г.			5,38 %	4,63 %	0,62 %
ПРОГНОЗ 2	ош. прогноза на 1 г.	0,21 %	1,45 %	0,27 %	0,23 %	0,09 %
ПРОГНОЗ 1	ош. прогноза на 2 г.			5,59 %	5,41 %	0,32 %
ПРОГНОЗ 2	ош. прогноза на 2 г.	1,45 %	0,27 %	0,23 %	0,09 %	0,29 %
ПРОГНОЗ 1	ош. прогноза на 3 г.			6,44 %	5,97 %	0,03 %
ПРОГНОЗ 2	ош. прогноза на 3 г.	0,27 %	0,23 %	0,09 %	0,29 %	0,60 %
ПРОГНОЗ 1	ош. прогноза на 4 г.			7,07 %	6,52 %	
ПРОГНОЗ 2	ош. прогноза на 4 г.	0,23 %	0,09 %	0,29 %	0,60 %	
ПРОГНОЗ 1	ош. прогноза на 5 л.			7,69 %		
ПРОГНОЗ 2	ош. прогноза на 5 л.	0,09 %	0,29 %	0,60 %		

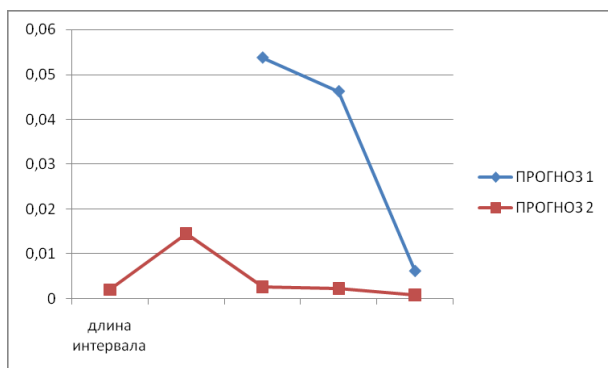


График зависимости ошибки прогноза на 1 год (в %) от длины интервала

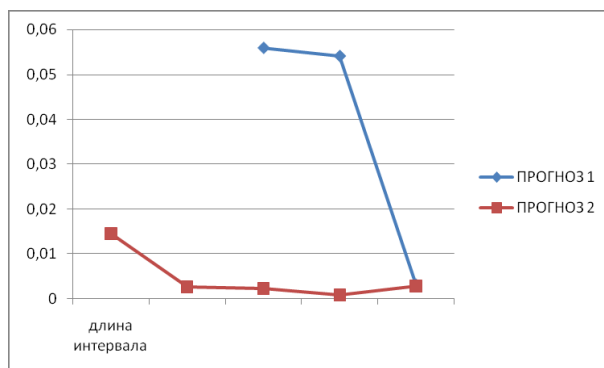


График зависимости ошибки прогноза на 2 года (в %) от длины интервала

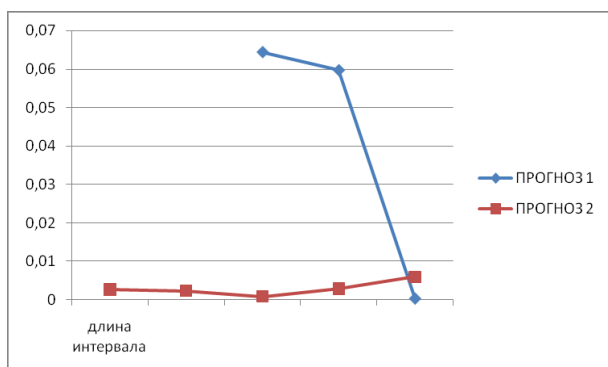


График зависимости ошибки прогноза на 3 года (в %) от длины интервала

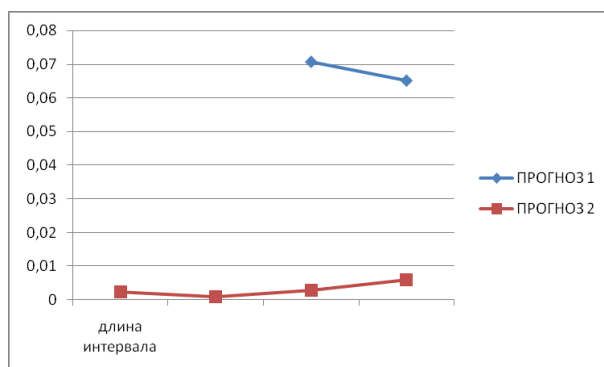


График зависимости ошибки прогноза на 4 года (в %) от длины интервала

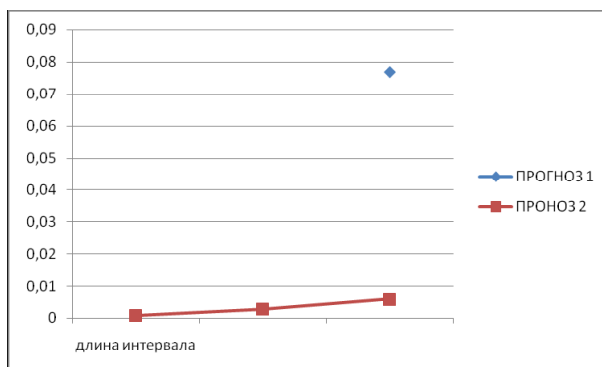


График зависимости ошибки прогноза на 5 лет (в %) от длины интервала

Рис. 3. Графики зависимости ошибки прогноза от длины интервала, полученных по результатам работы программ «ПРОГНОЗ1» и «ПРОГНОЗ2»

7. Голуб В.А. Программное обеспечение для прогнозирования динамики изменения демографических данных / В. А. Голуб, С. Н. Семенов, Н. А. Гладских / Системные проблемы надежности, качества, информационных и электронных технологий // Материалы X Международной конференции и Российской научной школы. — М.: Радио и связь, 2005. — Часть 5. — С. 124—128.

8. Есауленко И.Э. Разработка регрессионных моделей для прогнозирования динамики медико-демографических показателей / И. Э. Есауленко [и др.] // Системный анализ и управление в биомедицинских системах. — 2005. — Т. 4. — № 1. — С. 104—107.

Статья принята к опубликованию
25 декабря 2006 г.