

УДК 004.89

Д.М. Сапицкая, С.В. Хмелевой

Донецкий национальный технический университет г.Донецк
кафедра автоматизированных систем управления

ПРОГНОЗИРОВАНИЕ КОЛИЧЕСТВА ТРАВМ С ИСПОЛЬЗОВАНИЕМ МЕТОДА БОКСА-ДЖЕНКИНСА

Аннотация

Сапицкая Д.М., Хмелевой С.В. Прогнозирование количества травм с использованием метода Бокса-Дженкинса. Выполнен анализ входных данных. Получены значения сезонной составляющей. Вычислены значения первых и вторых разностей. Получены прогнозируемые данные. Выполнен анализ полученных данных.

***Ключевые слова:** прогнозирование, искусственный интеллект, метод ARIMA, метод Бокса-Дженкинса*

Постановка проблемы. Прогнозирование количества травм – задача, которой на данный момент занимаются в основном с целью повышения качества охраны труда на предприятии. Основной задачей прогнозирования травматизма является выявление закономерностей изменения и конкретной величины статистических его показателей в будущем, а также определение неблагоприятных тенденций основных показателей (травмирующих факторов) для принятия плановых решений.

Входными данными для анализа служат показатели количества обращений в Областную травматологию г. Донецка с 2004 по 2012 гг. Данные разделены по месяцам и представляют собой временной ряд.

Описание метода прогнозирования. Подход Бокса-Дженкинса является одним из лучших методов, позволяющих нам понять и прогнозировать временные ряды. Он имеет вид:

1. В семействе ARIMA-процессов Бокса-Дженкинса выбирается достаточно простой процесс, позволяющий получить данные, которые в целом выглядят примерно так же, как наш ряд (за исключением фактора случайности). Для этого необходимо выбрать конкретный тип модели и оценить требуемые параметры на основе своих данных. Из результирующей модели можно узнать в какой мере каждое наблюдение влияет на будущее и в какой мере каждое наблюдение содержит полезную новую информацию, позволяющую прогнозировать будущее.

2. Прогноз на любой момент времени представляет собой ожидаемое (т.е. среднее) будущее значение оцениваемого процесса в этот момент времени. Формула для прогноза позволяет быстро вычислить среднее значение для множества вариантов поведения ряда, начиная с исходных данных и экстраполируя их в соответствии с выбранной моделью.

3. Стандартная ошибка прогноза для любого момента времени представляет собой стандартное отклонение от всех возможных (допустимых) будущих значений для этого времени [1].

4. Границы прогноза простираются выше и ниже прогнозируемого значения так, что (если выбранная модель оказалась правильной) с вероятностью, например, 95% можно утверждать, что будущее значение для любого момента времени уложится в указанные границы прогноза. Эти границы прогноза формируются таким образом, чтобы для каждого будущего периода времени 95%, возможных (и допустимых) вариантов будущего поведения ряда укладывались в эти границы [2].

ARIMA-процессы Бокса-Дженкинса представляют собой семейство линейных статистических моделей, основанных на нормальном распределении, которые позволяют имитировать поведение множества различных временных рядов путем комбинирования процессов авторегрессии, интегрирования и скользящего среднего.

Процесс случайного шума состоит из случайной выборки (независимых наблюдений) из нормального распределения с постоянным средним и стандартным отклонением. Какие-либо тенденции (тренды) в этом случае отсутствуют, поскольку – по причине независимости – наблюдения не помнят о прошлом поведении ряда.

В соответствии с моделью случайного шума в момент времени t наблюдаемые данные, Y_t , будут состоять из константы, μ (долгосрочное среднее процесса), плюс случайный шум, ε_t , с 0-м средним значением.

Модель процесса случайного шума:

Данные = среднее значение + случайный шум

$$Y_t = \mu + \varepsilon_t \quad (1)$$

Долгосрочное среднее значение Y равно μ .

Любое наблюдение процесса авторегрессии представляет собой линейную функцию от предыдущего наблюдения плюс случайный шум. Процесс авторегрессии помнит о предыдущем состоянии и использует эту информацию для определения своего дальнейшего поведения.

Модель процесса авторегрессии:

Данные = $\delta + \varphi$ (предыдущее значение) + случайный шум

$$Y_t = \delta + \varphi Y_{t-1} + \varepsilon_t \quad (2)$$

Долгосрочное среднее значение Y равно $\delta / (1 - \varphi)$.

Процесс случайного среднего не помнит в точности своего прошлого, но помнит компонент случайного шума того состояния, в котором он (процесс) находился. Его память ограничена одним шагом в будущее; за пределами этого шага для процесса все начинается заново.

Модель процесса случайного среднего:

Данные = μ + случайный шум – θ (предыдущий случайный шум)

$$Y_t = \mu + \varepsilon_t - \theta \varepsilon_{t-1} \quad (3)$$

Долгосрочное среднее значение Y равно μ .

Процесс авторегрессии и скользящего среднего запоминает как свое предыдущее состояние, так и компонент случайного шума предыдущего состояния. Таким образом, его память сочетает в себе память процесса авторегрессии с памятью процесса скользящего среднего [3].

Если изменения или разности в ряде вырабатываются процессом авторегрессии скользящего среднего, то сам этот ряд соответствует процессу авторегрессионного интегрированного скользящего среднего (ARIMA)(autoregressive integrated moving-average process). Этот процесс знает, где он находится, помнит, как он попал в это состояние и помнит даже часть предыдущего шумового компонента. Следовательно, ARIMA-процесс можно использовать в качестве модели для совокупностей данных временного ряда, которые являются очень гладкими, с медленными изменениями направления. Эти ARIMA-процессы являются нестационарными из-за включения в них интегрированного компонента.

Модель процесса авторегрессионного интегрированного скользящего среднего (ARIMA) в разностной форме:

$$Y_t - Y_{t-1} = \delta + \varphi(Y_{t-1} - Y_{t-2}) + \varepsilon_t - \theta\varepsilon_{t-1} \quad (4)$$

Долгосрочное среднее значение изменения в Y равно $\delta/(1 - \varphi)$. Нельзя рассчитывать, что с течением времени Y останется достаточно близким к какому-нибудь долгосрочному среднему значению.

Практические результаты. Были получены значения линейного тренда и сезонной составляющей, рассчитаны значения тренда с поправкой на сезонность. Результаты представлены на рис. 1.

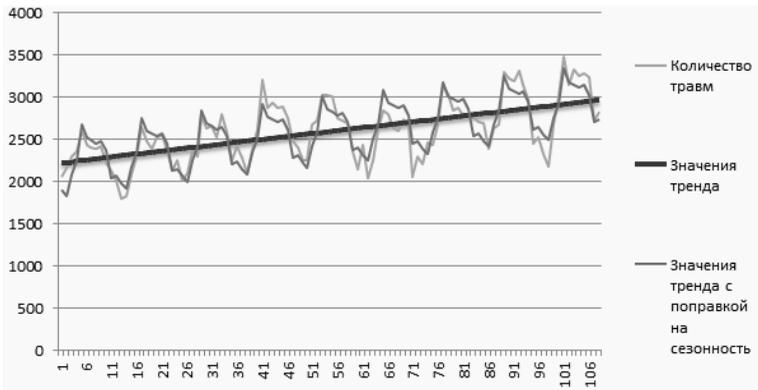


Рисунок 1 – Визуализация полученных результатов

В качестве проверочных значений для оценки качества прогноза были выбраны данные за 2012 год.

Таблица 1 – Практические значения на 2012 год

Количество грамм	Значения линейного тренда	Сезонная составляющая	Значения тренда с поправкой на сезонность	Погрешность
2327	2884,8702	-328,278	2556,592356	229,5924
2179	2891,7668	-397,174	2494,592356	315,5924
2679	2898,6634	-150,738	2747,925689	68,92569
3006	2905,56	8,032356	2913,592356	-92,4076
3476	2912,4566	417,5802	3330,0368	-145,963
3142	2919,3532	257,4614	3176,814578	34,81458
3325	2926,2498	220,4537	3146,703467	-178,297
3247	2933,1464	180,8904	3114,0368	-132,963
3273	2940,043	201,5494	3141,592356	-131,408
3225	2946,9396	78,98609	3025,925689	-199,074
2713	2953,8362	-254,133	2699,703467	-13,2965
2825	2960,7328	-235,14	2725,592356	-99,4076

Результаты применения метода ARIMA представлены в табл. 2 в виде числовых значений и на рис. 2 в виде графического представления практических и прогнозируемых данных.

Таблица 2 – Прогнозируемые значения на 2012 год

CaseNo.	Forecast	Lower 90,0%	Upper 90,0%	Observed	Residual
97	2491,924	2243,929	2767,327	2327	-164,924
98	2407,216	2143,751	2703,059	2179	-228,216
99	2618,239	2308,212	2969,906	2679	60,761
100	2784,018	2431,492	3187,654	3006	221,982
101	3246,224	2810,459	3749,555	3476	229,776
102	3129,217	2686,893	3644,358	3142	12,783
103	3048,816	2597,444	3578,626	3325	276,184
104	3040,836	2571,383	3595,996	3247	206,164
105	3025,254	2539,993	3603,222	3273	247,746
106	2907,483	2424,415	3486,803	3225	317,517
107	2553,412	2115,136	3082,503	2713	159,588
108	2570,667	2115,866	3123,226	2825	254,333

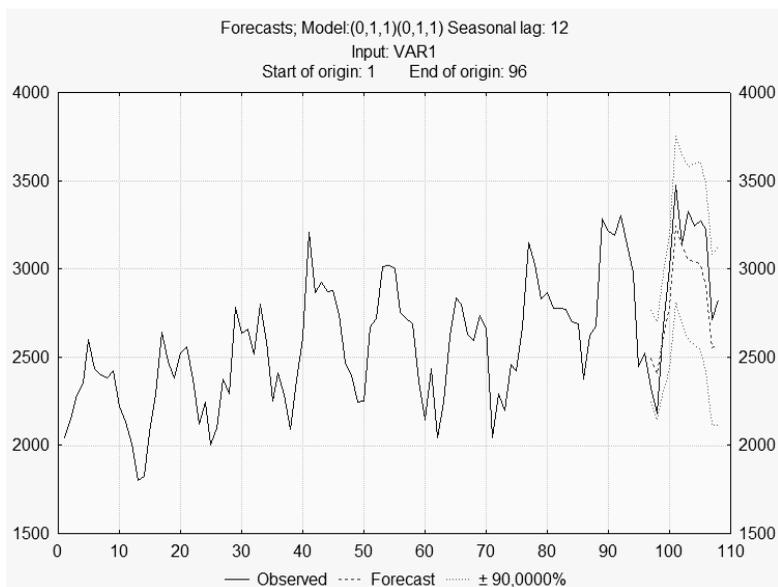


Рисунок 2 – Визуализация практических и прогнозируемых значений на 2012 год

Выводы. Произведен анализ имеющихся данных. На их основании получены значения линейного тренда и сезонной составляющей. Рассчитаны значения тренда с поправкой по сезонности. Получены прогнозируемые результаты. Анализ полученных данных и визуальная оценка позволяют судить о правильности полученных результатов.

Список литературы

1. Сигел Э. Практическая бизнес-статистика.: Пер. с англ. – М.: Издательский дом «Вильямс», 2002. – 1056 с.
2. Афанасьев В.Н. Эконометрика в пакете STATISTICA: учебное пособие по выполнению лабораторных работ / В.Н. Афанасьев, А.П. Цыпин, - Оренбург: ГОУ ОГУ, 2008. – 204с.
3. Кильдышев, Г.С., Френкель, А.А. Анализ временных рядов и прогнозирование. – М.: Статистика. 1973.
4. Боровиков, В.П. STATISTICS. Искусство анализа данных на компьютере : Дин профессионалов 2-е изд (+CD) – СПб.: Питер. 2003. – 668 с.