

УДК 004.622

Анализ данных для компьютеризированной системы прогнозирования уровня преступности

А.В. Сложеницын, С.Ю. Землянская
ГОУВПО «Донецкий национальный технический университет»
SlogenitsinAV@gmail.com, zsaa@ya.ru

Сложеницын А.В., Землянская С.Ю. Анализ данных для компьютеризированной системы прогнозирования уровня преступности. Определен наиболее оптимальный регион для поиска и анализа данных, теоретически характеризующих преступность. Выделены направления поиска показателей. Найден, сгруппирован и приведен к относительному виду набор параметров, теоретически характеризующий уровень преступности. Произведен корреляционный и множественный регрессионный анализ для определения зависимости между показателями-факторами и уровнем преступности. На основе анализа выбраны данные для прогнозирования преступности.

Ключевые слова: криминогенная обстановка, уровень преступности, корреляционный анализ, регрессионный анализ значимость факторов.

Введение

Прогнозирование всегда являлось неотъемлемой частью человеческой жизни. Все поступки человек совершает в результате прогноза «если я что-то сделаю, то что-то получу». И всегда это только вероятность. Человек не может с полной уверенностью сказать, что, придя на встречу, он встретится с тем, с кем собирался, или что если он устроится на работу, то получит желаемый результат. Но без планирования и доверия к своим маленьким прогнозам, человечество никогда бы ничего не добилось.

В наши дни, учитывая бурное развитие науки, техники, производства и общества в целом, роль прогнозирования возросла еще больше. Человечество все больше нуждается в стабильности, которую дает прогнозирование. Все больше аспектов повседневной жизни могут быть улучшены с помощью современных средств прогнозирования, которые ответят, чего ждать в этой области завтра.

Сегодня одной из проблемных областей прогнозирования является криминогенное прогнозирование или прогнозирование уровня преступности. Развитие этого направления идет медленными темпами. В первую очередь, это обусловлено сложностью прогнозирующей модели и большим количеством факторов, исследуемых для прогнозирования. В хорошей прогнозирующей модели должны корректно учитываться многие факторы, от которых может зависеть преступное поведение любого жителя города. Поэтому создание такого рода прогностических систем – довольно тяжелый

процесс.

Но все же криминогенное прогнозирование имеет большой потенциал. На основании данного прогноза возможно повышение эффективности деятельности правоохранительных органов, рациональное использование ресурсов для предотвращения преступной деятельности, сосредоточение внимания на наиболее актуальных проблемах, связанных с преступностью.

Первым шагом для прогнозирования криминогенной обстановки является поиск и анализ входных данных. Важно не только найти достоверные и точные данные, но и привести их к одной размерности, а также провести анализ их реальной значимости для прогнозирования преступлений.

Регион для поиска данных

Криминологические исследования последних лет выявили устойчивые различия в показателях преступности (уровень, структура, динамика) в различных регионах России. В частности, в диссертации Абызова К.Р. [1], доказываемая зависимость различных показателей преступности от региона. На основе этого был сделан вывод о целесообразности анализа отдельного региона, где существуют свои специфические факторы, детерминирующие преступность.

Наибольший интерес представляет анализ показателей для Донецкой области. Но в силу того, что на данной территории в последние годы фактически ведутся активные военные действия, достоверных данных по этой

территории нет. Этот вывод сделан на основании суждений, что во время военного конфликта не ведется полномерный сбор данных, а небольшие собранные данные не отражают действительность и преобразовываются для мнимого улучшения показателей жизни населения. Также во время военного конфликта изменяются модель поведения человека и система его жизненных приоритетов, вследствие чего применение собранных данных для прогнозирования на бесконфликтной территории будет иметь существенную погрешность.

Учитывая вышесказанное, для проводимого в дипломной работе анализа был выбран ближайший к Донецкой области регион Российской Федерации – Ростовская область, входящая в состав Южного федерального округа. Для Донецкой и Ростовской областей, помимо территориальной близости, характерна общность факторов экономического, социального, демографического, исторического и культурного порядка.

Искомые данные брались из открытых источников. Основные ресурсы, использованные для этого, – это базы данных, таблицы статистических данных и публикуемые справочники Федеральной Службы Государственной Статистики [2], Судебного департамента при Верховном Суде Российской Федерации [3], НИИ при Федеральной службе исполнительных наказаний Российской Федерации [4], Генеральной Прокуратуры Российской Федерации [5], Аналитического центра Юрия Левады [6], Социологической службы Фонд «Общественное мнение» [7], а также данные переписи населения [8].

Направления поиска

В учебном пособии «Психология преступника и расследования преступлений» указано: «Для криминологии стало аксиомой утверждение, что причиной преступного поведения является сложное взаимодействие внешних, объективных условий и внутренних, субъективных факторов, то есть среды и личности. Тем самым среда и личность получают равную криминологическую оценку, а поведение предстает лишь как следствие этого взаимодействия» [9]. Поэтому было принято решение учитывать и изучать не только статистические данные о лицах, совершающих преступления, но и данные, характеризующие состояние социальных явлений и процессов, так или иначе влияющих на преступность.

Таким образом, поиск показателей проводился в двух направлениях: поиск социальных показателей, которые можно сравнить с показателем уровня преступности, а

также поиск характеристик осужденных преступников, которые возможно сопоставить с этими же характеристиками у населения.

В первом случае найденные показатели впоследствии анализировались на предмет их связи с уровнем преступности. Если данные изменяются соизмеримо с уровнем преступности, то такие данные считаются значимыми для преступности в регионе и будут использоваться при построении модели.

Во втором случае поиск осуществлялся с расчетом на возможность сопоставления конкретных данных о преступниках с этими же данными у всего населения. Анализ зависимостей в этом случае сводился к анализу пар показателей: для населения и для преступников. Это более значимый анализ, так как при составлении портрета преступника или прогнозировании преступной деятельности, зависимость конкретных характеристик преступников от характеристик общества наиболее показательна.

Для обоих направлений поиск производился по годам с 2010 по 2015 включительно. Данные по 2016 году еще не представлены в полном объеме и появятся только осенью. Такой временной диапазон был выбран как наиболее приближенный к сегодняшним показателям. Поиск данных по годам был необходим, чтобы проследить динамику изменения показателей и впоследствии произвести проверку значимости отдельных показателей.

Большая часть данных для анализа была взята для Ростовской области. Однако, некоторая часть анализируемых данных была доступна исключительно для Южного федерального округа или Российской Федерации в целом. В частности, данные о характеристиках преступников представлены для Южного федерального округа, а для Ростовской области этих данных значительно меньше. Это не противоречит условиям анализа данных, так как данные были приведены к виду процентного соотношения и сравнивались с территориально соответствующими показателями преступности.

Начальная работа с данными

Насколько бы ни был качественно произведен поиск, всегда будет часть данных, которые отсутствуют или представлены в некорректном формате. Для решения этой проблемы были применены методы работы с отсутствующими данными:

– При отсутствии данных по региону (например, Ростовской области) за один год из шести, и наличии этих данных по более общему региону (например, Южному федеральному

округу), выявлялась зависимость изменения показателя по более общему региону. После этого, найденная зависимость использовалась для определения недостающих данных, на основе имеющихся данных по региону (за предыдущие или последующие периоды).

– При отсутствии данных за один год из шести как по региону, так и по более общему региону, находилось и использовалось среднее значение между предшествующим и последующим годами.

– При отсутствии показателя, который можно было рассчитать на основании других показателей, не учитываемых в анализе, производился расчет.

Для корректного анализа данных, большинство данных были приведены к процентному виду (удельному весу) либо отношению показателя к определенному количеству человек. Часть данных невозможно было представить в процентах или в относительном виде, поэтому эти данные были приведены к определенной единице измерения.

Для приведения показателей к относительным величинам, производился расчет, представленный на примере расчета процента «женщин, совершивших преступление». В этом примере, нам известно количество лиц, совершивших преступление (X), и количество женщин, совершивших преступление (Y). Для расчета относительного количества женщин (W), среди лиц, совершивших преступление, применим формулу (1).

$$W = Y/X * 100 \quad (1)$$

Часть данных была рассчитана на основании не используемых в анализе данных. Например, процент людей, имеющих судимость. Идея данного расчета была взята из статьи газеты Коммерсант от 22.09.2008: «По данным судебного департамента при Верховном суде РФ, в 2007 году осуждены 931 тыс. человек. При этом, по статистике МВД, 29% всех преступников являются рецидивистами, следовательно, первую в жизни судимость получили 661 тыс. человек. По некоторым экспертным оценкам, средний возраст российского преступника – 28 лет. Средняя продолжительность жизни – 67 лет. Получается, что поколение преступников полностью обновляется раз в 39 лет. За это время 25,8 млн граждан впервые преступят закон. Это число при допущении, что уровень преступности и численность населения — постоянные величины, и является количеством судимых граждан. Таким образом, уголовное прошлое имеют 18,2% населения: почти каждый второй взрослый мужчина или почти каждый шестой россиянин,

включая младенцев» [10].

Согласно расчетам, приведенным в этом источнике, приблизительный процент ранее судимых людей можно рассчитать, зная количество осужденных в этом году, количество рецидивистов среди осужденных, средний возраст преступника, среднюю продолжительность жизни. Результаты расчетов можно занести в показатель «Число людей, имеющих судимость».

Описание данных

После поиска и обработки данных для анализа, получили два набора данных. В первом наборе представлены характеристики осужденных преступников, которые возможно сопоставить с этими же характеристиками у населения. Результаты поиска и анализа первого набора данных представлены в таблице 1.

Таблица 1 – Пример результатов поиска и попарного анализа характеристик обычных людей и осужденных

Показатель	2010	...	2015	Корреляция
Имеют полное общее образование	20,1	...	20,8	0,2838
Количество осужденных имеющие полное общее образование	25,1	...	25,2	
Средний доход населения	15114	...	27004	0,9925
Средний доход осужденных	10652	...	17288	
Число людей, имеющих судимость	19,67		19,66	-0,9725
Число ранее судимых, совершивших преступление	23,68		25,59	
...	...	...	...	...

Как уже отмечалось, очень важно, помимо личности преступника, исследовать и социальную среду, в которой эта личность живет, воспитывается и в которой проявляет свое преступное поведение. Следует учитывать демографические, экономические, социальные и культурно-образовательные условия жизни людей в конкретном регионе.

Для анализа были выбраны такие демографические показатели Ростовской области, как половой состав населения (на основании численности населения всего, а также отдельно мужчин и женщин рассчитан половой состав в процентном выражении), доля

населения разных возрастных категорий. Одним из показателей экономической ситуации в обществе является уровень безработицы. Чтобы оценить материальный уровень жизни населения, было принято решение рассмотреть доходы, приходящиеся на 1 человека в месяц и количество легковых автомобилей на 1000 населения. Так как в исследуемом периоде отмечались колебания цен, для более достоверного анализа среднедушевые доходы были проиндексированы с учетом индекса потребительских цен и приведены к уровню 2010 года. Условно индекс 2010 года принят за единицу, а каждый последующий год индексировался нарастающим итогом. Также отдельно была выделена доля людей с доходом ниже прожиточного минимума и доля семей, нуждающихся в жилье (поставлены на учет в качестве нуждающихся в жилых помещениях), так как неудовлетворительное материальное и жилищное положение может послужить предпосылкой противозаконных действий.

Как отмечают некоторые исследователи, на формирование личности преступника влияет так называемое социально-психологическое отчуждение, которое развивается чаще всего в результате эмоционального отвержения родителями детей [9]. В связи с этим для анализа были выбраны и такие показатели, характеризующие неблагополучную ситуацию внутри семьи, как количество разводов и численность детей, родители которых лишены родительских прав.

Можно предположить, что на уровень преступности, особенно среди молодежи, оказывает влияние получение образования, занятие спортом, участие в культурной жизни общества. То есть, чем больше у людей заполнен их досуг, тем менее криминогенная обстановка в обществе. Поэтому в данную работу включены данные о доле среди населения студентов, получающих среднее и высшее образование, а также объем предоставления услуг на 1 человека в сфере туризма, культуры и спорта (последний показатель также проиндексирован с учетом индекса потребительских цен).

Все вышеперечисленные данные были найдены в результате поиска социальных показателей, которые можно сравнить с показателем уровня преступности. Они представлены в таблице 2.

Таблица 2 – Пример результатов поиска и начального анализа социальных показателей

Показатель	2010	...	2015	Корреляция
число преступлений, на 100000 чел.	1386	...	1469	

доля трудоспособного населения	60,8	...	57,2	-0,2147
уровень безработицы	7,7	...	6,1	0,3709
количество людей, страдающих алкоголизмом (стоят на учете), на 1000 чел.	1367,1	...	1076,2	-0,234
...	...	...	...	...

Собранные данные являются неполным перечнем характеристик, имеющих влияние на преступность. В первую очередь, это вызвано тем, что многие данные отсутствуют в открытом доступе. В качестве примера можно привести форму «N 2 Статистическая карточка на лицо, совершившее преступление» [11], в которой приведено около 35 обязательных к заполнению характеристик преступника. Данная карточка является лишь одной из перечня заполняемых для осужденного документов в обязательном порядке.

Помимо обязательных документов, есть ряд документов, которые предлагается заполнить заключенным, для сбора статистической информации Федеральной Службой Государственной Статистики и иными службами. В таких карточках количество собираемой информации существенно больше. Однако все эти данные, на текущий момент, не являются открытыми.

Анализ данных

Анализ данных основан на корреляционном анализе или выявлении коэффициента корреляции, а также на множественной регрессии. Согласно определению, приведенному в учебном пособии для вузов «Корреляционный анализ»: «Корреляционная зависимость – стохастическая зависимость между случайными величинами, при которой наблюдается функциональная зависимость между значениями одной величины и средними значениями другой величины» [12].

Иначе говоря, определив коэффициент корреляции, можно сказать, имеют ли два показателя статистическую взаимосвязь. Важно заметить, что «наличие корреляционной зависимости между переменными не всегда означает наличие непосредственной связи этих величин друг с другом» [12, с.5]. То есть, даже при больших уровнях корреляции, параметры могут быть не связаны между собой.

Для определения взаимосвязи между несколькими переменными используется

множественная регрессия. Довольно точное, но, в то же время, простое для понимания определение множественной регрессии дается в психологической энциклопедии, авторства Р. Корсини: «Множественная регрессия – метод многомерного анализа, посредством которого зависимая переменная (или критерий) Y связывается с совокупностью независимых переменных (или предикторов) X посредством линейного уравнения [13]:

$$Y' = a + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

Исходя из определения, можно сделать вывод, что с помощью множественной регрессии строится уравнение зависимости одной переменной от других. В этом случае, в отличие от корреляционного анализа, можно утверждать, что построенные зависимости обоснованы для множества переменных, а не для одной.

Для определения коэффициентов корреляции, была использована стандартная функция прикладного программного обеспечения Microsoft Excel – «КОРРЕЛ». Эта функция вычисляет коэффициент корреляции Пирсона для двух наборов данных.

Коэффициент корреляции всегда расположен в диапазоне $[-1;1]$. Если этот коэффициент равен 0, то анализируемые параметры не связаны. Если он равен 1 или -1, то параметры связаны, причем знак указывает на прямую или обратную связь. При вещественных значениях коэффициента говорят о силе корреляционной связи.

Согласно данным учебно-методического пособия «Корреляционный анализ». Использование MS Excel для расчета коэффициента корреляции» [14], сила корреляционной связи разделена на три уровня:

- сильная: $\pm 0,7$ до ± 1 ,
- средняя: $\pm 0,3$ до $\pm 0,699$,
- слабая: 0 до $\pm 0,299$.

Для текущего исследования допустимым уровнем корреляции будем считать силу коэффициента корреляции больше $\pm 0,3$, что соответствует среднему и сильному уровню связи. Если два набора данных коррелированы с силой более указанной, то данные параметры будем считать значимыми и используем их в последующих расчетах.

В результате проведенного анализа можно сказать, что большинство показателей первого набора данных являются значимыми для преступности и будут использованы при построении модели. Некоррелированными и, соответственно, не значимыми, оказались 4 пары значений из 22.

Показатели второго набора данных менее зависимы друг от друга. Среди них лишь 8 наборов из 19 оказались коррелированными с

показателем преступности и будут использованы при последующем анализе. Остальные данные, согласно результатам корреляционного анализа, недостаточно связаны с уровнем преступности в обществе.

Для данных из второго набора данных недостаточно корреляционного анализа, ведь все данные сравниваются не попарно, а с одним параметром. Следовательно, часть данных может перекрываться другими данными или зависимость двух показателей может быть обоснована третьим показателем. Для более корректного вывода, для второго набора данных рассчитывается множественная регрессия.

Для анализа данных с помощью множественной регрессии, использовался пакет прикладных программных средств STATISTIKA, фирмы StatSoft.

В общем случае для регрессионного анализа необходимо, чтобы количество наблюдений было больше количества независимых переменных. В другом случае невозможно обработать корреляционную матрицу. При анализе второго набора данных, количество наблюдений равно шести, а количество независимых переменных, оставшихся после корреляционного анализа, равно восьми. Поэтому, стандартный множественный регрессионный анализ невозможен.

Согласно справочным данным для программы STATISTIKA: «В случае, когда количество переменных больше количества наблюдений, единственным способом построения регрессионной модели является пошаговая регрессия с включением» [15].

Для описания пошаговой регрессии приведем определение, представленное в книге Себер Дж. «Линейный регрессионный анализ»: «Метод пошаговой регрессии состоит в том, что на каждом шаге производится либо включение в модель, либо исключение из модели какого-то одного регрессора» [16].

Таким образом, при пошаговой регрессии с включением для построения регрессионной матрицы используют не все переменные (регрессоры), а лишь часть. При этом значимые данные остаются, а незначимые удаляются, оставляя возможность анализа и сравнения новых включаемых переменных.

Анализ проводился по уже коррелированной таблице показателей. Результаты выполнения пошаговой регрессии с включением для независимых переменных приведены в таблице:

Таблица 3 – Результаты регрессионного анализа

N=6	Beta	B	t
Свободный член		385,39	3,3277
Переменная 8	0,29911	36,65	3,2305
Переменная 10	0,47690	43,32	2,7516
Переменная 15	-0,87452	-25,90	-2,4867
Переменная 17	0,57930	742,50	3,6301
Переменная 20	0,48205	0,43	-2,56

Значения коэффициента детерминации $R=0,96279207$, что является хорошим результатом.

Для определения значимости регрессии, используется F-коэффициент, который сравнивается с коэффициентом Фишера. Анализ показал, что он равен $F = 6,468997$. Коэффициент Фишера при уровне значимости $\alpha=0,05$ и $k_1=8$ и $k_2=92$ степеней свободы, равен 2,04. Так как $F=6,468997 > 2,04$, то можно сделать вывод, что построенная регрессия значимая.

По результатам проведенного регрессионного анализа, можно скорректировать набор значимых переменных. При регрессионном анализе коррелированных данных, можно выделить параметры 8, 10, 15, 17, 20, как значимые. Данный вывод сделан на основании сравнения коэффициентов beta с уровнем 0,1. При $\beta > 0,1$, данные считаются значимыми.

Для проверки результатов необходимо сравнить полученные результаты каждой переменной со значением коэффициента Стьюдента. Значение t-критерия (критерия Стьюдента) при заданном уровне значимости $\alpha=0,05$ и $k=92$ степенях свободы, равно 2,0003.

Сравнивая по абсолютной величине полученные значения t-критерия для независимых переменных с критерием Стьюдента, приходим к выводу о корректности полученных результатов.

В результате расчетов набор значимых показателей, выявленных при анализе второго набора данных, равен пяти. Эти социальные показатели будут использованы в прогнозирующей модели.

Заключение

Поиск и анализ данных для построения прогнозов является обязательным условием для обеспечения их достоверности. Для корректного построения прогноза уровня преступности, необходимы правильные входные данные.

Для обеспечения большей достоверности данные были отобраны из наиболее оптимального региона. Поиск данных велся в направлении сравнения количества преступников и количества обычных граждан, а также в направлении сравнения социальных

показателей с уровнем преступности.

Найденные данные были приведены к единой размерности или относительному виду. После этого, данные были проанализированы с помощью инструментов статистического анализа, таких как расчет коэффициентов корреляции и построение множественной регрессии.

Был сделан вывод о включении определенных показателей в последующую модель для прогнозирования криминогенной обстановки. Из результатов первого анализа включены такие параметры: количество безработных, несовершеннолетних из общего количества людей, людей трудоспособного возраста, людей старше трудоспособного возраста, не имеющих начального образования, имеющих начальное общее образование, имеющих основное общее образование, имеющих среднее профессиональное образование, имеющих высшее образование, количество людей состоящих в браке, количество людей, потерявших супруга/супругу, средний доход населения, число людей, имеющих судимость, количество людей, относящих себя к православию, количество атеистов, процент людей, стоящих на учете с диагнозом алкоголизм.

По результатам второго анализа из девятнадцати параметров, были включены только пять: уровень безработицы; удельный вес численности населения с доходом ниже прожиточного минимума; численность детей (0-17), родители которых лишены родительских прав; количество студентов (среднее специальное образование); проиндексированный объем предоставления услуг на 1 человека (туризм, культура, физ.культура и спорт).

Литература

1. Проблемы криминологического прогнозирования и предупреждения региональной преступности: По материалам Сибирского федерального округа, диссертационная работа по ВАК 12.00.08, кандидат юридических наук Абызов К.Р.

2. Федеральная Служба Государственной Статистики, Интерактивная витрина, на сайте представлены срезы из базы данных госстатистики [Электронный ресурс]. – Режим доступа: <http://cbsd.gks.ru/>

3. Судебный департамент при Верховном Суде Российской Федерации, Данные судебной статистики [Электронный ресурс]. – Режим доступа: <http://www.cdep.ru/index.php?id=79&item=3212>

4. Характеристика осужденных [Текст] – аналитический материал / А. Н. Антипов [и

др.]; под общ. ред. начальника ФКУ НИИ ФСИН России, профессора А. В. Быкова; ФКУ НИИ ФСИН России. – Тверь – 47 с.

5. Генеральная Прокуратура Российской Федерации, Портал правовой статистики, Социальный портрет преступника по регионам России [Электронный ресурс]. – Режим доступа: http://crimestat.ru/social_portrait

6. Аналитический центр Юрия Левады, ежегодник «Общественное мнение» 2016. Общественное мнение – 2016. М.: Левада-Центр, 2017 – 272 с.

7. Сайт независимой социологической службы Фонд «Общественное мнение», результаты опроса «Ценности: религиозность» [Электронный ресурс]. – Режим доступа: <http://fom.ru/obshchestvo/10953>

8. Федеральная Служба Государственной Статистики, территориальный орган, Приложение 2 к докладу "Об итогах Всероссийской переписи населения 2010 года", Москва, 2011.

9. Антонян Ю.М., Еникеев М.И., Эминов В.Е. Психология преступника и расследования преступлений, М., 1996.

10. Александр Куколевский, 18,2% россиян имеют судимость / Газета коммерсант, 22.09.2008, с 3-4.

11. Форма N 2 Статистическая карточка на лицо, совершившее преступление [Электронный ресурс]. – Режим доступа: <http://www.zakonprost.ru/content/base/part/467065>

12. Харченко М.А., Корреляционный анализ, Учебное пособие для вузов, Воронеж: ВГУ, 2008, 31 с.

13. Психологическая энциклопедия. 2-е изд. / Под ред. Р. Корсини, А. Ауэрбаха. — СПб.: Питер, 2006. – 1096 с

14. Корреляционный анализ. Использование MS Excel для расчета коэффициента корреляции, Учебно-методическое пособие для студентов, ГБОУ ВПО «КГМУ», Казань 2011.

15. Документация на сайте компании StatSoft, Описание множественной регрессии [Электронный ресурс]. – Режим доступа: http://statsoft.ru/products/STATISTICA_Base/multiple-regression.php

16. Себер Дж. Линейный регрессионный анализ. М.: Мир, 1980. — 456 с.

Сложеницын А.В., Землянская С.Ю. Анализ данных для компьютеризированной системы прогнозирования уровня преступности. Определен наиболее оптимальный регион для поиска и анализа данных, теоретически характеризующих преступность. Выделены направления поиска показателей. Найден, сгруппирован и приведен к относительному виду набор параметров, теоретически характеризующий уровень преступности. Произведен корреляционный и множественный регрессионный анализ для определения зависимости между показателями-факторами и уровнем преступности. На основе анализа выбраны данные для прогнозирования преступности.

Ключевые слова: криминогенная обстановка, уровень преступности, корреляционный анализ, регрессионный анализ, значимость факторов.

Slozhenitsyn A.V., Zemlyansky S.Y. Data analysis for computerized forecasting of the crime level. The most optimal region for searching and analysis of the data theoretically characterizing the crime is determined. Directions of crime indicators searching are selected. The set of parameters is defined, grouped and reduced to relative units of measurement. The defined parameters set theoretically characterizes the level of crime. Correlation and multiple regression analysis is performed for the definition of the relationships between factors and the crime level. The significance of the factors is determined. The data for crime prediction based on the analysis are selected.

Key words: crime, crime rate, correlation analysis, regression analysis, the significance of the factors.

Статья поступила в редакцию 20.5.2017
Рекомендована к публикации д-ром физ.-мат. наук А.С. Миненко