

Б.М. Пранов

(Академия ГПС МЧС России; e-mail: info@academygps.ru)

МЕТОДЫ МНОГОМЕРНЫХ СТАТИСТИЧЕСКИХ ИССЛЕДОВАНИЙ В ПРОБЛЕМАХ ТЕХНОСФЕРНОЙ БЕЗОПАСНОСТИ

Показано, что для эффективного решения важных проблем техносферной безопасности необходимо применять статистические методы эконометрики.

Ключевые слова: статистические исследования, эконометрика.

B.M. Pranov

MULTIDIMENSIONAL STATISTICAL METHODS IN PROBLEMS OF TECHNOSPHERIC SAFETY

It is shown that for the effective solution of important problems of technospheric safety one must use the statistical methods of econometrics.

Key words: statistical research, econometrics.

Статья поступила в редакцию Интернет-журнала 24 сентября 2013 г.

1. Статистические методы исследований на протяжении последних десятилетий широко использовались при анализе и прогнозировании социально-экономических явлений в пожарной охране, в области динамики и прогнозирования преступности, техносферных и природоохранных явлений. Первоначальная информация в социально-экономических исследованиях чаще всего характеризуется рядом признаков, факторов (показателей). В качестве объектов исследований могут быть страны, регионы, предприятия, респонденты и т.д., а в качестве признаков – различные показатели социально-экономической структуры изучаемых объектов. В настоящее время вычислительные системы и компьютерные сети позволяют накапливать большие массивы данных для решения задач обработки и анализа данных.

Как показывает опыт анализа массовых явлений, число объектов может достигать многих десятков и сотен; число признаков может исчисляться десятками. Очевидно, что непосредственный (визуальный) анализ исходных данных при большом количестве объектов и признаков практически малоэффективен – можно лишь выявить отдельные особенности изучаемой структуры, извлечь иллюстративные частные примеры.

В качестве инструментального средства для реализации методов многомерной статистики используется распространенная **программа обработки статистической информации** – **SPSS** (аббревиатура от *Statistical Package for the Social Science*). Существуют и другие пакеты программ, однако этот продукт наиболее доступен, поэтому его применяют при решении многих прикладных задач и в качестве вспомогательного средства в дисциплинах, связанных с подготовкой данных при принятии решений.

В настоящее время самой последней является 19-я версия программы. Есть также еще две программы, имеющие сравнимые возможности по обработке статистического материала – "STATISTICA" и "StatGraph". Однако, эти программы не локализованы для русскоязычных пользователей, и потому не имеют широкого распространения в России.

SPSS содержит хорошее аналитическое программное обеспечение, предоставляет решения в области "добычи знаний" (Data Mining), позволяет выявить скрытые связи данных, находящихся в базах и хранилищах данных. Программные продукты SPSS помогают решать прикладные задачи в различных областях: от классификации и профилирования совокупностей данных о пожарах и экстремальных ситуациях до анализа кредитного риска, управления контролем качества и повышения производительности персонала, занимающегося продажами.

Механизм анализа данных и прогнозирования, который еще принято называть Data Mining, то есть "извлечение информации", "добыча данных", "добыча знаний", позволяет использовать в прикладных решениях инструменты для выявления закономерностей, которые обычно скрываются за большими объемами информации. Математической основой инструментария анализа данных является многомерный статистический анализ, бурное развитие которого началось в конце XX века в период массового распространения компьютеров.

В настоящее время методы и модели *многомерного статистического анализа (МСА)* используются в составе новых информационных технологий, называемых "интеллектуальным анализом данных".

Можно выделить три основные задачи, решаемые в рамках МСА:

- исследование характера явных и неявных зависимостей между объектами или признаками;
- классификация объектов или признаков как при задании профиля групп, так и при его отсутствии;
- снижение размерности пространства признаков за счет выявления внутренней структуры в заданной совокупности.

В экономическом анализе применимо большинство методов МСА. Например, важнейшей составной частью экономических исследований оказывается задача выявления и анализа силы, характера и формы взаимосвязей показателей, выбранных для описания того или иного процесса. При наличии количественной информации, записанной в виде матрицы исходных данных, для решения указанных задач может быть привлечен *корреляционный анализ*, затем он подкрепляется *множественным регрессионным анализом*.

Статистический анализ задач на получение экстремальных значений предназначен для ответов на вопросы двух типов:

- не выпадает ли за разумно ожидаемые границы какое-либо отдельное наблюдение в выборке из совокупности, распределение которой предполагается известным?
- проявляет ли последовательность экстремальных значений регулярное поведение?

В обоих случаях слова "разумно ожидаемые" и "регулярное поведение" должны быть определены строго как математические понятия. Такого рода задачи часто возникают при анализе и прогнозировании чрезвычайных ситуаций – примерами могут служить лесные и степные пожары, наводнения и паводки.

В экономических исследованиях актуальными являются также различные **процедуры классификации объектов**. При этом наблюдается тенденция к комплексному использованию методов факторного анализа и автоматической классификации. Сначала с использованием **методов классификации** получают группировку объектов в исходном многомерном пространстве признаков. Затем с использованием **факторного анализа** выявляют небольшое количество основных факторов (лучше, если их будет два). В результате каждый объект представляется точкой в пространстве главных факторов, а каждый кластер – группой точек. Расположение этих групп на плоскости факторов определяет "статус" каждой из них, а также дает наглядную интерпретацию компактности полученных групп, их объёма, относительной удаленности друг от друга и т.п.

2. В научных исследованиях для противопожарной службы статистические методы занимают значительное место. Прежде всего, необходимо отметить труды Н.Н. Брушлинского и его школы. В этих работах отражается как мировая пожарная статистика, её анализ и прогнозирование [1, 2], так и прикладные аспекты статистических исследований, связанные с имитационным моделированием деятельности пожарной охраны [3]. В последние годы появились работы [4-6], в которых применяется более обширный арсенал статистических методов исследования. Однако, в этих работах затронут анализ только мировой статистики пожаров и оставлена без внимания отечественная статистика.

Эконометрические методы и модели представлены в фундаментальном труде [7], а также во множестве русскоязычных публикаций [8], среди которых хотелось бы отметить книгу [9], в которой излагается применение многомерного статистического анализа с использованием программы SPSS. Рассмотрим некоторые группы задач, которые можно решать с использованием методов МСА.

Категорирование пожаров. Здесь можно применять методы кластерного анализа с предварительным использованием факторного анализа (выделение главных компонент). Число значимых главных компонент определяется обычно таким их числом, что добавление следующей компоненты не увеличивает существенно объясненную дисперсию ("метод каменистой осыпи"). После этих процедур можно уже обратиться к построению многомерных регрессионных зависимостей, в том числе с использованием запаздывающих аргументов (как самой прогнозируемой величины, так и независимых параметров – такие модели называются авторегрессионными). В полном объёме анализ такого рода еще не проводился.

Прогнозирование динамических рядов наблюдений. Задачи такого рода похожи на предыдущие, но у них имеется ряд существенных отличий. Общим является элиминирование мультиколлинеарных параметров (то есть таких,

взаимный коэффициент корреляции между которыми достаточно велик). Новой существенной особенностью является гетероскедастичность. При проведении регрессионного анализа, основанного на *методе наименьших квадратов (МНК)*, следует внимательно анализировать проблемы, связанные с выполнимостью свойств случайных отклонений модели. Свойства оценок коэффициентов регрессии зависят от свойств случайного члена в уравнении регрессии. Качественные оценки моделей получаются при выполнении некоторых условий (условия Гаусса-Маркова), так как при их нарушении МНК может давать оценки с плохими статистическими свойствами. Одной из самых важных предпосылок МНК является условие постоянства дисперсии случайных отклонений. Казалось бы, эти требования могут показаться малозначительными, но в 2003 г. ученым Роберту Инглу (США) и Клайву Грэнджеру (Великобритания) была присуждена Нобелевская премия по экономике с формулировкой "За разработку метода анализа временных рядов в экономике на основе математической модели с авторегрессионной условной гетероскедастичностью (ARCH)". Чаще всего гетероскедастичность возникает в моделях, основанных на перекрестных выборках, но встречается и во временных рядах.

Исследования в области управления персоналом ("Управление человеческими ресурсами") являются актуальными для любой области экономики нашей страны. Особое значение они приобретают в специальных службах (МЧС, МВД и др.). Все эти методы основаны на статистических исследованиях и моделировании. Более того, в основу информационных технологий HRM (human resource management) обычно кладут соответствующие модели и разработки. К информационным HRM-технологиям относятся:

- информационно-поисковые справочные и консультационные юридические системы;
- документационные системы, отражающие отдельные стороны документооборота организации (по заключенным договорам и их движению, трудовым контрактам, организационно-распорядительным актам руководства организации и пр.);
- системы, отражающие текущую оперативную деятельность (аудиторские, криминалистические, социально-статистические и др.);
- идентификационные системы, а также подсистемы в комплексных информационных системах, ориентированные на конкретную прикладную сферу.

Если раньше программа SPSS широко использовалась в таких "классических" областях науки и бизнеса, как биология, социология, психология, управление качеством производства, общие маркетинговые исследования и экономическое прогнозирование, то сейчас новую версию можно с успехом применять в таких актуальных специализированных областях, как анализ техноферных опасностей, основанный на использовании баз данных, Data Mining, Data Warehousing и другие. Внимания заслуживает тот факт, что изменения, внесенные в модуль SPSS Regression Models, позволяют использовать SPSS при реше-

нии задач управления лояльностью клиентов (CRM) (в том числе клиентов, обслуживаемых силами МЧС при чрезвычайных ситуациях). Отметим, что данная тема представляет собой один из наиболее популярных разделов современного практического управления.

Отдельного упоминания заслуживает то факт, что большинство наиболее популярных статистических методов прогнозирования, включенных в модуль SPSS Regression Models, позволяют работать с большим объемом недоступной информации. В математике в таком случае говорят о повышении робастности метода, то есть его устойчивости по отношению к неопределенностям и существенным отклонениям от диапазона параметров, для которого разрабатывался метод. Такое повышение робастности весьма желательно в маркетинговых исследованиях и в социологии, где всегда большой объем недостающих или недостоверных данных.

Особенно повышение быстродействия чувствительно в случае, когда речь идет о методах кластерного анализа, широко используемого в социологии, психологии и медицине, которые иногда требовали многочасовых расчётов даже на мощных компьютерах, для чего в предыдущих версиях SPSS был предусмотрен пакетный режим выполнения задач.

Случаи мошенничества, злоупотреблений и нецелевого расходования средств ежегодно обходятся органам государственного управления и государственным программам в миллиарды рублей. Выявлять и предотвращать эти негативные явления очень и очень непросто и требует привлечения и без того ограниченных ресурсов органов государственного управления. Многие органы государственного управления сталкиваются в своей работе с такими препятствиями:

- ограниченность финансовых и людских ресурсов;
- огромное количество обращений, транзакций и прочей информации, потенциально доступной для анализа;
- примитивные методы, упускающие новые тенденции и необычные явления;
- отсутствие экспертизы и подготовки в сфере анализа данных.

Мощные аналитические и прогностические технологии SPSS позволяют **проводить расследование**. Аналитика SPSS проникает глубоко в скрытые структуры данных для обнаружения взаимосвязей и создания моделей для различных типов мошенничества, злоупотреблений и нецелевого расходования средств. Затем при помощи построенных моделей анализируются поступающие данные. Программное обеспечение оповещает, когда поступающие данные будут соответствовать моделям мошенничества, злоупотреблений или нецелевого расходования средств, что позволит в реальном времени проверять подозрительные обращения, транзакции и запросы.

Проведение опросов и обследований. Надёжная информация – это ключевой элемент в процессе принятия верных решений. Это особенно актуально для органов государственного управления, ведь качество обслуживания насе-

ления во многом зависит от понимания и удовлетворения их потребностей. Опросы и обследования совершенно необходимы в государственном управлении, так как они позволяют получать ценную информацию непосредственно от населения. Результаты опросов и обследований используются для принятия обоснованных решений практически в любой сфере государственного управления. Оценка удовлетворенности населения или работников, выявление предпочтений избирателей, создание и функционирование системы показателей работы учреждения – опросы и обследования позволяют получать точные, достоверные данные, обеспечивающие эффективную поддержку работы органов государственного управления.

SPSS помогает ведущим организациям во всем мире превращать данные в капитал при помощи масштабируемых, рентабельных аналитических технологий. Руководители, задумывающиеся о перспективе своих организаций, обращаются к SPSS для анализа своих данных и получают доступ к широкому диапазону лучших аналитических возможностей SPSS, обеспечивающих принятие действенных аналитических решений. Затем, используя средства внедрения, они заставляют прогностическую аналитику работать на улучшение качества принимаемых решений, развитие основанных на прогнозах стратегий взаимодействия с клиентами и совершенствование взаимоотношений с клиентами. Прогнозирование помогает компаниям максимизировать отдачу от клиентов, привлекать новых выгодных клиентов, совершенствовать взаимоотношения с клиентами и сохранять клиентов, приносящих прибыль, как можно дольше.

IBM предложила два готовых конфигурируемых решения – для отрасли страхования (IBM SPSS Decision Management for Claims) и для маркетинговых служб компаний из любых отраслей (IBM SPSS Decision Management for Customer Interactions). Первое решение позволяет выявлять мошеннические запросы на страховые выплаты, ежегодно приносящие огромные убытки. Напомним, что, по данным организации Insurance Information Institute, мошеннические действия является причинами 10 % убытков и расходов на урегулирование убытков в сфере страхования собственности/страхования от несчастных случаев, что составляет около 30 млрд долл. в год. Второе решение поддерживает управление взаимодействиями с клиентами или потенциальными клиентами, обращающимися в компании через контакт-центры, Интернет, точки продаж или электронную почту. Конфигурируемый Web-интерфейс позволяет бизнес-пользователям буквально тремя щелчками мышью создать прогнозную модель и провести анализ на базе сценариев what if ("что, если") для сравнения и проверки результатов, прежде чем реализовать эту модель в действующей системе. Бизнес-пользователи получают полный контроль над процессом анализа, что позволяет им принимать точные решения в реальном времени, с учётом меняющихся стратегий, моделей поведения клиентов или рыночных условий. Любая организация может конфигурировать и настраивать это ПО для решения специфических проблем бизнеса, характерных для конкретной отрасли, будь то управление рисками в сфере финансовых услуг, прогнозирование отказов конструктивных элементов в сфере производства или борьба с преступностью в государственном секторе.

Литература

1. **Брушлинский Н.Н., Вагнер П., Соколов С.В., Холл Д.Р.** Мировая пожарная статистика. М.: Академия ГПС МЧС России, 2004. 126 с.
2. **Брушлинский Н.Н., Соколов С.В., Кленко Е.А., Попков С.Ю.** Оценка пожарных рисков в муниципальных образованиях Калужской области // Технологии техносферной безопасности: интернет-журнал. Вып. № 2 (36). 2011.
3. **Брушлинский Н.Н., Соколов С.В., Алехин Е.М. и др.** Безопасность городов: имитационное моделирование городских процессов и систем. М.: Изд-во "ФАЗИС", 2004. 172 с.
4. **Третьяков Н.П.** Компонентный анализ мировой статистики пожаров// Технологии техносферной безопасности: интернет-журнал. Вып. № 3 (25). 2009.
5. **Третьяков Н.П.** К вопросу о методологии прогнозирования мировой статистики пожаров// Технологии техносферной безопасности: интернет-журнал. Вып. № 6 (34). 2010.
6. **Третьяков Н.П.** Применение кластерного анализа к мировой статистике пожаров. // Технологии техносферной безопасности: интернет-журнал. Вып. № 2 (24). 2009.
7. **Greene W.H.** Econometric analysis, 5-th edition, Prentice Hall. 2003. 1093 p.
8. **Кремер Н.Ш., Путко Б.А.** Эконометрика: учебник для вузов. М.: ЮНИТИ-ДАНА, 2002. 311 с.
9. **Многомерный** статистический анализ в экономических задачах: компьютерное моделирование в SPSS: учеб. пособие / Под ред. Орловой И.В. М.: Вузовский учебник, 2009. 310 с.