

УДК 519.24+519.81

МАТЕМАТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА: УЧЕБНОЕ ПОСОБИЕ

Лялькина Г.Б.¹, Бердышев О.В.¹

¹ГОУ ВПО «Пермский национальный исследовательский политехнический университет», Пермь, Россия (614000, г. Пермь, Комсомольский пр-т, 29), e-mail: ol_brine@mail.ru

Учебное пособие Лялькиной Г.Б. и Бердышева О.В. «Математическая обработка результатов эксперимента» по дисциплине «Планирование эксперимента и обработка экспериментальных данных» предназначено студентам направления «Техносферная безопасность». Может быть полезно студентам и аспирантам технического университета, а также специалистам, использующим методы статистического анализа. В краткой форме изложены основные принципы планирования эксперимента, включающие различные процедуры сбора и первичной обработки статистических данных, проверку нормальности распределения и сопоставления числовых характеристик нормально распределенных статистических совокупностей, а также ряд процедур линейного, нелинейного и множественного корреляционно-регрессионного анализа. Детально расписаны все процедуры построения регрессионной модели, начиная с выбора ее вида, включая возможную линеаризацию в случае нелинейности искомой зависимости, оценку параметров на основе статистических гипотез и заканчивая проверкой адекватности построенной модели исходным статистическим данным, а также заключительную оценку точности. Приведены примеры использования процедур совместного исследования двух и более нормально распределенных статистических совокупностей с целью мониторинга изменения состояния безопасности техносферы и управления ее поведением.

Ключевые слова: математическая обработка результатов эксперимента, первичная обработка данных, корреляционно-регрессионный анализ.

MATHEMATICAL PROCESSING OF THE EXPERIMENTAL RESULTS: TRAINING MANUAL

Lyalkina G.B.¹, Berdyshev O.V.¹

¹Perm National Research Polytechnical University, Perm, Russia (614000, Perm, Komsomolski avenue, 29), e-mail: ol_brine@mail.ru

Training manual "Mathematical treatment of experimental data" written by Lyalkina G.B. and Berdyshev O.V. is intended to expound the corresponding academic discipline for the students of the direction "Technosphere safety". The book can be useful for the students and post-graduates of technical university as well as for those specialists who are using methods of statistical analysis. Main principles of the experiment planning are presented. Different procedures of collecting and initial roughing of statistical data are expounded. The procedure of comparing numerical characteristics of the normally distributed sets of statistical data is also under consideration. Certain procedures of linear, non-linear and plural regressive analysis are expounded. All construction procedures of regressive model are described in detail. Procedures begin from the choice of the analytical model form and include possible procedures of linearization (in the case of non-linearity) for the searching form of dependence. Procedures include the estimation of model parameters with a help of statistical hypotizis. Procedure of verification of the model and estimation of its exactness concludes constuction of the model. Examples of joint research of two and more normally distributed sets are given. The aim of statistical analysis is monitoring of changing of safety state in techno sphere.

Keywords: mathematical treatment of experimental results, primary analysis of experimental data, correlation and regression analysis.

Введение

С целью обеспечения безопасности человека в техносфере необходимо управлять протекающими в ней процессами. Результаты текущего мониторинга состояния безопасности в системах «человек – техносфера - окружающая среда» описываются массивами статистических данных, и специалисту в области техносферной безопасности

необходимо владеть методами их обработки. Учебное пособие Лялькиной Г.Б. и Бердышева О.В. [1] предназначено студентам направления «Техносферная безопасность», может быть полезно студентам и аспирантам технических университетов и входит в систему методического обеспечения учебной дисциплины «Планирование эксперимента и обработка статистических данных» и комплекса смежных учебных дисциплин.

Целью настоящей статьи является представление пособия [1], позволяющего с теоретической и практической позиций изучить базовые математические методы статистической обработки опытных данных и представить методики их применения. Приведенные примеры в основном ориентированы на возможность изучения проблем техносферной безопасности и принятия соответствующих управленческих решений, в том числе на основе методов статистического анализа [2; 3].

Опытные числовые данные исследователь может получить либо активными методами путем измерений, выполненных в ходе специально спланированного эксперимента, меняя по мере необходимости его условия, либо с помощью так называемого пассивного эксперимента, условия которого исследователь не может менять по своему усмотрению. В ходе пассивного эксперимента исследователь только регистрирует опытные значения или выбирает их из совокупности известных статистических данных.

В результате эксперимента получаются выборочные совокупности измеряемых величин, по своей сути имеющие случайный характер. Поэтому модели, построенные на их основе, всегда будут иметь вероятностный характер. Отсюда следует, что сбор и обработка статистических данных должны опираться на строгие методы математической статистики, методологической основой которых является теория вероятностей.

Цель математической обработки совокупности экспериментальных данных – это построение полезной аналитической модели исследуемого явления или процесса на основе конечной выборочной совокупности экспериментальных данных.

Прежде чем строить модель, необходимо выяснить, имеет ли вообще смысл делать это на основе имеющейся совокупности. Может оказаться, что нарушены процедуры сбора данных, и полученная выборочная совокупность не отражает исследуемые свойства генеральной совокупности значений случайной величины в достаточной степени. В частности, может оказаться, что объем выборочной совокупности мал, или сами данные имеют недостаточную точность, или данных достаточно и они достаточно точны, но при этом они неудачно сгруппированы и представляют свойства только части генеральной совокупности. Наконец, может оказаться, что выдвинутая в начале исследования гипотеза о существовании зависимости между случайными величинами неверна и устанавливать вид этой зависимости бессмысленно.

Корректность математической модели и возможность ее применения на практике зависят от того, насколько грамотно спланирован эксперимент, насколько корректно с точки зрения математической статистики проведены процедуры сбора, обработки и анализа результатов эксперимента, и, наконец, насколько корректно выполнена интерпретация полученных результатов. При этом следует отметить, что любые аналитические модели, которые мы собираемся строить на основе опытных данных, не могут иметь точность, превышающую точность измерения самих данных.

Интерпретация полученных результатов и оценка точности прогнозов, которые можно выполнить с помощью полученных регрессионных моделей, завершает процедуры обработки экспериментальных данных.

Настоящее пособие также содержит основную и дополнительную литературу, которая может оказаться востребованной для более глубокого изучения методов обработки экспериментальных данных.

По окончании каждого из разделов учебного пособия приводится список вопросов, предназначенных для самоконтроля адекватности восприятия информации, представленной в текущем разделе.

Во введении рассматриваемого учебного пособия актуализирована проблема статистического моделирования процессов обеспечения безопасности в техносфере и представлены основные принципы планирования эксперимента.

Первый раздел – «Основные понятия теории случайных величин».

Здесь приводится понятие случайной величины и законов ее распределения. В качестве законов распределения рассматриваются функция распределения, функция плотности распределения, полигон распределения и таблица вероятностей. Подробно рассмотрены свойства функции распределения одномерной случайной величины. Также здесь приводятся формулы для вычисления математического ожидания, дисперсии и среднего квадратичного отклонения дискретной и непрерывной случайных величин. Представлена проблема корреляционной зависимости, а также формулы для вычисления коэффициентов парной корреляции. Описаны законы распределения (Пирсона, Стьюдента, Фишера-Снедекора и др.), наиболее используемые при статистической обработке опытных данных.

Второй раздел – «Процедуры сбора опытных данных».

Грамотно осуществленный сбор исходных данных – необходимое условие качественного исследования надежности технических систем и процессов обеспечения безопасности в техносфере [4].

Одно из важнейших требований к выборочной совокупности экспериментальных данных – её репрезентативность. Репрезентативность (представительность) выборочной совокупности означает, что собранные данные должны отражать исследуемые свойства генеральной совокупности корректно и с достаточной полнотой, а процедуры сбора данных, предшествующие их статистическому анализу, должны удовлетворять условиям системности.

В настоящем разделе кратко изложены основные правила начального этапа планирования эксперимента, а именно приведены методы сбора данных и определения необходимого объема формируемой статистической совокупности.

Третий раздел пособия – «Методы первичной обработки опытных данных», посвящен методам обработки собранных данных.

Недостаток качества исходных данных может привести к неверным выводам, что является особо опасным при изучении проблем защиты человека [2]. В совокупности методы, представленные в третьей части, обеспечивают корректность дальнейших процедур анализа статистических данных и возможность использования их.

Здесь рассматриваются проблемы, связанные с возможностью использования и интерпретации данных, полученных в ходе экспериментов. Первичная обработка позволяет выявить так называемые ложные данные и исключить их из дальнейшего рассмотрения.

К источникам ложных данных относятся: грубые (недостаточно точные) измерения; нарушение условий эксперимента; использование неисправного оборудования; ошибки при обработке информации и др.

Случайным путем ложные результаты проникают в исходную совокупность опытных данных из-за действия достаточно большого числа случайных причин, которые не удается заранее предусмотреть при проведении измерений или при сборе статистических сведений. Например, оператор, собиравший данные, сделал опisku. Или прибор-измеритель мог случайно подвергнуться постороннему воздействию именно в момент измерения. Наконец, нельзя исключать случай, когда ошибки в статистические данные вносятся злонамеренно, например с целью искажения или сокрытия истинного положения вещей.

Процедура исключения ложных данных позволяет выявить такие данные в статистической совокупности с большой степенью вероятности.

В настоящем разделе указываются также процедуры проверки оставшихся данных на случайность и независимость.

Эксперимент считается нарушенным, если внутри какой-либо из исследуемых совокупностей опытных данных обнаруживается закономерность, связывающая эти данные. Это может произойти, если на измеряемые величины оказывают влияние посторонние

неслучайные факторы. Поэтому прежде, чем применять методы статистического анализа, исследователь должен убедиться в отсутствии такого влияния. Случайность и независимость опытных данных – необходимое условие репрезентативности выборочной совокупности.

В разделе приведен метод, основанный на медиане выборочной совокупности, а также метод «восходящих» и «нисходящих» серий.

В четвертом разделе – «Статистический анализ выборочных совокупностей», представлены формулы для статистических оценок числовых характеристик случайных величин, процедуры подбора законов распределения и исследования совместного поведения двух и более случайных величин. Выделена процедура проверки нормальности закона распределения генеральной совокупности как одного из наиболее важных и часто используемых на практике.

В разделах, посвященных оценке числовых параметров предполагаемых законов распределения, представлены две задачи о проверке статистических гипотез:

- 1) проверка предположения о том, что исправленная выборочная дисперсия $S^2(X)$ и предполагаемая (гипотетическая) дисперсия DX генеральной совокупности отличаются статистически незначимо;
- 2) проверка предположения о статистической незначимости отличия среднего выборочного значения от гипотетического значения генеральной средней.

Отметим, что эти задачи возникают достаточно часто, например при обработке данных о числе несчастных случаев на производстве, числе отказов оборудования, числе профессиональных заболеваний и т.п. [2; 4].

Кроме того, в пособии описаны также следующие процедуры, часто используемые в процессах принятия решений на основании обработки статистических данных:

- 1) сравнение исправленных выборочных дисперсий двух нормальных генеральных совокупностей;
- 2) сравнение генеральных средних двух нормальных совокупностей, дисперсии которых известны;
- 3) сравнение генеральных средних двух нормальных совокупностей, имеющих неизвестные дисперсии.

Отдельно выделена процедура проверки гипотезы о значимости выборочного коэффициента корреляции, как один из важнейших этапов корреляционно-регрессионного анализа.

Пятый раздел – «Корреляционно-регрессионный анализ».

Принятие решений на основе анализа опытных данных опирается на использование вероятностных законов распределения и корреляционно-регрессионного анализа.

Линейный корреляционно-регрессионный анализ выборочных совокупностей состоит из двух основных этапов.

1. Корреляционный анализ. На первом этапе на основании анализа выборочных совокупностей, то есть с помощью выборочного коэффициента или индекса корреляции необходимо выяснить наличие корреляционной зависимости между исследуемыми случайными величинами.

2. Регрессионный анализ. В случае обнаружения корреляционной зависимости переменных можно приступать к регрессионному анализу, то есть к подбору уравнений регрессии, связывающих коррелирующие величины количественными соотношениями.

В случае если не предполагается линейность связи выборочных совокупностей, первый этап в общем случае пропускается.

Здесь представлены основные этапы разработки регрессионных моделей на основе метода наименьших квадратов. Рассмотрены алгоритмы построения линейных и линеаризуемых нелинейных моделей, а также более общие процедуры для существенно нелинейных моделей. В заключение приводятся формулы оценки их точности.

Изложение материала начинается с описания метода наименьших квадратов и построения с его помощью парной линейной регрессионной модели.

Далее рассматривается случай нелинейности статистической связи между исследуемыми совокупностями экспериментальных данных. Приводится общий вид функции, подлежащей минимизации в соответствии с методом наименьших квадратов. Уточняется, что построение нелинейных уравнений регрессии в отдельных случаях можно свести к отысканию линейных зависимостей с помощью подходящих замен переменных. Представлены примеры замен и их реализация.

Подробно рассмотрена реализация метода наименьших квадратов при построении множественной линейной регрессионной модели. Приведен общий вид соответствующей системы равных нулю частных производных минимизируемой функции. Указаны условия, необходимые для практического использования построенной регрессионной модели:

- 1) все коэффициенты уравнения регрессии должны быть значимы;
- 2) само уравнение регрессии должно быть значимо;
- 3) точность уравнения регрессии должна быть достаточной для решения задачи в условиях проводимого эксперимента.

Для проверки значимости коэффициентов уравнения регрессии используется t -статистика Стьюдента.

Для проверки значимости уравнения регрессии в целом используется F -статистика Фишера.

Оценка точности построенного уравнения регрессии выполняется с помощью среднеквадратической погрешности.

Список литературы

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика в задачах и упражнениях. – М. : ЮНИТИ, 2001. – 270 с.
2. Бердышев О.В. Методы статистического анализа : практическое руководство по дисциплине «Математическая статистика». – Пермь : Перм. обл. ин-т повышения квалификации работников образования, 2005. – 36 с.
3. Лялькина Г.Б. Математические основы теории принятия решений : учебное пособие. – Пермь : Изд-во Перм. нац. исслед. политехн. ун-та, 2012. – 118 с.
4. Лялькина Г.Б. Надежность технических систем и техногенный риск. Ч. 1. Надежность технических систем : учебное пособие. – Пермь : Изд-во Перм. гос. техн. ун-та, 2011. – 90 с.
5. Математическая обработка результатов эксперимента : учеб. пособие / Г.Б. Лялькина, О.В. Бердышев. – Пермь : Изд-во Перм. нац. исслед. политехн. ун-та, 2013. – 78 с.

Рецензенты:

Яковлев В.И., д.ф.-м.н., профессор, зав. кафедрой «Процессы управления и информационная безопасность», Пермский государственный национальный исследовательский университет, г. Пермь.

Цаплин А.И., д.т.н., профессор, зав. кафедрой «Общая физика», Пермский национальный исследовательский политехнический университет, г. Пермь.