

УДК 330.5 : 519.2

UDC 330.5 : 519.2

5.2.2. Математические, статистические и инструментальные методы в экономике (физико-математические науки, экономические науки)

5.2.2. Mathematical, statistical and instrumental methods of economics (physical and mathematical sciences, economic sciences)

ОСНОВНЫЕ ИДЕИ КОНТРОЛЛИНГА СТАТИСТИЧЕСКИХ МЕТОДОВ

THE MAIN IDEAS OF STATISTICAL METHODS CONTROLLING

Орлов Александр Иванович
д.э.н., д.т.н., к.ф.-м.н.,
профессор

Orlov Alexander Ivanovich
Dr.Sci.Econ., Dr.Sci.Tech., Cand.Phys-Math.Sci.,
professor

РИНЦ SPIN-код: 4342-4994

RSCI SPIN-code: 4342-4994

prof-orlov@mail.ru

Московский государственный технический университет им. Н.Э. Баумана, Россия, 105005, Москва, 2-я Бауманская ул., 5,

Bauman Moscow State Technical University, Moscow, Russia

В прикладной статистике необходимы правила проверки адекватности используемых расчётных методов реальной ситуации. Разработкой таких правил занимается контроллинг статистических методов. Настоящая статья посвящена обсуждению его базовых идей. Он устанавливает основные требования к методам статистической обработки данных, их характеристики, которые должны быть отражены в нормативно-технической, методической и справочной документации (далее – НТД), а также требования к оформлению результатов обработки данных. В прикладной статистике используют вероятностно-статистические методы и методы анализа данных. Вероятностно-статистические методы должны удовлетворять следующим основным требованиям: наличие адекватной вероятностной модели явления; использование результатов математической статистики; адекватность относительно допустимых преобразований шкал измерения; возможность установления точности получаемых результатов и выводов. Если метод обработки данных не удовлетворяет хотя бы одному из перечисленных требований, он называется методом анализа данных. При разработке НТД по статистическим методам обработки данных необходимо установить и указать следующие характеристики используемых методов: степень выполнения основных требований; устойчивость выводов по отношению к допустимым отклонениям исходных данных и предпосылок модели; возможность последовательного применения методов статистической обработки данных; оптимальность (в указанном в НТД смысле); трудоёмкость; наглядность. Обсуждаются требования к построению и обоснованию адекватной вероятностной модели явления; к результатам математической статистики; к адекватности

In applied statistics, rules are needed to verify the adequacy of the calculation methods used in the real situation. The development of such rules is carried out by controlling statistical methods. This article is devoted to discussing his basic ideas. It sets out the basic requirements for statistical data processing methods, their characteristics, which should be reflected in regulatory, technical, methodological and reference documentation (hereinafter referred to as NTD), as well as requirements for the design of data processing results. Applied statistics uses probabilistic statistical methods and methods of data analysis. Probabilistic and statistical methods must meet the following basic requirements: the availability of an adequate probabilistic model of the phenomenon; the use of mathematical statistical results; adequacy with respect to acceptable transformations of measurement scales; the ability to establish the accuracy of the results and conclusions. If a data processing method does not meet at least one of the listed requirements, it is called a data analysis method. When developing statistical data processing methods, it is necessary to establish and specify the following characteristics of the methods used: the degree of fulfillment of the basic requirements; the stability of conclusions in relation to the permissible deviations of the initial data and the prerequisites of the model; the possibility of consistent application of statistical data processing methods; optimality (in the sense indicated in the statistical data processing); labor intensity; visibility. The requirements for the construction and substantiation of an adequate probabilistic model of the phenomenon are discussed; for the results of mathematical statistics; to adequacy regarding acceptable transformations of measurement scales; to regulatory, technical and reference documentation on methods of data processing using computers; to the design of data processing results. The characteristics of the accuracy and stability of conclusions; consistent application of

относительно допустимых преобразований шкал измерения; к нормативно-технической и справочной документации по методам обработки данных с помощью компьютеров; к оформлению результатов обработки данных. Рассмотрены характеристики точности и устойчивости выводов; последовательного применения методов обработки данных; оптимальности методов обработки данных; трудоёмкости и наглядности и требования к ним. Рекомендации контроллинга статистических методов предназначены для сотрудников организаций и предприятий (объединений) всех отраслей народного хозяйства, разрабатывающих НТД, связанную со статистической обработкой данных, в том числе с использованием компьютеров

data processing methods; optimal data processing methods; labor intensity and clarity, and requirements for them are considered. The recommendations for controlling statistical methods are intended for employees of organizations and enterprises (associations) of all sectors of the national economy developing statistical data processing related to statistical data processing, including the use of computers

Ключевые слова: КОНТРОЛЛИНГ, ВЕРОЯТНОСТНО-СТАТИСТИЧЕСКИЕ МОДЕЛИ, МЕТОДЫ АНАЛИЗА ДАННЫХ, НЕПАРАМЕТРИЧЕСКАЯ СТАТИСТИКА, ТЕОРИЯ ИЗМЕРЕНИЙ, НЕЧЕТКОСТЬ

Keywords: CONTROLLING, PROBABILISTIC AND STATISTICAL MODELS, DATA ANALYSIS METHODS, NONPARAMETRIC STATISTICS, MEASUREMENT THEORY, FUZZINESS

<http://dx.doi.org/10.21515/1990-4665-210-033>

Введение

В прикладной статистике необходимы правила проверки адекватности используемых расчётных методов реальной ситуации. Разработкой таких правил занимается контроллинг статистических методов, его результаты применимы при анализе статистических данных в любой области. Одно из правил – анализ данных должен начинаться с выбора (построения) и обоснования вероятностной модели порождения данных. Так, распространённая ошибка состоит в необоснованном принятии модели нормального распределения элементов выборки. Хорошо известно, что распределения реальных данных, как правило, не являются нормальными (гауссовскими). Поэтому необходимо тщательное обоснование возможности использования критериев Стьюдента и Фишера. Основные идеи контроллинга статистических методов опубликованы в наших работах [1 – 3].

<http://ej.kubagro.ru/2025/06/pdf/33.pdf>

Истоки этого направления наших исследований относятся к 1980-х годам. Ряд базовых идей контроллинга статистических методов содержится в нормативно-техническом документе [4]. Эта работа выполнена в русле научно-общественного движения, приведшего к созданию Всесоюзного центра статистических методов и информатики Центрального правления Всесоюзного экономического общества (1989) и Всесоюзной статистической ассоциации (1990). Хотя перечисленные организации погибли в результате развала СССР, но научные результаты вошли в учебники XXI в. (см. сводку [5]).

Согласно контроллингу статистических методов, непараметрическая статистика предпочтительнее параметрической. Потому были разработаны и изучены различные непараметрические оценки плотности распределения вероятностей в пространствах общей природы. Необходимо учитывать размытость математических объектов, применять статистику интервальных данных, статистику нечисловых данных, теорию нечёткости. Статистические методы должны соответствовать шкалам измерения. Требуется обсуждения использование компьютеров при анализе статистических данных. И т.д.

Настоящая статья посвящена обсуждению базовых идей контроллинга статистических методов. Он устанавливает основные требования к методам статистической обработки данных, их характеристики, которые должны быть отражены в нормативно-технической, методической и справочной документации (далее – НТД), а также требования к оформлению результатов обработки данных. Его рекомендации предназначены для сотрудников организаций и предприятий (объединений) всех отраслей народного хозяйства, разрабатывающих НТД, связанную со статистической обработкой данных, в том числе с использованием компьютеров. Примером конкретизации

требований контроллинга статистических методов в конкретных ситуациях служит статья [6], в которой анализ идёт на примере задач классификации.

Вероятностно-статистические методы и методы анализа данных

Применение рекомендаций контроллинга статистических методов позволяет повысить научно-технический уровень НТД, дать критерии сравнения методов обработки данных и выбора их наиболее соответствующего конкретной задаче, обеспечить сопоставимость результатов проведённого и повторного прикладного статистического исследования. Рекомендации применяют:

- при выборе методов статистической обработки данных для включения в НТД;
- при описании этих методов в НТД;
- при описании результатов их применения для обработки конкретных данных.

В прикладной статистике используют вероятностно-статистические методы и методы анализа данных со следующими составляющими:

- модель исходных данных (вероятностная или детерминированная);
- цель обработки данных и вид принимаемого решения;
- алгоритм расчётов (правила получения выводов).

Каждая из составляющих методов обработки данных является относительно самостоятельной. Например, рассмотрим вероятностную модель исходных данных, согласно которой результаты наблюдений являются реализациями независимых одинаково распределённых случайных величин, с функцией распределения, входящей в параметрическое семейство. Для этой модели рассматривают задачи точечного или интервального оценивания (значения параметра, математического ожидания, квантиля порядка 0,1 и т.д.) и задачи проверки

гипотез (например, гипотез о значении параметра или о его попадании в некоторую область). В каждом из этих случаев цели обработки данных различны, равно как и вид принимаемых решений. Если же цель обработки данных фиксирована, то для её достижения могут применяться различные методы (алгоритмы). Например, точечную оценку параметра можно определять методами моментов, максимального правдоподобия, одношаговых оценок, несмещённых оценок и др. (Орлов, 2006а, 2022д).

Вероятностно-статистические методы должны удовлетворять следующим основным требованиям:

- наличие адекватной вероятностной модели явления;
- использование результатов математической статистики;
- адекватность относительно допустимых преобразований шкал измерения;
- возможность установления точности получаемых результатов и выводов.

Вероятностно-статистические методы применяют для получения обоснованных научных и практических выводов.

Если метод обработки данных не удовлетворяет хотя бы одному из перечисленных выше основных требований, он является методом анализа данных. Такие методы применяются для предварительного (разведочного) анализа и (или) предварительной формулировки принимаемого решения. Выводы предварительного анализа и (или) принимаемое решение должны быть обоснованы с помощью профессиональных методов рассматриваемой отрасли народного хозяйства (например, на основе опыта эксплуатации соответствующего технологического процесса) или с помощью вероятностно-статистических методов обработки данных. Примерами методов анализа данных являются графические методы, методы кластер-анализа, многомерного шкалирования, ряд методов анализа качественных признаков (например, изложенные в [7]).

Целесообразность использования методов анализа данных для предварительного анализа основана на их большей наглядности и меньшей трудоёмкости. Применение методов анализа данных оправдано, если нет необходимости переносить выводы с обследованной совокупности на более широкую. В настоящее время ряд методов обработки данных приходится рассматривать как методы анализа данных, поскольку свойства этих методов недостаточно изучены.

Если для обработки конкретных данных можно использовать метод анализа данных и вероятностно-статистический метод, то на стадии обоснования статистического вывода следует применять вероятностно-статистический метод.

При разработке НТД по статистическим методам обработки данных необходимо установить и указать следующие характеристики используемых методов:

- степень выполнения основных требований (см. выше);
- устойчивость выводов по отношению к допустимым отклонениям исходных данных и предпосылок модели;
- возможность последовательного (друг за другом) применения методов статистической обработки данных;
- оптимальность (в указанном в НТД смысле) метода статистической обработки данных;
- трудоёмкость;
- наглядность.

Как исключение, указывают на отсутствие сведений об определённой характеристике, например: «Устойчивость не изучалась».

Разработку НТД следует рассматривать как научно-исследовательскую работу, при выполнении которой метод обработки данных изучается теоретически и экспериментально (например, с помощью метода статистических испытаний). В частности, если для

решения прикладной задачи имеются лишь методы анализа данных, то целесообразно изучить их свойства на основе вероятностной модели с целью разработки вероятностно-статистических методов решения этой задачи.

Требования к построению и обоснованию адекватной вероятностной модели явления

В настоящее время теория вероятностей рассматривается как часть математики. Общепринятой является аксиоматика теории вероятностей, предложенная А.Н. Колмогоровым в 1933 г. [8]. Аксиоматическая теория вероятностей пришла на смену концепции Р. Мизеса, в которой теория вероятностей рассматривалась как естественнонаучная дисциплина, вероятность определялась как предел частоты. Понятия «статистического ансамбля», «статистической однородности», «статистической устойчивости» относятся к частотной концепции. Аксиоматическая теория вероятностей не требует для своего применения обязательной возможности проведения «большого числа опытов». В частности, вероятностной моделью может быть описано поведение уникального объекта.

Приведём пример уникального объекта, поведение которого описывается вероятностной моделью. Поместим пылинку в воду. Теория броуновского движения предсказывает, что движение пылинки должно описываться так называемым винеровским процессом. Пусть $w(t)$ – функция, выражающая зависимость одной из координат пылинки от времени t . Можно ли по одной траектории $w(t)$ проверить вероятностную модель? Ответ – да. Для этого достаточно вычислить $X_1 = w(1)$, $X_2 = w(2) - w(1)$, $X_3 = w(3) - w(2)$, ... Если вероятностная модель верна, то $X_1, X_2, X_3, \dots$ – независимые случайные величины со стандартным нормальным распределением (с математическим ожиданием 0 и дисперсией 1).

Нормальность распределения проверяют с помощью критериев согласия (см., например, [9]).

Наличие «статистического ансамбля», «статистической однородности», «статистической устойчивости» достаточно для возможности применения вероятностных методов, но не является необходимым. Условие применимости вероятностно-статистических методов – не «статистическая устойчивость», а наличие адекватной вероятностной модели явления.

Вероятностную модель реального явления (функционирования технологического оборудования, получения результатов эксперимента, течения заболевания и т.д.) следует считать заданной, если рассматриваемые величины и связи между ними выражены с помощью понятий теории вероятностей.

При необходимости переноса выводов с обследованной совокупности объектов на более широкую совокупность следует построить адекватную вероятностную модель и применять вероятностно-статистические методы.

Вероятностные модели могут задавать как правила отбора единиц совокупности в выборку, так и вероятностное описание свойств отобранных единиц. Более широкая совокупность, на которую переносятся сделанные по выборке выводы, может быть как реально существующей (например, партия изделий), так и используемой в вероятностной модели явления или процесса (например, совокупность возможных значений контролируемого параметра).

В вероятностной модели выборки результаты наблюдений $x_1, x_2, \dots, x_n$ рассматриваются как реализации независимых одинаково распределённых случайных величин $X_1, X_2, \dots, X_n$. При этом результаты наблюдений $x_1, x_2, \dots, x_n$ могут быть числами, векторами (элементами конечномерных пространств), функциями, объектами нечисловой природы

(бинарными отношениями – ранжировками, разбиениями, толерантностями и др.); множествами; нечёткими множествами; измерениями в качественных шкалах (шкалах наименований и порядка; и т.д.).

Примечание. Специфика исследуемого реального явления может потребовать использования более сложной модели выборки, когда выборочные значения $X_1, X_2, \dots, X_n$ зависимы и (или) неодинаково распределены и (или) цензурированы, и т.д.

Математическая статистика основана на теории вероятностей, основные понятия которой используем в соответствии со справочниками [9, 10] и энциклопедией [11]. В этих книгах рассмотрены понятия случайной величины со значениями в пространстве произвольной природы, её распределения, независимости случайных величин и т.д. Подчеркнём, что значениями случайных величин могут быть не только действительные числа, но и вектора, функции, множества и различные виды объектов нечисловой природы.

Запросы прикладных исследований привели к необходимости обработки статистических данных неклассического вида, т.е. к необходимости развития статистики нечисловых данных, известной также как нечисловая статистика и статистика объектов нечисловой природы [12]. Под объектами нечисловой природы понимают элементы нелинейных пространств, в частности, бинарные отношения (толерантности, разбиения, ранжировки и др.), результаты парных и множественных сравнений, множества, нечёткие множества [13], измерения в шкалах, отличных от абсолютных, тексты и т.д. В статистике нечисловых данных такие задачи математической статистики, как описание данных, оценивание, проверка гипотез, рассматриваются для результатов наблюдений, являющихся объектами нечисловой природы. Статистика нечисловых данных находит применение в различных областях технических исследований, в

экономике, управлении, организационных системах, медицине, психологии, социологии и других областях [12].

Различным аспектам разработки и применения статистических методов посвящены миллионы публикаций, в том числе сотни работ автора настоящей монографии [5]. Укажем здесь лишь наиболее важную, на наш взгляд, монографию по классической математической статистике [14].

Для обоснования вероятностной модели выборки необходимо проверить с помощью статистических критериев независимость и одинаковую распределённость результатов наблюдений, либо обосновать эти свойства наблюдений условиями их проведения. Статистические критерии независимости и одинаковой распределённости рассмотрены, например, [15]. Выбор конкретного критерия зависит от вида альтернативной гипотезы. Вероятностную модель выборки целесообразно обосновывать, исходя из условий проведения наблюдений – независимости их проведения (здесь «независимость» в общенаучном смысле, а не в смысле теории вероятностей), одинаковых условий для всех наблюдений. Следует иметь в виду, что по ограниченному числу результатов наблюдений невозможно достоверно установить независимость или одинаковую распределённость, поэтому не рекомендуется применять статистические критерии, если соображения прикладной области приводят к сомнению о справедливости рассматриваемой вероятностной модели выборки.

Вероятностные модели выборки, в которых функция распределения результатов наблюдений принадлежит определённому параметрическому семейству, применяют, когда вид параметрического семейства обоснован теоретически (из соображений прикладной области) и согласие опытного (эмпирического) распределения с параметрическим семейством проверено

с помощью статистических критериев. Распределения реальных данных, как правило, не являются нормальными (гауссовскими).

Установление того, что в некоторой прикладной области можно рассматривать как выборку из определённого параметрического семейства, является фундаментальным результатом, имеющим большое научное и практическое значение, поскольку это позволяет применять более эффективные методы оценивания и более мощные критерии проверки статистических гипотез. Так, в многих случаях изучаемую случайную величину можно рассматривать как совокупный результат действия многих малых причин. Если эти причины действуют аддитивно, то в силу Центральной предельной теоремы распределение изучаемой случайной величины близко к нормальному. Если мультипликативно – то к логарифмически нормальному. Длительность промежутка между последовательными разладками станка (или отказами аппаратуры) целесообразно приближать экспоненциально распределённой случайной величиной, если есть основания полагать, что имеется много независимых между собой причин отказа. Распределение Вейбулла-Гнеденко появляется, когда поведение системы определяется минимумом или максимумом описывающих её параметров [16]. Исходя из некоторой системы аксиом, А.Н. Колмогоров установил, что распределение размеров частиц при дроблении является логарифмически-нормальным [17].

По имеющемуся объёму экспериментальных данных отнюдь не всегда удаётся отличить одно семейство распределений от другого. Так, максимум модуля разности функций нормального и логистического распределений не превосходит 0,01, а потому для различения этих распределений требуется несколько тысяч наблюдений [18].

Вероятностная модель является параметрической, если все используемые в ней распределения вероятностей задаются k действительным числами, причём k , называемое размерностью модели, не

зависит от объёма выборки. Если распределения вероятностей нельзя задать указанным способом, то вероятностная модель является непараметрической, а основанные на ней статистические методы – непараметрическими. Применения непараметрических методов в параметрических моделях обычно нецелесообразно, поскольку параметрические методы, использующие специфику модели, обычно имеют более хорошие статистические свойства. Изучение параметрических методов в непараметрических моделях необходимо, в частности, для определения степени устойчивости статистических выводов по отношению к допустимым отклонениям исходных данных и предпосылок модели.

Наряду с непараметрическими моделями проверки гипотез разработаны непараметрические методы оценивания. В частности, непараметрической оценкой математического ожидания является выборочное среднее арифметическое. В силу закона больших чисел эта оценка является состоятельной. Асимптотические доверительные интервалы могут быть построены на основе Центральной предельной теоремы, а остаточный член оценён с помощью неравенства Берри-Эссеена [10].

Если нет достаточных оснований для принятия параметрической модели, следует применять непараметрические статистические методы. При неизвестном виде распределения данных не допускается применять распределение конкретного вида (нормальное, равномерное и др.). Исключение составляют случаи, когда минимаксными или иными методами удаётся найти распределение, которое в рассматриваемой статистической ситуации ведёт к наибольшим средним потерям.

Специальный вид альтернативной гипотезы при проверке статистических гипотез должен быть строго обоснован. При отсутствии

подобного обоснования следует рассматривать альтернативную гипотезу общего вида.

Так, для проверки гипотезы однородности двух независимых выборок, т.е. гипотезы о том, что функции распределения элементов двух выборок совпадают, иногда используют критерий Вилкоксона (Манна-Уитни). Этот критерий обоснован для модели, в которой функции распределения двух независимых выборок отличаются только сдвигом. Возможность применения этой модели при обработке конкретных данных должна быть обоснована. Если обоснования нет, то для обработки данных необходимо использовать не критерий Вилкоксона, а какой-либо критерий, состоятельный при альтернативе общего вида, например, двухвыборочный критерий Н.В. Смирнова или критерий типа омега-квадрат Лемана-Розенблатта [19].

Требования к результатам математической статистики

Основаниями для результатов (правил, алгоритмов) математической статистики, связанных с распределениями статистик, являются:

- предельные теоремы;
- предельные теоремы с оценкой остаточных членов;
- точные распределения при конечных объёмах выборок.

Предельные теоремы необходимо использовать тогда, когда неизвестно распределение рассматриваемой статистики при конечном объёме выборки и выбор статистического метода опирается на асимптотическую теорию. В частности, оценки параметров должны быть состоятельными и, по возможности, несмещёнными и эффективными.

Требование несмещённости не является необходимым. Широко используемые оценки максимального правдоподобия и одношаговые оценки, как правило, являются смещёнными, но при увеличении объёма выборки смещение стремится к 0.

Методы, опирающиеся только на предельные теоремы, лишь незначительно превосходят по обоснованности методы анализа данных. При разработке НТД целесообразно изучить скорость сходимости методом статистических испытаний. Если это невозможно, то в НТД в связи со скоростью сходимости должны быть даны точные ссылки на публикации либо даны явные указания на то, что о точности аппроксимаций нет сведений.

Предельные теоремы с оценкой остаточных членов используют для установления точности аппроксимаций. Показателем точности аппроксимаций для методов проверки гипотез является относительное отклонение уровня значимости от номинального. Показателем точности аппроксимаций для методов оценки параметров является относительное отклонение доверительной вероятности от номинальной.

Номинальный уровень значимости – уровень, который используется в предельной теории и задаётся в НТД. Он может отличаться от уровня значимости критерия, в частности, из-за того, что распределение статистики этого критерия при конечном объёме выборки может отличаться от предельного распределения. Уровень значимости критерия может отличаться от номинального также из-за дискретности распределения статистики критерия. Так, в [14] приведены уровни значимости критерия Смирнова, соответствующие обычно применяемым уровням значимости.

Наиболее предпочтительными для использования при обработке данных и включения в НТД являются вероятностно-статистические методы, основанные на точных распределениях рассматриваемых статистик при конечных объёмах выборок.

Алгоритмы расчётов, используемые при обработке данных вероятностно-статистическими методами, должны быть обоснованы по крайней мере одним из указанных выше способов.

Требования к адекватности относительно допустимых преобразований шкал измерения

Метод обработки данных является адекватным относительно допустимого преобразования шкал измерения, если полученные выводы не меняются при переходе от измерений в исходной шкале к измерениям в преобразованной шкале.

Рассматриваемые результаты наблюдений – действительные числа (а не вектора или объекты иной природы). Группа допустимых преобразований шкал задаёт тип шкалы измерения (порядковая, интервалов, отношений и др.). Термин «группа» используется в смысле, принятом в высшей алгебре. Метод обработки данных является адекватным в шкале определённого типа, если он является адекватным относительно любого допустимого преобразований шкалы, входящего в группу, задающую рассматриваемый тип шкалы измерения.

Выделение типа шкалы, по которой измерена определённая переменная – задача прикладной области. В случае разногласий рекомендуется использовать методы, адекватные относительно более широкой группы допустимых преобразований. С прогрессом науки и техники может меняться тип шкалы, по которой измеряется определённая переменная. Так, температура вначале измерялась в порядковой шкале. С изобретением термометров Цельсия, Фаренгейта, Реомюра тип шкалы стал интервальным, так как переход от одной из указанных шкал к другой описывается линейным преобразованием. С открытием абсолютного нуля температур и введением шкалы Кельвина появилась возможность измерять температуру по шкале отношений.

При вычислении средних величин для совокупностей результатов наблюдений следует использовать средние, результат сравнения которых для двух совокупностей является адекватным в шкале, в которой получены

рассматриваемые результаты наблюдений. Так, из средних по Колмогорову можно использовать в шкале интервалов только среднее арифметическое, а в шкале отношений – только степенные средние.

Для обработки данных, измеренных в порядковой шкале, применяют ранговые методы математической статистики. Методы обработки данных, в которых используется т.н. «оцифровка» (приписывание числовых значений градациям порядковой или номинальной шкал), следует рассматривать как методы анализа данных, если нет строгого обоснования принципиальной возможности измерения по количественным шкалам.

В наших учебниках, как правило, имеются главы «Теория измерений и средние величины» (см., например, [18, 19]), к которым и рекомендуется обратиться за подробностями по рассматриваемой тематике..

Требования и характеристики точности и устойчивости выводов

Точность методов оценивания выражается с помощью доверительных интервалов.

Точность методов проверки гипотез выражается с помощью уровней значимости и соответствующих им процентных точек, задающих критические области.

Для обоснования выводов о точности методов обработки данных и выбора наиболее точного метода должно быть изучено влияние погрешностей наблюдений на окончательные выводы. Конкретные алгоритмы разработаны в статистике интервальных данных как части прикладной статистики.

Для установления точности расчётов и выбора наиболее точного алгоритма должно быть изучено влияние ошибок округления на результаты расчётов, в частности, с помощью компьютеров. Надо иметь в виду, что равносильные преобразования формул могут привести к

принципиально иному значению ошибок округления. Примером является формула для выборочной дисперсии:

$$s^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - (\bar{x})^2.$$

Если результаты наблюдений велики, но имеют малый разброс, то ошибки округления для первого выражения много меньше, чем для второго. Это связано с тем, что в правой части рассматриваемой формулы стоит разность двух больших чисел.

Как ранее сказано, должна быть установлена степень устойчивости статистических выводов, получаемых с помощью определённого метода обработки данных, относительно допустимых отклонений исходных данных и предпосылок используемой модели. При изучении устойчивости статистических процедур на основе модели выборки рекомендуется использовать методы робастной статистики. Они основаны на модели малых отклонений функций распределения элементов выборки. Частным её случаем является модель засорения, идейно близкая к модели выбросов, в рамках которой разработаны методы отбраковки грубых ошибок. Проблемам устойчивости организационно-экономических моделей и методов посвящены наши монографии [20, 21] и многочисленные статьи (см. сводку [5]).

Если результаты статистической обработки результатов наблюдений используются как исходные данные в дальнейших рассуждениях и расчётах, то необходимо учитывать при дальнейшем анализе точность и устойчивость этих результатов. В частности, в выборочных исследованиях с ответами типа «да» - «нет» надо указывать ошибку выборки и значимость различия долей. Не допускается преувеличивать точность статистических выводов. Если точечные оценки параметров использовать как точные значения этих параметров, то это может привести к неверным выводам. Необходимо различать, в том числе в обозначениях, выборочное

среднее арифметическое и математическое ожидание, а также выборочную дисперсию и теоретическую дисперсию.

Характеристики последовательного применения методов обработки данных

При последовательном применении нескольких методов обработки данных необходимо обеспечить проверку условий применения каждого последующего метода, в частности, когда первым применяется алгоритм, осуществляющий:

- заполнение пропусков в таблицах «объект – признак»;
- преобразование переменных;
- классификацию многомерных наблюдений;
- оценивание размерности и структуры модели в многомерном статистическом анализе.

Если при заполнении пропусков в таблицах «объект – признак» для восстановления значения признака некоторого объекта используется информация о других объектах, то в результате работы алгоритма заполнения пропусков случайные вектора, соответствующие отдельным объектам, не являются независимыми. Следовательно, дальнейшее применение методов, основанных на предположении, что наблюдения представляют собой выборку, даёт методы анализа данных, а не вероятностно-статистические.

Если параметры преобразования переменных определяются по результатам наблюдений, то преобразованные данные нельзя рассматривать как выборку, а потому методы, основанные на применении к преобразованным данным алгоритмов анализа выборки, следует считать методами анализа данных.

Элементы кластеров, выделенных с помощью методов классификации, зависимы и имеют распределения, отличные от нормальных. Поэтому методы многомерного статистического анализа, основанные на нормальной теории, применительно к элементам таких кластеров следует считать методами анализа данных. Методы дискриминантного анализа позволяют выделить кластеры с помощью дискриминантных поверхностей, которые при увеличении объёмов выборок стремятся к предельным поверхностям. Совместное распределение любого конечного числа элементов такого кластера стремится к распределению независимых случайных величин, имеющих соответствующее усечённое распределение (но не нормальное, даже если исходные классы порождались многомерными нормальными распределениями). Элементы кластера, выделенного с помощью дискриминантного анализа, асимптотически независимы, поэтому применительно к элементам такого кластера методы, не опирающиеся на предположение нормальности распределения, например, методы робастной регрессии, являются вероятностно-статистическими.

Методы оценивания размерности и структуры модели в многомерном статистическом анализе включают методы:

- оценки степени полинома в регрессионном анализе;
- выделения информативного подмножества признаков в многомерной регрессии и дискриминантном анализе;
- определения числа элементов смеси при расщеплении смеси при распознавании образов без учителя;
- определения размерности пространства в факторном анализе, методе главных компонент и многомерном шкалировании, и т.д.

Перечисленные методы в ряде случаев не являются вероятностно-статистическими, так как дают несостоятельные оценки соответствующих параметров и множеств. В этих случаях все методы обработки данных,

применяемые вслед за перечисленными, являются методами анализа данных. Если же методы оценивания размерности и структуры модели дают состоятельные оценки соответствующих параметров и множеств, то их следует рассматривать как вероятностно-статистические. В этом случае выполнены условия применимости последующих вероятностно-статистических методов регрессионного анализа, дискриминантного анализа и др.

Нами показано, что распространённые методы оценки степени полинома в регрессионном анализе являются несостоятельными, и предложены состоятельные оценки [18]. Большинство задач многомерного статистического анализа может быть представлено в виде задач оптимизации. В них могут быть использованы общие результаты о поведении решений экстремальных статистических задач [12]. Анализ методов оценивания размерности вероятностно-статистической модели проведён в [22].

Характеристики оптимальности методов обработки данных

Понятие оптимальности, применяемое при выборе и описании метода обработки данных, следует определять и описывать в тексте НТД.

Результаты статистического анализа данных применяют в двух ситуациях:

- результаты используются для принятия решений в полностью описанной процедуре и последствия тех или иных решений известны;
- процедура принятия решений не задана или описана частично.

В первой ситуации метод обработки данных оценивают по его соответствию процедуре принятия решений, оптимальность понимают как максимизацию эффекта или минимизацию затрат. Свойства метода изучают с помощью теории статистических решений.

Во второй ситуации оптимальность метода обработки данных понимают в соответствии с методологией прикладной статистики. Так, под показателем качества статистической оценки $\hat{\beta}$ некоторого параметра β следует понимать средний квадрат ошибки $M(\hat{\beta} - \beta)^2$, т.е. оптимальным является метод, для которого средний квадрат ошибки минимален.

В качестве примера рассмотрим задачу дискриминантного анализа (синоним – диагностики). Если известны потери от неправильной классификации и априорные вероятности классов, то решающее правило находят из условия минимизации математического ожидания потерь. Если же указанная информация отсутствует, то задачу дискриминантного анализа рассматривают согласно теории проверки статистических гипотез с использованием таких понятий, как ошибки первого и второго рода, и выбор уровня значимости производится специалистом прикладной области субъективно. В качестве показателя качества алгоритма диагностики следует использовать т.н. «прогностическую силу» [18]. Средний квадрат ошибки является суммой дисперсии оценки и квадрата её смещения, поэтому он лучше характеризует качество статистической оценки, чем каждое из слагаемых в отдельности. Ряд критериев оптимальности в случае оценивания многомерного параметра приведён в [23].

При проверке статистических гипотез оптимальность критерия зависит от вида альтернативной гипотезы. Если нет оснований для выбора определённого вида альтернативной гипотезы, то понятие оптимальности в большинстве статистических задач применять нецелесообразно, следует использовать понятия допустимости, состоятельности, несмещённости и т.п.

Представление ряда задач прикладной статистики как задач минимизации некоторых функций от результатов наблюдений позволяет изучать свойства соответствующих методов обработки данных и

разрабатывать алгоритмы расчётов, но не может служить решающим аргументом при выборе статистического метода.

Характеристики трудоёмкости и наглядности

Для обеспечения возможности распространения и использования метода обработки данных в НТД должны указываться характеристики его трудоёмкости. Часто трудоёмкость статистического исследования пропорциональна объёму выборки. Правила выбора необходимого объёма выборки систематизированы в [24]. Как показывает опыт применения статистических методов, доверительные интервалы, как правило, шире, чем это интуитивно кажется. Поэтому вычисление доверительных границ необходимо, чтобы исключить преувеличение точности метода.

С точки зрения прикладной статистики целесообразно рассматривать следующие характеристики трудоёмкости (в зависимости от объёма данных):

- затраты времени на подготовку данных к расчётам;
- затраты времени на осуществление расчётов, а также характеристики вычислительных средств и количество требуемого машинного времени;
- затраты времени на анализ и интерпретацию результатов расчётов, на подготовку итогового документа для потребителя (заказчика);
- прочие затраты различных видов ресурсов, необходимых для проведения расчётов.

Третий из перечисленных видов затрат времени связан с тем, что итоговый компьютерный документ может требовать более или менее трудоёмкой ручной обработки с целью получения научных и практических выводов. Например, на печать выдаётся матрица выборочных коэффициентов корреляции, а затем вручную отбираются коэффициенты, значимо отличающиеся от 0. При проведении ряда однотипных расчётов

целесообразно выводить данные на печать непосредственно в виде итогового документа, направляемого заказчику.

С целью минимизации трудозатрат целесообразно осуществлять ввод данных в компьютер непосредственно со средств измерения, разрабатывать программы печати итоговых документов и т.п.

При проведении единичных и (или) простых обработок статистических данных ручной счёт может оказаться менее трудоёмким, чем разработки и (или) применение программных продуктов.

Трудоёмкость и наглядность являются менее важными характеристиками метода обработки данных, чем устойчивость и оптимальность, особенно на этапе обоснования статистического решения. Так, при оценивании параметров целесообразно использовать оценки максимального правдоподобия и одношаговые оценки, хотя они и более трудоёмки, чем оценки метода моментов.

При подготовке НТД и итоговых документов для потребителя (заказчика) следует обращать внимание на наглядность, облегчающую восприятие результатов, а также получение выводов, их формулировку в терминах конкретной прикладной области.

При анализе данных целесообразно использовать не только расчётные, но и наглядные методы, в частности, графические и методы визуализации данных. Сравнение аналитических и графических методов оценки параметров распределений вероятностей по объективности, точности, наглядности и трудоёмкости позволяет в каждом конкретном случае обоснованно выбирать те или иные методы. Графические методы целесообразно использовать при предварительном (разведочном) анализе, а аналитические – при массовой обработке данных.

Требования к нормативно-технической и справочной документации по методам обработки данных с помощью компьютеров

При описании в нормативно-технической или справочной документации нового метода обработки данных, реализованного в виде программного продукта, или результатов апробации известных методов, следует придерживаться правил настоящего раздела. Они развивают положения единой системы программной документации применительно к реализованным на компьютере методам прикладной статистики. Описания ранее реализованных методов обработки данных приводят в соответствие с требованиями настоящего раздела при пересмотре документации.

При использовании компьютерного метода обработки данных в НТД приводят:

- программы, реализующие данный метод, составленные на одном из языков высокого уровня;
- точность представления чисел в компьютере;
- информацию об отклонениях значений минимизируемой функции от её глобального минимума, определяемых при решении тестовых задач с заранее известными ответами. Целесообразно приводить не только окончательные результаты расчётов, но и промежуточные.

Для облегчения программной реализации метода обработки данных и обеспечения возможности сравнения результатов различных пользователей в НТД приводят дополнительную информацию о компьютерной реализации метода. В частности, для задач оптимизации приводят согласно [25]:

- тип компьютера, длину машинного слова, сведения о трансляторе и т.д.;
- различные характеристики точности приближения (отклонение приближения от истинного значения, т.е. расстояние от приближения до истинного значения; невязку в ограничениях и выполнении условий экстремума и т.п.; в задачах малой (до 10) размерности целесообразно приводить и сами приближения);

- сведения о трудоёмкости вычислений (число итераций, число вычислений значений функции, её градиента, затраты машинного времени, необходимый объем памяти и т.д.).

Для обеспечения правильности принимаемых пользователем решений о выборе того или иного метода для обработки конкретных данных необходимо указывать области применимости метода. Если для решения конкретных задач можно использовать ряд методов, необходимо обосновать необходимость использования нового метода и указать области применимости существующих методов. Понятие «область применимости метода» включает в себя:

- вероятностную модель, на которой основан метод, или, для методов анализа данных, вид исходных данных (числа, вектора, множества и т.д.);

- вид получаемого решения (принятие или отклонение гипотез, точечная оценка числового параметра, доверительная область для функции, разбиение множества объектов на кластеры и т.д.);

- прикладные области, в которых рассматриваемый метод позволяет получать научные и практические выводы (управление качеством, надёжность в технике, клиническая медицина, заводская социология и т.д.), а также тип выводов в терминах прикладной области (а не в терминах математической статистики). Перечень прикладных областей, как правило, отражает определённый этап внедрения метода и с течением времени расширяется.

Для возможности сравнения результатов различных исследований необходимо приводить точную формулировку задачи, для которой проводились расчёты на компьютере (включая все параметры задачи). В число параметров задачи входит и начальное приближение.

Для обеспечения потребностей специалиста, использующих НТД по методам обработки данных, необходимо давать подробное описание применяемого алгоритма или отсылать к его публикации. Применяемый

алгоритм должен быть описан на естественном языке и оформлен в виде программы на одном из принятых алгоритмических языков публикаций. Программный текст может быть заменён ссылкой на источник, где эта программа содержится. Необходимо указать, является метод вероятностно-статистическим или методом анализа данных, дать необходимые обоснования целесообразности его применения либо отослать к источникам. Кроме того, в НТД следует указывать перечисленные выше характеристики метода и правила прекращения вычислений, а также, являются эти правила эвристическими или теоретически обоснованными.

При использовании итерационного алгоритма должно быть отмечено, доказана ли его сходимость теоретически или же она обоснована эвристически. В последнем случае должны быть приведены обоснования или ссылка на соответствующий литературный источник. Следует учесть, что даже если сходимость теоретически доказана, то практически из-за накопления ошибок при вычислении значений функций, округления чисел в компьютере и т.д. она может нарушаться. Поэтому, кроме теоретического обоснования, необходима проверка практической сходимости.

Не рекомендуется разрабатывать итеративные процедуры нахождения оценок максимального правдоподобия. Вместо подобных оценок следует использовать одношаговые. (Процедуры рекуррентного оценивания, применяемые при возрастании объёма выборки, не относятся к итеративным процедурам.) Одношаговые оценки параметров распределений вероятностей имеют те же асимптотические свойства, что и оценки максимального правдоподобия, но находятся за одну итерацию. Их использование существенно снижает трудоёмкость расчётов и избавляет от необходимости изучения сходимости итерационных процедур [18].

Если оценки максимального правдоподобия не выражаются в явном виде, целесообразно использовать одношаговые оценки. Итеративные

процедуры нахождения оценок максимального правдоподобия следует использовать только тогда, когда установлена их более высокая точность при конечных объёмах выборки.

В НТД приводят информацию о влиянии погрешностей компьютерных вычислений на точность получаемых результатов. Влияние погрешностей вычислений на точность результатов расчётов должно быть установлено в ходе выполнения научно-исследовательской работы по созданию НТД или определено на основе ранее проведённых исследований. Чтобы изучить влияние погрешностей компьютерных вычислений на точность получаемых результатов, необходимо применение численных алгоритмов.

Требования к оформлению результатов обработки данных

При описании результатов обработки конкретных данных указывают используемую вероятностную модель реального явления и обосновывают возможность её применения. Необходимо тщательно различать выборочные и теоретические характеристики (выборочное среднее арифметическое и математическое ожидание, выборочную и теоретическую дисперсии, и т.д.).

Для возможности сравнения результатов различных исследований указывают объёмы выборок, по которым определяют те или иные параметры и (или) характеристики (средние, выборочные дисперсии, доли и т.д.). Указывают также погрешности измерения исходных данных.

Используемый алгоритм расчёта должен быть полностью описан, либо дана точная ссылка на источник. Употребление выражений типа «расчёты проведены по общепринятой методике» не допускается. Если литературный источник практически недоступен пользователю НТД (например, выпущенная за рубежом монография), то алгоритм расчёта необходимо подробно описать в тексте НТД.

Для оцениваемых параметров и характеристик должны быть указаны не только точечные оценки, но и доверительные границы. Следует сделать ссылку на используемый метод вычисления доверительных границ. С появлением теории интервального оценивания и статистики интервальных данных только доверительные области (а не точечные оценки) следует рекомендовать для практического использования.

При описании результатов проверки статистических гипотез допустимо приводить заключение о том, принята или нет гипотеза на определённом уровне значимости, или значение достигаемого уровня значимости. Достигаемый уровень значимости – это значение уровня значимости, при котором происходит переход от отклонения к принятию рассматриваемой статистической гипотезы при обработке конкретных статистических данных (см. [9, 15]). Достигаемый уровень значимости – случайная величина. Если для статистики X статистического критерия используют критические области $\{X > c\}$ с уровнем значимости $P(X > c) = F(c)$, то достигаемый уровень значимости – это $F(x_0)$, где x_0 – значение статистики X , определённое по конкретным данным, т.е. $F(x_0) = P(X > x_0)$.

При описании результатов компьютерных расчётов должны быть указаны название языка программирования, тип компьютера, точность вычислений, затраты машинного времени. Перечисленные характеристики указывают в документации, посвящённой результатам обработки данных. Документацию по методам обработки данных оформляют согласно правилам предыдущего раздела.

При использовании метода статистического моделирования (синонимы – метода статистических испытаний, метода Монте-Карло) необходимо указывать датчик псевдослучайных чисел, число испытаний, доверительные интервалы для параметров, оцениваемых методом статистического моделирования. При описании датчика псевдослучайных

чисел целесообразно указывать также исходное псевдослучайное число (с которого начинается работа датчика).

Необходима дальнейшая разработка контроллинга статистических методов.

Литература

1. Orlov A. I. Basic requirements for statistical methods of data analysis // Polythematic Online Scientific Journal of Kuban State Agrarian University. 2022. N 181. P. 316-343.
2. Орлов А.И. Контроллинг статистических методов // Контроллинг. 2022. №86. С. 2-11.
3. Орлов А.И. О требованиях к статистическим методам анализа данных (обобщающая статья) // Заводская лаборатория. Диагностика материалов. 2023. Т.89. №11. С. 98-106.
4. Рекомендации. Прикладная статистика. Методы обработки данных. Основные требования и характеристики / Орлов А.И., Миронова Н.Г., Фомин В.Н., Черчинцев А.Н. - М.: ВНИИСтандартизации, 1987. - 62 с.
5. Орлов А.И. Шестьдесят лет в мире формул (1964 - 2023): Комментарии к списку научных и методических трудов. Изд. 4, испр. и доп. – М.: Институт высоких статистических технологий и эконометрики МГТУ им. Н.Э. Баумана, 2024б. – 524 с. [Электронный ресурс]. URL: <http://ukros.ru/archives/37928>, <https://ibm2.ru/library/> (дата обращения 31.03.2025).
6. Орлов А.И. Основные требования к методам анализа данных (на примере задач классификации) / Научный журнал КубГАУ. 2020г. №159. С. 239–267.
7. Миркин Б.Г. Анализ качественных признаков и структур. – М.: Статистика, 1980. – 320 с.
8. Колмогоров А.Н. Основные понятия теории вероятностей. Изд. стереотип. – М.: Ленанд, 2025. – 120 с.
9. Орлов А.И. Вероятность и прикладная статистика: основные факты: справочник. — М.: КноРус, 2009. — 192 с.
10. Прохоров Ю.В., Розанов Ю.А. Теория вероятностей. Основные понятия. Предельные теоремы. Случайные процессы. Изд. 3-е. – М.: Наука, 1987. – 400 с.
11. Вероятность и математическая статистика: Энциклопедия / Гл. ред. Ю.В. Прохоров. – М.: Большая российская энциклопедия, 1999. – 910 с.
12. Орлов А.И. Искусственный интеллект: нечисловая статистика. — М.: Ай Пи Ар Медиа, 2022а. — 446 с.
13. Орлов А.И. Задачи оптимизации и нечёткие переменные. - М.: Знание, 1980. - 64 с.
14. Большев Л.Н., Смирнов Н.В. Таблицы математической статистики. – М.: Наука, 1983. – 416 с.
15. Гаек Я., Шидак З. Теория ранговых критериев. – М.: Наука, 1971. – 376 с.
16. Гнеденко Б.В., Коваленко И.Н. Введение в теорию массового обслуживания. – М.: URSS, 2021. – 400 с.
17. Колмогоров А.Н. О логарифмически-нормальном законе распределения размеров частиц при дроблении // Доклады АН СССР. 1041. Т.XXXI. №2. С. 99-101.

18. Орлов А.И. Прикладной статистический анализ. — М.: Ай Пи Ар Медиа, 2022. — 812 с.
19. Орлов А.И. Эконометрика. — М.: Ай Пи Ар Медиа, 2024. — 525 с.
20. Орлов А.И. Устойчивость в социально-экономических моделях. — М.: Наука, 1979. — 296 с.
21. Орлов А.И. Устойчивые экономико-математические методы и модели : монография. — М.: Ай Пи Ар Медиа, 2022. — 337 с.
22. Орлов А.И. Оценивание размерности вероятностно-статистической модели // Научный журнал КубГАУ. 2020. №162. С. 1–36.
23. Налимов В.В., Голикова Т.И. Логические основания планирования эксперимента. Изд. 2-е. — М.: Металлургия, 1981. — 152 с.
24. Орлов А.И. Какой объем выборки целесообразно использовать? // Научный журнал КубГАУ. 2024. №03(197). С. 123–149.
25. Поляк Б.Т. Введение в оптимизацию. Изд. 3-е. — М.: URSS, 2023. - 404 с.

References

1. Orlov A. I. Basic requirements for statistical methods of data analysis // Polythematic Online Scientific Journal of Kuban State Agrarian University. 2022. N 181. P. 316-343.
2. Orlov A.I. Kontrolling statisticheskikh metodov // Kontrolling. 2022. №86. S. 2-11.
3. Orlov A.I. O trebovaniyah k statisticheskim metodam analiza dannyh (obobshchayushchaya stat'ya) // Zavodskaya laboratoriya. Diagnostika materialov. 2023. T.89. №11. S. 98-106.
4. Rekomendacii. Prikladnaya statistika. Metody obrabotki dannyh. Osnovnye trebovaniya i harakteristiki / Orlov A.I., Mironova N.G., Fomin V.N., ChErchincev A.N. - M.: VNIISStandartizacii, 1987. - 62 s.
5. Orlov A.I. SHest'desyat let v mire formul (1964 - 2023): Kommentarii k spisku nauchnyh i metodicheskikh trudov. Izd. 4, ispr. i dop. — М.: Institut vysokih statisticheskikh tekhnologij i ekonometriki MGTU im. N.E. Bauman, 2024b. — 524 s. [Elektronnyj resurs]. URL: <http://ukros.ru/archives/37928>, <https://ibm2.ru/library/> (data obrashcheniya 31.03.2025).
6. Orlov A.I. Osnovnye trebovaniya k metodam analiza dannyh (na primere zadach klassifikacii) / Nauchnyj zhurnal KubGAU. 2020g. №159. S. 239–267.
7. Mirkin B.G. Analiz kachestvennyh priznakov i struktur. — М.: Statistika, 1980. — 320 s.
8. Kolmogorov A.N. Osnovnye ponyatiya teorii veroyatnostej. Izd. stereotip. — М.: Lenand, 2025. — 120 s.
9. Orlov A.I. Veroyatnost' i prikladnaya statistika: osnovnye fakty: spravochnik. — М.: KnoRus, 2009. — 192 s.
10. Prohorov YU.V., Rozanov YU.A. Teoriya veroyatnostej. Osnovnye ponyatiya. Predel'nye teoremy. Sluchajnye processy. Izd. 3-e. — М.: Nauka, 1987. — 400 s.
11. Veroyatnost' i matematicheskaya statistika: Enciklopediya / Gl. red. YU.V. Prohorov. — М.: Bol'shaya rossijskaya enciklopediya, 1999. — 910 s.
12. Orlov A.I. Iskusstvennyj intellekt: nechislovaya statistika. — М.: Aj Pi Ar Media, 2022a. — 446 с.
13. Orlov A.I. Zadachi optimizacii i nechyotkie peremennye. - М.: Znanie, 1980. - 64 s.
14. Bol'shev L.N., Smirnov N.V. Tablicy matematicheskoy statistiki. — М.: Nauka, 1983. — 416 s.

15. Gaek YA., SHidak Z. Teoriya rangovyh kriteriev. – M.: Nauka, 1971. – 376 s.
16. Gnedenko B.V., Kovalenko I.N. Vvedenie v teoriyu massovogo obsluzhivaniya. – M.: URSS, 2021. – 400 s.
17. Kolmogorov A.N. O logarifmicheski-normal'nom zakone raspredeleniya razmerov chastic pri droblenii // Doklady AN SSSR. 1041. T.XXXI. №2. S. 99-101.
18. Orlov A.I. Prikladnoj statisticheskij analiz. — M.: Aj Pi Ar Media, 2022. — 812 c.
19. Orlov A.I. Ekonometrika. — M.: Aj Pi Ar Media, 2024. — 525 c.
20. Orlov A.I. Ustojchivost' v social'no-ekonomicheskikh modelyah. — M.: Nauka, 1979. — 296 s.
21. Orlov A.I. Ustojchivye ekonomiko-matematicheskie metody i modeli : monografiya. — M.: Aj Pi Ar Media, 2022. — 337 c.
22. Orlov A.I. Ocenivanie razmernosti veroyatnostno-statisticheskoy modeli // Nauchnyj zhurnal KubGAU. 2020. №162. S. 1–36.
23. Nalimov V.V., Golikova T.I. Logicheskie osnovaniya planirovaniya eksperimenta. Izd. 2-e. – M.: Metallurgiya, 1981. – 152 s.
24. Orlov A.I. Kakoj ob'em vyborki celesoobrazno ispol'zovat'? // Nauchnyj zhurnal KubGAU. 2024. №03(197). S. 123–149.
25. Polyak B.T. Vvedenie v optimizaciyu. Izd. 3-e. – M.: URSS, 2023. - 404 s.