ДОНСКОЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

УПРАВЛЕНИЕ ЦИФРОВЫХ ОБРАЗОВАТЕЛЬНЫХ ТЕХНОЛОГИЙ

Кафедра «Робототехника и мехатроника»

Учебное пособие

«Методы компьютерной обработки статистических данных»
по дисциплине

«Теория эксперимента в исследовании систем»

Авторы
Лукьянов Е. А., Ивацевич Ю. Б.,
Изюмов А. И., Васильева Е. В.

Ростов-на-Дону, 2020

Аннотация

Учебное пособие предназначено для студентов очной формы обучения направлений 15.04.06«Мехатроника и робототехника» и 09.04.01 «Информатика и вычислительная техника».

Авторы

к.т.н., доцент кафедры «Робототехника и мехатроника» Лукьянов Е.А.,
к.т.н., доцент кафедры «Робототехника и мехатроника» Ивацевич Ю.Б.,
к.т.н., доцент кафедры «Робототехника и мехатроника» Изюмов А.И.,
магистрант Васильева Е.В.

Оглавление

ВВЕДЕНИЕ

Владение основами статистических методов необходимо специалистам, работающим в различных областях науки и техники. В настоящее время разработано большое количество статистических пакетов- программ, разделенных на две основных группы: специализированные пакеты и пакеты общего назначения Среди универсальных статистических пакетов одним из лидеров является SPSS (SuperiorPerfomanceSoftwareSystem, что переводится, как «Система программного обеспечения высшей производительности») [1]. Пакет отличается гибкостью и мощностью применения для всех видов статистических расчетов. В России существует представительство компании SPSS, которое распространяет русифицированную версию пакета. На русском языке создан электронный учебник по применению данного пакета, изданы работы по его применению [2]. Среди универсальных систем статистического анализа данных широкое распространение получил пакет STATISTICA [3,4,5]. Пакет относится к числу базовых пакетов вузов России. Фирма разработчик имеет российское представительство [6], на сайте которого размещен электронный учебник по статистике, а также студенческая версия программы, распространяемая бесплатно. К числу достаточно мощных универсальных статистических пакетов относится также STATGRAPHICSPLUS [7], важнейшим достоинством пакета считаются хорошая интеграция математико-статистического аппарата обработки данных с современной интерактивной графикой и его динамичная эволюция с учетом развития компьютерных технологий.

Несмотря на разнообразие статистического программного обеспечения в России чаще всего используется программный комплекс (приложение) Microsoft Excel [4, 8, 9]. Это объясняется широким распространением русскоязычной версии данного ПО для персональных компьютеров. В программной среде MSOffice приложение MSExcel выполняет функции электронной таблицы с достаточно мощной математической поддержкой решения задач, в которой определенные статистические процедуры являются дополнительными встроенными формулами. Существует также макрос-дополнение XLSTAT-Pro[10, 11] для приложения MSExcel, включающее в себя более 50 статистических процедур.

Таким образом, благодаря современному уровню развития информационных технологий в распоряжении исследователей различных сфер науки и техники, экономики и производства, а также образовательных учреждений имеются доступные научные

и статистические пакеты программ, удовлетворяющие разнообразные потребности пользователей.

ГЛАВА 1. ТЕХНОЛОГИИ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ

Общая технология статистического анализа данных с использованием типового статистического пакета включает в себя следующие основные этапы:

- ввод исходных данных в электронную таблицу и их предварительное преобразование перед анализом;
- визуализация данных при помощи того или иного типа графиков;
- определение подходящих методов статистической обработки;
- применение конкретного метода статистической обработки;
- вывод результата анализа в виде графиков и электронных таблиц с численной и текстовой информацией;
- подготовка, печать и сохранение отчета.

1.1. Определение целей и типа экспериментальных исследований

Согласно общепринятому определению [12], экспериментом называется постановка опытных исследований, включающая наблюдение за протекающим исследованием в точно учитываемых условиях. Эксперимент должен обладать свойством точного воссоздания его при повторении условий. Исходя из этого, можно сделать вывод, что само понятие «эксперимент» означает действие по созданию требуемых условий, чтобы осуществить конкретное явление, при минимизации сторонних воздействий на него со стороны других явлений.

Цель эксперимента заключается в определении свойств исследуемых объектов, проверке обоснованности гипотез и изучении темы научного исследования, основываясь на полученных данных. Постановка и организация экспериментального исследования определяется его назначением.

Различают разные типы экспериментов:

1. по способу формирования условий (естественные и искусственно созданные условия);

2. по цели исследования:преобразующие, констатирующие, поисковые, контролирующие, решающие;
3. по организации проведения: лабораторные, производственные;
4. по структуре изучаемых объектов и явлений: простые, сложные;
5. по характеру внешних воздействий на объект исследования: вещественные, энергетические, информационные;
6. по характеру взаимодействия средства экспериментального исследования с объектом исследования: обычный и модельный;
7. по числу варьируемых факторов (однофакторный и многофакторный);
8. по характеру изучаемых объектов или явлений (технологические, социометрические) и т.п.

Для классификации могут быть использованы и другие признаки, определяющиеся техническими условиями.

1.1.1. Обработка статистических данных при помощи MSOfficeExcel

ПрограммаMSOfficeExcel применяется в тех случаях, когда требуется быстрая обработка больших объёмов данных. Она полезна операций статистической обработкиданных, их последующего анализа, решения задач оптимизации, построения выходных диаграмм и графиков. Для такого рода задач применяют как основные средства программного продуктаMSOfficeEXCEL, так и дополнительные надстройки

Для расчетного анализа данных используются отдельные библиотеки модулей. Под модулем понимается внешняя процедура или программа на языке программирования высокого уровня, удовлетворяющая некоторым дополнительным ограничениям, наиболее важными из которых являются:

- ограничения на способ аварийного завершения работы модуля;
- на способы связи по информации, например, на допустимость переменных внешнего типа и использование общей области памяти;
- на возможность передачи управления между модулями с помощью операторов вызова, расположенных в теле модуля;
- на использование операторов ввода-вывода.

Наиболее типовые расчетные модули современных статических пакетов условно делятся на три группы:

- описательная статистика и разведочный анализ исходных данных;
- статистическое исследование зависимостей;
- вспомогательные программы.

При помощи модуля описательной статистики и разведочного анализа исходных данных можно решить ряд следующих задач:

- анализ резко выделяющихся наблюдений;
- проверка статистической независимости рядов наблюдений;
- определение основных числовых характеристик и частотная обработка исходных данных (построение гистограмм, полигонов частот, вычисление выборочных средних, дисперсий и т.д.);
- расчет критериев однородности (средних, дисперсий, законов распределения и т.д.);
- определение критериев согласия (хи-квадрат, Колмогорова- Смирнова и др.);
- статистическое оценивание параметров;
- вычисление наиболее распространенных законов распределения вероятностей (нормального, Пуассона, хи-квадрат и некоторых других);
- визуализация анализируемых многомерных статистических данных.

Модуль статистического исследования зависимостей занимает достаточно объемную частью типового статистического пакета. Такой модуль включает решение следующих задач:

- корреляционно регрессионный анализ;
- дисперсионный анализ;
- планирование регрессионных экспериментов, выборочных обследований и др.

Вспомогательные программы расширяют возможности статистических пакетов и реализуют, в частности, оптимизационные алгоритмы, вычислительные процедуры, основанные на нейросетях, и генетических алгоритмах, задачи статистического моделирования на ЭВМ, которые являются полезными составными элементами компьютерных имитационных экспериментов, используемых при анализе сложных реальных систем.

1.1.2. Определение числовых характеристик экспериментальных данных при помощи пакета MSExcel

Необходимо выполнить следующую последовательность действий:

1. Запишите на рабочем листе Excel экспериментальные данные, согласно ячейкам: в столбец или в строку, после чего установите курсор на той ячейке,в которую будет занесено рассчитанное значение функции. Удобнее всего установить его в том же столбце, через ячейку ниже от введённых данных
2. При помощи кнопки «Вставка функции»на стандартной панели инструментов вызовем диалоговое окно «Мастер функций»(рисунок 1).

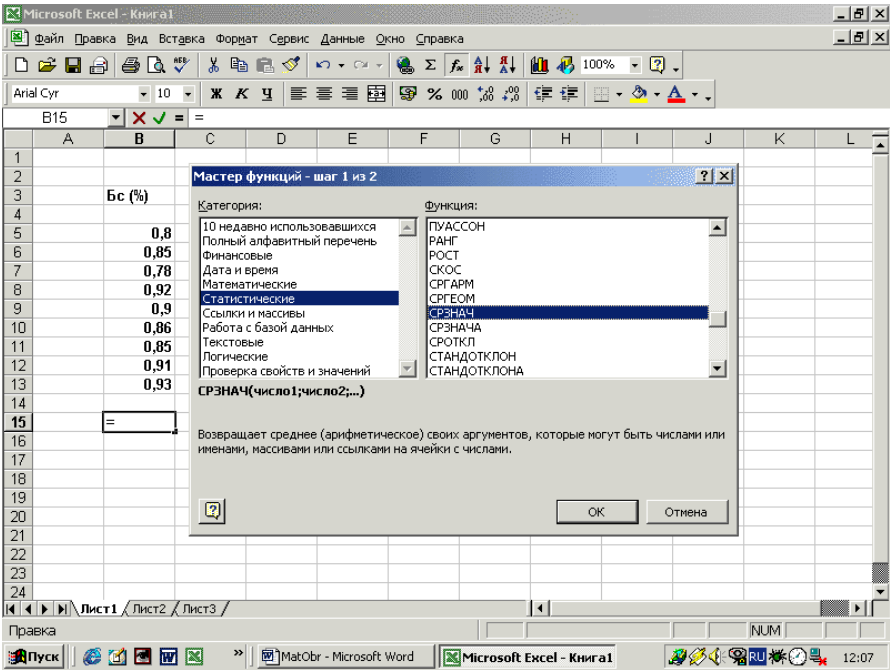


Рисунок 1 – Диалоговое окно «Мастер функций»

3. В диалоговом окне выберете категорию «Статистические».
4. Выберите необходимую функцию из списка, которая будет использоваться для последующей обработки данных. После подтверждения появится диалоговое окно (рисунок 2).

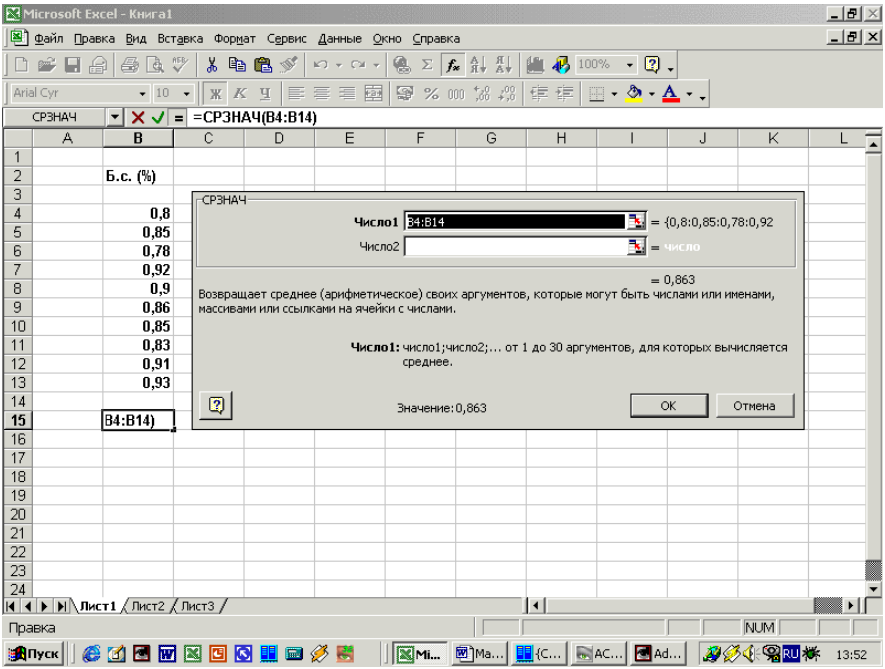


Рисунок 2 – Параметры настройки функции

5. В окне выбранной функции в нужный ряд чисел вводят данные. Для этого данные выделяют на листе, после чего те автоматически заносятся в требуемый ряд чисел.
6. В нижней части окна выводится рассчитанная числовая характеристика. При подтверждении она выводится на рабочий лист.

1.1.3. Построение экспериментального графика и диаграммы

Графическое представление позволяет облегчить восприятие и интерпретацию экспериментальных данных. Такое представление часто упрощает анализ и сравнение данных эксперимента. Результаты представлены диаграммами в виде полос, линий, столбиков, секторов, точек и в иной форме. Пакет Excel позволяет создавать диаграммы в виде внедрённых диаграмм и диаграммных страниц.

Внедрённые диаграммы – это диаграммы, наложенные на рабочий лист, содержащий таблицы с данными. Такие диаграммы

сохраняются вместе с таблицей в одном файле. Для диаграммных страниц, создаются отдельные графические файлы. Внедрённые диаграммы создаются при помощи«Мастера диаграмм» (рисунок 3).

Мастер Диаграмм - это последовательность диалоговых окон, которая позволяет сделать все необходимые шаги для создания новой диаграммы или для изменения установок уже существующей диаграммы. Мастер Диаграммы вызывают с помощью кнопки расположенной на стандартной панели инструментов.

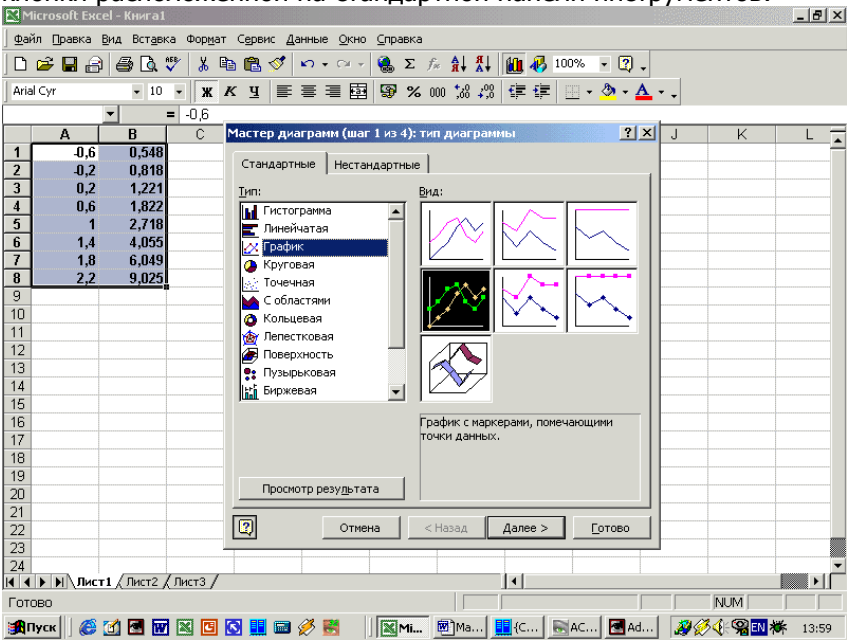


Рисунок 3 – Диалоговое окно «Мастер диаграмм»

Построение диаграмм включает последовательность от двух до пяти шагов, в зависимости от области выделенных данных. Если используется уже существующая диаграмма, то реализуются два шага.Если выделены данные на рабочем листе, то выполняются все пять шагов, поскольку в этом случае создаётся новая диаграмма.

1. На начальном этапе выбирают форму диаграммы. Доступные формы перечислены в списке «Тип»на вкладке «Стандартные»(Рисунок 3). Для выбранного типа диаграммы справа указываются несколько вариантов представления данных в палитре «Вид», из которых следует выбрать наиболее подходящий условиям эксперимента.

2. На следующем этапе выделяют диапазон данных, по которым будет строиться график. В случае если используется тип диаграммы«График», то для установления зависимости (Рисунок 4) следует установить параметр«Ряды в столбцах».Вкладка«Ряд»в диалоговом окне позволяет убрать выделенный курсором ненужный ряд. Следует учитывать, что, например, при выборе типа диаграммы«Точечная»данные процедуры не нужны.

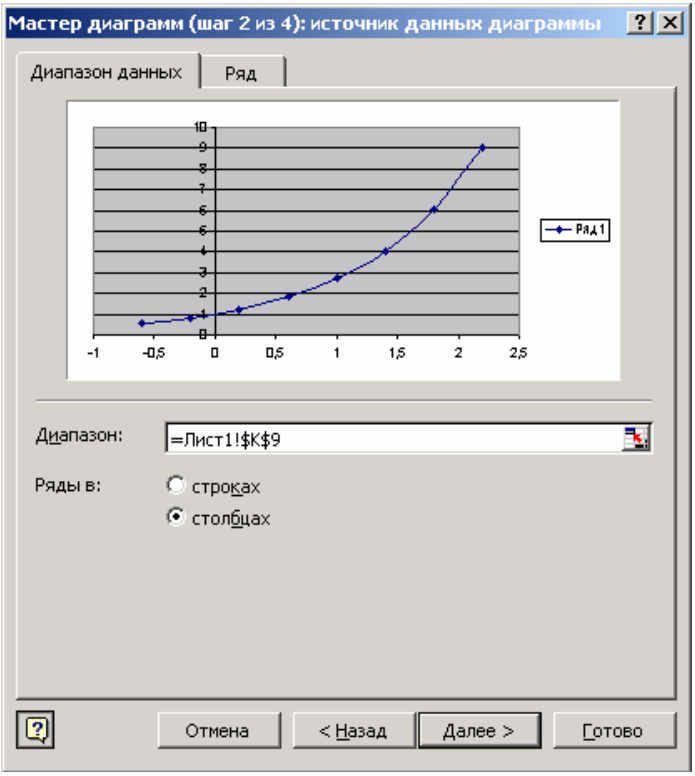


Рисунок 4 – Источник данных диаграммы

3. После этого необходимо перейти в окно «Параметры диаграммы» (Рисунок 5), щелкнув по кнопке «Далее». Вкладки данного окна позволяют оформить диаграмму.

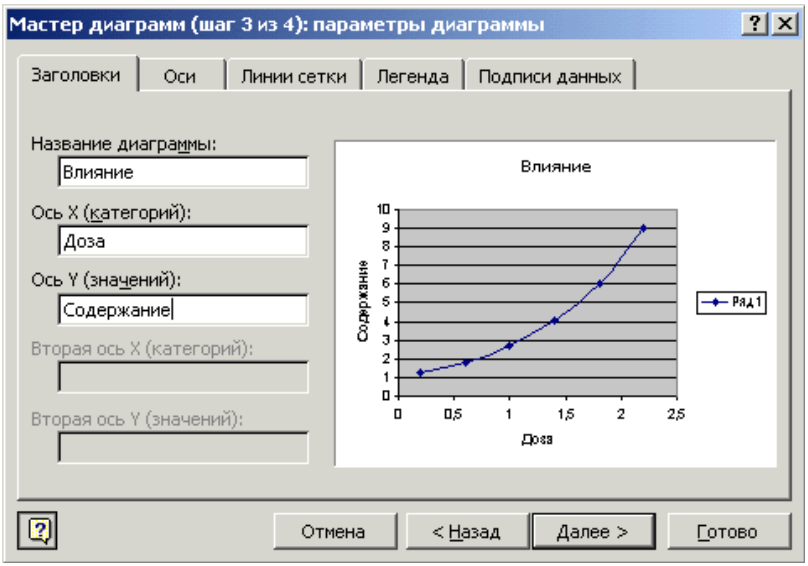


Рисунок 5 – Параметры диаграммы

4. Для установления функциональной зависимости экспериментальных данных необходимо выделить построенную диаграмму. При этом в меню приложения Excel появляется новый пункт – «Диаграмма». Затем необходимо вызвать команду «Диаграмма/Добавить линию тренда».

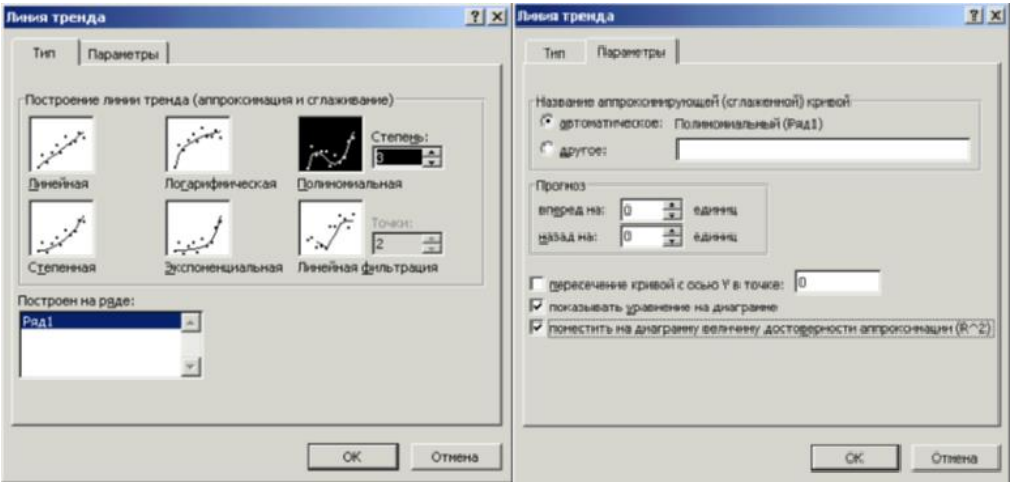
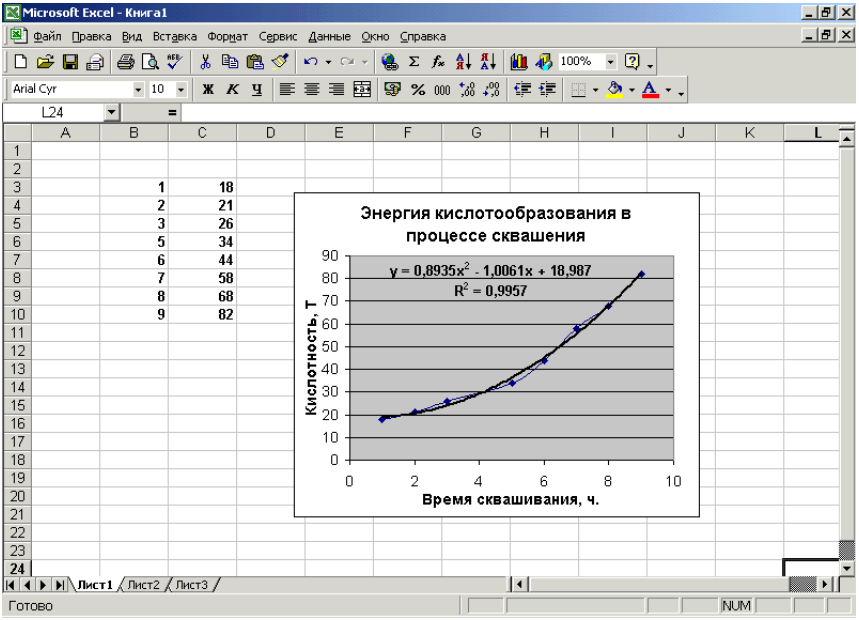


Рисунок 6 – Параметры линии тренда

5. Из предлагаемых кривых во вкладке «Тип», выбирается функция, которая наилучшим образом описывает исследуемый процесс. Для вывода уравнения функции во вкладке «Параметры»отмечаются пункты «Показать уравнение на диаграмме»и «Поместить величину достоверности аппроксимации R^2».



1	18
2	21
3	26
5	34
6	44
7	58
8	68
9	82

Рисунок 7 – Диаграмма на основе экспериментальных значений

Величина R^2, должна иметь значение не менее 0,88 (при меньших значениях выбрать другой тип, приближающий R^2 к 1.

1.2. Обработка статистических данных при помощи программного пакета «Statistica»

Statistica – программный пакет, предназначенный для обработки статистических данных. Данный продукт имеет ряд встроенных функций для расчёта основных статистических характеристик, что значительно облегчает работу над большими массивами данных.

Порядок работы с пакетом Statistica:

1. Внести в лист данных параметры вариационного ряда.

Во вкладке «Статистика» выбрать «Статистика данных блоков», «Столбцы блока» (рисунок 8):

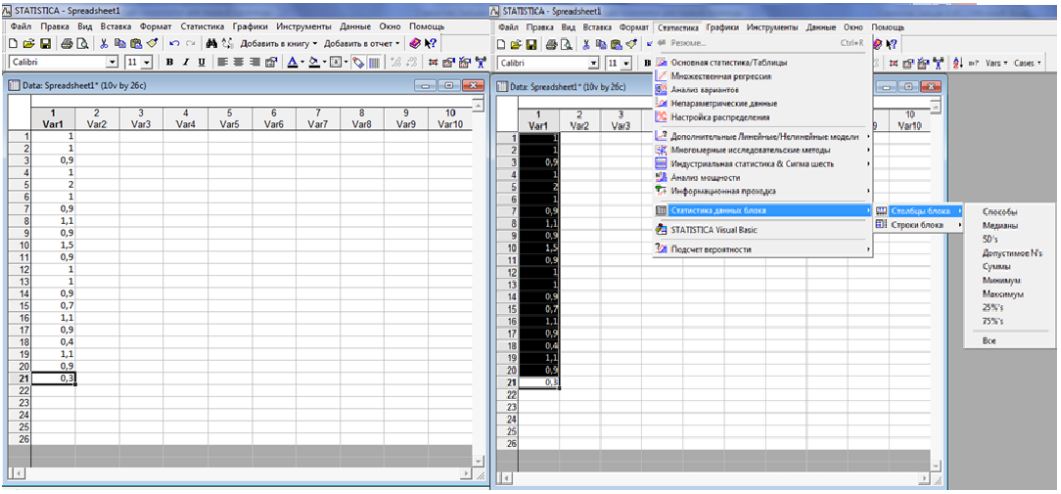


Рисунок 8 – Параметры вариационного ряда

2. Для расчёта медианы необходимо последовательно выбрать команды: «Статистика» / «Статистика данных блоков» / «Столбцы блока» / «Медианы». Для расчёта суммы необходимо последовательно выбрать команды: «Статистика» / «Статистика данных блоков» / «Столбцы блока» / «Суммы» (рисунок 9)

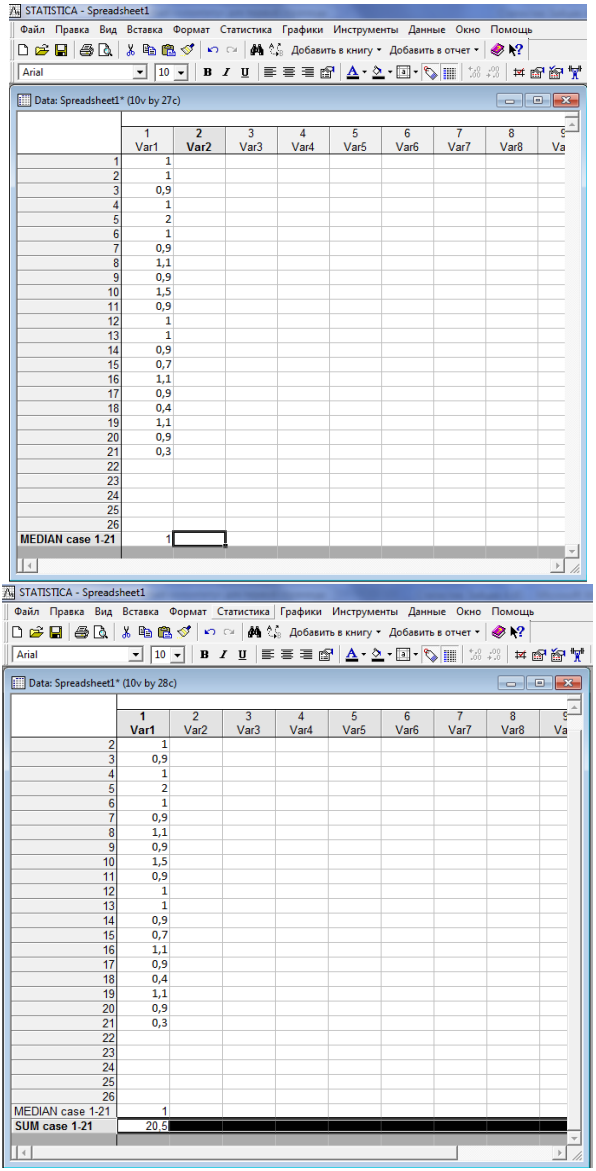


Рисунок 9 – Порядок расчета медианы вариационного ряда

3. Для расчёта минимального значения необходимо последовательно выбрать команды: «Статистика» / «Статистика данных блоков» / «Столбцы блока» / «Минимум».

Для расчёта максимального значения необходимо последовательно выбрать команды: «Статистика» / «Статистика данных блоков» / «Столбцы блока» / «Максимум» (Рисунок 10).

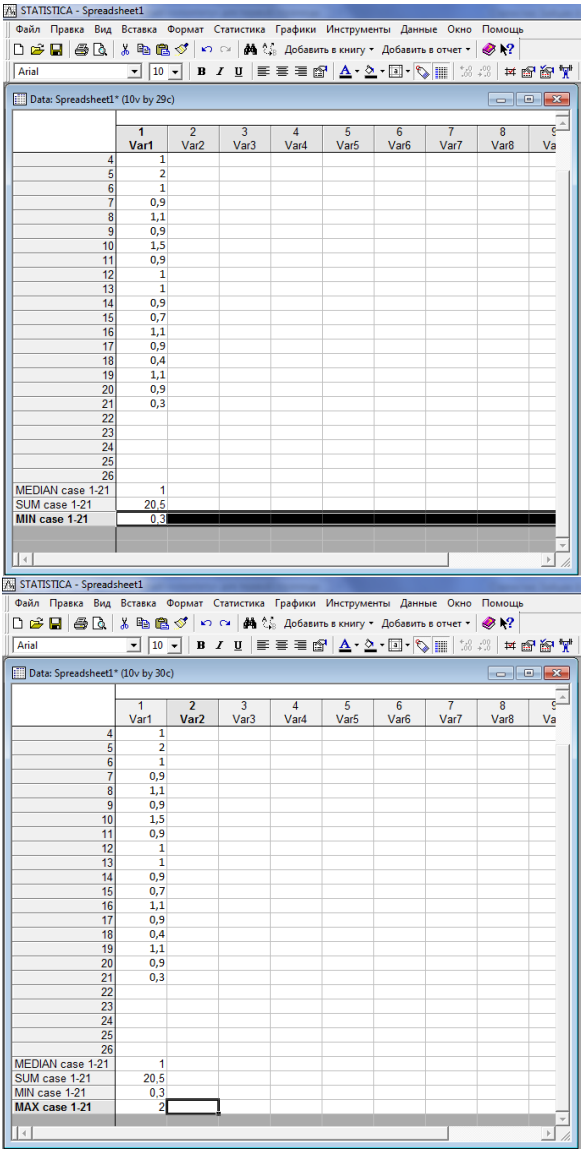
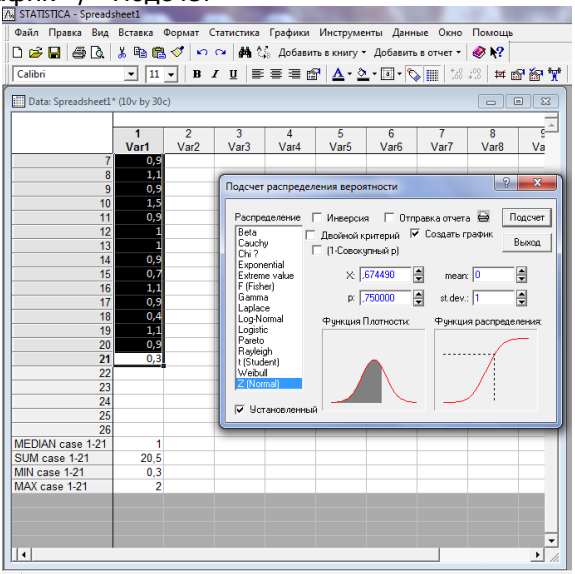


Рисунок 10 – Порядок расчета минимального и максималь-

ного значений вариационного ряда

4. Для расчёта и построения графика нормального распределения случайной величины необходимопоследовательно выбрать команды: «Статистика» / «Подсчёт вероятностей» / «Распределения» / «Z-Normal» / «Создать график» / «Подсчёт»



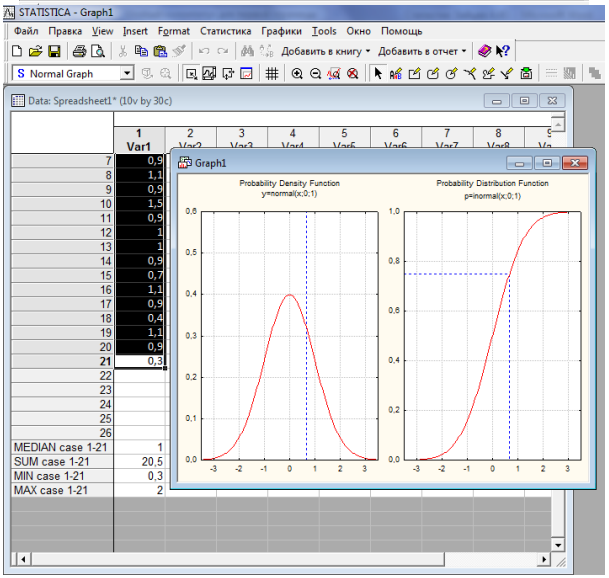
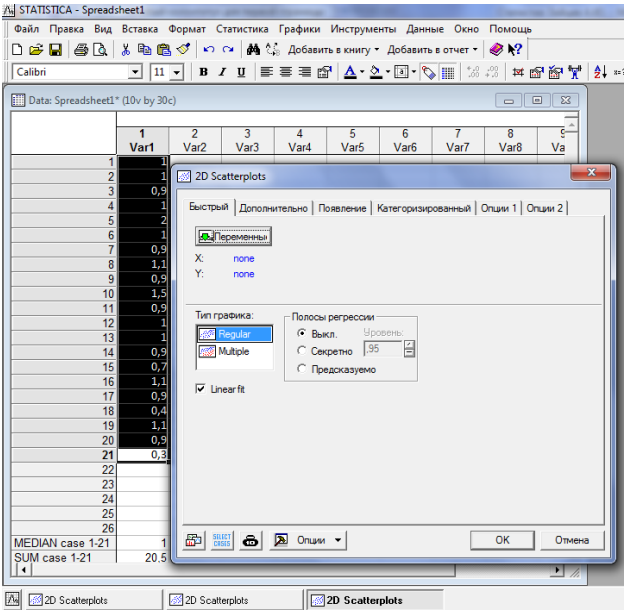


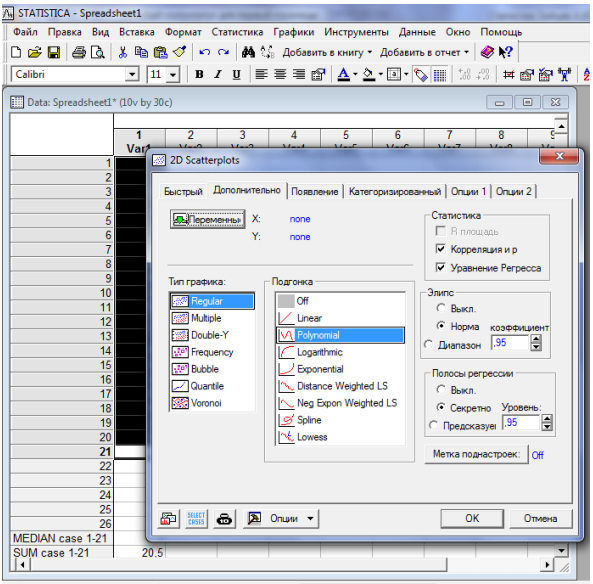
Рисунок 11 – Нормальное распределение случайной величины

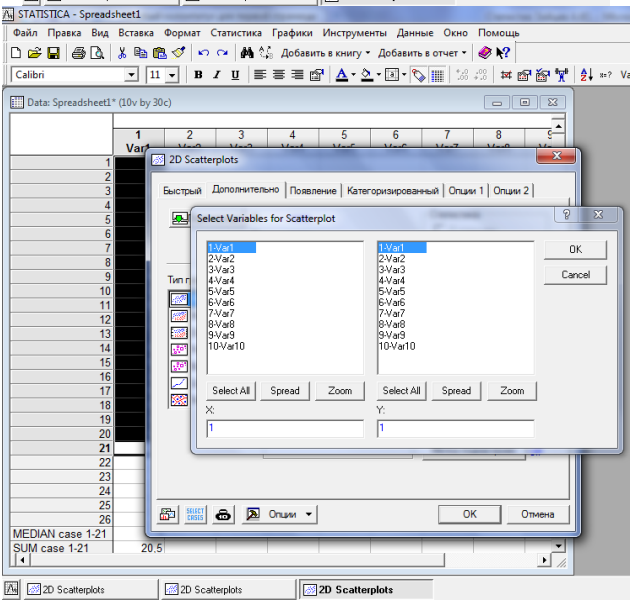
5. Для расчёта и построения графика рассеяния случайной величины необходимо последовательно выбрать команды: «Графики» / «2-D Графики» / «Графики рассеяния» / «Дополнительно» / «Regular» / «Polinomial» / «Корреляция» /«Уравнение регрессии» / «Элипс – норма» /«Полосы регрессии – секретно».

6. Во всплывающем окне выбрать: «1-var1 справа и слева» **/** «Ок».

Результат построения представлен на Рисунке 12.




STATISTICA - Spreadsheet1
Файл Правка Вид Вставка Формат Статистика Графики Инструменты Данные Окно Помощь
Data: Spreadsheet1* (10v by 30c)
2D Scatterplots
Быстрый Дополнительно Появление Категоризированный Опции 1 Опции 2
Переменные
X: none
Y: none
Статистика
Корреляция и р
Уравнение Регресса
Тип графика:
Regular
Multiple
Double-Y
Frequency
Bubble
Quantile
Voronoi
Подгонка
Off
Linear
Polynomial
Logarithmic
Exponential
Distance Weighted LS
Neg Expon Weighted LS
Spline
Lowess
Элипс
Выкл.
Норма коэффициент
Диапазон .95
Полосы регрессии
Выкл.
Секретно Уровень:
Предсказуем .95
Метка подчастроек: Off
Опции
OK
Отмена
MEDIAN case 1-21
SUM case 1-21 20.5
2D Scatterplots
2D Scatterplots
2D Scatterplots


STATISTICA - Spreadsheet1
Файл Правка Вид Вставка Формат Статистика Графики Инструменты Данные Окно Помощь
Data: Spreadsheet1* (10v by 30c)
2D Scatterplots
Быстрый Дополнительно Появление Категоризированный Опции 1 Опции 2
Select Variables for Scatterplot
1-Var1
2-Var2
3-Var3
4-Var4
5-Var5
6-Var6
7-Var7
8-Var8
9-Var9
10-Var10
OK
Cancel
Select All
Spread
Zoom
X:
Y:
Опции
OK
Отмена
MEDIAN case 1-21
SUM case 1-21 20.5
2D Scatterplots
2D Scatterplots
2D Scatterplots

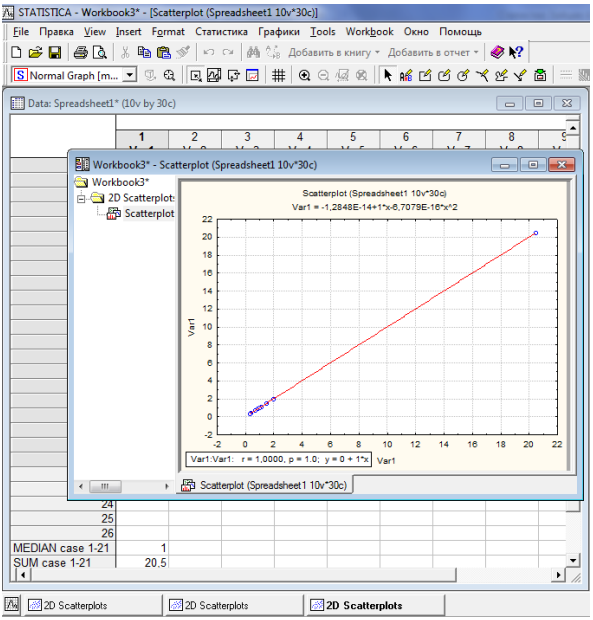


Рисунок12 – Порядок построения графика рассеяния случайной величины

1.3. Обработка статистических данных при помощи программного пакета «Matlab»

Для статистических вычислений и статистической обработки данных при помощи программного пакета Matlab используется пакет прикладных программ StatisticsandMachineLearningToolbox, который реализует множество статистических функций:

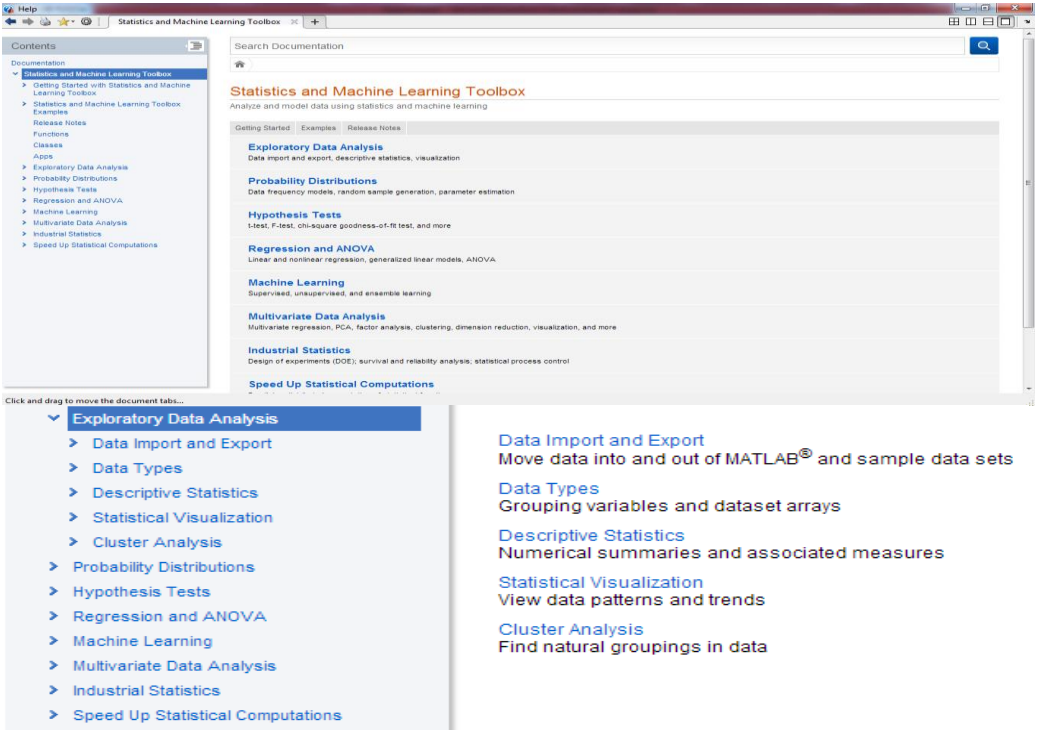
- описательная статистика;
- распределения вероятностей;
- оценка параметров и аппроксимация;
- проверка гипотез;
- множественная регрессия;
- интерактивная пошаговая регрессия;
- моделирование Монте-Карло;
- аппроксимация на интервалах;
- статистическое управление процессами;
- планирование эксперимента;

- моделирование поверхности отклика;
- аппроксимация нелинейной модели;
- анализ главных компонент;
- статистические графики.

Графический интерфейс пользователя приведен на рисунке 13.

Статистические вычисления в Matlab включают:

- исследовательский анализ данных;
- функции вероятности распределения;
- гипотетические тесты;
- регрессионный и дисперсионный анализы;
- машинное обучение;
- многомерный анализ данных;
- промышленная статистика



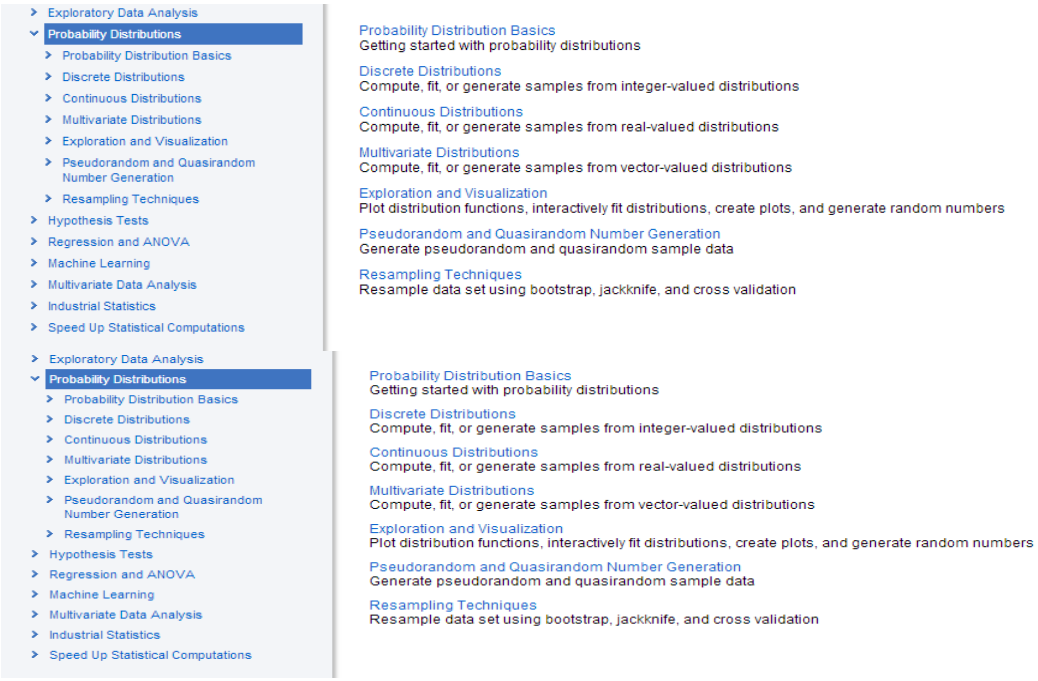
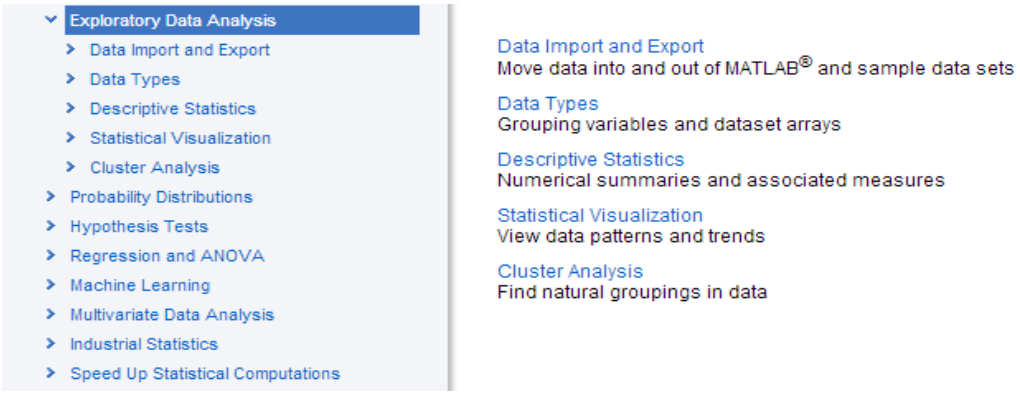


Рисунок 13 – Графический интерфейс пользователя.

1.3.1. Исследовательский анализ данных



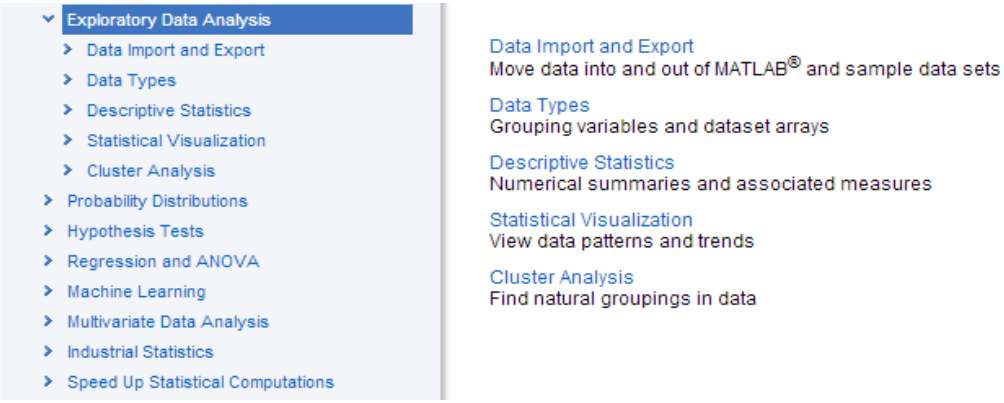


Рисунок 14 – Исследовательский анализ данных

StatisticsandMachineLearningToolbox позволяет просматривать наши данные численно, создавая сводную статистику, включая измерения центральной тенденции, дисперсии, формы и корреляции.

Исследовательский анализ данных содержит в себе:

- Импорт и экспорт данных. Перемещение данных в и из MATLAB и наборов выборочных данных.
- Группировка переменных и массивов наборов данных.
- Описательная статистика. Численные резюме и связанные с ними меры.
- Статистическая визуализация. Просмотр шаблонов данных и трендов.
- Анализ кластеров. Поиск естественных группировок в данных

1.3.2. Функции вероятности распределений

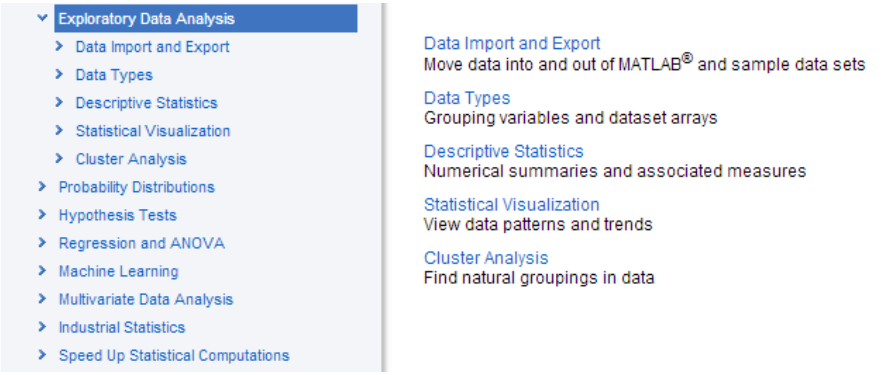


Рисунок 15 – Функции вероятности распределений

StatisticsandMachineLearningToolbox позволяет работать с дискретными, непрерывными и непараметрическими распределениями вероятностей.

Функции вероятности распределений содержат:

- Основы распределения вероятностей: начало работы с распределением вероятностей.
- Дискретные распределения: вычислить, сопоставить или сгенерировать образцы из целочисленных распределений.
- Непрерывные распределения: вычислить, сопоставить или сгенерировать образцы из вещественных распределений.
- Многомерные распределения: вычислить, сопоставить или сгенерировать образцы из векторных распределений.
- Исследование и визуализация: функции распределения графиков, интерактивное сопоставление распределений, создание графиков и генерация случайных чисел.
- Генерация псевдослучайных и квазислучайных чисел: генерировать псевдослучайные и квазислучайные данные выборки.
- Методыповторнойвыборки.

1.3.3. Гипотетические тесты

Рисунок 16 –Гипотетические тесты

StatisticsandMachineLearningToolbox предоставляет множество параметрических и непараметрических тестов гипотез.Мы можем протестировать заданное распределение, среднее, медианное или дисперсионное значение (дисперсия).

Гипотетические тесты содержат в себе:

- Тесты распределения:данные тестирования против определенного распределения или независимости.
- Тесты местоположения:параметрические и непараметрические тесты для сред, тесты для эффектов столбцов.
- Дисперсионные тесты:параметрические и непараметрические тесты на отклонения.

1.3.4. Регрессионный и дисперсионный анализы

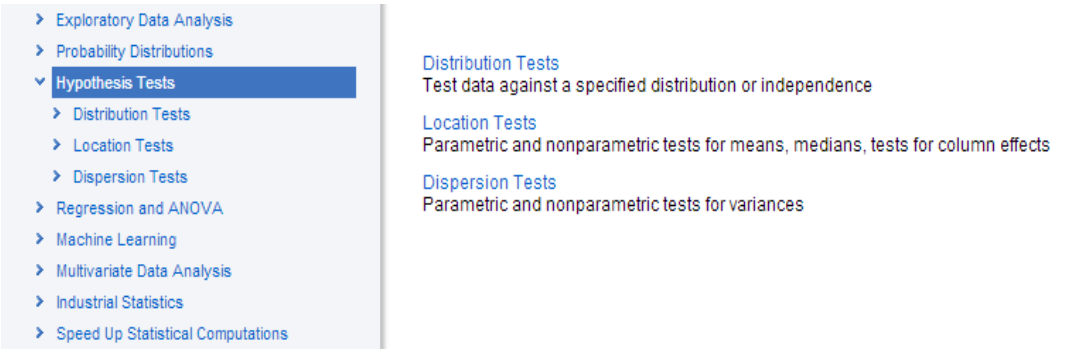


Рисунок 17 –Регрессионный и дисперсионный анализы

Модели регрессии описывают взаимосвязь между переменной ответа и одной или несколькими предикторными переменными.

Анализ дисперсии (ANOVA) - это процедура определения того, возникает ли изменение в переменной ответа внутри или среди разных групп населения.

Регрессионный и дисперсионный анализы содержат в себе:

- Линейная регрессия: Множественные, ступенчатые, многомерные модели регрессии и многое другое.
- Обобщенные линейные модели: Логистическая регрессия, многочленная регрессия, регрессия Пуассона и многое другое.
- Нелинейная регрессия: Нелинейные модели регрессии с фиксированным и смешанным эффектами.
- Дисперсионный анализ: Анализ дисперсии и ковариации, одномерный и многомерный дисперсионный анализ, повторные измерения дисперсионного анализа

1.3.5. Машинное обучение

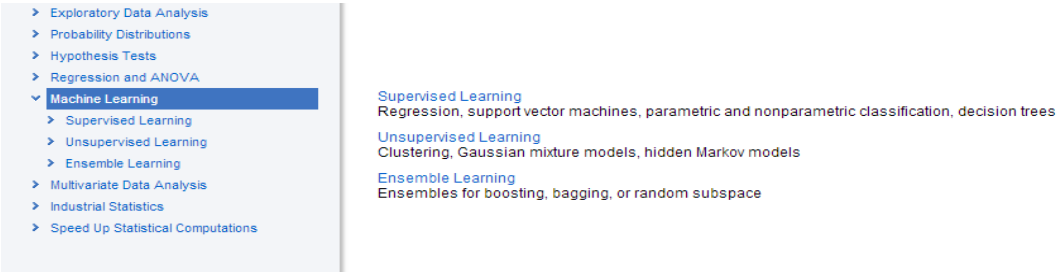


Рисунок 18 – Машинное обучение

Целью машинного обучения является построение модели, которая принимает решения на основе доказательств в условиях неопределенности. Поскольку адаптивные алгоритмы идентифицируют шаблоны в данных, компьютер «учится» из наблюдений. При использовании дополнительных наблюдений компьютер улучшает эффективность принятия решений.

Машинное обучение содержит в себе:

- Контролируемое обучение (с учителем): Регрессия, опорные векторные машины, параметрическая и непараметрическая классификация, древа решений.
- Неконтролируемое обучение (без учителя): Кластеризация, смеси гауссовских моделей, скрытые марковские модели.
- Ансамбль обучений: Ансамбли для ускорения, упаковки

или случайного подпространства.

1.3.6. Многомерный анализ данных

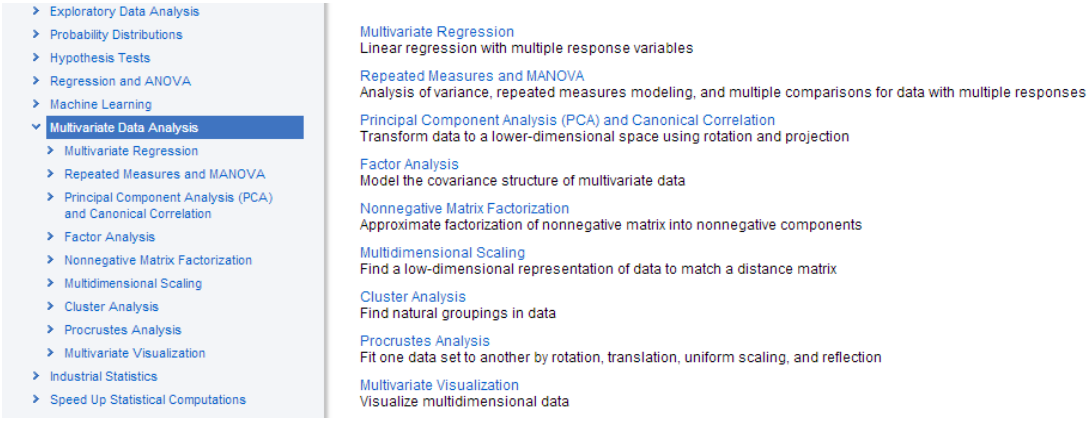


Рисунок 19 – Многомерный анализ данных

Многовариантная статистика изучает взаимосвязь между множественными измерениями, наблюдаемыми у субъекта, и прогностическими переменными. StatisticsandMachineLearningToolbox предоставляет ряд методов для подгонки моделей к нескольким ответам, уменьшения размеров высокоразмерных данных и определения естественных группировок среди точек данных.

Многомерный анализ данных включает в себя:

- Многомерная регрессия: линейная регрессия с несколькими переменными ответа.
- Повторные измерения и MANOVA: анализ дисперсии, моделирование повторяющихся мер и множественное сравнение данных с несколькими ответами.
- Анализ основных компонентов (PCA) и каноническая корреляция: преобразование данных в более низкое пространство с использованием вращения и проецирования.
- Факторный анализ: модель ковариационной структуры многомерных данных.
- Неотрицательная матричная факторизация: приближенная факторизация неотрицательной матрицы в неотрицательные компоненты.
- Многомерное масштабирование: найти низкоразмерное представление данных в соответствии с матрицей расстояний.
- Анализ кластеров: поиск естественных группировок в

данных

- Прокрустов анализ: установите один набор данных в другой путем вращения, трансляции, равномерного масштабирования и отражения.
- Многомерная визуализация: визуализировать многомерные данные.

1.3.7. Промышленная статистика

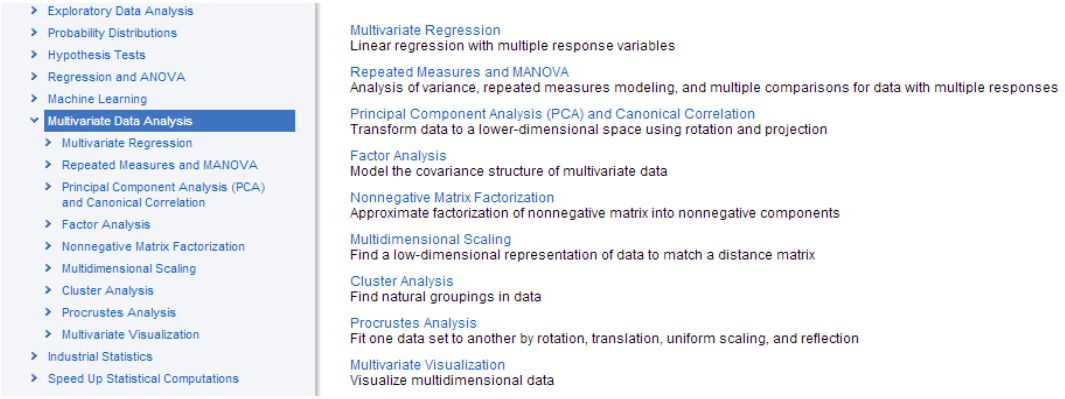


Рисунок 20 – Промышленная статистика

StatisticsandMachineLearningToolbox предоставляет инструменты для проектирования экспериментов, анализа надежности и данных о выживании, контроля качества процесса и наблюдения данных.

Промышленная статистика включает в себя:

- Разработка экспериментов: Планирование экспериментов с систематическим сбором данных.
- Анализ данных о продолжительности жизни: Непараметрические и полупараметрические методы анализа данных о надежности и выживаемости.
- Статистическое управление процессами: Статистические методы контроля качества и контроля производственных процессов.

1.3.8. Контейнер table

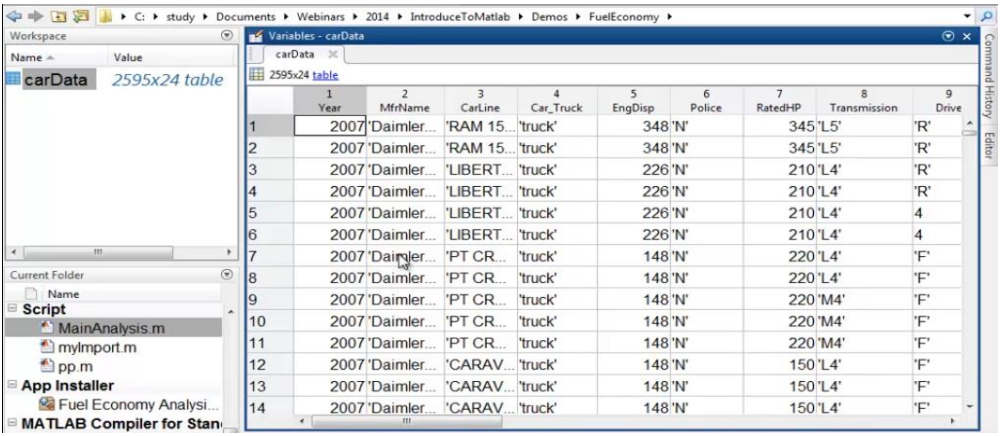


Рисунок 21 – Контейнер table

StatisticsandMachineLearningToolbox поддерживает все типы данных, которые поддерживает сам Matlab, кроме того он содержит свои типы данных.

Один из удобных контейнеров, которые можно использовать для статистических вычислений – это тип table, который может включать в себя любой тип данных. Он содержит численный тип данных, строковый и т.д.

1.3.9. Работа с распределениями

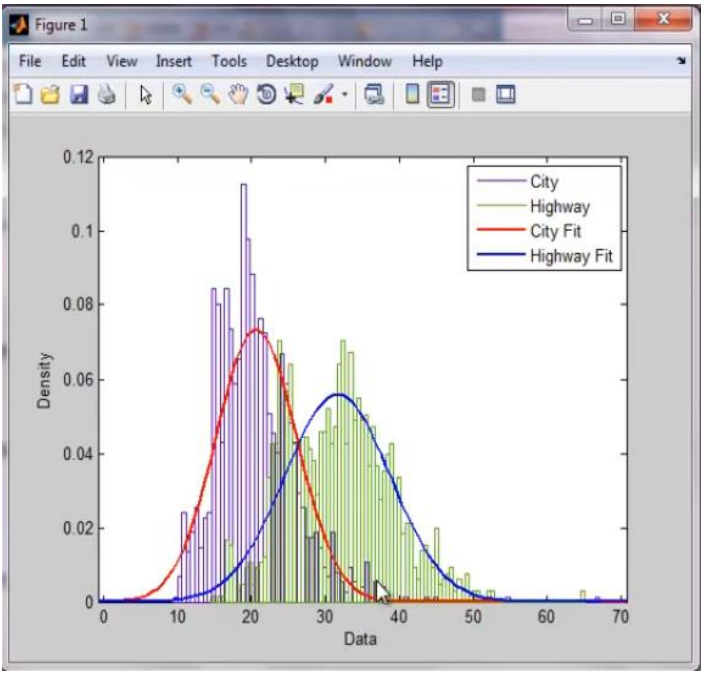


Рисунок 22 –Пример исследовательского анализа данных

На основе данных, которые у нас уже есть, проведем небольшой исследовательский анализ.

Обращение к любому столбцу и любым элементам происходит с помощью круглых скобок. Выделяем значения, которые нам нужны. Выделяем с 1 по 5 элементы колонки MPG.

Для получения сводной статистики из таблицы применяем команду summary. Из контейнера данных выделяем колонки, которые нас интересуют. Двоеточие обозначает, что мы выделяем все строки данных в этих колонках. Таким образом мы получаем нашу статистику. Здесь показывается сколько переменных, какие они: минимальные, средние – медиана и максимум.

1.3.10. Визуализация данных

Визуализация данных происходит путем обращения к колонкам данных из таблицы. Применяем команду plotMPG. На рисунке 23 показывается расход топлива в зависимости от мощности двигателя.

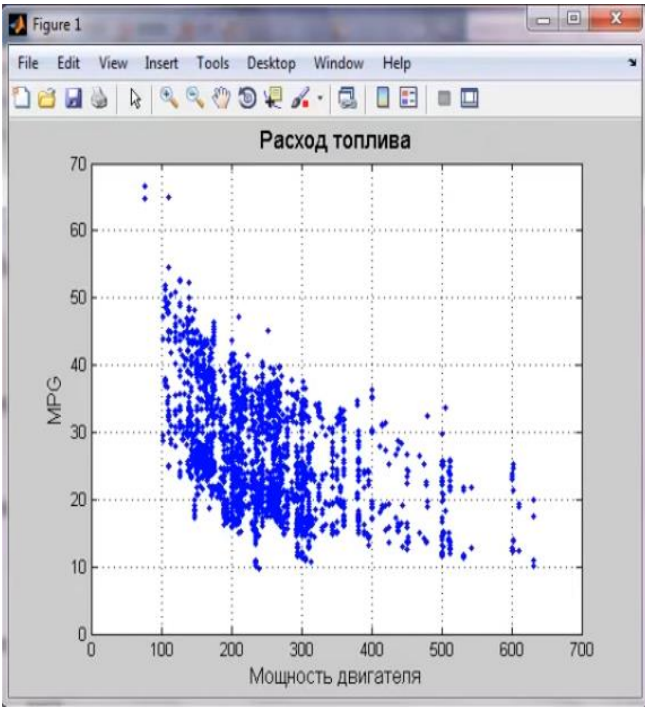


Рисунок 23 – Пример визуализации данных

1.4. Обработка статистических данных с помощью программного пакета «Statistica»

Statistica – программный пакет, предназначенный для обработки статистических данных [7, 13]. Данный продукт имеет ряд встроенных функций для расчёта основных статистических характеристик, что значительно облегчает работу над большими массивами данных.

Рассмотрим несколько примеров расчетов в пакете «Statistica». Для начала необходимо внести вариационный ряд, как показано на рисунке 24.

1. Внести вариационный ряд:

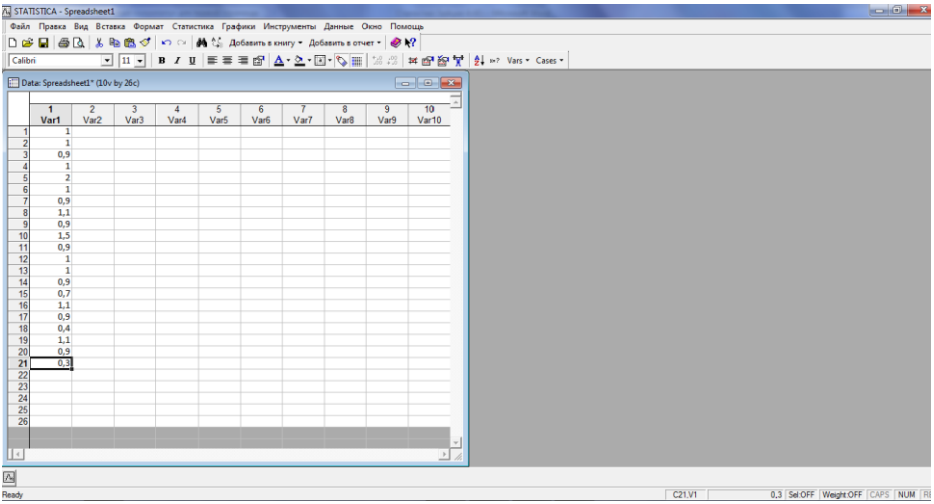


Рисунок 24 – Вариационный ряд

2. Во вкладке статистика выбрать статистика данных блоков, столбцы блока на рисунке 25:

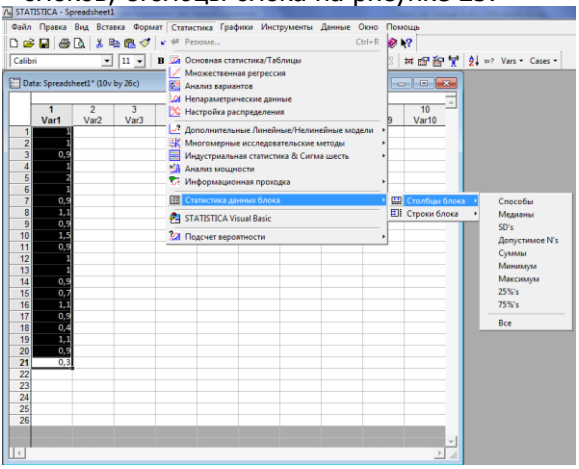


Рисунок 25 – Статистика данных блока

3. Для расчёта медианы необходимо выбрать: **Статистика/Статистика данных блоков/Столбцы блока/Медианы** на рисунке 26:

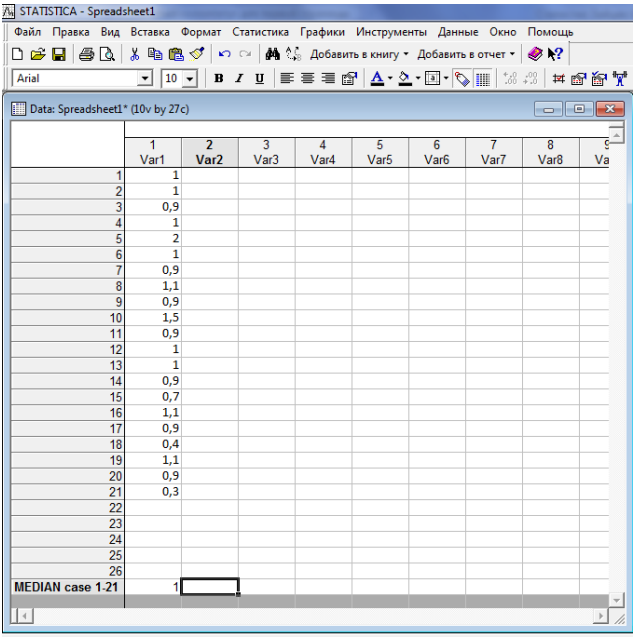


Рисунок 26 – Статистика/Статистика данных блоков/ Столбцы блока/Медианы

4. Для расчёта суммы необходимо выбрать: **Статистика/Статистика данных блоков/Столбцы блока/ Суммы** на рисунке 27:

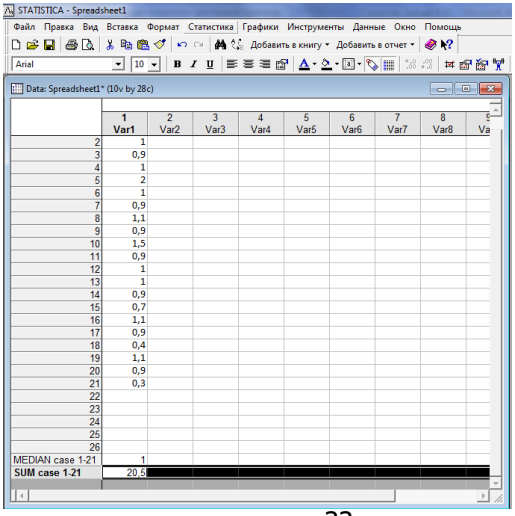


Рисунок 27 – Статистика/Статистика данных блоков/ Столбцы блока/Суммы

5. Для расчёта мин. значения необходимо выбрать: **Статистика/Статистика данных блоков/Столбцы блока/Минимум**(рисунок 28)

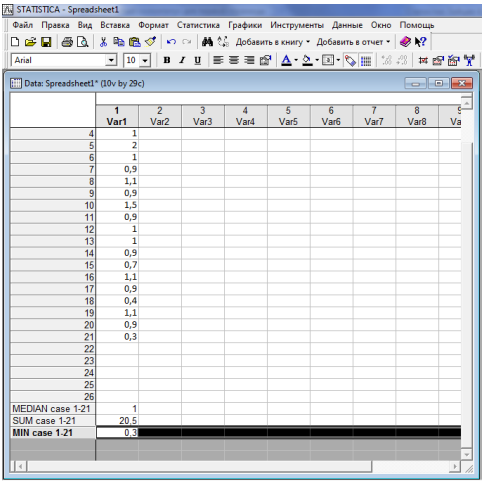


Рисунок 28 – Статистика/Статистика данных блоков/Столбцы блока/Минимум

6. Для расчёта макс. значения необходимо выбрать: **Статистика/Статистика данных блоков/Столбцы блока/Максимум** (рисунок 29)

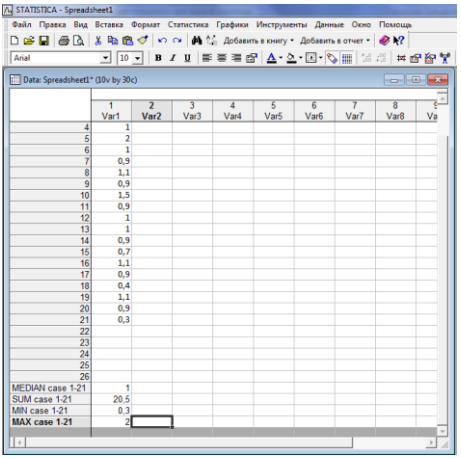


Рисунок 29 – Статистика/Статистика данных блоков/Столбцы блока/Максимум

7. Расчёт и построение графика нормального распределения (рисунок 30,31): **Статистика/Подсчёт вероятностей/Рспределения/Z-Normal/Создать график (**поставить галочку**)/Подсчёт**

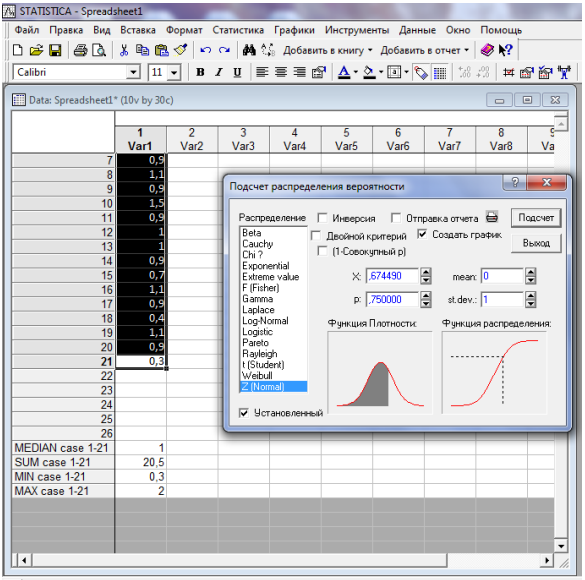


Рисунок 30 – Статистика/Подсчёт вероятностей/Распределения/
Z-Normal/Создать график (поставить галочку)/Подсчёт

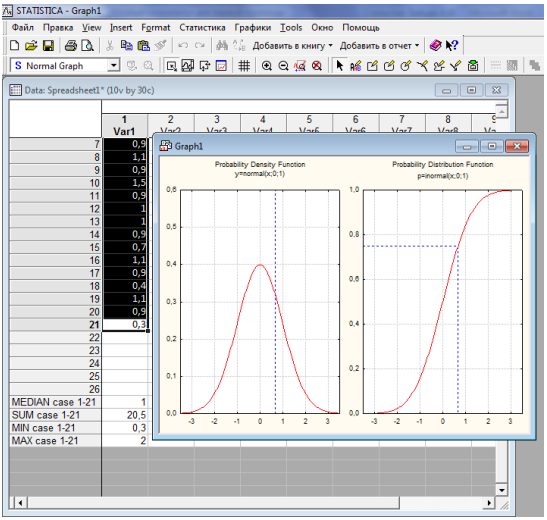


Рисунок 31 – Статистика/Подсчёт вероятностей/Распределения/ Z-Normal/Создать график (поставить галочку)/Подсчёт

8. Построение графика рассеяния (рисунок 32): **Графики/2-D Графики /Графики рассеяния:**

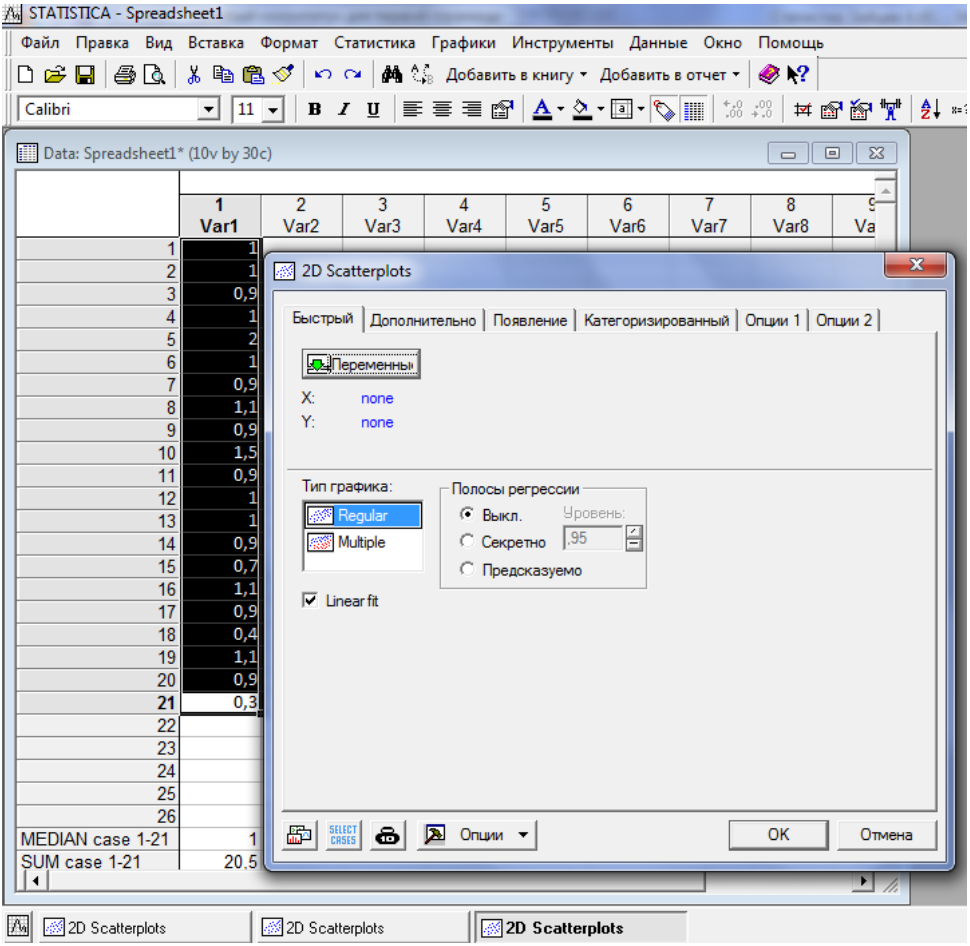


Рисунок 32 – Графики/2-D Графики /Графики рассеяния

…/Дополнительно/Regular/Polinomial/Корреляция, Уравнение регрессии (поставить галочки**)/Элипс – норма; Полосы регрессии – секретно/Ок** (рисунок 33)**/…**

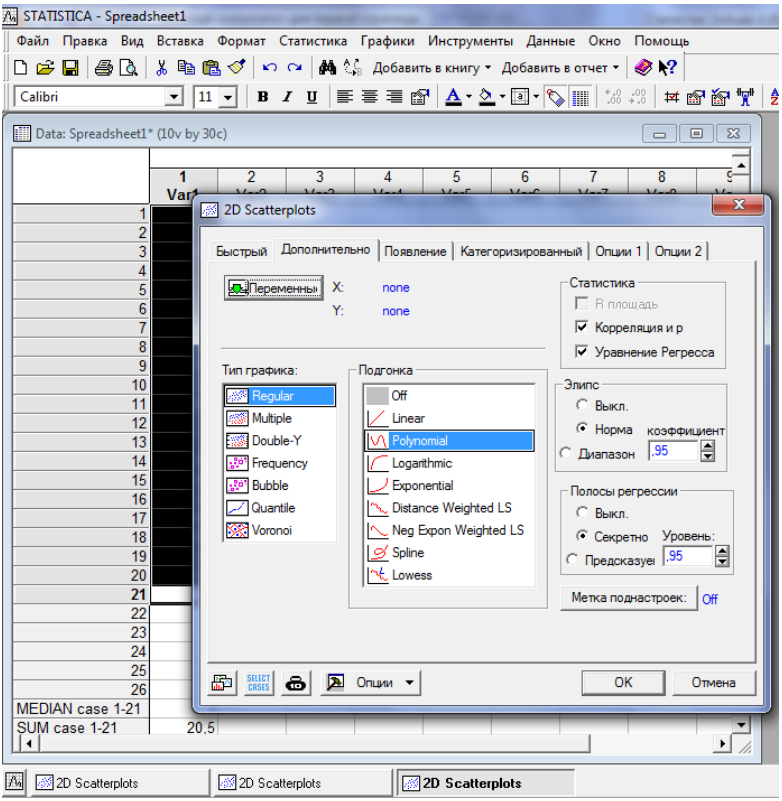


Рисунок 33 – **.../Дополнительно/Regular/Polinomial/Корреляция, Уравнение регрессии (**поставить галочки**)/Элипс – норма; Полосы регрессии – секретно/Ок**

Во всплывающем окне выбрать: 1-var1 справа и слева (рисунок11**)/Ок**

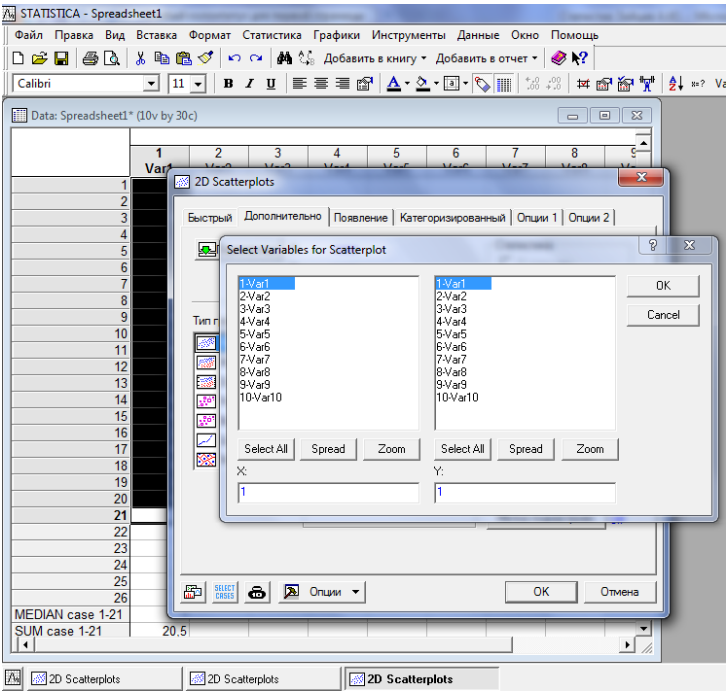


Рисунок34 – Всплывающее окно

Результат построения представлен на рисунке 35.

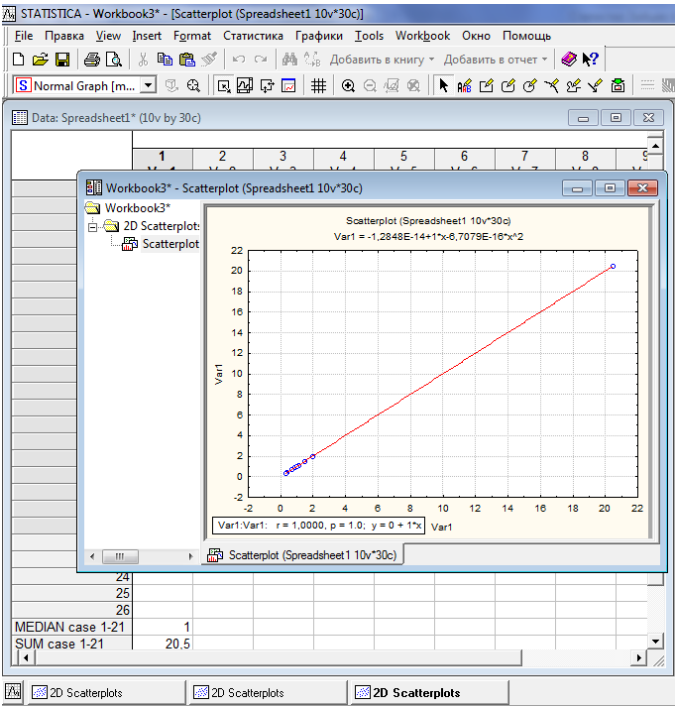


Рисунок 35 – Результат построения

1.4.1. Многомерная групповая визуализация

Визуализация на различных осях графиков, также различных данных. Применяется команда gplotmatrix.

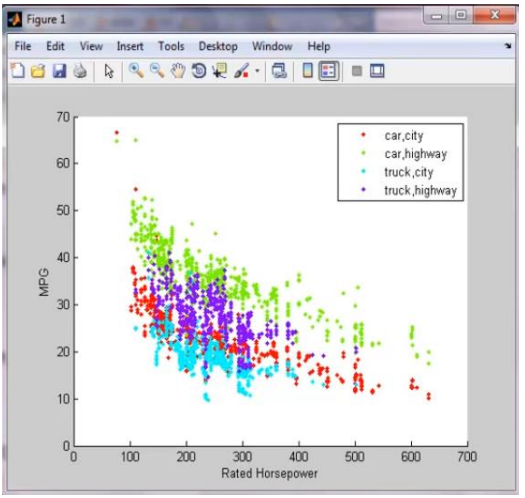


Рисунок 36 – Пример графика

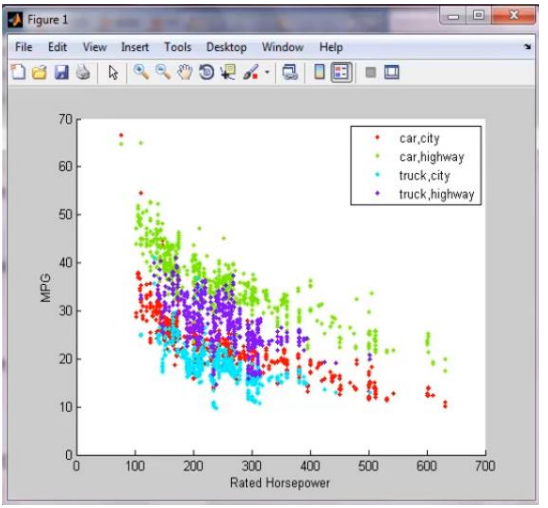


Рисунок37– Пример графика

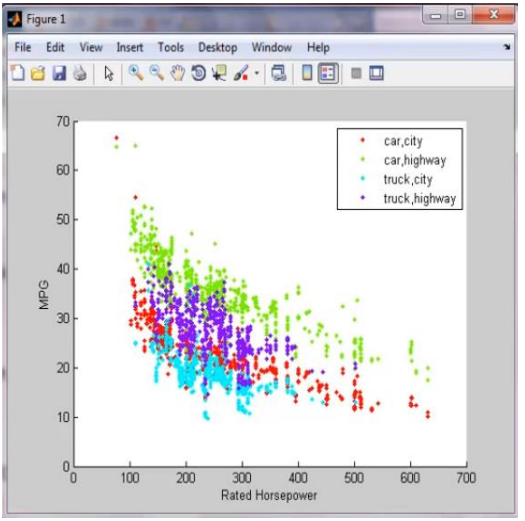


Рисунок 38 – Пример графика

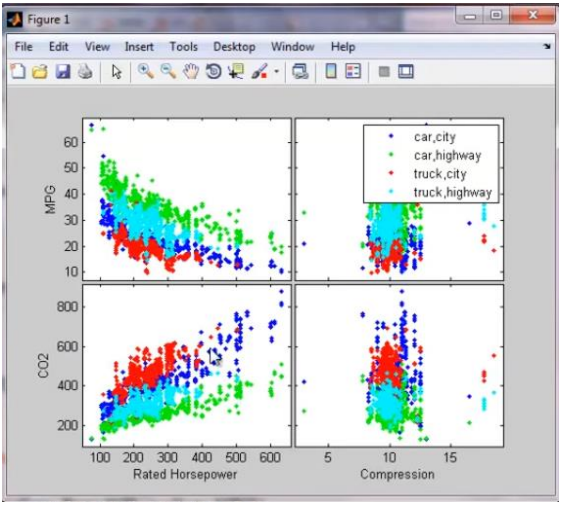


Рисунок 39 – Пример графика

1.4.2. Исследование многомерных данных

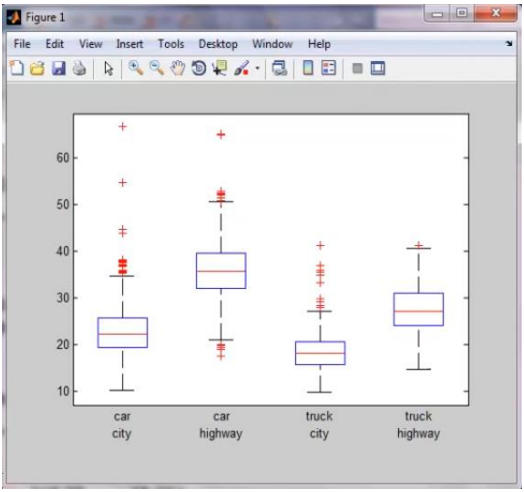


Рисунок 40– График статистики для разных переменных

Применяется команда boxplot, которая строит график статистики для разных переменных. Она отражает для конкретной переменной значения квантилей: 25% и 75%, среднее значение и крайне значение, которое отображает значение 3σ.

1.4.2. Работа с распределениями

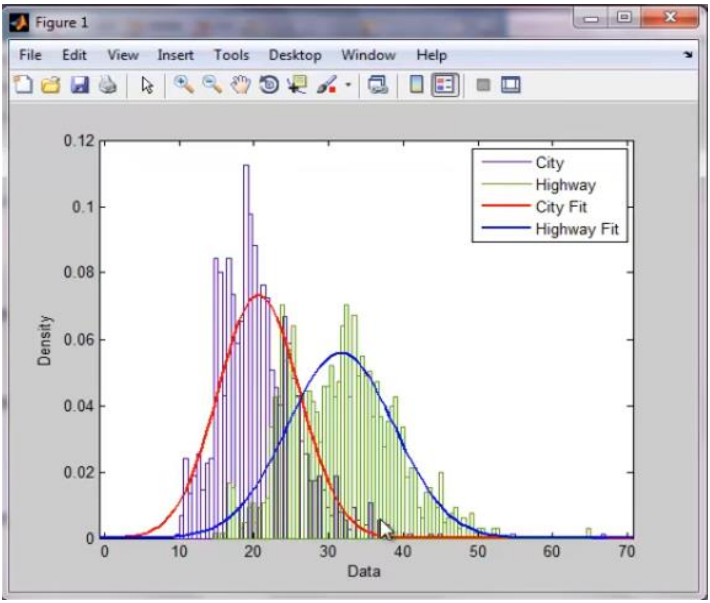


Рисунок 41 – График распределения

Различные возможности для работы с распределениями позволяют подобрать из StatisticsandMachineLearningToolbox нужные распределения под имеющиеся данные.Для вывода графика применяем команду mpgDistribution.

ГЛАВА 2. ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА И ОБРАБОТКА РЕЗУЛЬТАТОВ

2.1. Проверка статистических гипотез

2.1.1. Сущность задачи проверки статистических гипотез

Проверка статистических гипотез складывается из следующих этапов [14]:

– формулируется в виде статистической гипотезы задача исследования;

– выбирается статистическая характеристика гипотезы;

– выбираются нулевая H_0 и альтернативная H_1 гипотезы на основе анализа возможных ошибочных решений и их последствий;

– выбирается приемлемый уровень значимости α ;

– выбирается критерий проверки гипотезы H_0;

– вычисляется фактическое значение статистического критерия;

– определяется критическое значение статистического критерия по соответствующей таблице;

– проверяется нулевая гипотеза на основе сравнения фактического и критического значений критерия, в зависимости от результатов проверки гипотеза либо отклоняется, либо не отклоняется.

При проверке гипотез по одному из критериев возможны два ошибочных решения:

– неправильное отклонение нулевой гипотезы – ошибка первого рода;

– неправильное принятие нулевой гипотезы – ошибка второго рода.

Статистическая гипотеза представляет собой некоторое предположение о законе распределения случайной величины или о параметрах этого закона, формулируемое на основе выборки.

Примерами статистических гипотез являются предположения: генеральная совокупность распределена по экспоненциальному закону; математические ожидания двух экспоненциально распределенных выборок равны друг другу.

В первой из них высказано предположение о виде закона распределения, а во второй – о параметрах двух распределений. Гипотезы, в основе которых нет никаких допущений о конкретном виде закона распределения, называют непараметрическими, в противном случае –параметрическими. Гипотезу, утверждающую, что различие между сравниваемыми характеристиками отсутствует, а наблюдаемые отклонения объясняются лишь случайными колебаниями в выборках, на основании которых производится сравнение, называют нулевой(основной) гипотезой и обозначают H_0.

Наряду с основной гипотезой рассматривают и альтернативную (конкурирующую, противоречащую) ей гипотезу H_1. и если нулевая гипотеза будет отвергнута, то будет иметь место альтернативная гипотеза.

Различаютпростые и сложные гипотезы. Гипотезу называют простой, если она однозначно характеризует параметр распределения случайной величины. Например, если l является параметром экспоненциального распределения, то гипотеза H_0 о равенстве l=10 – простая гипотеза.

Сложной называют гипотезу, которая состоит из конечного

или бесконечного множества простых гипотез. Сложная гипотеза H_0 о неравенстве l>10 состоит из бесконечного множества простых гипотез H_0 о равенстве l = b_i, где b_i – любое число, большее 10. Гипотеза H_0 о том, что математическое ожидание нормального распределения равно двум при неизвестной дисперсии, тоже является сложной.

Сложной гипотезой будет предположение о распределении случайной величины X по нормальному закону, если не фиксируются конкретные значения математического ожидания и дисперсии.

Проверка гипотезы основывается на вычислении некоторой случайной величины – критерия, точное или приближенное распределение которого известно. Обозначим эту величину через z, ее значение является функцией от элементов выборки $z=z(x_1, x_2, \ldots, x_n)$.

Процедура проверки гипотезы предписывает каждому значению критерия одно из двух решений – принять или отвергнуть гипотезу. Тем самым все выборочное пространство и соответственно множество значений критерия делятся на два непересекающихся подмножества S_0 и S_1. Если значение критерия z попадает в область S_0, то гипотеза принимается, а если в областьS_1, то гипотеза отклоняется.

Множество S_0 называется областью принятия гипотезы или областью допустимых значений, а множество S_1 – областью отклонения гипотезы или критической областью. Выбор одной области однозначно определяет и другую область.

Принятие или отклонение гипотезы H_0 по случайной выборке соответствует истине с некоторой вероятностью и, соответственно, возможны два рода ошибок. Ошибка первого рода возникает с вероятностью a тогда, когда отвергается верная гипотеза H_0 и принимается конкурирующая гипотеза H_1.

Ошибка второго рода возникает с вероятностью b в том случае, когда принимается неверная гипотеза H_0, в то время как справедлива конкурирующая гипотезаH_1.

Доверительная вероятность – это вероятность не совершить ошибку первого рода и принять верную гипотезу H_0.

Вероятность отвергнуть ложную гипотезу H_0 называетсямощностью критерия. Следовательно, при проверке гипотезы возможны четыре варианта исходов, представленных в таблице 1:

Таблица 1 – Возможные варианты исхода по выдвигаемым гипотезам

Гипотеза H_0	***Решение***	***Вероятность***	***Примечание***
Верна	Принимается	$1 - \alpha$	Доверительная вероятность
	Отвергается	α	Вероятность ошибки 1-го рода
Неверна	Принимается	b	Вероятность ошибки 2-го рода
	Отвергается	1 - b	Мощность критерия

Например, рассмотрим случай, когда некоторая несмещенная оценка параметраθ вычислена по выборке объема n, и эта оценка имеет плотность распределения f(θ), рисунок 42.

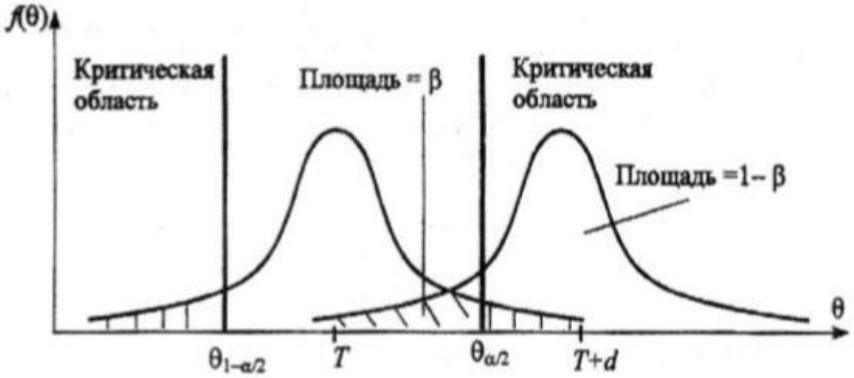


Рисунок 42 – Области принятия и отклонения гипотезы

Предположим, что истинное значение оцениваемого параметра равно T. Если рассматривать гипотезу H_0 о равенстве θ=T, то насколько велико должно быть различие между θиT, чтобы эту гипотезу отвергнуть. Ответить на данный вопрос можно в статистическом смысле, рассматривая вероятность достижения некоторой заданной разности между q иТна основе выборочного распределения параметра

Целесообразно полагать одинаковыми значения вероятности выхода параметра θ за нижний и верхний пределы интервала. Такое допущение во многих случаях позволяет минимизировать доверительный интервал, т.е. повысить мощность критерия проверки. Суммарная вероятность выхода параметра θ за пределы

интервала с границами $\theta_{1-a/2}$ и $\theta_{a/2}$, составляет величину a. Эту величину следует выбирать такой, чтобы выход за пределы интервала был маловероятен. Если оценка параметра попала в заданный интервал, то нет основания подвергать сомнению проверяемую гипотезу. Гипотезу равенства θ=T можно принять. Но если после получения выборки окажется, что оценка выходит за установленные пределы, то в этом случае есть серьезные основания отвергнуть гипотезу H_0. Отсюда следует, что вероятность допустить ошибкупервого рода равна a (равна уровню значимости критерия).

Если предположить, например, что истинное значение параметра в действительности равно T+d, то согласно гипотезе H_0 о равенстве θ=T – вероятность того, что оценка параметра θ попадет в область принятия гипотезы, составит b, рисунок 42.

При заданном объеме выборки вероятность совершения ошибки первого рода можно уменьшить, снижая уровень значимости a. Однако при этом увеличивается вероятность ошибки второго рода b (снижается мощность критерия).Аналогичныерассуждения можно провести для случая, когда истинное значение параметра равно T–d.

Единственный способ уменьшить обе вероятности состоит в увеличении объема выборки (плотность распределения оценки параметра при этом становится более "узкой"). При выборе критической области руководствуются правилом Неймана – Пирсона: следует так выбирать критическую область, чтобы вероятность a была мала, если гипотеза верна, и велика в противном случае. Однако выбор конкретного значения a относительно произволен. Употребительные значения лежат в пределах от 0,001 до 0,2.

В целях упрощения ручных расчетов составлены таблицы интервалов с границами $\theta_{1-a/2}$ и $\theta_{a/2}$ для типовых значений ai различных способов построения критерия. При выборе уровня значимости необходимо учитывать мощность критерия при альтернативной гипотезе. Иногда большая мощность критерия оказывается существеннее малого уровня значимости, и его значение выбирают относительно большим, например 0,2. Такой выбор оправдан, если последствия ошибок второго рода более существенны, чем ошибок первого рода. Например, если отвергнуто правильное решение "продолжить работу пользователей с текущими паролями", то ошибка первого рода приведет к некоторой задержке в нормальном функционировании системы, связанной со сменой паролей.

Если же принято решения не менять пароли, несмотря на

опасность несанкционированного доступа посторонних лиц к информации, то эта ошибка повлечет более серьезные последствия.

В зависимости от сущности проверяемой гипотезы и используемых мер расхождения оценки характеристики от ее теоретического значения применяют различные критерии. К числу наиболее часто применяемых критериев для проверки гипотез о законах распределения относят критерии хи-квадрат Пирсона, Колмогорова, Мизеса, Вилкоксона, о значениях параметров – критерии Фишера, Стьюдента.

2.2. Типовые распределения

При проверке гипотез широкое применение находит ряд теоретических законов распределения [15]. Наиболее важным из них является нормальное распределение. С ним связаны распределения хи-квадрат, Стьюдента, Фишера, а также интеграл вероятностей. Для указанных законов функции распределения аналитически не представимы. Значения функций определяются по таблицам или с использованием стандартных процедур пакетов прикладных программ. Указанные таблицы обычно построены в целях удобства проверки статистических гипотез в ущерб теории распределений – они содержат не значения функций распределения, а критические значения аргумента z(α). Для односторонней критической области $z(\alpha)=z_{1-a}$, т.е. критическое значение аргумента z(α) соответствует квантилиz_{1-a} уровня 1–α, рисунок 43, так как:

$$\int_z^{\infty} f(z)dz = \alpha = 1-\int_{-\infty}^{z(\alpha)} f(z)dz$$

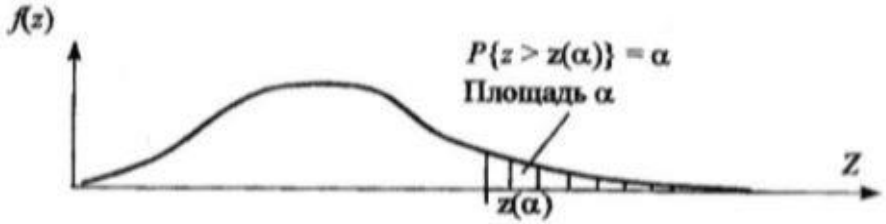


Рисунок 43 – Односторонняя критическая область

Для двусторонней критической области, с уровнем значимости α, размер левой области α_2, правой α_1 ($\alpha_1+\alpha_2=\alpha$) рисунок 44. Значения z(α_2) и z(α_1) связаны с квантилями распределения соотношениями

$$z(\alpha_1) = z_{1-a1},\ z(\alpha_2) = z_{a2}$$

так как:

$$\int_{-\infty}^{z(\alpha_1)} f(z)\,dz = 1 - \alpha_1,$$

$$\int_{-\infty}^{z(\alpha_2)} f(z)dz = 1 - \alpha_2$$

Для симметричной функции плотности распределения f(z) критическую область выбирают из условия $\alpha_1 = \alpha_2 = \alpha/2$ (обеспечивается наибольшая мощность критерия). В таком случае левая и правая границы будут равны $|z(\alpha/2)|$.

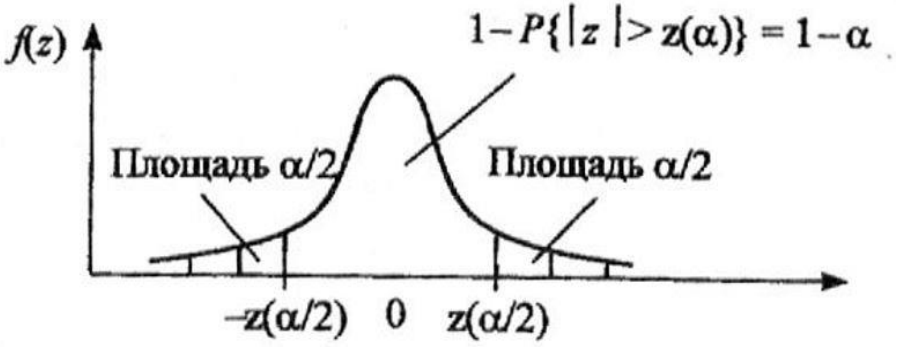


Рисунок 44 - Двусторонняя критическая область

2.2.1. Нормальное распределение

Этот вид распределения является наиболее важным в связи с центральной предельной теоремой теории вероятностей: распределение суммы независимых случайных величин стремится к нормальному с увеличением их количества при произвольном законе распределения отдельных слагаемых, если слагаемые обладают конечной дисперсией. Кроме того, А.М. Ляпунов доказал, что распределение параметра стремится к нормальному, если на параметр оказывает влияние большое количество факторов и ни один из них не является превалирующим. Функция плотности нормального распределения

$$f(x) = \frac{1}{\sqrt{2\pi m_2}} exp\left[-\frac{(x-m_1)^2}{2m_2}\right]$$

– унимодальная, симметричная, аргумент x может прини-

мать любые действительные значения, рисунок 43, 45.

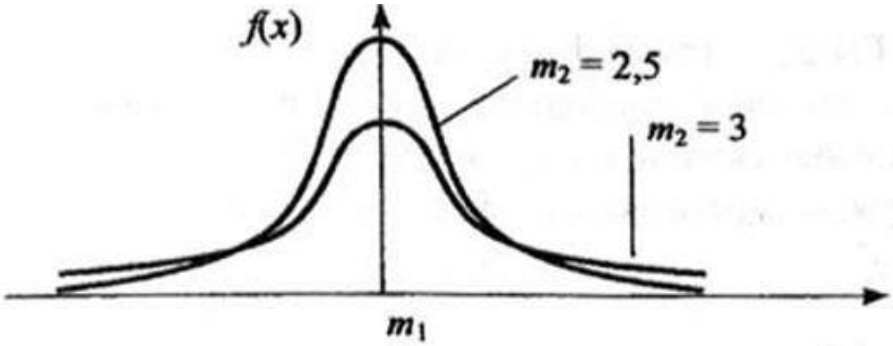


Рисунок 45 - Плотность нормального распределения

Функция плотности нормального распределения стандартизованной величины u имеет вид:

$$f(u) = \frac{1}{\sqrt{2\pi}} exp\left[-\frac{u^2}{2}\right]$$

Вычисление значений функции распределения Φ(u) для стандартизованного неотрицательного аргумента u(u ≥ 0) можно произвести с помощью полинома наилучшего приближения:

$$\Phi(u)= 1– 0,5(1 + 0,196854u+ 0,115194u^2 + 0,000344u^3 + 0,019527u^4) – 4$$

Такая аппроксимация обеспечивает абсолютную ошибку не более 0,00025. Для вычисления Φ(u) в области отрицательных значений стандартизованного аргумента u(u<0) следует воспользоваться свойством симметрии нормального распределения

$$\Phi(u) = 1 – \Phi(–u).$$

Интеграл вероятностей связан с функцией нормального распределения стандартизованной величиныисоотношением

$$\Phi(u) = 0,5 +F(u).$$

2.2.2. Распределение хи-квадрат

Распределению хи-квадрат (χ^2 - распределению) с k сте-

пенями свободы соответствует распределение суммы

$$\chi = \sum_{i=1}^{n} u_j^2$$

квадратов n стандартизованных случайных величин u_j, каждая из которых распределена по нормальному закону, причем k из них независимы, n≥k . Функция плотности распределения хи-квадрат сkстепенями свободы

$$f(x) = \left[2^{k/2}\Gamma\left(\frac{k}{2}\right)\right](x)^{\frac{k}{2}-1} e^{-k/2},\ x \geq 0$$

где Г(k/2) – гамма-функция.

Функция плотности при k, равном одному или двум, – монотонная, а при k>2 – унимодальная, несимметричная, рисунок 46.

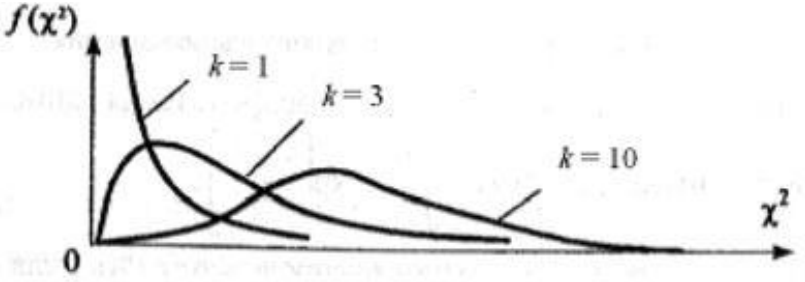


Рисунок 46 –Плотность распределения хи-квадрат

Математическое ожидание и дисперсия величины χ^2 равны соответственно k и 2k. Распределение хи-квадрат является частным случаем более общего гаммараспределения, а величина, равная корню квадратному из хи-квадрат с двумя степенями свободы, подчиняется распределению Рэлея.

С увеличением числа степеней свободы (k>30) распределение хи-квадрат приближается к нормальному распределению с математическим ожиданием k и дисперсией 2k.

2.2.3. Распределение Стьюдента

Распределение Стьюдента (t-распределение, предложено в 1908 г. английским статистиком В. Госсетом, публиковавшим

научные труды под псевдонимом Student) характеризует распределение случайной величины

$$t = \frac{u_0}{\sqrt{\left(u_1^2 + u_2^2 + \ldots + u_k^2\right)/k}},$$

гдеu_0,u_1, …, взаимно независимые нормально распределенные случайные величины с нулевым средним и конечной дисперсией. Аргументtне зависит от дисперсии слагаемых. Функция плотности распределения Стьюдента

$$f(t) = \frac{\Gamma\left[(k+1)/2\right]}{\sqrt{\pi k}\,\Gamma(k/2)}\left[1 + \frac{t^2}{k}\right]^{-\frac{(k+1)}{2}}$$

Величина k характеризует количество степеней свободы. Плотность распределения – унимодальная и симметричная функция, похожая на нормальное распределение, рисунок 47.

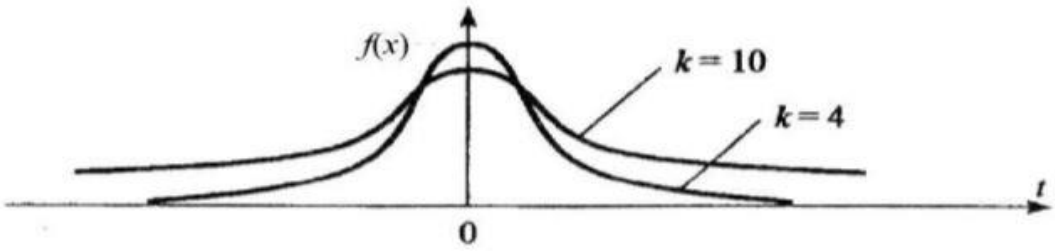


Рисунок 47 – Плотность распределения Стьюдента

Область изменения аргумента t от минус до плюс бесконечности. Математическое ожидание и дисперсия равны 0 и k/(k–2) соответственно, при k>2. По сравнению с нормальным распределение Стьюдента более пологое, оно имеет меньшую дисперсию. Это отличие заметно при небольших значенияхk, что следует учитывать при проверке статистических гипотез (критические значения аргумента распределения Стьюдента превышают аналогичные показатели нормального распределения). Таблицы распределения содержат значения для односторонней (пределы ин-

тегрирования от r(k;α) до бесконечности

$$\int_{r(k,\alpha)}^{\infty} f(t)\,dt = \alpha$$

или двусторонней (пределы интегрирования от –r(k;α) до r(k;α)

$$\int_{-r(k,\alpha)}^{r(k,\alpha)} f(t)\,dt = \alpha$$

критической области.

Распределение Стьюдента применяется для описания ошибок выборки при k<30. При k, превышающем 100, данное распределение практически соответствует нормальному, для значений k из диапазона от 30 до 100 различия между распределением Стьюдента и нормальным распределением составляют несколько процентов.

Поэтому относительно оценки ошибок малыми считаются выборки объемом не более 30 единиц, большими – объемом более 100 единиц. При аппроксимации распределения Стьюдента нормальным распределением для односторонней критической области вероятность

$$P\{t>t(k;\alpha)\} = u_{1-a}(0,k/(k-2)),$$

Где $u_{1-a}(0, k/(k–2))$ – квантиль нормального распределения. Аналогичное соотношение можно составить и для двусторонней критической области.

2.2.4. Распределение Фишера

Распределению Р.А. Фишера (F-распределению Фишера – Снедекора) подчиняется случайная величина

$$X = [(y1/k1)/(y2/k2)],$$

равная отношению двух случайных величине y_1 и y_2, имеющих хи-квадрат распределение с k_1 и k_2 степенями свободы. Область изменения аргумента x от 0 до бесконечности. Плотность распределения

$$f(x)=\left[\frac{k_1}{k_2}\right]^{k_1/2}\frac{\Gamma[(k_1+k_2)/2]}{\Gamma(k_1/2)\Gamma(k_2/2)}x^{(k_1-2)/2}\left[1+\frac{k_1}{k_2}x\right]^{-\frac{(k_1+k_2)}{2}}$$

В этом выражении k_1 обозначает число степеней свободы величины y_1 с большей дисперсией, k_2 – число степеней свободы величины y_2 с меньшей дисперсией. Плотность распределения – унимодальная, несимметричная, рисунок 48

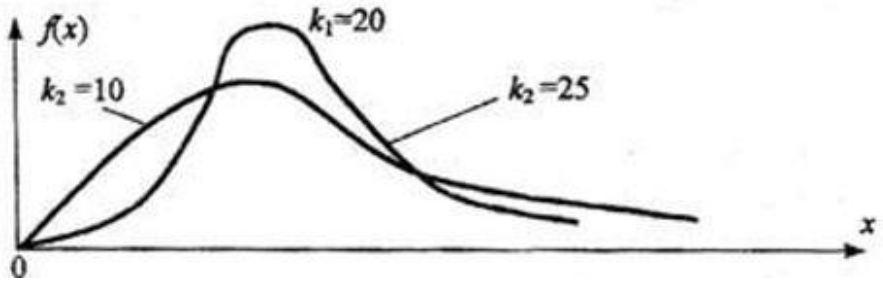


Рисунок 48 – Плотность распределения Фишера

Математическое ожидание случайной величины x

$m_1 = k_2/(k_2-2)$ при $k_2>2$,

дисперсия $m_2 = [2k_2^2(k_1+k_2-2)]/[k_1(k_2-2)^2(k_2-4)]$ при $k2 > 4$.

При $k_1>30$ и $k_2>30$ величина x распределена приближенно нормально: с центром распределения $(k_1-k_2)/(2k_1k_2)$ и дисперсией $(k_1+k_2)/(2k_1k_2)$.

2.3. Проверка гипотез о законе распределения

Обычно сущность проверки гипотезы о законе распределения ЭД заключается в следующем: Имеется выборка ЭД фиксированного объема, выбран или известен вид закона распределения генеральной совокупности. Необходимо оценить по этой выборке параметры закона, определить степень согласованности ЭД и выбранного закона распределения, в котором параметры заменены их оценками. Пока не будем касаться способов нахождения

оценок параметров распределения, а рассмотрим только вопрос проверки согласованности распределений с использованием наиболее употребительных критериев.

2.3.1. Критерий хи-квадрат К. Пирсона

Использование этого критерия основано на применении такой меры (статистики) расхождения между теоретическим $F(x)$ и эмпирическим распределением $F_n(x)$, которая приближенно подчиняется закону распределения χ^2.Гипотеза H_0 о согласованности распределений проверяется путем анализа распределения этой статистики. Применение критерия требует построения статистического ряда.

Итак, пусть выборка представлена статистическим рядом с количеством разрядов y. Наблюдаемая частота попаданий в i-й разряд n_i. В соответствии с теоретическим законом распределения ожидаемая частота попаданий в i-й разряд составляет F_i. Разность между наблюдаемой и ожидаемой частотой составит величину $(n_i–F_i)$. Для нахождения общей степени расхождения между $F(x)$ и $F_n(x)$ необходимо подсчитать взвешенную сумму квадратов разностей по всем разрядам статистического ряда:

$$\chi^2 = \sum_{i=1}^{n} \frac{(n_i - F_i)^2}{F_i}$$

Величина χ^2 при неограниченном увеличении n имеет распределение хи-квадрат (асимптотически распределена как хи-квадрат). Это распределение зависит от числа степеней свободы k, т.е. количества независимых значений слагаемых в выражении. Число степеней свободы равно числу y минус число линейных связей, наложенных на выборку. Одна связь существует в силу того, что любая частота может быть вычислена по совокупности частот в оставшихся y–1 разрядах. Кроме того, если параметры распределения неизвестны заранее, то имеется еще одно ограничение, обусловленное подгонкой распределения к выборке. Если по выборке определяются f параметров распределения, то число степеней свободы составит $k = y - f - 1$.

Очевидно, что чем меньше расхождение между теоретическими и эмпирическими частотами, тем меньше величина критерия. Область принятия гипотезы H_0 определяется условием $\chi2<\chi2(k,\alpha)$ – критическая точка распределения хи-квадрат с уровнем значимости α .

Вероятность ошибки первого рода равна α, вероят-

ность ошибки второго рода четко определить нельзя, потому что существует бесконечно большое множество различных способов несовпадения распределений. Мощность критерия зависит от количества разрядов и объема выборки. Критерий рекомендуется применять при n>200, допускается применение при n>40, именно при таких условиях критерий состоятелен (как правило, отвергает неверную нулевую гипотезу). Теоретическое значение вероятности F_i попадания случайной величины в i-й интервал определяется по формуле

$$F_i = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_{i-1}}^{x_i} exp\left(-\frac{(x-\mu_1)^2}{2\sigma^2}\right)dx.$$

Для нормального закона возможные значения случайной величины лежат в диапазоне от минус до плюс бесконечности, поэтому при расчетах оценок вероятностей крайний левый и крайний правый интервалы расширяются до минус и плюс бесконечности соответственно. Вычислить значения функции нормального распределения можно, воспользовавшись стандартными функциями табличного процессора или полиномом наилучшего приближения.

2.3.2. Критерий А.Н. Колмогорова

Для применения критерия А.Н. Колмогорова ЭД требуется представить в виде вариационного ряда (ЭД недопустимо объединять в разряды). В качестве меры расхождения между теоретической F(x) и эмпирической $F_n(x)$ функциями распределения непрерывной случайной величины X используется модуль максимальной разности

$$d_n = max|F(x) - F_n(x)|$$

А.Н. Колмогоров доказал, что какова бы ни была функция распределения F(x) величины X при неограниченном увеличении количества наблюдений n функция распределения случайной величины $d_n\sqrt{n}$ асимптотически приближается к функции распределения

$$K(\lambda) = P\left(d\sqrt{n} < \lambda\right) = \sum_{k=-\infty}^{\infty}(-1)^k\, exp\left(-2k^2\lambda^2\right).$$

Иначе говоря, критерий А.Н. Колмогорова характеризует

вероятность того, что величина $d_n\sqrt{n}$ не будет превосходить параметр l для любой теоретической функции распределения. Уровень значимости α выбирается из условия

$$P\left(d_n\sqrt{n} > \lambda\right) = \alpha = 1 - K(\lambda),$$

в силу предположения, что почти невозможно получить это равенство, когда существует соответствие между функциями F(x) и $F_n(x)$. Критерий А.Н. Колмогорова позволяет проверить согласованность распределений по малым выборкам, он проще критерия хи-квадрат, поэтому его часто применяют на практике. Но требуется учитывать два обстоятельства:

1. В соответствии с условиями его применения необходимо пользоваться следующим соотношением

$$d_n = \max(d_n^+, d_n^-),$$

$$d_n^+ = \max_{1 \le j \le n} \left|\frac{j}{n} - F(x)\right|, d_n^- = \max_{1 \le j \le n} \left|F(x) - \frac{j-1}{n}\right|$$

2. Условия применения критерия предусматривают, что теоретическая функция распределения известна полностью – известны вид функции и значения ее параметров. На практике параметры обычно неизвестны и оцениваются по ЭД. Но критерий не учитывает уменьшение числа степеней свободы при оценке параметров распределения по исходной выборке. Это приводит к завышению значения вероятности соблюдения нулевой гипотезы, т.е. повышается риск принять в качестве правдоподобной гипотезу, которая плохо согласуется с ЭД (повышается вероятность совершить ошибку второго рода). В качестве мерыпротиводействия такому выводу следует увеличить уровень значимости α, приняв его равным 0,1 – 0,2, что приведет к уменьшению зоны допустимых отклонений.

2.3.3. Критерий Р. Мизеса

В качестве меры различия теоретической функции распределения F(x) и эмпирической $F_n(x)$ по критерию Мизеса (критерию - ω^2) выступает средний квадрат отклонений по всем значениям аргумента x:

$$\omega_n^2 = \int_{-\infty}^{+\infty} [F_n(x) - F(x)]^2 \, dF(x)$$

Статистика критерия:

$$n\omega_n^2 = \frac{1}{12n} + \sum_{i=1}^{n} \left[F(x_i) - \frac{i - 0{,}5}{n} \right]^2$$

При неограниченном увеличении n существует предельное распределение статистики $n\omega_n^2$. Задав значение вероятности α можно определить критические значения $n\omega_n^2$ (α). Проверка гипотезы о законе распределения осуществляется обычным образом: если фактическое значение $n\omega_n^2$ окажется больше критического или равно ему, то согласно критерию Мизеса с уровнем значимостиα гипотеза H_0 о том, что закон распределения генеральной совокупности соответствует F(x), должна быть отвергнута.

Достоинством критерия Мизеса является быстрая сходимость к предельному закону, для этого достаточно не менее 40 наблюдений в области часто используемых на практике больших значений $n\omega_n^2$ (а не несколько сот, как для критерия хиквадрат).

Сопоставляя возможности различных критериев, необходимо отметить следующие особенности:

1. Критерий Пирсона устойчив к отдельным случайным ошибкам в ЭД. Однако его применение требует группирования данных по интервалам, выбор которых относительно произволен и подвержен противоречивым рекомендациям.

2. Критерий Колмогорова слабо чувствителен к виду закона распределения и подвержен влиянию помех в исходной выборке, но прост в применении.

3. Критерий Мизеса имеет ряд общих свойств с критерием Колмогорова: оба основаны непосредственно на результатах наблюдения и не требуют построения статистического ряда, что повышает объективность выводов; оба не учитывают уменьшение числа степеней свободы при определении параметров распределения по выборке, а это ведет к риску принятия ошибочной гипотезы. Их предпочтительно применять в тех случаях, когда параметры закона распределения известны априори, например, при проверке датчиков случайных чисел.

При проверке гипотез о законе распределения следует помнить, что слишком хорошее совпадение с выбранным законом

распределения может быть обусловлено некачественным экспериментом («подчистка» ЭД) или предвзятой предварительной обработкой результатов (некоторые результаты отбрасываются или округляются).

Выбор критерия проверки гипотезы относительно произволен. Разные критерии могут давать различные выводы о справедливости гипотезы, окончательное заключение в таком случае принимается на основе неформальных соображений. Точно также нет однозначных рекомендаций по выбору уровня значимости.

Рассмотренный подход к проверке гипотез, основанный на применении специальных таблиц критических точек распределения, сложился в эпоху "ручной" обработки ЭД, когда наличие таких таблиц существенно снижало трудоемкость вычислений. В настоящее время математические пакеты включают процедуры вычисления стандартных функций распределений, что позволяет отказаться от использования таблиц, но может потребовать изменения правил проверки.

2.4. Методы оценки параметров распределения

2.4.1 Точечная оценка параметров распределения

Точечная оценка предполагает нахождение единственной числовой величины, которая и принимается за значение параметра. Такую оценку целесообразно определять в тех случаях, когда объем ЭД достаточно велик. Причем не существует единого понятия о достаточном объеме ЭД, его значение зависит от вида оцениваемого параметра (к этому вопросу вернёмся при изучении методов интервальной оценки параметров, а предварительно будем считать достаточной выборку, содержащую не менее чем 10 значений). При малом объеме ЭД точечные оценки могут значительно отличаться от истинных значений параметров, что делает их непригодными для использования.

Задача точечной оценки параметров в типовом варианте постановки состоит в следующем:

Имеется выборка наблюдений (x_1, x_2, …, x_n) за случайной величиной X. Выборка должна быть представительной. Объем выборки n фиксирован. Известен вид закона распределения величины X, например, в форме плотности распределения f(T, x), где T – неизвестный (в общем случае векторный) параметр распределения. Параметр является неслучайной величиной.

Требуется найти оценку θ параметра T закона распределения.

Существует несколько методов решения задачи точечной оценки параметров, самыми распространёнными из которых являются методымаксимального (наибольшего) правдоподобия, моментов и квантилей.

2.4.2. Метод максимального правдоподобия

Метод был предложен Р. Фишером в 1912 г. Метод основан на исследовании вероятности получения выборки наблюдений (x_1, x_2, …x_n) за некоторым количественным признаком X. Значения выборки случайные, поэтому будем рассматривать их как независимые случайные величины X_1, X_2, …X_n, имеющие одинаковые распределения. Если количественный признак X непрерывный, то случайные величины X_1, X_2, …X_n имеют одинаковые плотности распределения вероятностей, зависящих от x и от вектора параметров распределения $\bar{\theta} = (\theta_1, \theta_2, \dots \theta_n)$, т.е. плотности $f(x, \bar{\theta})$.

Суть метода максимального правдоподобия заключается в свойстве случайной величины реализовывать в эксперименте в основном те свои значения(x_1, x_2, …x_n), вероятность которых максимальна. В этом случае совместная плотность распределения вероятностей$f(x_1, x_2, \dots x_n, \bar{\theta})$ случайных величин X_1, X_2, …X_n должна быть максимальной для непрерывного признака или совместное распределение вероятностей $P(X_1 = x_1, X_2 = x_2, \dots X_n = x_n,)$случайных величин X_1, X_2, …X_nдолжно быть максимальным для дискретного признака.

Ввиду независимости случайных величинX_1, X_2, …X_n их совместная плотность распределения будет равна произведению плотностей распределения, т.е.

$$f(x_1, x_2, \dots x_n, \bar{\theta}) = f(x_1, \bar{\theta}) \cdot f(x_2, \bar{\theta}) \cdot \dots \cdot f(x_n, \bar{\theta}),$$

а совместная вероятность равна произведению вероятностей событий

$$P(X_1 = x_1, X_2 = x_2, \dots X_n = x_n, \bar{\theta}) = p_{x_1}(\bar{\theta}) \cdot \dots \cdot p_{x_n}(\bar{\theta}).$$

Функцией правдоподобия $L(x_1, x_2, \dots x_n, \bar{\theta})$ называется функция, которая в случае непрерывного количественного признака определяется по формулам: для непрерывного признака

$$L(x_1, x_2, \ldots x_n, \bar{\theta}) = f(x_1, \bar{\theta}) \cdot f(x_2, \bar{\theta}) \cdot \ldots \cdot f(x_n, \bar{\theta}).$$

для дискретного признака

$$L(x_1, x_2, \ldots x_n, \bar{\theta}) = p_{x_1}(\bar{\theta}) \cdot \ldots \cdot p_{x_n}(\bar{\theta}).$$

В качестве оценки $\bar{\theta}_1$, …$\bar{\theta}_n$неизвестных параметров распределения $\theta_1, \ldots \theta_n$ берут те значения, при которых функция правдоподобия достигает максимума.

Технически задача поиска максимального значения функции правдоподобия облегчается, если рассмотреть не саму функцию, а натуральный логарифм от неё, т.е. функцию $\ln L(x_1, x_2, \ldots x_n, \bar{\theta})$. За оценки неизвестного вектора параметров q берётся решение $\bar{\theta}_1$, …$\bar{\theta}_n$ уравнения правдоподобия

$$\frac{\partial \ln L(x_1, x_2, \ldots x_n, \bar{\theta})}{\partial \theta_k} = 0; \quad k = \overline{1, n}.$$

Для проверки того, что точка оптимума соответствует максимуму функции правдоподобия, необходимо найти вторую производную от этой функции. И если вторая производная в точке оптимума отрицательна, то найденные значения параметров максимизируют функцию.

Итак, нахождение оценок максимального правдоподобия включает следующие этапы: построение функции правдоподобия (ее натурального логарифма); дифференцирование функции по искомым параметрам и составление системы уравнений; решение системы уравнений для нахождения оценок; определение второй производной функции, проверку ее знака в точке оптимума первой производной и формирование выводов.

Метод максимального правдоподобия позволяет получить состоятельные, эффективные (если таковые существуют, то полученное решение даст эффективные оценки), достаточные, асимптотически нормально распределенные оценки. Этот метод может давать как смещенные, так и несмещенные оценки. Смещение удается устранить введением поправок. Метод особенно полезен при малых выборках. Если функция максимального правдоподобия имеет несколько максимумов, то из них выбирают глобальный.

2.4.3. Метод моментов

Метод предложен К. Пирсоном в 1894 г.

Будем считать, что вид функции распределения изучаемого количественного признака известен, но параметры этого распределения неизвестны. Нужно оценить эти параметры по выборке. Сущность метода моментов заключается в том, что по конкретной выборке x_1, x_2, …x_n генеральной совокупности X, распределение которой известно с точностью до параметров $\theta_1, \dots \theta_m$ выбирается столько эмпирических моментов, сколько требуется оценить неизвестных параметров распределения. Желательно применять моменты младших порядков, так как погрешности вычисления оценок резко возрастают с увеличением порядка момента.

Вычисленные по ЭД оценки моментов приравниваются к теоретическим моментам; параметры распределения определяются через моменты, и составляются уравнения, выражающие зависимость параметров от моментов, в результате получается система уравнений. Решение этой системы дает оценки параметров распределения генеральной совокупности.

2.4.4. Интервальная оценка параметров распределения

Интервальный метод оценивания параметров распределения случайных величин заключается в определении интервала (а не единичного значения), в котором с заданной степенью достоверности будет заключено значение оцениваемого параметра.

Интервальная оценка характеризуется двумя числами – концами интервала, внутри которого предположительно находится истинное значение параметра. Иначе говоря, вместо отдельной точки для оцениваемого параметра можно установить интервал значений, одна из точек которого является своего рода "лучшей" оценкой. Интервальные оценки являются более полными и надежными по сравнению с точечными, они применяются как для больших, так и для малых выборок. Совокупность методов определения промежутка, в котором лежит значение параметра T, получила название методов интервального оценивания. К их числу принадлежит метод Неймана.

Постановка задачи интервальной оценки параметровзаключается в следующем: Имеется выборка наблюдений x_1, x_2, …x_n, за случайной величинойX. Объем выборки nфиксирован. Предположим, что статистическая характеристика $\hat{\theta}$, рассчитанная по данным выборки, является выборочной оценкой неизвест-

ного параметра θ генеральной совокупности, причёмθ - это постоянное число. Оценка$\tilde{\theta}$характеризуетпараметрθ тем точнее, чем меньше абсолютная величина разности$|\theta - \tilde{\theta}|$.

Если $|\theta - \tilde{\theta}|<\delta$, где $\delta>0$, то числоδ называется точностью оценки $\tilde{\theta}$.

Нельзя утверждать, что оценкаабсолютно$\tilde{\theta}$удовлетворяет неравенству$|\theta - \tilde{\theta}|<\delta$. Однако можно задать вероятностьγ, с которой это неравенство осуществляется.

Надёжность или доверительная вероятность оценки$\tilde{\theta}$ - это вероятность γ, с которой осуществляется неравенство $|\theta - \tilde{\theta}|<\delta$.

Принято надёжность оценки задавать перед процессом оценивания параметра генеральной совокупности. В качестве вероятности γ берут число, близкое к 1 (от 0,95 до 0,999).

Пусть $P = (|\theta - \tilde{\theta}|<\delta)=\gamma$. Если неравенство $|\theta - \tilde{\theta}|<\delta$ заменить равносильным ему двойным неравенством:

$$-\delta < \theta - \tilde{\theta} < \delta, \text{ или } \tilde{\theta} - \delta < \theta < \tilde{\theta} + \delta,$$

то получим:

$$P\left(\tilde{\theta} - \delta < \tilde{\theta} + \delta\right) = \gamma .$$

Вероятностный смысл данного отношения таков: вероятность того, что интервал $(\tilde{\theta} - \delta; \tilde{\theta} + \delta)$ покрывает неизвестный параметр$\tilde{\theta}$, равна γ .

Интервал $(\tilde{\theta} - \delta; \tilde{\theta} + \delta)$который покрывает оцениваемый неизвестный параметр $\tilde{\theta}$ с заданной вероятностьюγ называетсядоверительным интервалом.

Границы$\left(\tilde{\theta} - \delta\right)$ и $(\tilde{\theta} + \delta)$ называются доверительными границами интервала. Они определяются на основе выборочных данных и являются функциями от случайных величин X_1, X_2, …X_n , и, следовательно, сами являются случайными величинами.

2.4.5. Общий метод построения доверительных интервалов

Доверительный интервал для оценки математического ожидания нормального распределения при известном среднем квадратичном отклонении.

Будем считать, что количественный признак X генеральной совокупности подчиняется нормальному закону распределения с параметрами α и σ . Среднее квадратичное данного распределения σ считается известным. Нужно оценить неизвестное математическое ожидание α генеральной совокупности по выборочной средней $\bar{x}$, т.е. необходимо определить доверительный интервал, который покрывает параметр α с заданной надёжностью γ .

Имеет место следующее утверждение: если случайная величина X подчиняется нормальному закону распределения, то выборочная средняя $\bar{X}$, найденная по независимым наблюдениям, также подчиняется нормальному закону распределения с параметрами:

$$M(\bar{X}) = \alpha; \quad \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}.$$

Для того, чтобы определить доверительный интервал, необходимо, чтобы выполнялось равенство $P(|\bar{X} - \alpha| < \delta) = \gamma$.

Для нормально распределённой случайной величины X имеет место равенство:

$$P\left(\left|\bar{X} - \alpha\right| < \delta\right) = 2\Phi\left(\frac{\delta}{\sigma(\bar{X})}\right).$$

$\bar{X}$ имеет нормальное распределениеи$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}$. Поэтому

$$P\left(\left|\bar{X} - \alpha\right| < \delta\right) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = 2\Phi(t); \quad t = \frac{\delta\sqrt{n}}{\sigma}$$

Из равенства $t = \frac{\delta\sqrt{n}}{\sigma}$выразим$\delta = \frac{\sigma}{\sqrt{n}}$и получим:

$$P\left(\left|\bar{X} - \alpha\right| < t\frac{\sigma}{\sqrt{n}}\right) = 2\Phi(t).$$

Так как вероятность P задана и равна γ, окончательно получаем:

$$P\left(\bar{x}-t\frac{\sigma}{\sqrt{n}}<\alpha<\bar{x}+t\frac{\sigma}{\sqrt{n}}\right)=2\Phi(t)=\gamma.$$

Таким образом, с доверительной вероятностью γ можно утверждать, что доверительный интервал($\bar{x}$ - $t\frac{\sigma}{\sqrt{n}}$<α<$t\frac{\sigma}{\sqrt{n}}$) покрывает неизвестный параметр α с точностью $\delta = t\frac{\sigma}{\sqrt{n}}$.

$\bar{x}$ - находим по данным выборки, n – это объём выборки, σ- величина, известная заранее.

Число t находится из равенства 2Φ(t)=γ, или$\Phi(t)=\frac{\gamma}{2}$. По таблице функции Лапласа определяется для заданной доверительной вероятности γ величина t и затем находится величина δ.

2.4.6. Методика Вальда проверки гипотезы о свойствах случайной величины

Рассматривается следующая задача: имеются экспериментальные данные о некоторой случайной величине ξ; известно аналитическое выражение для плотности распределения этой случайной величины - некоторое выражение $f(x,0)$, где x- аргумент, а θ - некоторый параметр, принимающий одно из двух возможных значений $-\theta=\theta_1$ или $\theta=\theta_2$, причем $\theta_1<\theta_2$. Какое именно из этих двух значений принимает параметр θ, - неизвестно. Именно это и надо определить по значениям случайной величины ξ, которые можно разыгрывать неограниченно много.

Примером такой плотности может служить плотность нормального распределения:

$$f(x,\theta)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\theta)^2}{2\sigma^2}},$$

где, кроме параметраθ, участвует так же еще один параметр σ.

Опишем методику Вальда решения поставленной задачи.

Фиксируем достаточно малые числаα иβ, например$\alpha,\beta\in(0,1)$; фиксируем, далее, числа

$$A=\frac{1-\beta}{\alpha},\quad B=\frac{\beta}{1-\alpha};$$

при достаточно малыхα, β между числами A, B имеет место неравенство: $A > B$ именно для этого неравенства должны быть достаточно малы числаα, β. Пусть, далее, $\xi_1, \xi_2, \ldots, \xi_k, \ldots$ - значения случайной величины ξ , получаемые экспериментально. Построим суммы:

$$\Lambda_t = \sum_{i=1}^{t} \ln \frac{f(\xi_i, \theta_2)}{f(\xi_i, \theta_1)}.$$

Основная еорема Вальда: если $\Lambda_t > \ln A$, то с вероятностью ошибки β; если же $\Lambda_t > \ln B$, то $\theta = \theta_1$ с вероятностью ошибки α.

Принято рассматривать в описанной ситуации две величины - n и Δg_t, определяемые так:

то минимальное значение t, при котором выполняется одно из неравенств$\Lambda_t > \ln A$или $\Lambda_t > \ln B$ - является случайной величиной, зависящей от той или иной реализации случайной величины$\xi_1, \xi_2, \ldots, \xi_k, \ldots$; именно это минимальное значение t и является случайной величиной n;

учитывая смысл только что введенного обозначения n, введем величину

$$\Delta g_t = \frac{1}{n} \sum_{i=1}^{n} \ln \frac{f(\xi_i, \theta_2)}{f(\xi_i, \theta_1)};$$

очевидно, и эта величина - также случайная. На практике часто имеют смысл математические ожидания последних двух случайных величин; их находят как обычные средние арифметические значения реализаций этих величин, конструируемых по различным реализациям $\xi_1, \xi_2, \ldots, \xi_k, \ldots$

ЗАКЛЮЧЕНИЕ

Как можно было заметить из вышеперечисленных методов планирования эксперимента и обработки статистических данных, большинство из них сосредоточены на концепции однофакторного эксперимента [16].

Однофакторный, или классический, эксперимент базируется на допущении о том, что исследователь имеет возможность варьировать факторы, участвующие в исследовательской ситуации, по одному. Из этого следует, что экспериментатор способен выделить изучаемую зависимость в чистом виде, может чётко вычленять воздействующие на зависимые переменные факторы (может, скажем, как-то упорядочить их во времени и пространстве, «включать» и «выключать» их по своему усмотрению и т.п.). Однако на самом деле исследовательские ситуации часто оказываются гораздо более сложными.

Выход к более утончённой методологии, имеющей дело с комплексным, принципиально неразделимым действием факторов, был осуществлён прежде всего под влиянием работ английского учёного Рональда Фишера (1890-1962), посвящённых агробиологическим экспериментам в 1925г. В сложных системах факторы, воздействующие на изучаемый объект, действуют не изолированно и не независимо друг от друга, как это предполагала концепция классического эксперимента, а довольно сложным, взаимосвязанным способом. Они зачастую сцеплены между собой таким образом, что попытка варьировать одну независимую переменную автоматически приводит к некоему замысловатому изменению и других факторов. Это означает, что исследователю, приходится иметь дело с особой комплексной организацией этих факторов. Кроме того, исследователя может интересовать действие не изолированных факторов, которое в реальности не встречается, а именно влияние различных возможных комбинаций факторов.

Идея многофакторного эксперимента (иногда используют упрощённое название факторный эксперимент) состоит в следующем. Исследователь может варьировать независимые переменные как комплекс, т.е. одновременно сразу несколько; после серии экспериментов полученные результаты должны быть подвергнуты специальному статистическому анализу, где каждый участвующий фактор будет оценён по результатам всех опытов данной серии. Используя соответствующие схемы и обрабатывая данные по особым статистическим методикам, позволяющим изу-

чать эффективность совместного полифакторного воздействия (методики дисперсионного анализа), исследователь получает картину, отражающую вклад каждого фактора в изменяющихся условиях. В итоге экспериментатор имеет возможность изучать самые сложные комбинации факторов. Причём это осуществляется достаточно экономичным способом, т.к. информативность экспериментов зависит в данном случае не от их количества в серии, а от концептуальной организации исследований.

Многофакторный эксперимент - мощное средство современной науки. К его достоинствам относятся: эффективность использования времени и средств (ведь проведение ряда экспериментов с отдельными пофакторными модификациями требует значительных затрат), что выражается прежде всего в сокращении числа опытов, необходимых для решения исследовательской задачи; значительная информативность эксперимента (т.к. получаемый результат показывает удельный вес каждого фактора в их совокупном действии); высокая степень достоверности данных (в то время как при попытке использовать методологию классического эксперимента результаты могут оказаться неудовлетворительными из-за воздействий неподконтрольных факторов).

Как показывает практика, часто исследователи в своих работах рассматривают многофакторный эксперимент как совокупность однофакторных [17]. Обработка данных в большинстве случаев затруднительна, поэтому фактически исследуемая система представляется посредством «черного ящика». Обработка данных осуществляется при помощи современных вычислительных средств и программных продуктов, в полной мере реализующих методы математического моделирования, описанные в данном пособии.

СПИСОК ЛИТЕРАТУРЫ

1. Официальный русскоязычный сайт компании SPSS – http://www/spss.com, http://www.predictivesolutions.ru/software/

2. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА-М, 2003. – 544

3. Боровиков, В.П. Прогнозирование в программе STATISTICA в среде Windows: Основы теории и интенсивная практика на компьютере: учеб. пособие / В.П. Боровиков, Г.И. Ивченко. – М.: Финансы и статистика, 2006. – 368 с.

4. Вуколов, Э.А. Основы статистического анализа. Практи-

кум по статистическим методам и исследованию операций с использованием пакетов Statistica и Excel: учеб. пособие / Э.А. Вуколов. – М.: ИНФРА-М, 2004. – 462 с.

5. Кацко, И.А. Практикум по анализу данных на компьютере / И.А. Кацко, Н.Б. Паклин; под ред. Г.В. Гореловой. – М.: КолосС, 2009. – 278 с.

6. http://www.statsoft.ru/Электронный ресурс: Российское представительство компании StatSoft. Дата обращения: 23.12.2019.

7. http://info.statgraphics.com/ Электронный ресурс: Сайт статистического пакета STATGRAPHICSPLUS. Дата обращения: 23.12.2019.

8. Орлова И.В. Экономико-математическое моделирование: практическое пособие по решению задач. – М.: Вузовский учебник, 2007. – 144 с. 12. Орлова, И.В. Экономико-математические методы и модели: компьютерное моделирование: учеб. пособие. – М.: Вузовский учебник, 2009. – 144 с.

9. Орлова, И.В. Экономико-математические методы и модели: компьютерное моделирование: учеб. пособие. – М.: Вузовский учебник, 2009. – 365 с.

10. http://www.xlstat.com/Электронный ресурс: Макрос-дополнение XLSTAT-Pro для MSExcel. Дата обращения: 23.12.2019.

11. http://www.exponenta.ru/Электронный ресурс: Образовательный математический сайт.Дата обращения: 23.12.2019.

12. https://studfile.net/preview/3507786/Электронный ресурс: Дата обращения: 23.12.2019.

13. https://www.malavida.com/ru/soft/statgraphics/Электронный ресурс: Statgraphics, программа для статистики. Дата обращения: 23.12.2019.

14. Реброва И.А. Планирование эксперимента: учебное пособие. – Омск: СибАДИ, 2010. – 105 с.

15. Рогов В.А., Поздняк Г.Г. Методика и практика технических экспериментов: учеб. пособие для студ. высш. учеб. заведений /В.А. Рогов, Г.Г. Поздняк. – М.: Издательский центр «Академия», 2005. – 288 с.

16. https://vuzlit.ru/817990 Электронный ресурс. Дата обращения: 30.01.2019.

17. В.Н. Шкляр. Планирование эксперимента и обработка результатов. Конспект лекций для магистров по направлению 220200 «Автоматизация и управление в технических (мехатронных) системах». Издательство Томского политехнического университета, 2010.