СПЕЦИАЛИЗИРОВАННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Практикум
для магистров, обучающихся по направлению подготовки
35.04.06 «Электрооборудование и электротехнологии»



Зерноград – 2025

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА РОССИЙСКОЙ ФЕДЕРАЦИИ

ДЕПАРТАМЕНТ НАУЧНО-ТЕХНОЛОГИЧЕСКОЙ ПОЛИТИКИ И ОБРАЗОВАНИЯ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ «ДОНСКОЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО Донской ГАУ)

АЗОВО-ЧЕРНОМОРСКИЙ ИНЖЕНЕРНЫЙ ИНСТИТУТ – ФИЛИАЛ
ФЕДЕРАЛЬНОГО ГОСУДАРСТВЕННОГО БЮДЖЕТНОГО ОБРАЗОВАТЕЛЬНОГО
УЧРЕЖДЕНИЯ ВЫСШЕГО ОБРАЗОВАНИЯ
«ДОНСКОЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ» В Г. ЗЕРНОГРАДЕ
(Азово-Черноморский инженерный институт ФГБОУ ВО Донской ГАУ)

Кафедра математики и биоинформатики

СПЕЦИАЛИЗИРОВАННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Практикум

Зерноград – 2025

© Грачева Н.Н., Руденко Н.Б.,
Литвинов В.Н., Украинцев М.М., 2025
© Азово-Черноморский инженерный
институт ФГБОУ ВО Донской ГАУ, 2025

Об издании – 1, 2
Содержание

УДК 004.4:167.2

Публикуется по решению методического совета
энергетического факультета
Азово-Черноморского инженерного института – филиала
Федерального государственного бюджетного образовательного учреждения
высшего образования «Донской государственный аграрный университет»
в г. Зернограде

Рецензенты:
к.т.н., доцент кафедры М и БИ Лебедев К.Н.

Специализированное программное обеспечение для научных исследований [Электронный ресурс]: практикум / Н.Н. Грачева, Н.Б. Руденко, В.Н. Литвинов, М.М. Украинцев. – 2-е изд., перераб. – Электрон. дан. – Зерноград: Азово-Черноморский инженерный институт ФГБОУ ВО ДГАУ, 2025. – 256 с. – Режим доступа: Локальная сеть Библиотеки Азово-Черноморского инженерного института ФГБОУ ВО Донской ГАУ.

Учебное пособие содержит практикум для магистрантов, обучающихся по направлению подготовки – 35.04.06 «Агроинженерия», магистерская программа – «Электрооборудование и электротехнологии» и изучающих дисциплину «Специализированное программное обеспечение для научных исследований». Материал, изложенный в пособии, может быть использован для магистрантов очной и заочной форм обучения, аспирантов.

Рассмотрено и одобрено на заседании кафедры М и БИ.
Протокол № 9 от 22 апреля 2025 г.

Рассмотрено и одобрено методическим советом энергетического факультета.
Протокол № 4 от 19 мая 2025 г.

© Грачева Н.Н., Руденко Н.Б.,
Литвинов В.Н., Украинцев М.М., 2025
© Азово-Черноморский инженерный
институт ФГБОУ ВО Донской ГАУ, 2025

Содержание

Введение 4

Практическое занятие № 1.
Изучение возможностей и интерфейса программы для обработки экспериментальных данных 6

Практическое занятие № 2.
Представление экспериментальных данных при помощи графиков. Настройка графиков 29

Практическое занятие № 3.
Законы распределения вероятностей. Проверка статистических гипотез в практике ученого. Подбор теоретического распределения к эмпирическим данным с использованием критерия χ^2 55

Практическое занятие № 4.
Оценка различий между дисперсиями по критерию Фишера. Однофакторный дисперсионный анализ. Многофакторный дисперсионный анализ 79

Практическое занятие № 5.
Линейная регрессия и нелинейная парная регрессии. Множественная регрессия 107

Практическое занятие № 6.
Контрольная работа № 1. *«Применение методов и способов решения исследовательских задач. Дисперсионный и регрессионный анализ»* 138

Практическое занятие № 7.
Планирование научного эксперимента. Планы первого и второго порядков 140

Практическое занятие № 8.
Реализация плана многофакторного эксперимента. Обработка и анализ результатов эксперимента. Анализ регрессионной модели. 178

Практическое занятие № 9.
Контрольная работа № 2. *«Прогнозирование изменения состояний объекта исследования. Получение научного результата, имеющего практическое значение»* 201

Самостоятельная работа 202

Практическое занятие № 10.
Настройка интерфейса программы Statistica 202

Практическое занятие № 11.
Первичная обработка массива данных 214

Практическое занятие № 12.
Корреляция 228

Литература 242

Приложения 245

Введение

Формой существования и развития науки является научное исследование. Научное исследование – процесс познания окружающего нас мира для последующего практического использования изученных закономерностей. Одним из основных способов получения новых знаний являются экспериментальные исследования.

Эксперимент – это метод научного познания, который помогает исследовать окружающие нас процессы и явления в определённых (заданных исследователем) воспроизводимых условиях. Это основной способ научного познания и часто – критерий истинности научного знания в целом.

Планирование научного эксперимента – важная часть научных исследований, от которой зависит качество научного познания. Важным этапом в подготовке молодых исследователей является обучение правилам планирования научного эксперимента, методам выбора факторов, влияющих на поведение исследуемого объекта или развитие исследуемого явления.

Практикум «Специализированное программное обеспечение для научных исследований» соответствует Федеральному государственному образовательному стандарту высшего образования (ФГОС ВО) и является частью учебно-методического комплекса дисциплины «Специализированное программное обеспечение для научных исследований».

Пособие предназначено для магистрантов, обучающихся по направлению подготовки – 35.04.06 «Агроинженерия», направленность – «Электрооборудования и электротехнологии» и изучающих дисциплину «Специализированное программное обеспечение для научных исследований», может быть использован для обучения аспирантов.

Практикум составлен таким образом, чтобы магистранты осваивали дисциплину не только на аудиторных занятиях, а также могли изучить необходимые разделы дисциплины самостоятельно.

В результате изучения дисциплины «Специализированное программное обеспечение для научных исследований» магистрант должен:

– знать основные направления развития программного обеспечения, используемого для математической обработки экспериментальных данных;

– уметь выполнять статистическую обработку данных, полученных в результате научного эксперимента, в программе Statistica;

– уметь планировать научный эксперимент в программе Statistica;

– проводить корреляционный и дисперсионный анализ в программе Statistica;

– строить регрессионные модели по экспериментальным данным в программе Statistica.

Символика статистических терминов и характеристик

$Y_i(y_i)$ – отдельное значение признака (варианта, отклика, зависимой переменной);
$X_i(x_i)$ – фактор (независимая переменная);
$N(n)$ – объем выборки или число наблюдений (повторность в опыте);
$\bar{y}$ – средняя выборочная: среднеарифметическая или средневзвешенная;
SS – сумма квадратов отклонений;
df – число степеней свободы;
C – корректирующий (вычитаемый) член, или поправка;
S^2 – дисперсия (средний квадрат отклонений);
S – среднеквадратическое, или стандартное, отклонение (показатель изменчивости);
σ – среднее квадратическое отклонение;
S_d – ошибка разности средних (долей признака);
S_r – стандартная ошибка коэффициента корреляции;
H_0 – нулевая гипотеза;
H_1 – альтернативная гипотеза;
α – уровень значимости или риска и ошибка I рода;
t – фактическое значение критерия Стьюдента;
$t_{05}(t_{01})$ – табличное значение, или критический квантиль критерия t на 5% (1%) i-го уровня риска;
F – фактическое значение критерия Фишера;
F_α – расчетный (табличный) критерий существенности Фишера;
$\chi^2(\chi^2_\alpha)$ – расчетный (табличный) критерий существенности хи-квадрат;
p – уровень доверия или надежности (доверительная вероятность);
HCP – наименьшая существенная разность, критерий существенности при сравнении средних;
E_x – коэффициент эластичности;
S_x и S_y – стандартное отклонение для признака *x* и *y* соответственно;
S_{xy} – общее стандартное отклонение для *x* и *y*, или ковариация в среднем;
r – линейный коэффициент корреляции (ранг);
$B_i(b_i)$ – коэффициент уравнения регрессии *y* по *x* (и *x* по *y*);
R^2 – коэффициент детерминации, %;
ε – ошибка опыта.

Практическое занятие № 1

Тема: **ИЗУЧЕНИЕ ВОЗМОЖНОСТЕЙ И ИНТЕРФЕЙСА ПРОГРАММЫ STATISTICA**

1 ЦЕЛЬ РАБОТЫ

Изучить:

- структуру программы;
- назначение отдельных модулей;
- назначение отдельных элементов интерфейса;
- назначение основных команд главного меню;
- структуру, типы и форматы данных таблиц SpreadSheet.

Научиться:

- задавать параметры переменных при помощи окна спецификаций переменных;
- использовать операции редактирования переменных, случаев и блоков данных.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Модульная структура программы Statistica

Statistica – программный пакет для статистического анализа, разработанный компанией StatSoft, реализующий функции анализа данных, управления данными, визуализации данных с привлечением статистических методов

Программа Statistica реализована по модульному принципу. Все методы статистической обработки разбиты на группы логически связанных задач. Каждая группа задач реализована в виде отдельного модуля.

Модуль представляет собой законченную программу. В программе Statistica 5.5 всего около тридцати модулей. Взаимодействия между модулями организовано с помощью переключателя модулей Module Switcher. Назначение некоторых модулей представлено в таблице 1.1.

Работать с данными в программе можно в двух основных режимах:

1 Режим одного открытого статистического модуля.

2 Режим нескольких открытых модулей. Это значит, что Statistica предоставляет возможность работать с одной и той же таблицей данных в нескольких модулях одновременно. Данное обстоятельство позволяет проводить одновременно различные статистические тесты над данными и сравнивать их результаты, что дает возможность «видеть предмет одновременно с разных сторон». Количество одновременно активированных модулей ограничивается системными ресурсами компьютера. Современные компьютеры позволяют открыть все имеющиеся модули одновременно. Изменения, вне-

сенные в таблицу данных в одном модуле, отражаются в другом модуле только после сохранения файла с изменениями.

Таблица 1.1 – Назначение модулей программы Statistica

Название модуля	Имя файла	Назначение модуля
Basic Statistics and Tables	Sta_bas	Содержит основные описательные статистики, методы статистического анализа различных таблиц, инструментарий для разведочного анализа данных
Nonparametric Statistics	Sta_non	Непараметрические статистики и подгонка распределений
ANOVA/MANOVA	Sta_man	Статистика измерений с несколькими повторениями, ковариация, регрессионный анализ, дискриминантный анализ
Multiple Regression	Sta_reg	Множественная регрессия, оценка коэффициентов регрессии
Experimental Design	Sta_exp	Планирование эксперимента, анализ результатов эксперимента, оценка коэффициентов регрессии
Time Series/ Forecasting	Sta_tim	Анализ временных рядов и прогнозирование
Module Switcher	Sta_win	Переключатель модулей

2.2 Типы документов программы Statistica

Для работы с данными в программе Statistica используются четыре типа документов:

– таблица данных SpreadSheet – предназначена для ввода исходных данных, подвергаемых статистическому анализу;

– таблица результатов ScrollSheet – предназначена для вывода числовых и текстовых данных, полученных в результате статистического анализа;

– график – документ в графическом формате, предназначенный для визуализации числовой информации;

– отчет – документ в формате RTF (Rich Text Format – расширенный текстовый формат), предназначенный для вывода текстовой, числовой и графической информации о результатах анализа.

2.3 Рабочие книги

В ходе анализа данных, размещенных в таблице SpreadSheet, получают таблицы результатов (ScrollSheet) и графики (Graphs). Часто бывает необходимо сохранить полученные результаты вместе с исходной таблицей для по-

следующих сеансов анализа данных. Для этой цели служат рабочие книги (Workbook). После создания рабочей книги открытие таблицы данных книги приводит к автоматическому отображению таблиц результатов и графиков.

2.4 Запуск программы Statistica

Запустить программу Statistica можно несколькими способами. Рассмотрим некоторые из них.

1 С помощью меню **Пуск** (Пуск – Программы – Statistica – Statistica). При этом на экране появляется переключатель модулей (Statistica Module Switcher), в котором выбирается необходимый модуль. Модуль Basic Statistics and Tables можно открыть непосредственно из меню **Пуск**, выбрав последовательно Пуск – Программы – Statistica – Basic Statistics.

2 При помощи ярлыка на рабочем столе. При этом запускается модуль Basic Statistics and Tables.

2.5 Переключатель модулей Module Switcher

Вызов модулей производится с помощью переключателя модулей Module Switcher (рисунок 1.1). Переключатель можно вызвать одним из следующих способов:

– при помощи кнопки Переключатель модулей на панели инструментов ;

– двойным щелчком левой кнопки мыши на свободном месте рабочей области Statistica;

– из меню Analysis – Other Statistics (Анализ – Другие статистики).

В левой части диалогового окна Переключателя модулей расположен список модулей, а в правой части дается краткое описание статистических процедур, доступных в выбранном модуле.

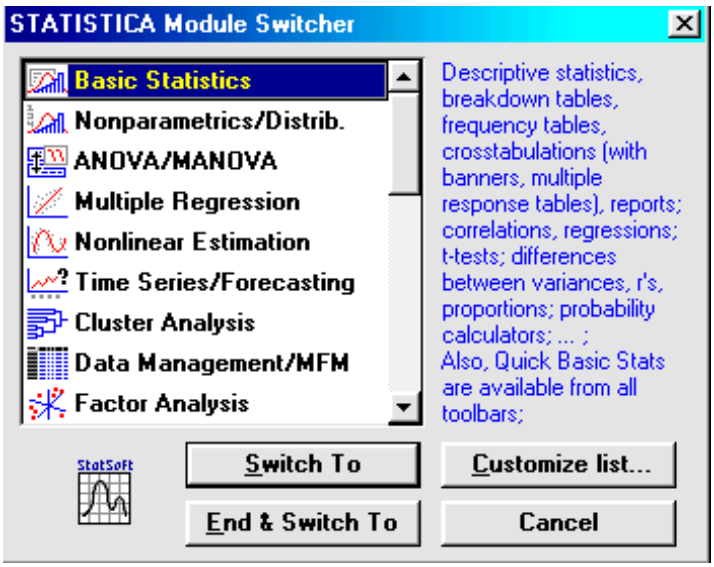


Рисунок 1.1 – Переключатель модулей (**Module Switcher**)

Выбор модуля может быть осуществлен с помощью кнопок:

– Switch To (Переключить) – открывается дополнительное окно выбранного модуля при этом окно текущего модуля остается открытым;

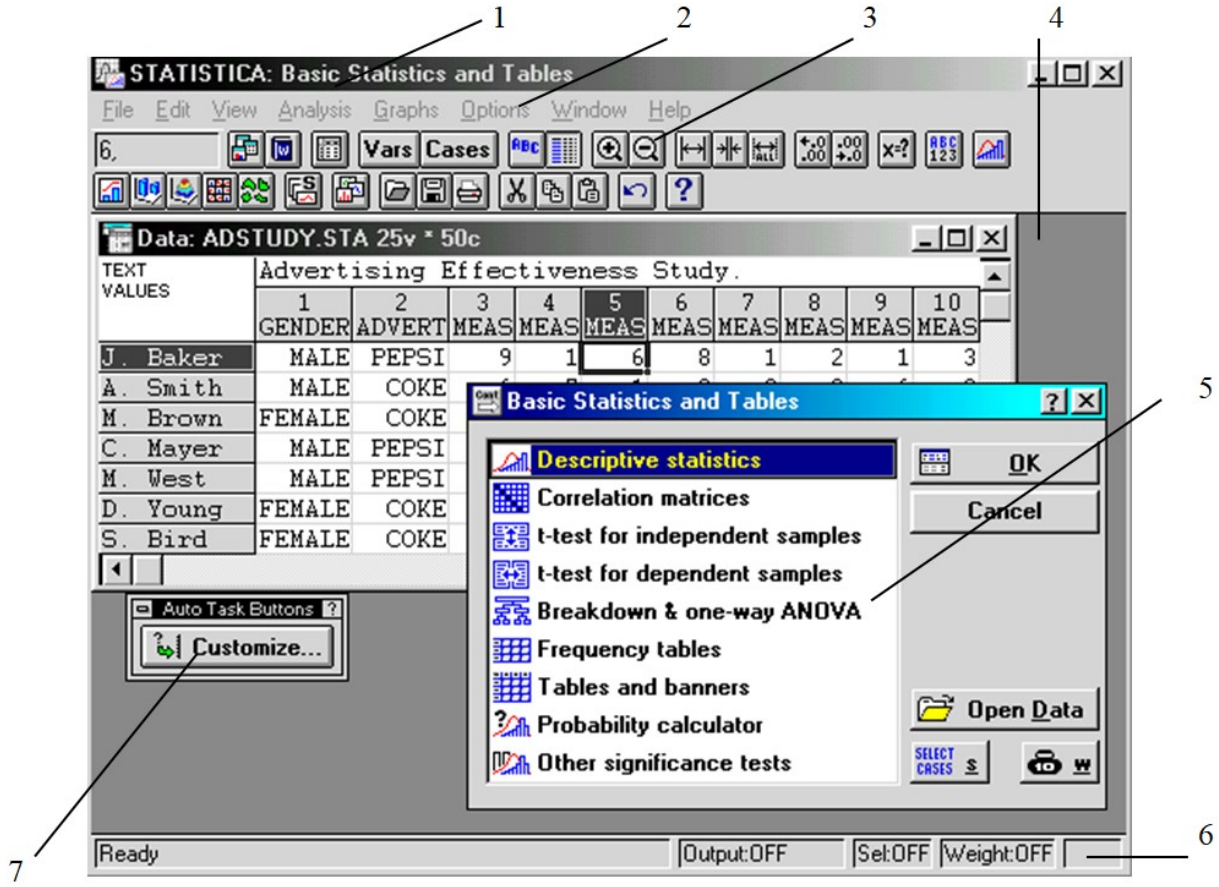
– End & Switch To (Закончить и переключить) – закрывается окно текущего модуля и открывается окно выбранного модуля.

Двойной щелчок мыши на названии модуля в левой части Переключателя модулей аналогичен выделению этого модуля и нажатию на кнопку Switch To (Переключить).

2.6 Окно программы Statistica

Окно программы состоит из следующих элементов (рисунок 1.2):

– строка заголовка 1;
– главное меню 2;
– панели инструментов 3;
– рабочая область 4;
– стартовая панель 5;
– строка состояния 6;
– кнопка автозадач 7.



1 – строка заголовка, 2 – главное меню, 3 – панели инструментов, 4 – рабочая область, 5 – стартовая панель, 6 – строка состояния, 7 – кнопка автозадач

Рисунок 1.2 – Окно программы Statistica

2.6.1 Строка заголовка

В программе Statistica используется стандартная для Windows-приложений строка заголовка. Она содержит название программы, название модуля и кнопки управления состоянием окна.

2.6.2 Главное меню

Главное меню содержит команды программы, разбитые на группы по функциональному признаку и представляет собой набор выпадающих меню. Количество пунктов главного меню и набор команд, включенных в эти пункты, отличается для различных типов документов.

Переключение между окнами различных типов документов в одном модуле ведет к автоматическому изменению набора команд Главного меню.

Главное меню для документов ScrollSheet и SpreadSheet состоит из следующих выпадающих меню:

- File (Файл);
- Edit (Правка);
- View (Вид);
- Analysis (Анализ);
- Graphs (Графики);
- Options (Сервис);
- Window (Окно);
- Help (Помощь).

Назначение основных опций главного меню для наиболее часто используемых документов SpreadSheet и ScrollSheet приведено в таблице 1.2. Часть команд главного меню для документов Graphics (Графики) будет рассмотрена далее в работах, посвященных построению и настройке графиков.

Таблица 1.2 – Назначение опций главного меню для таблиц данных и результатов

Раздел меню	Опция меню	Назначение
1	2	3
File (Файл)	Содержит опции открытия, сохранения, печати файлов; импорта, экспорта данных	
	New Data (Новые данные)	Создание нового файла данных (создается таблица размером 10×10)
	Import Data (Импортировать данные)	Импорт данных из других программ
	Open Data (Открыть данные)	Открытие файла данных

Продолжение таблицы 1.2

1	2	3
File (Файл)	Open Other (Открыть другие)	Открытие файлов графика (Graph), результатов (ScrollSheet), отчета (Text/Output)
	Save (Сохранить), Save As (Сохранить как)	Сохранение файла и сохранение с выбором параметров
	Print (Печать)	Распечатка данных на принтере
	Page/Output Setup (Параметры страницы)	Настройка параметров страницы при выводе на печать
Edit (Правка)	Содержит опции для редактирования параметров таблицы; для работы с буфером обмена; средства автоматизированного ввода и преобразования данных	
	Cut (Вырезать)	Вырезание блока данных в буфер обмена
	Copy (Копировать)	Копирование блока данных в буфер обмена
	Paste (Вставить)	Вставка данных из буфера обмена
	Undo (Отмена)	Отмена последнего действия
	Cell (Ячейка)	Группа команд для работы с ячейками
	Variables (Переменные)	Группа команд для работы с переменными
	Cases (Случаи)	Группа команд для работы со случаями
	Find (Найти)	Поиск имен переменных и значений переменных
	Replace (Заменить)	Поиск и замена имен переменных и значений переменных
	Transpose (Транспонировать)	Транспонирование блока данных (блок должен быть квадратным)
	Fill/Standardize Block (Заполнить стандартизовать блок)	Группа команд для заполнения и стандартизации блоков
	Block Stats (Блоковые статистики)	Вычисление основных статистик для строк и столбцов

Продолжение таблицы 1.2

1	2	3
View (Вид)	Содержит опции, задающие параметры отображения элементов окна программы	
	Tool Bar (Панель инструментов)	Задание вида панели инструментов
	Status Bar (Строка состояния)	Включение/выключение отображения строки состояния
	Auto Task Buttons (Кнопка автозадач)	Включение/выключение отображения кнопки автозадач
	Font (Шрифт)	Задание параметров шрифта
	Colors (Цвета)	Задание цветов различных элементов
	Text/Output Window (Окно отчета)	Создание окна отчета
Analysis (Анализ)	Содержит команды, необходимые непосредственно для статистического анализа данных. Набор команд этого меню отличается для различных модулей	
	Startup Panel (Стартовая панель)	Вызов стартовой панели
	Resume Analysis (Резюме анализа)	Вызов стартовой панели, если уже производился анализ данных в модуле
	Other Statistics (Другие статистики)	Вызов переключателя модулей Module Switcher
Graphs (Графики)	Содержит команды для построения графиков различных типов	
Options (Сервис)	Содержит команды, задающие настройки программы	
	General (Глобальные)	Глобальные настройки программы
	Display (Отображение)	Настройка отображения на экране различных элементов интерфейса
	Print (Печать)	Настройка параметров печати
	Stat Graphs (Статистические графики), User Graphs (Графики пользователя)	Настройка исходных параметров вновь создаваемых графиков
	Select (Выбор)	Задание условий отбора данных для анализа

Окончание таблицы 1.2

1	2	3
Options (Сервис)	Weight (Вес)	Задание весов переменных
	Macro (Макрос)	Автоматизация работы: программы на языке Statistica Command Language – SCL, программы на языке Statistica Basic; макросы (Keyboard Macro, Recorded Show Macro), кнопка автозадач (Auto Task Buttons)
Window (Окно)	Содержит команды управления количеством и видом окон	
	ScrollSheet Manager (Менеджер результатов)	Управление окнами таблиц результатов
	Graph Manager (Менеджер графиков)	Управление окнами графиков
	Cascade, Tile, Close all	Команды управления расположением окон: расположить каскадом, расположить рядом, закрыть все
	Список открытых окон	Переключение между окнами (активное окно отмечено галочкой)
Help (Помощь)	Содержит команды, предназначенные для поиска справочной информации. Справочная информация в программе Statistica представлена на английском языке	

2.6.3 Панель инструментов

Панель инструментов содержит кнопки, дублирующие наиболее часто используемые команды главного меню. Так же, как и в случае Главного меню, вид Панели инструментов зависит от типа активного в данный момент документа.

Панель инструментов может располагаться в верхней и в левой части экрана (по умолчанию), вверху и справа, в одну или две строки. Настройка вида панели инструментов осуществляется с помощью команды View – Tool Bar или с помощью контекстного меню, вызываемого щелчком правой кнопки мыши на панели инструментов.

2.6.4 Рабочая область

Рабочая область содержит окна различных типов, в которых отображены исходные данные и результаты анализа (SpreadSheet – таблица исходных данных, ScrollSheet – таблица результатов, Graphs – графики, Text/Output Window – окно отчета).

2.6.5 Стартовая панель

Стартовая панель обеспечивает быстрый доступ ко всем статистическим процедурам модуля, дублирует часть команд меню Analysis и позволяет уменьшить количество манипуляций при анализе в различных модулях.

Стартовая панель может быть вызвана из Главного меню командой Analysis – Startup Panel или Analysis – Resume Analysis, если уже производился анализ данных в модуле. Щелчок мышью за пределами стартовой панели, приводит к ее сворачиванию, и она оказывается в левом нижнем углу окна модуля.

2.6.6 Строка состояния

Строка состояния содержит информацию о состоянии программы Statistica или назначении выбираемых команд главного и контекстного меню; информацию о включении (ON) или выключении (OFF) окна отчета (Output), условий отбора данных (Select) и их весов (Weight). Двойной щелчок мыши на полях соответствующих состояний вызывает окно настройки отчета, условий отбора или весов.

2.6.7 Кнопка автозадач

Кнопка автозадач служит для настройки пользовательской панели инструментов. На панели могут быть размещены кнопки запуска макрокоманд или программ, написанных на языках, встроенных в программу.

Вызвать/убрать кнопку автозадач можно с помощью команды меню View – Auto Task Buttons или сочетания клавиш Ctrl+M.

2.7 Работа в таблице с исходными данными SpreadSheet

2.7.1 Организация данных в электронной таблице

Электронная таблица в программе Statistica состоит из строк и столбцов. В отличие от обычных электронных таблиц (например, Ms Excel), где строки и столбцы равноправны, в Statistica они имеют разные смысловые значения.

Столбцы называют Variables (Переменные), а строки – Cases (Случаи). Переменными обычно являются исследуемые величины, а случаями – повторения испытаний. Файл с исходными данными имеет расширение **.sta**.

2.7.2 Имена переменных и случаев

Каждой переменой и каждому случаю присваивается уникальное имя, которое отображается в заголовке столбца и строки соответственно. Имена переменных и случаев отображаются на сером фоне и не перемещаются при прокручивании таблицы, не помещающейся на экране.

Например, в таблице **Adstudy.sta** (см. рисунок 1.2) именами переменных

являются: **Gender**, **Advert**; **Measur1…Measur23**, а именами случаев – **I. Neil**; **G. Boss**; **T. Bush** и т.д. При создании новой таблицы переменным присваиваются имена **Var1**, **Var2** и т.д.

Изменить имя переменной можно двойным щелчком левой кнопки мыши на поле с именем, либо нажав кнопку Vars (Переменные) на панели инструментов и выбрав в выпадающем меню команду Current Specs (Текущие спецификации) для редактирования имени текущей переменной или команду All Specs (Все спецификации) для редактирования имен всех переменных таблицы.

Изменить имя случая можно двойным щелчком левой кнопки мыши на поле с именем, либо нажав кнопку Cases (Случаи) на панели инструментов и выбрав команду Names (Имена) в выпадающем меню.

2.7.3 Типы данных таблицы SpreadSheet

Ячейки данных могут иметь следующие типы:

- текстовый (Text);
- числовой (Number);
- дата (Date);
- время (Time);
- экспоненциальный (Scientific);
- денежный (Currency);
- процентный (Percentage).

2.7.4 «Двойная запись» текстовых значений

В программе Statistica для переменных, принимающих текстовые значения, реализован механизм «двойной записи». Это означает, что каждому текстовому значению переменной ставится в соответствие некоторое число.

Механизм «двойной записи» позволяет в удобной форме вводить текстовые записи и, кроме того, выполнять любой статистический анализ над текстовыми данными так, как над числовыми.

Переключить форму представления на экране текстовых данных (текстовый или числовой вид) можно с помощью кнопки  на панели инструментов или команды Text Values (Текстовые значения) меню View (Вид). Текстовые переменные содержат текст длиной не более восьми символов.

2.7.5 Метки

В Statistica имеется возможность каждой текстовой переменной сопоставить длинный текст (более восьми символов), т.е. присоединить так называемую метку (Label), поясняющую значение текстовой переменной.

Для управления значениями текстовых переменных (как текстовых значений, так и соответствующих им числовых значений) и метками, служит

Менеджер текстовых значений (Text Values Manager), который можно активировать нажатием кнопки на панели инструментов.

2.7.6 Спецификации переменных

Задать имена переменных, определить их тип и формат отображения, установить соответствие текстовых значений числовым, вставить метки, связи, формулы можно при помощи окна спецификаций переменной (рисунок 1.3).

Окно спецификаций можно открыть двойным щелчком левой кнопки мыши на поле с именем переменной, либо нажав кнопку Vars (Переменные) на панели инструментов и выбрав в выпадающем меню команду Current Specs (Текущие спецификации). Назначение опций окна спецификаций приведено в таблице 1.3.

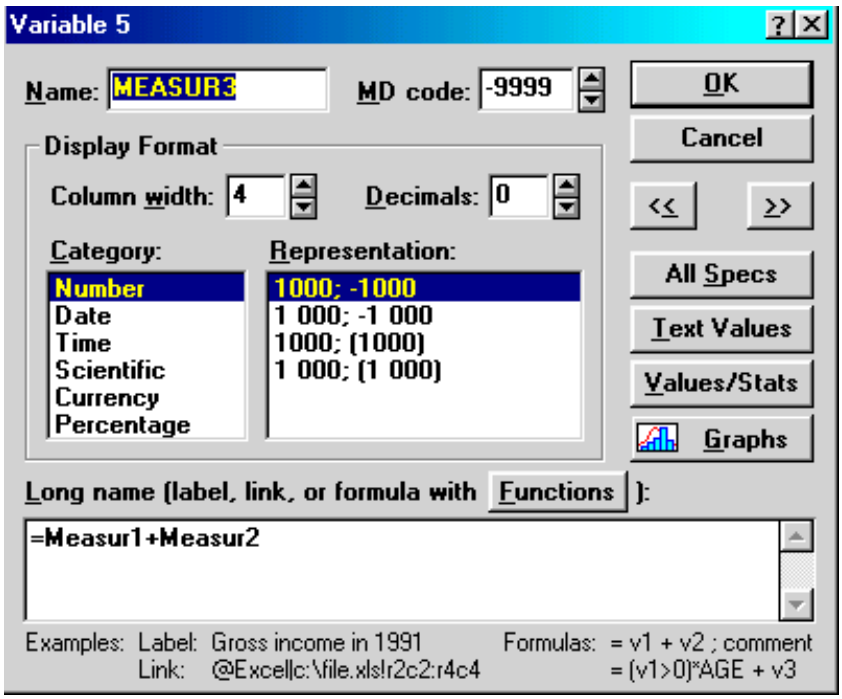


Рисунок 1.3 – Окно спецификаций переменной

Таблица 1.3 – Назначение опций окна спецификаций

Элемент окна	Назначение элемента
1	2
Name (Имя)	Задание имени переменной
MD code (ПД код)	Задание кодов пропущенных данных
Display Format (Формат отображения)	Задание формата отображения данных

Продолжение таблицы 1.3

1	2
Column width (Ширина колонки)	Задание ширины колонки (переменной)
Decimals (Десятичные знаки)	Задание количества десятичных знаков переменной
Category (Тип)	Задание типа данных переменной
Representation (Представление)	Задание формата представления данных
Long name (label, link or formula with Functions) (Длинное имя (метка, связь или формула с функцией))	Вставка меток, связей и функций
<<	Переход к предыдущей переменной
>>	Переход к следующей переменной
All Specs (Все спецификации)	Редактирование спецификаций всех переменных
Text Values (Текстовые значения)	Задание соответствий текстовых и числовых значений
Values/Stats (Значения/Статистики)	Отображение значений и основных статистик переменной
Graphs (Графики)	Построение быстрых статистических графиков для переменной

2.7.7 Использование формул для вычисления значений переменных

Ввод формулы производится в окне спецификаций переменной в поле Long name (label, link or formula with Functions) – Длинное имя (метка, связь или формула с функцией) (см. рисунок 1.3).

Правила ввода формул в окне спецификаций:

1. Ввод формулы должен начинаться со знака =.
2. Переменные, являющиеся операндами в формуле, могут быть записаны как в виде имен переменных, так и в виде номеров переменных.

Например, выражение **=Measur1+Measur2** и выражение **=V3+V4** дадут одинаковый результат при вычислении переменной **V5 (Measur3)** в файле **Adstudy.sta**.

3. При написании формул могут быть использованы стандартные функции. Для этого необходимо нажать кнопку Functions. Двойной щелчок на имени функции приведет к ее вставке в окно редактирования формулы.

Список функций представлен по категориям: General (Основные), Operators (Операторы), Math (Математические), Distributions (Распреде-

ления), All (Все функции).

Информацию о правилах написания выражений и назначении функций можно получить, нажав на кнопку Syntax (Синтаксис).

Для подтверждения ввода формулы нажимают клавишу Enter или кнопку OK. При этом появляется окно с запросом на пересчет значений переменной. Следует позволить пересчитать значения.

2.7.8 Редактирование исходных данных

Все команды, используемые для редактирования данных доступны из Главного меню Edit (Редактирование). Часть команд меню Edit дублируется в выпадающих меню при нажатии на кнопки Vars (Переменные) и Cases (Случаи) на панели инструментов. В этих меню содержатся команды редактирования переменных и случаев соответственно.

Получить доступ к вышеупомянутым командам можно так же из контекстного меню применительно к ячейке, имени случая или переменной, а так же к выделенному блоку.

2.7.9 Основные операции редактирования переменных

Основные команды, применяемые для редактирования переменных, и их назначение приведены в таблице 1.4.

Таблица 1.4 – Команды редактирования переменных

Название команды	Назначение команды	Параметры команды
1	2	3
Add Variables (Добавить переменные)	Добавление пустых столбцов в электронную таблицу	Number of Variables to add (число добавляемых переменных); Insert After Variable (вставить после переменной)
Move Variables (Переместить переменные)	Перемещение переменных с сохранением всех спецификаций: имени переменной, формата данных и параметров отображения на экране. Используется для изменения порядка переменных в таблице	From Variable (От переменной); To Variable (До переменной); Insert After (Вставить после)

Продолжение таблицы 1.4

1	2	3
Copy Variables (Копировать переменные)	Копирование переменных вместе со всеми спецификациями	From Variable (От переменной); To Variable (До переменной); Insert After (Вставить после)
Delete Variables (Удалить переменные)	Удаление столбцов из таблицы данных	From Variable (От переменной); To Variable (До переменной)
Current Specs (Текущие спецификации)	Задание параметров переменной	Name (Имя переменной); Column Width (Ширина столбца); Decimals (Количество десятичных знаков); Category (Тип данных); Representation (отображение данных на экране); Label (метки); Formula (формула); Links (динамические связи с другими файлами)
All Specs (Все спецификации)	Редактирование спецификаций всех переменных таблицы	Номер переменной; Name (имя переменной); MD Code (код пропущенных значений); Format (формат) – ширина столбца и число десятичных знаков, Long Name (label, formula or link) (длинные имена, (метки, формулы или связи))

Продолжение таблицы 1.4

1	2	3
Text Values (Текстовые значения)	Вызов Менеджера текстовых значений. В его окне приводится соответствие текстовых, числовых значений и соответствующих им меток	
Date Values (Значения дат)	Создание даты из двух ли трех переменных, в которых хранятся значения дня месяца или года. Преобразование даты в текстовый формат	
Recalculate Variables (Пересчитать переменные)	Пересчет данных, вычисляемых по формулам при изменении исходных данных.	Auto-recalculate when the data change (Автоматический пересчет значений)
Shift (Lag) Variables (Сдвиг переменных)	Сдвиг значений переменных на заданное число случаев.	Variable (имя переменной); Lag (шаг); Forward (Сдвиг вниз); Backward (Сдвиг вверх)
Rank Variables (Ранжировать переменные)	Ранжирование переменных. При выполнении на место исходной переменной записываются значения рангов	
Recode Variables (Перекодировать переменные)	Перекодировка переменных, т.е. замена исходного значения переменной новым значением в соответствии с заданным условием	

2.7.10 Основные операции редактирования случаев

Основные команды, применяемые для редактирования случаев, и их назначение приведены в таблице 1.5.

Таблица 1.5 – Команды редактирования случаев

Название команды	Назначение команды	Параметры команды
Add Cases (Добавить случаи)	Добавление случаев (пустых строк) в электронную таблицу	Number of Cases to add (число добавляемых случаев); Insert After Case (вставить после случая)
Move Cases (Переместить случаи)	Перемещение строк таблицы вместе с именами случаев	From Case (От случая); To Case (До случая); Insert After (Вставить после)
Copy Cases (Копировать случаи)	Копирование строк таблицы вместе с именами случаев	From Case (От случая); To Case (До случая); Insert After (Вставить после)
Delete Cases (Удалить случаи)	Удаление строк из таблицы данных	From Case (От случая); To Case (До случая)
Case Names (Имена случаев)	Задание имен случаев	

2.7.11 Операции с выделенным блоком

В электронных таблицах программы Statistica для работы с блоками данных реализованы все стандартные операции с буфером обмена. Назначение команд для работы с блоками данных меню Edit (Правка) приведено в таблице 1.6.

Таблица 1.6 – Команды редактирования блоков данных

Название команды	Сочетание клавиш, кнопка	Назначение команды
Cut (Вырезать)	Ctrl+X, Shift+Del	Перемещение выделенного блока в буфер обмена. При этом на месте выделенных значений остаются пустые ячейки. Команда используется для перемещения данных в другую часть документа или другое приложение
Copy (Копировать)	Ctrl+C Ctrl+Insert	Помещение в буфер обмена копии выделенного блока. При этом копируются не только данные, но и имена случаев и переменных
Copy Raw (Копировать исходные)		Помещение в буфер обмена только значений переменных выделенного блока
Paste (Вставить)	Ctrl+V, Shift+Insert	Вставка содержимого буфера обмена в место, указываемое положением курсора
Clear (Очистить)	Delete	Удаление содержимого ячеек
Paste Link (Вставить связь)		Вставка данных из другого файла или приложения. При этом данные вставляются вместе со связями, т.е. при изменении данных в исходном файле происходит изменение и вставленных данных
Transpose (Транспонировать)		Транспонирование (замена строк столбцами) данных блока или файла. При транспонировании блока количество строк и столбцов в нем должно совпадать
Fill/Standardize Block (Заполнить/ Стандартизовать блок)		Заполнение блока случайными переменными (Random Values) или заданным в первой ячейке блока значением – Copy Down (Копировать вниз) и Copy Right (Копировать вправо). Стандартизация столбцов (Columns) и строк (Rows) производится по формуле: Нов.знач=(Исход.знач–сред.) /стандарт.откл.
Block Stats (Блоковые статистики)		Вычисление блоковых статистик столбцов (строк): среднего (Means), медианы (Medians), стандартного отклонения (SD's), количества значений (Valid N's), суммы (Sums), минимума (Min's), максимума (Max's), 25% и 75% квантилей (25%'s и 75%'s).

2.7.12 Дополнительные команды редактирования данных

В программе Statistica имеются дополнительные команды, применяемые для повышения удобства редактирования данных, а так же выполняющие специфические функции. Список этих команд и их назначение приведены в таблице 1.7.

Таблица 1.7 – Дополнительные команды редактирования данных

Название команды	Сочетание клавиш, кнопка	Назначение команды
Undo (Отменить)	Ctrl+Z, Alt+Bksp 	Отмена последнего действия. Можно несколько раз последовательно воспользоваться этой командой для отмены нескольких действий
Find (Найти)	Ctrl+F, Ctrl+Alt+F	Поиск заданного значения или имени случая. Задается направление поиска (Direction) – вверх (Up) или вниз (Down), предмет поиска – значения (Values) или имена случаев (Case Names)
Replace (Заменить)		Поиск и замена значений или имен случаев. Диалоговое окно команды аналогично окну предыдущей команды. В нем добавлены строка для ввода новых значений Replace With (Заменить на) и кнопка Replace All (Заменить все)
Links... (Связи)		Редактирование связей данных с другими файлами или приложениями
Screen Cather (Захват экрана)		Копирование в буфер обмена прямоугольной области экрана. Выбор команды приводит к изменению формы указателя мыши. Нажатием на левую кнопку мыши и перемещением указателя выделяется необходимая область экрана. При отпускании кнопки выделенная область помещается в буфер обмена

2.7.13 Использование технологии Drag-and-Drop (Перетащить и отпустить)

Операции Drag-and-Drop доступны при работе с выделенным блоком.

Выделение блока. Выделить блок можно с помощью мыши. Для этого необходимо установить указатель мыши в угловую ячейку блока и, нажав левую кнопку мыши, переместить указатель по диагонали в противоположный угол блока. При отпускании левой кнопки блок остается выделенным.

Для выделения блока можно воспользоваться так же клавишей Shift. При этом необходимо активизировать ячейку в одном из углов выделяемого блока. После нажатия клавиши Shift следует переместить указатель в противоположный угол блока либо при помощи мыши, либо при помощи кнопок клавиатуры ←,→,↑,↓.

Перемещение блока. Для перемещения выделенного блока необходимо установить указатель мыши на его границу. При этом указатель изменит форму с «перекрестия» на «стрелку». Если нажать левую кнопку мыши и переместить указатель, то на экране появляется пустая перемещаемая рамка. При отпускании кнопки перемещаемые данные заполнят пустую рамку. При этом ячейки с исходным блоком очищаются.

Копирование блока. Если выполнить действия, указанные в предыдущем методе при нажатой клавише Ctrl, то произойдет копирование блока. Т.е. данные из выделенного блока будут помещены в указанное место, но исходный блок останется в том же виде (ячейки не очистятся).

Примечание. Функции Drag-and-Drop могут быть настроены (в меню Options – General) таким образом, что при нажатой клавише Ctrl происходит перемещение, а при отпущенной – копирование. Отличить операцию копирования от перемещения всегда можно по наличию «плюса» рядом со «стрелкой» указателя мыши.

Вставка блока между ячейками. При необходимости блок может быть вставлен не в существующие ячейки, а в добавленные в соответствии с размером блока. Для этого необходимо при перемещении или копировании удерживать нажатой клавишу Shift.

Очистка блока. Очистить выделенный блок или его часть можно, поместив указатель мыши на квадрат, расположенный в нижнем правом углу блока (при этом курсор принимает вид «жирного плюса»), и перемещая его внутрь выделенного блока.

Автозаполнение. Если необходимо ввести некоторую последовательность чисел (например, 1,2,3…50), то нет необходимости набирать ее с помощью клавиатуры. Следует ввести два первых числа, затем выделить их. При помещении в нижний правый угол выделенного блока указатель мыши примет вид «жирного плюса» – маркер автозаполнения. Перемещение указателя с нажатой левой кнопкой мыши вызовет автоматическое заполнение выделенных ячеек.

Если автозаполнение применяется не для двух выделенных ячеек, а для одной, то выделенный блок заполнится значениями этой ячейки.

Просмотр частей таблицы. При работе с большими таблицами бывает необходимо видеть различные ее части одновременно. Statistica позволяет разделить окно просмотра на четыре части. В верхней части вертикальной и в левой части горизонтальной полос прокрутки имеется темный прямоугольник. При его перетаскивании на полосу прокрутки последняя разделяется на две независимые части. При этом документ разделяется на независимо прокручиваемые части.

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

3.1 Знакомство с интерфейсом программы Statistica

Упражнение 1

В этом упражнении вы познакомитесь с командами Главного меню.

1 Запустите модуль Basic Statistics and Tables любым известным вам способом.

2 Откройте файл **Adstudy.sta**. Этот файл находится в папке **Examples** (**C:\Program Files\Stat\Examples**).

3 Активизируйте Module Switcher.

4 При помощи кнопки End & Switch To перейдите в модуль Multiple Regression. При этом модуль Basic Statistics and Tables закроется.

5 Вернитесь в модуль Basic Statistics and Tables, не закрывая модуль Mutiple Regression. Для этого активизируйте Module Switcher двойным щелчком левой кнопки мыши на свободном месте рабочей области и нажмите кнопку Switch To.

6 Используя меню File, сохраните файл **Adstudy.sta** в личной папке с именем **MyAdstudy.sta**. Для этого выполните команду File – Save As.... Обратите внимание на изменения в строке заголовка таблицы данных.

7 При помощи команды Open Other – Graph File откройте файл графика **Surface.stg** из папки **Examples**. Разверните его окно в полный экран. Обратите внимание на изменения в главном меню и на панели инструментов.

8 Сохраните график в личной папке с именем **MySurface.stg**. Обратите внимание на изменения в строке заголовка.

9 При помощи команды View – Text/Output Window создайте файл отчета. Запишите в нем свою фамилию, имя, отчество и текущую дату. Сохраните файл отчета в личной папке с именем **Otchet**.

10 Не сворачивая окно графика, перейдите последовательно к окну таблицы данных, отчету, и затем к окну стартовой панели. Используйте для этого список открытых окон меню Window.

11 Закройте все окна.

Упражнение 2

В этом упражнении вы научитесь работать в таблице с исходными данными.

1 Откройте файл **MyAdstudy.sta** из личной папки.

2 Запишите в отчете имена трех первых переменных и трех первых случаев. Укажите, какие переменные являются текстовыми, а какие числовыми.

3 При помощи кнопки  на панели инструментов включите числовую форму представления данных переменной **Advert**. Запишите в отчет соответствие числовых и текстовых значений. Включите текстовую форму пред-

ставления при помощи меню View. Для этого выполните команду View – Text values – Text values.

4 Скопируйте переменные от **MEASUR11** до **MEASUR13**, разместив их после всех переменных. **Обратите внимание, что в названии столбцов вверху стоит число, которое обозначает номер столбца, а не номер переменной.** Перед выполнением команды Copy Variables, переменные для копирования лучше выделить, тогда параметры команды From Variable и To Variable будут автоматически заполнены. Для копирования переменных щелкните левой кнопкой мыши по кнопке Vars на панели инструментов и в раскрывшемся списке выберите команду Copy.... На экране появится диалоговое окно Copy Variables. Нажмите OK.

5 В окне спецификаций переменной измените имена скопированных переменных на **MEASUR24** – **MEASUR26**.

6 Удалите последнюю переменную **MEASUR26** при помощи команды Delete Variables. Для удаления переменной выделите ее в таблице и щелкните левой кнопкой мыши по кнопке Vars на панели инструментов. В раскрывшемся списке выберите команду Delete.... На экране появится диалоговое окно Delete Variables. Нажмите OK.

7 С помощью команды Add Cases добавьте три случая в конце таблицы. Для этого в таблице выделите последнюю строку и щелкните левой кнопкой мыши по кнопке Cases на панели инструментов. В раскрывшемся списке выберите команду Add.... На экране появится диалоговое окно Add Cases. Нажмите OK.

8 Задайте имена добавленных случаев **D. Taylor**, **S. Brown**, **F. Malken**.

9 С помощью буфера обмена скопируйте значения первых трех случаев в ячейки добавленных случаев.

10 Переместите случай **F. Malken** таким образом, чтобы он располагался перед случаем **S. Brown**. Для этого выделите случай **F. Malken** и щелкните левой кнопкой мыши по кнопке Cases на панели инструментов. В раскрывшемся списке выберите команду Move.... На экране появится диалоговое окно Move Cases и в параметре Insert After набрать **51** – номер случая **S. Brown**.

11 Удалите случай **S. Brown**.

12 При помощи методов Drag-and-Drop заполните переменную **MEASUR24** значениями **1, 2, 3…52**. Для этого заполните первую и вторую ячейку значениями **1**, **2** соответственно. Выделите обе ячейки и протяните вниз за маркер автозаполнения.

13 При помощи методов Drag-and-Drop скопируйте значения переменной **MEASUR24** в переменную **MEASUR25**.

14 После переменной **Advert** добавьте новую переменную командой Add Variables. Для добавления переменной выделите **Advert** в таблице и щелкните левой кнопкой мыши по кнопке Vars на панели инструментов.

В раскрывшемся списке выберите команду Add.... На экране появится диалоговое окно Add Variables. Нажмите OK.

15 Откройте окно спецификаций новой переменной. Задайте имя переменной **Percent**. Установите процентный формат отображения с точностью до двух десятичных знаков.

16 Введите формулу для расчета переменной **Percent**:

=MEASUR1/(MEASUR1+MEASUR2+MEASUR3+MEASUR4+MEASUR5+MEASUR6+MEASUR7+MEASUR8+MEASUR9+MEASUR10).

Она показывает, какой процент от суммы значений переменных **Measur1** – **Measur10** составляет значение переменной **Measur1**.

17 Сохраните файл **MyAdstudy.sta**. Завершите работу с программой Statistica.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Создайте новый файл данных. Сохраните его в личной папке с именем **Visota.sta**.
2. Создайте структуру таблицы и введите данные согласно таблице 1.8. Значения переменной **Повтор** вводите при помощи автозаполнения.

Таблица 1.8 – Высота растений ячменя

Повтор	Опыт1	Опыт2	Контроль	Опыт3	Опыт4
Повтор1	80	83	78	83	80
Повтор1	88	77	85	81	86
Повтор1	89	82	86	83	87
Повтор1	92	87	75	77	85
Повтор1	84	77	85	77	79
Повтор1	89	84	80	75	91
Повтор1	76	77	86	69	85
Повтор1	93	81	81	86	84
Повтор1	81	91	79	77	82
Повтор1	87	78	78	85	83
Повтор2	91	80	82	81	90
Повтор2	89	80	81	75	81
Повтор2	87	84	87	80	86
Повтор2	87	72	85	80	83
Повтор2	92	80	85	81	94
Повтор2	89	81	70	80	82
Повтор2	92	75	81	83	87
Повтор2	80	80	77	75	86
Повтор2	92	81	84	81	91
Повтор2	92	81	81	80	86

Продолжение таблицы 1.8 на следующей странице.

Продолжение таблицы 1.8

Повтор3	79	80	81	90	82
Повтор3	90	89	70	91	90
Повтор3	82	82	74	85	70
Повтор3	79	93	84	87	85
Повтор3	81	89	81	85	82
Повтор3	85	94	75	86	86
Повтор3	84	76	86	92	78
Повтор3	78	87	80	84	82
Повтор3	91	80	83	90	81
Повтор3	77	90	75	80	81

5 СОДЕРЖАНИЕ ОТЧЕТА

Отчет оформляется в соответствии с требованиями, принятыми в АЧИИ. Отчет должен содержать:

- ответы на контрольные вопросы.

6 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие модули программы Statistica Вы знаете?
2. С какими типами документов работает программа Statistica?
3. Какие существуют способы запуска программы Statistica?
4. Что такое Module Switcher? Как его активизировать?
5. Назовите основные элементы окна программы Statistica?
6. Что такое стартовая панель модуля? Для чего она нужна?
7. Что такое кнопка автозадач? Ее назначение.
8. Перечислите назначение опций главного меню File, Edit, View.
9. Перечислите назначение опций главного меню Analysis, Options, Window.
10. Перечислите назначение модулей Basic Statistics and Tables, Multiple Regression, ANOVA/ MANOVA.
11. Как отобразить на экране стартовую панель модуля?
12. Перечислите назначение способы переключения между окнами рабочей области Statistica.
13. Какие операции для редактирования выделенного блока имеются в программе Statistica?
14. Какой командой добавляют переменные, случаи? Какие параметры имеет эта команда?
15. Как копировать, переместить, удалить переменные или случаи? Назовите параметры соответствующих команд.
16. Как копировать данные из одного столбца (переменной) в другой?
17. Как выполнить автозаполнение?
18. Как вставить формулу для расчета значений переменной?

Практическое занятие № 2

Тема: **ПРЕДСТАВЛЕНИЕ ЭКПЕРИМЕНТАЛЬНЫХ ДАННЫХ ПРИ ПОМОЩИ ГРАФИКОВ. НАСТРОЙКА ГРАФИКОВ**

1 ЦЕЛЬ РАБОТЫ

Изучить:

- методы визуализации данных;
- назначение различных типов графиков Графической галереи Statistica;
- основные элементы статистических графиков.

Научиться:

- настраивать параметры отображения элементов статистических графиков при помощи команд главного и контекстных меню.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 График как метод представления данных

Графики являются важным средством выражения и анализа статистических данных. С точки зрения их восприятия графики являются самой эффективной формой представления данных.

Графическое изображение позволяет осуществить контроль достоверности исследуемых показателей, так как представленные на графике они делают более очевидными имеющиеся отклонения, связанные либо с наличием ошибок наблюдения, либо с сущностью изучаемого явления.

С помощью графиков достигается:

- наглядность изображенного анализируемого материала;
- сопоставление размеров совокупности;
- выявление наиболее характерных соотношений и взаимосвязей явлений;
- определение основных тенденций развития;
- определение пределов изменений показателей;
- определение сравнительной скорости изменения различных показателей, их варьирования.

График – это чертеж, на котором совокупности анализируемых данных, характеризуемые определенными показателями, описываются с помощью условных геометрических образов или знаков.

Представление данных в виде графика позволяет:

- лучше осмыслить результаты экспериментальных наблюдений;
- правильно истолковать результаты анализа;

– значительно облегчить понимание накопленного материала;

– сделать материал наглядным, выразительным, лаконичным, доступным.

2.2 Элементы графиков

График состоит из следующих элементов:

– графический образ;

– поле графика;

– пространственные ориентиры (направления осей, координатные сетки);

– масштабные ориентиры (шкала осей);

– пояснения.

Графический образ (основа графика) – это геометрические знаки, совокупность точек, линии, фигуры, с помощью которых и представляют данные. Эти знаки образуют собственно языковую ткань графика, его основу. Правильный выбор графического образа должен соответствовать цели графика.

Геометрические знаки, рисунки или образы, применяемые в статистических графиках, многообразны. К ним относят точки, отрезки прямых линий, фигуры различной формы, штриховки или окраски. Эти знаки применяются для сравнения величин, изображающих абсолютные и относительные размеры сравниваемых совокупностей.

Сравнение на графике производится по некоторым измерениям: площади или длине одной из сторон фигуры; местонахождению точек; их густоте; густоте штриховки;·интенсивности или цвету окраски.

Графические компьютерные программы имеют большие наборы геометрических знаков (одинарных и двойных, сплошных и прерывистых линий различной толщины и цвета, иных обозначений и символов), позволяющих изображать графические фигуры так, чтобы они легко отличались одна от другой.

Поле графика – это часть плоскости, пространство, в котором размещаются геометрические знаки. Поле графика характеризуется его форматом, т.е. размером и пропорциями (соотношениями сторон). В зависимости от целей и задач графического изображения поле графика может быть чистым или заштрихованным.

Пространственные ориентиры – определяют размещение геометрических знаков на поле. Задаются в виде системы координатных сеток или контурных линий, которые делят это поле на части.

Обычно применяется система прямоугольных (декартовых) координат в двухмерном или трехмерном изображении. Такие графики используются для сравнения нескольких однотипных совокупностей.

Нередко встречаются и круговые графики, построенные по принципу полярных координат. В полярной системе координат одной координатой яв-

ляется угол поворота луча относительно заданного положения, а второй координатой считается расстояние от центра сетки, называемое радиусом. Данные графики могут быть использованы для отображения изменений некоторой величины за определенный промежуток времени, например за год.

Масштабные ориентиры определяются масштабом и системой масштабных шкал, т.е. дают геометрическим знакам количественную определенность.

Масштабом графика называется условная мера перевода числовой величины в графическую. Чем меньше масштаб, тем гуще располагаются на шкале точки, имеющие одно и то же значение.

Шкала состоит из линии (носителя шкалы) и ряда помеченных на ней точек, расположенных в определенном порядке.

Шкалы могут быть *равномерными* и *неравномерными*. Одним из видов неравномерных шкал является логарифмическая шкала. На ней отрезки пропорциональны не изображаемым величинам, а их логарифмам.

Шкала имеет достаточно большое значение в графике. В ней различают:

– линию (или носитель шкалы);

– некоторое число точек, которые расположены на носителе шкалы;

– цифровое обозначение чисел, соответствующих помеченным точкам.

Носитель шкалы может быть представлен прямой или кривой линией. Поэтому шкалы называются *прямолинейными* и *криволинейными* (например, круговые и дуговые).

Пояснения – это словесные описания содержимого графика. Пояснения включают:

– название графика, которое в краткой форме передает его содержание;

– заголовок графика, в котором указывается наименование изображенного показателя и его единицы измерения;

– подписи вдоль масштабных шкал (название отложенных на них единиц измерения);

– пояснения к отдельным частям графика;

– легенду (геометрические образы и соответствующие им показатели).

Без пояснений график нельзя прочитать и понять. Пояснительные тексты могут располагаться в пределах графического образа или рядом с ними, а также выноситься за его пределы.

Легенда для помещенных на графике геометрических образов, различных по их конфигурации, штриховке или цвету, позволяют мысленно перейти от геометрических образов к явлениям и процессам, изображенным на графике.

2.3 Графические средства программы Statistica

В программе Statistica имеются следующие категории графиков:

– статистические графики (Stats Graphs);

– быстрые статистические графики (Quick Stats Graphs);

– пользовательские графики (Custom Graphs);

– блоковые статистические графики (Block Stats Graphs).

Все графики программы Statistica представлены в графической галерее (рисунок 2.1), которую можно активировать кнопкой на панели инструментов, и в главном меню Graphs (Графики). Каждая категория включает несколько групп графиков, имеющих десятки разновидностей.

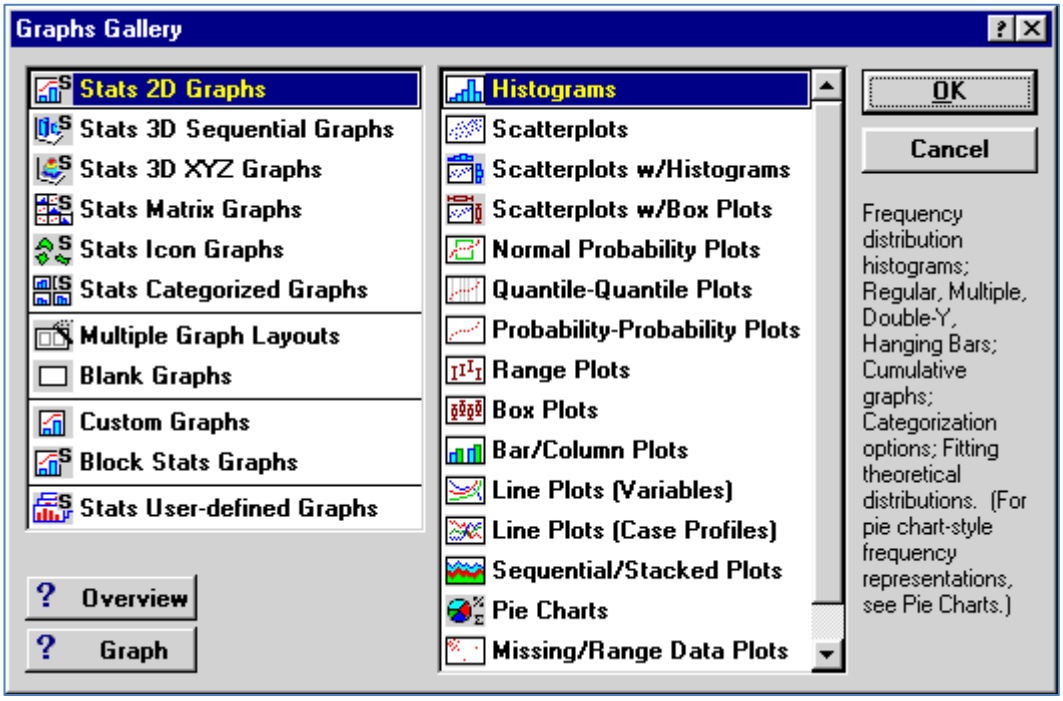


Рисунок 2.1 – Графическая галерея Statistica

Статистические графики (Stats Graphs) предназначены для визуализации всех значений переменных (столбцов) электронной таблицы. Быстрые статистические графики (Quick Stats Graphs) включают в себя наиболее распространенные статистические графики, имеющие типовые настройки. Быстрые статистические графики можно построить так же с помощью контекстного меню к выделенному блоку данных или меню, вызываемого кнопкой на панели инструментов.

Пользовательские графики (Custom Graphs) и блоковые статистические графики (Block Stats Graphs) предназначены для визуализации значений из предварительно выделенного блока данных. Часть кнопок для построения пользовательских графиков вынесена на панель инструментов . Эти графики можно построить с помощью контекстного меню к выделенному блоку данных.

Так как быстрые статистические графики (Quick Stats Graphs), пользовательские графики (Custom Graphs) и блоковые сстатистические графики (Block Stats Graphs) являются модификациями статистических графиков (Stats Graphs), то рассмотрим подробно лишь последние.

Среди статистических графиков (Stats Graphs) выделяют:

– двумерные графики (Stats 2D Graphs);

– трехмерные последовательные графики (Stats 3D Sequential Graphs);
– трехмерные графики в пространстве (Stats 3D XYZ Graphs);
– матричные графики (Stats Matrix Plots);
– пиктографики (Stats Icon Plots);
– категоризованные графики (Stats Categorized Plots).

2.3.1 Статистические двумерные графики

Назначение основных групп статистических двумерных графиков (Stats 2D Graphs) представлено в таблице 2.1.

Таблица 2.1 – Назначение статистических двумерных графиков

Пиктограмма и название	Назначение
1	2
Histograms (Гистограммы)	Отображение распределения переменных по интервалам группировки
Scatterplots (Диаграммы рассеяния или Точечные графики)	Визуальный анализ зависимости между двумя переменными X и Y
Scatterplots w/Histograms (Диаграммы рассеяния с гистограммами)	Составной график с зависимостью между двумя переменными и распределениями частот для каждой из них
Scatterplots w/Box Plots (Диаграммы рассеяния с диаграммами размаха)	Составной график с зависимостью между двумя переменными и распределениями значений для каждой из двух выборок (включая выбросы, экстремальные значения и т.д.)
Normal Probability Plots (Графики на нормальной вероятностной бумаге)	Визуальное исследование нормальности распределения
Quantile-Quantile Plots (Графики квантиль-квантиль)	Визуальная подгонка вероятностных распределений
Probability-Probability Plots (Графики вероятность-вероятность)	Сравнение вероятностного распределения переменной с теоретическим распределением
Range Plots (Диаграммы диапазона значений)	Отображение диапазонов изменения наблюдаемой величины, относящиеся к определенной точке
Box Plots (Диаграммы размаха)	Отображение диапазонов значений выбранных переменных и их основных статистик
Bar/Column Plots (Столбчатые диаграммы)	Отображение последовательности значений переменной в виде столбцов
Line Plots (Variables) (Линейные графики (перемен-	Изображение значений переменной для каждого случая в виде точек, соединен-

ных))	ных прямыми линиями

Продолжение таблицы 10.1

1	2
Line Plots (Cases Profiles) (Линейные графики (профили наблюдений))	Используются для представления последовательности значений различных признаков для каждого наблюдения
Sequential/Stacked Plots (Последовательные/ наложенные графики)	Используются для визуализации последовательностей значений нескольких переменных путем суммирования значений переменных по каждому случаю
Pie Charts (Круговые диаграммы)	Представление пропорций (долей) от общего числа
Missing/Range Data Points (Графики пропущенных значений и интервалов)	Исследование расположения пропущенных данных и точек «вне диапазона» текущего множества
Custom Function Plots (Графики функций, заданных пользователем)	Построение графиков функций, заданных формулой, которая вводится пользователем (пользовательский график)

Двумерные графики наиболее часто используются для визуализации экспериментальных данных и результатов их анализа. Подробное описание двумерных графиков программы Statistica находится в Приложении А.

2.3.2 Трехмерные последовательные графики

Трехмерные последовательные графики предназначены для визуализации последовательностей значений нескольких переменных в трехмерной перспективе. Данные графики сходны с составными линейными двухмерными графиками. Данные для каждой из переменных представляются в виде объемных фигур и изображаются в трехмерной перспективе. Благодаря этому изображения данных для различных переменных не накладываются друг на друга. При этом имеется возможность выбрать направление обзора путем вращения трехмерной системы координат.

Назначение основных групп статистических трехмерных последовательных графиков (Stats 3D Sequential Graphs) представлено в таблице 2.2.

Таблица 2.2 – Назначение трехмерных последовательных графиков

Пиктограмма и название	Назначение
1	2
Raw Data Plots (Графики предварительных данных)	Набор объемных диаграмм
Box Plots (Столбчатая диаграмма)	Отображение последовательных значений переменных для каждого случая в виде трехмерных столбцов (Аналогичны двумерным

	столбчатым диаграммам)

Продолжение таблицы 2.2

1	2
Bivariate Histograms (гистограммы двух переменных)	Визуализация распределения двух переменных (методы категоризации данных аналогичны таковым для двумерных гистограмм)
Range Plots (Диаграммы диапазона значений)	Отображение диапазонов изменения или ошибок переменной, относящиеся к определенной точке (Аналогичны двумерным диаграммам диапазонов). Различают Точечные диапазоны (Point Ranges), Граничные диапазоны (Border-style Ranges), Диапазоны ошибок (Error Bar-style Ranges), Диапазоны двойных лент (Double Ribbon Ranges), Летящие ящики (Flying Boxes) и Летящие блоки (Flying Blocks)
Scatterplots (Диаграммы рассеяния или Точечная диаграмма)	Визуальный анализ зависимости между тремя переменными X, Y и Z
Surface Plots (График поверхности)	К точкам исходных данных подгоняется сглаженная сплайнами поверхность
Contour Plots (Карта линий уровня)	Двумерная проекция сглаженной сплайнами поверхности
Trace Plots (Трассировочные графики)	Представление точек переменных в трехмерных координатах с соединением их отрезками в порядке следования случаев
Ternary Plots (Тернарные графики)	Это треугольные графики, на которых изображают набор данных с тремя переменными, сумма которых равна постоянной величине. Обычно данные представлены в процентах или в эквивалентной десятичной форме
Categorized XYZ Plots (Категоризованные трёхмерные графики)	Построение поверхностей для данных, разделённых на группы для сравнения структуры получившихся подмножеств
Custom Function Plots (Графики функций, заданных пользователем)	Построение графиков функций, заданных формулой, которая вводится пользователем (пользовательский график)

2.3.3 Трехмерные графики в пространстве

Трехмерные графики в пространстве (Stats 3D XYZ Graphs) предназначены для визуального представления зависимостей нескольких переменных в трехмерном пространстве. Назначение статистических трехмерных графиков в пространстве, наиболее часто используемых при анализе данных, представлено в таблице 2.3.

Таблица 2.3 – Назначение трехмерных графиков в пространстве

Пиктограмма и название	Назначение
Scatterplots (Диаграммы рассеяния)	Визуальный анализ зависимости между тремя переменными X, Y и Z
Surface Plots (Графики поверхностей)	К точкам исходных данных диаграммы рассеяния подгоняется сглаженная поверхность. Сглаживание может быть Линейным (Linear Smooth), Квадратичным (Quadr. Smooth), Методом наименьших квадратов (Least Smooth), Экспоненциально взвешенным с отрицательным показателем (Neg. Exp/Wght), Сплайнами (Spline Smooth), Пользовательской функцией (Custom Function)
Contour Plots (Карты линий уровня)	Двумерная проекция сглаженной трехмерной поверхности
Trace Plots (Трассировочные графики)	Представление точек переменных в трехмерных координатах с соединением их отрезками в порядке следования случаев
Custom Function Plots (Графики функций, заданных пользователем)	Построение трехмерных графиков функций, заданных формулой

2.4 Методы настройки элементов графиков

При отображении на экране или выводе на печать график программы Statistica выглядит как единое целое, однако, каждый график состоит из множества элементов:

- области построения, осей (Axes);
- сетки (Grid), линий графиков (Plot Lines);
- заголовков (Titles), легенды (Legend);
- меток данных (Data Labels) и т.д.

Иногда в одном графическом окне строится несколько логически взаимосвязанных графиков (Plots), которые так же состоят из отдельных элементов.

Настраивать элементы графика можно с помощью меню Options (Сервис) графического окна, кнопок панели инструментов или кон-

текстных меню, вызываемых к области построения графика либо к элементам графика.

Для настройки отдельных элементов графиков можно использовать двойной щелчок левой кнопки мыши на элементе графика. При этом открывается окно настройки наиболее часто используемых параметров выбранного элемента.

В меню Options (Сервис) наиболее целесообразно настраивать глобальные параметры, которые можно использовать по умолчанию.

Для настройки отдельных элементов графиков наиболее удобно использовать контекстные меню, вызываемые к требуемым элементам.

Возможен и такой вариант, когда с помощью контекстных меню настраивается внешний вид графика, а затем сохраняются все настройки для использования по умолчанию при последующих сеансах работы.

2.4.1 Настройка графиков с помощью главного меню Options

Меню Options (Сервис) предназначено для настройки графиков, сохранения файлов настроек и загрузки настроек из файлов настроек. Назначение опций главного меню Options (Сервис) приведено в таблице 2.4.

Таблица 2.4 – Назначение опций главного меню Options (Сервис)

Опция меню	Назначение опции
Global Defaults (Общие установки по умолчанию)	Открытие диалогового окна наиболее важных настроек отображения графиков
Save Defaults (Сохранить установки по умолчанию…)	Сохранение новых установок в качестве используемых по умолчанию. Все установки сохраняются в файле настройки Statist.ini
Save Defaults As (Сохранить установки по умолчанию как…)	Сохранение новых установок в качестве используемых по умолчанию. Все установки сохраняются в файле настройки, создаваемом пользователем
Retrieve Defaults (Применить установки по умолчанию)	Применение к редактируемому графику настроек по умолчанию из файла Statist.ini
Retrieve Defaults From (Применить установки по умолчанию из)	Применение к редактируемому графику настроек по умолчанию из указанного пользователем файла настройки
General Layout Dialog (Общий диалог отображения)	Определение параметров отображения графиков

2.4.2 Настройка установок по умолчанию с помощью диалога Global Defaults

При выборе опции Global Defaults (Общие установки по умолчанию) из главного меню Options (Сервис) открывается одноименное диалоговое окно (рисунок 2.2).

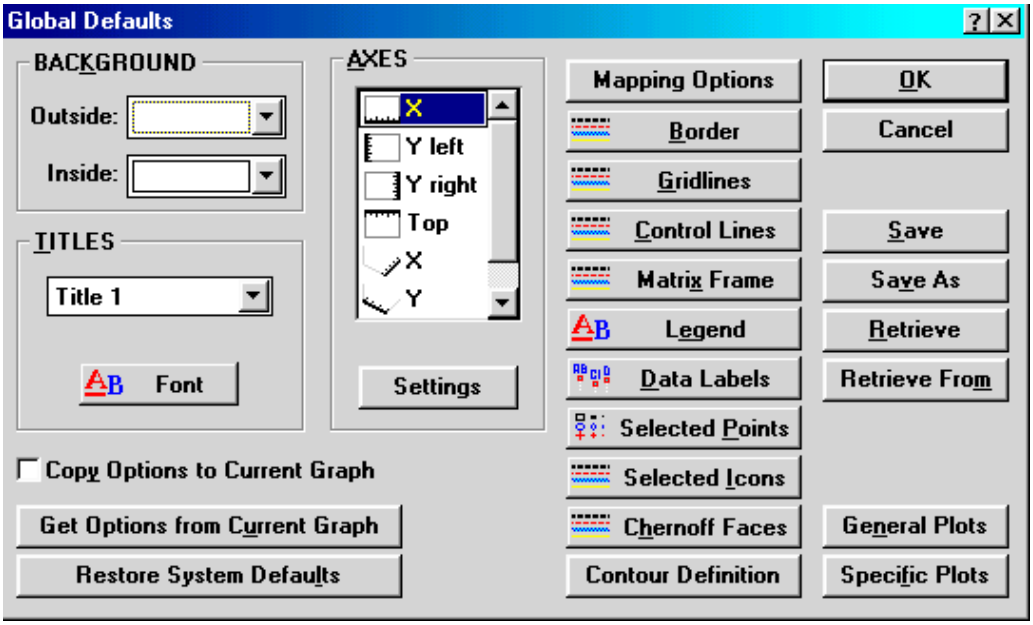


Рисунок 2.2 – Диалоговое окно Global Defaults (Общие установки по умолчанию)

В этом диалоговом окне можно задать настройки элементов графика с помощью опций, приведенных в таблице 2.5.

Таблица 2.5 – Назначение опций диалогового окна Global Defaults

Элемент окна	Назначение элемента окна
1	2
Background (Фон)	Группа опций для настройки фона
Inside (Внутри)	Настройка фона области построения
Outside (Снаружи)	Настройка фона внешней области окна графика
Titles (Заголовки)	Группа опций для настройки шрифтов заголовков
Title1 (Заголовок1)	Выпадающее меню для выбора уровня заголовка
Font (шрифт)	Кнопка для задания параметров шрифта выбранного заголовка

Продолжение таблицы 2.5

1	2
Axes (Оси)	Группа опций для настройки осей
X, Y, Z	Прокручиваемый список для выбора настраиваемой оси
Setting (Установка)	Настройка параметров оси
Mapping Options (Расположение)	Настройки позиции и масштаба графика в графическом окне и при выводе на принтер
Border (Рамка)	Настройка вида линий рамки
Gridlines (Линии сетки)	Настройка вида линий сетки
Control Lines (Контрольные линии)	Настройка вида контрольных линий
Matrix Frame (Кадр матричного графика)	Настройка вида линий матричного графика
Legend (Легенда)	Настройка параметров шрифта легенды
Data Labels (Метки данных)	Настройка параметров шрифта меток данных
Selected Points (Выбранные точки)	Настройка вида точек на графике
Selected Icons (Выбранные пиктограммы)	Настройка вида линий для различных категорий данных
Chernoff Faces (Лица Чернова)	Настройка вида линий для графиков типа «Лица Чернова»
Contour Definition (Определение контура)	Настройка вида линий контуров и цвета зон при построении трехмерных графиков и их двумерных сечений
Copy Options to Current Graph (Применить установки к текущему графику)	Применение изменений настроек по умолчанию к текущему графику
Get Options from Current Graph (Установить опции по умолчанию из текущего графика)	Использование настроек текущего графика в качестве настроек по умолчанию

Окончание таблицы 2.5

1	2
Restore System Defaults (Восстановить установки по умолчанию)	Сброс всех изменений в настройках графиков и установка первоначальных системных настроек
Save (Сохранить)	Сохранение новых установок в качестве используемых по умолчанию в файле настройки Statist.ini
Save As (Сохранить как)	Сохранение новых установок в качестве используемых по умолчанию в файле настройки, создаваемом пользователем
Retrieve (Применить)	Применение к редактируемому графику настроек по умолчанию из файла Statist.ini
Retrieve From (Применить из)	Применение к редактируемому графику настроек по умолчанию из указанного пользователем файла настройки
General Plots (Основные графики)	Настройка линий и заполнения основных графиков
Specific Plots (Специальные графики)	Настройка линий и заполнения специальных графиков

2.4.3 Диалоговое окно General Layout (Общая разметка)

Большая часть настроек графика доступна в диалоговом окне General Layout (Общая разметка) (см. рисунок 2.3). Это окно можно вызвать при помощи меню Layouts – General Layout (Общая разметка) или двойным щелчком мыши на свободной части графического окна.

Вид диалогового окна General Layout (Общая разметка) зависит от типа настраиваемого графика. Например, это окно будет значительно отличаться для двумерных (2D) и трехмерных (3D) графиков. Однако в данном окне есть настройки, общие для всех типов графиков. Рассмотрим на примере 3D графика назначение опций диалогового окна General Layout (таблица 2.6).

Таблица 2.6 – Назначение опций диалогового окна General Layout

Элемент окна	Назначение элемента окна
1	2
Окно просмотра	Отображение всех изменений при настройке
Graph Type (Тип графика)	Прокручиваемый список для выбора типа графика. В списке присутствуют все графики для указанных данных
Titles (Заголовки)	Группа опций для редактирования и настройки заголовков

Продолжение таблицы 2.6

1	2
Выпадающее меню	Выбор настраиваемого заголовка. Может быть выбрано пять уровней заголовков графиков и по два уровня заголовка для каждой из осей
Строка редактора	Редактирование выбранного заголовка
Font (Шрифт)	Кнопка для настройки шрифта заголовков
Scales (Оси)	Группа опций для настройки параметров осей
Список осей	Выбор настраиваемой оси
Scaling (Разметка)	Выпадающее меню для выбора автоматического (Auto) либо ручного (Manual) типа разметки осей
Type (Тип)	Выбор типа разметки оси
Linear (Линейный)	Линейный тип разметки оси
Logarithmic (Логарифмический)	Логарифмический тип разметки оси
Log-decimal (Логарифмический десятичный)	Логарифмический десятичный тип разметки оси
Min (Минимум)	Установка минимального значения шкалы
Max (Максимум)	Установка максимального значения шкалы
Step (Шаг)	Установка шага разметки шкалы
3D Text (3М Текст)	Опция для отображения заголовков осей в трехмерной перспективе
Scale Options (Настройка осей)	Кнопка для вызова окна дополнительных настроек осей. В этом окне можно задать формат отображения шкалы (Scale Value) – числовой (Numeric), дата (Data), или текстовый (Text Label); число и положение маркеров
Fit (Сглаживание)	Группа опций для выбора функции, подгоняющей отображаемую поверхность к исходным данным

Окончание таблицы 2.6

1	2
Number of cuts (Число разбиений)	Группа опций задания числа участков, на которые разбивается интервал при построении графика
Raw data spices (Отображать исходные данные)	Опция для отображения точек исходных данных на графике
Axes (Оси)	Кнопка для настройки вида осей графика
Gridlines (Линии сетки)	Кнопка для настройки линий сетки
Data Label (Метки данных)	Кнопка для отображения на графике меток данных (например, имен случаев)
Borders (Рамка)	Кнопка для задания параметров рамки вокруг графика

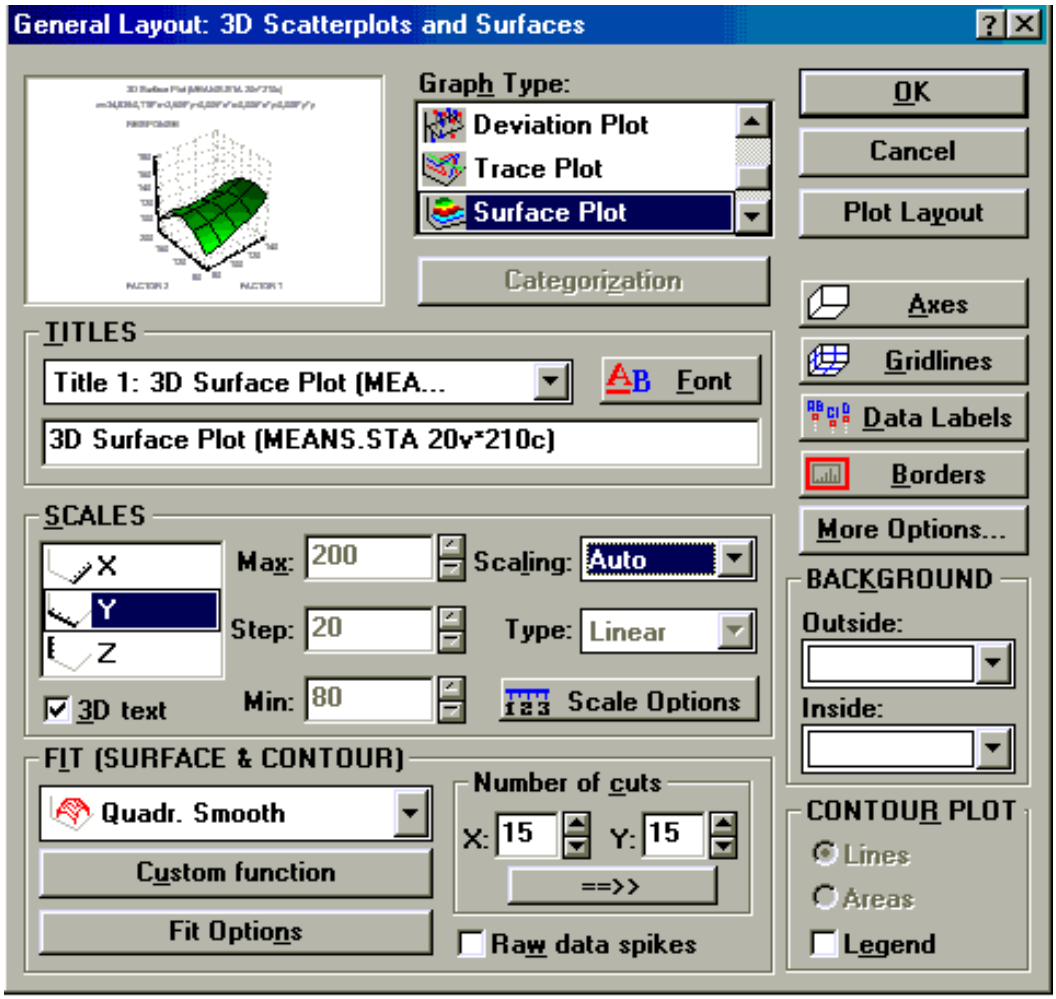


Рисунок 2.3 – Диалоговое окно General Layout

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

Упражнение 1

В этом упражнении вы научитесь строить статистические графики.

1. Запустите модуль Basic Statistics and Tables.
2. Откройте файл данных **Visota.sta**.
3. Для выбранной переменной, например **Опыт3**, постройте гистограмму. Для этого выполните команду Graphs – Stats 2D Graphs – Histograms... На экране появится диалоговое окно для построения гистограммы (рисунок 2.4). Выберите переменную **Опыт3**, нажав на кнопку Variables:.
4. Настройте процентное представление шкалы **Y** и отображение процентных значений над каждым столбцом. Для этого в разделе Graph Type: установите флажок Show Percentages и в списке Y axis: выберите **%** (рисунок 2.4).
5. Включите наложение на гистограмму кривой нормального теоретического распределения вероятностей. Для этого в разделе Fit Type выберите Normal. Сравните полученный график с рисунком 2.5. Обратите внимание, что для правильного отображения имени переменной – **Опыт3**, необходимо выбрать шрифт **Ms Sans Serif**. Сохраните график в файле **Histogramma1**.

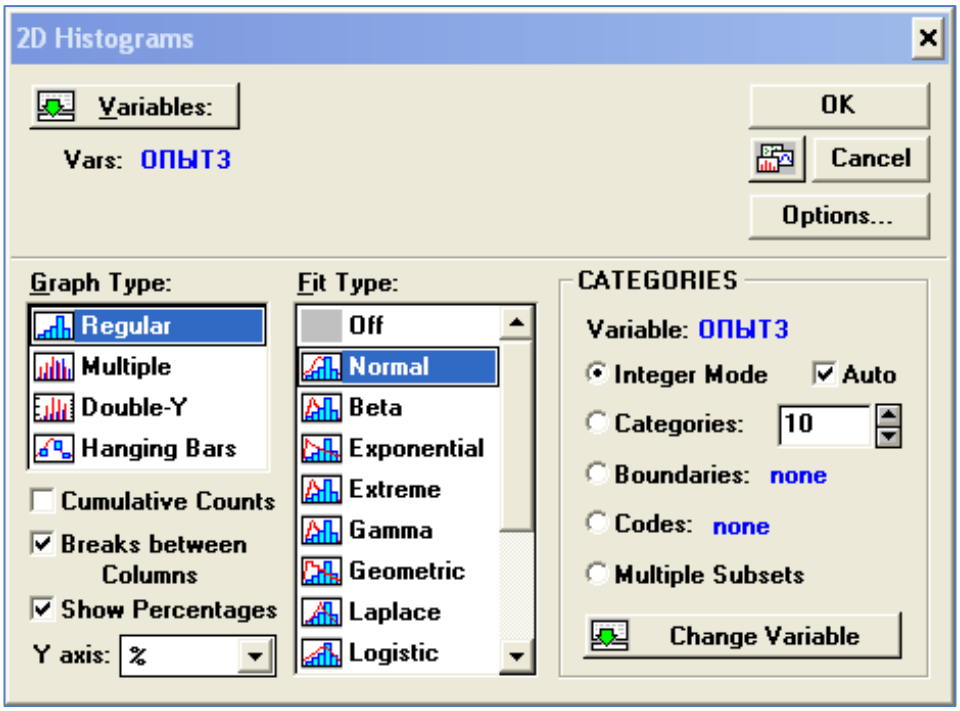


Рисунок 2.4 – Диалоговое окно для построения гистограммы

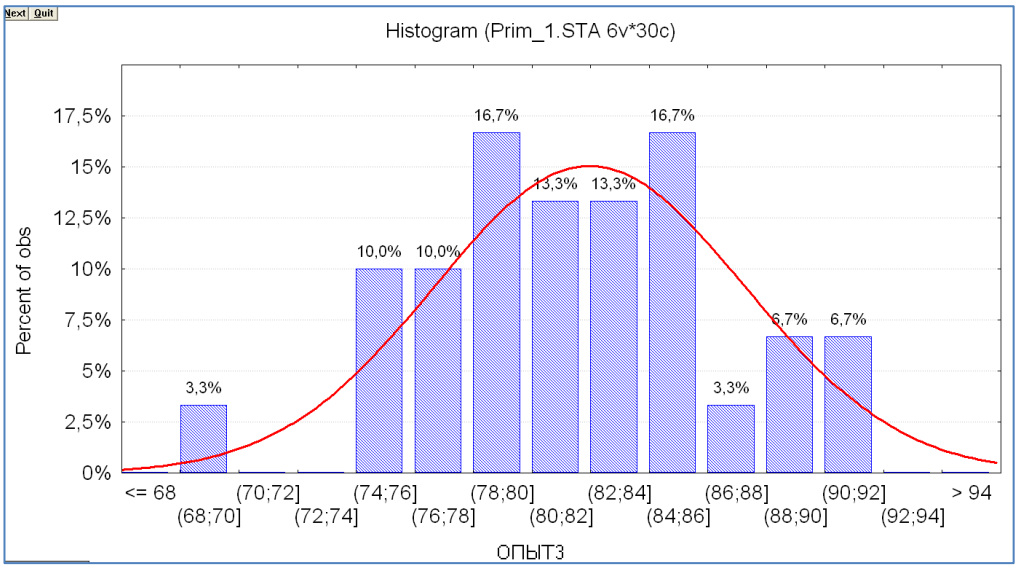


Рисунок 2.5 – Гистограмма для переменной Опыт3

6. Представьте данные переменной **Опыт3** в виде кумулятивной кривой (графика накопленных частот). Для этого в разделе Graph Type: установите флажок Cumulative Counts. Полученный график сравните с рисунком 2.6. Сохраните график в файле **Cumulata**.

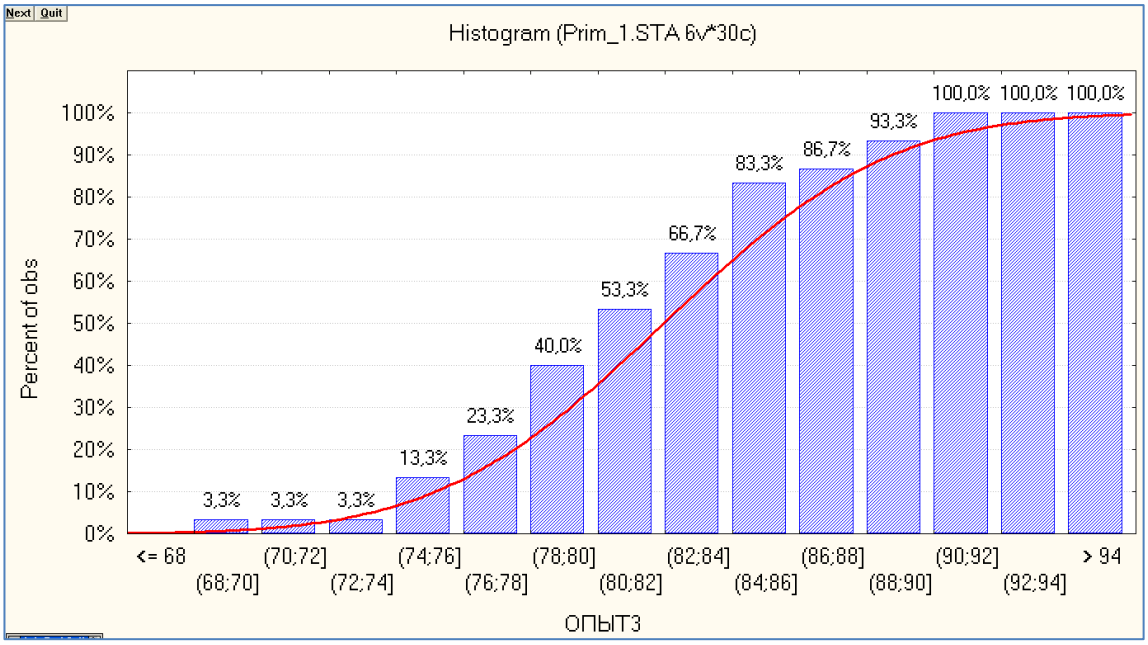


Рисунок 2.6 – Кумулятивная кривая для переменной Опыт3

7. Постройте составную гистограмму, отображающую данные в абсолютных величинах для переменных **Опыт**3 и **Контроль** (рисунок 2.7). Для этого в разделе Graph Type: выберите Multiple.

8. Для разбиения на интервалы задайте следующие границы интервалов: **75 80 85 90**. Границы вводятся через пробел. Обратите внимание, что при задании интервалов необходимо в разделе Fite Type: выбрать Off, а в разделе CATEGORIES выбрать Boundaries:. Затем нажать на кнопку Specify Bounda-

ries и ввести границы интервалов через пробел. Сравните полученный график с рисунком 2.8. Сохраните гистограмму в файле **Histogramma2**.

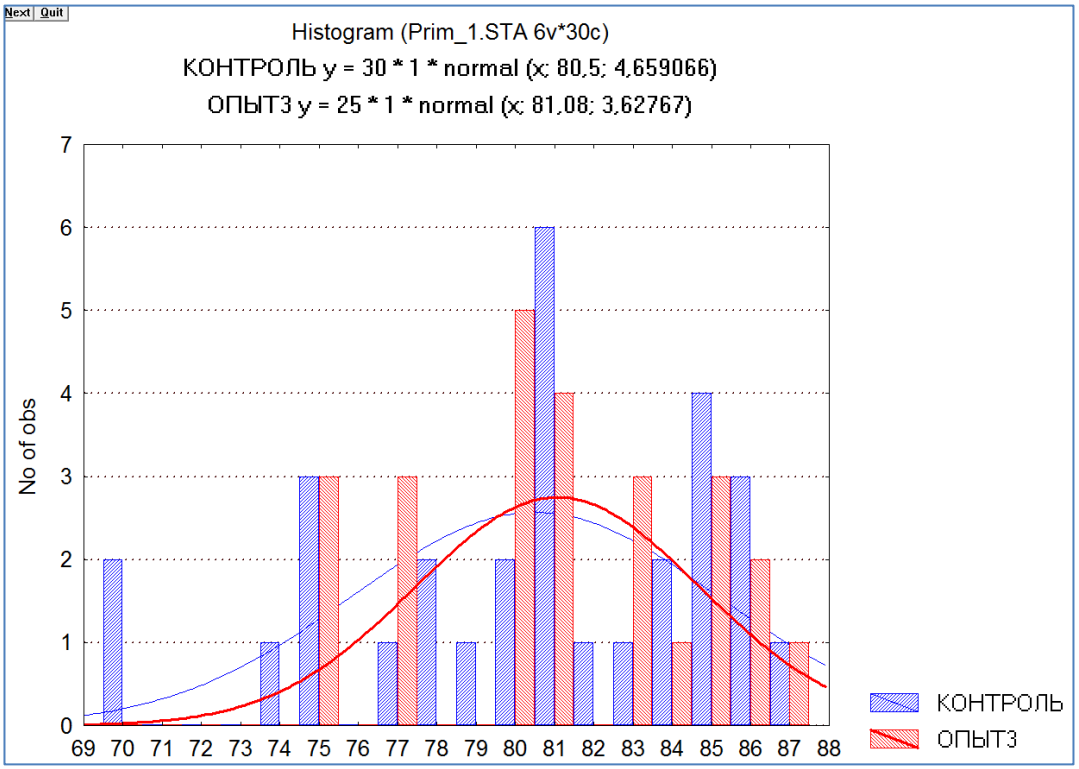


Рисунок 2.7 – Составная гистограмма

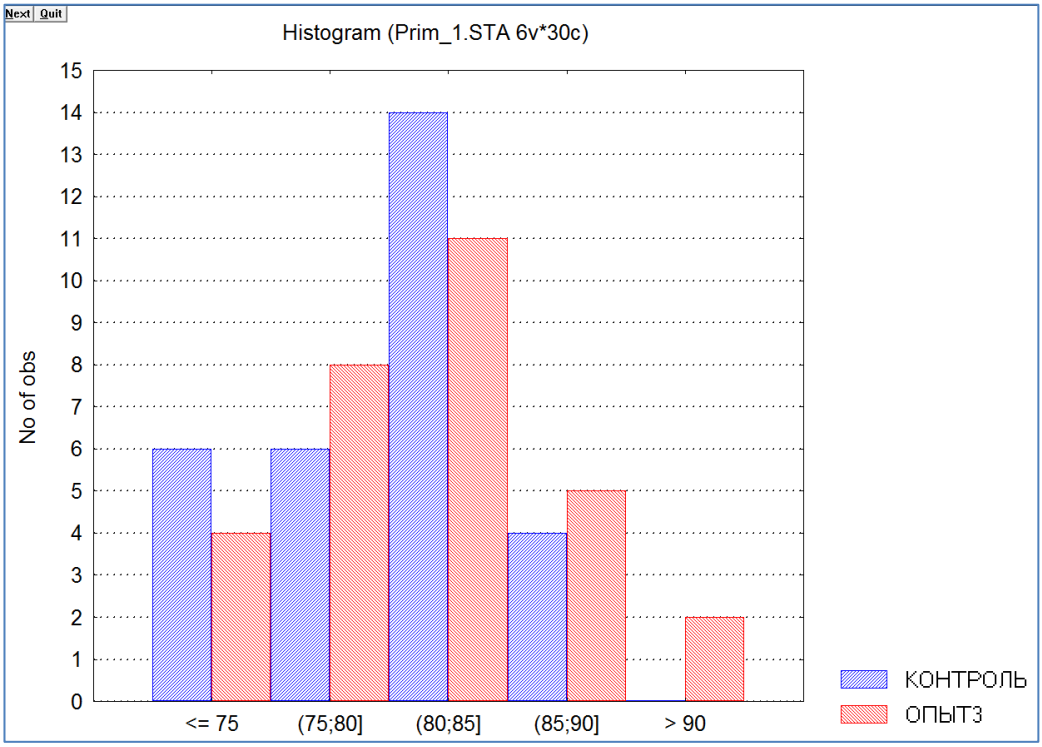


Рисунок 2.8 – Разбиение гистограммы на интервалы

9. На составной гистограмме опробуйте группировку данных в заданном количестве интервалов. Просмотрите, как изменится график при изменении числа интервалов с пяти до девяти. Закройте график без сохранения.

10. Постройте для переменной **Опыт3** графики вероятностей на нормальной (Normal) и полунормальной (Half-Normal) вероятностной бумаге. Для этого выполните команду Graphs – Stats 2D Graphs – Normal Probability Plots... В появившемся диалоговом выберите переменную и тип графика. Результаты работы сравните с рисунками 2.9 и 2.10. Закройте все созданные графики и файл данных.

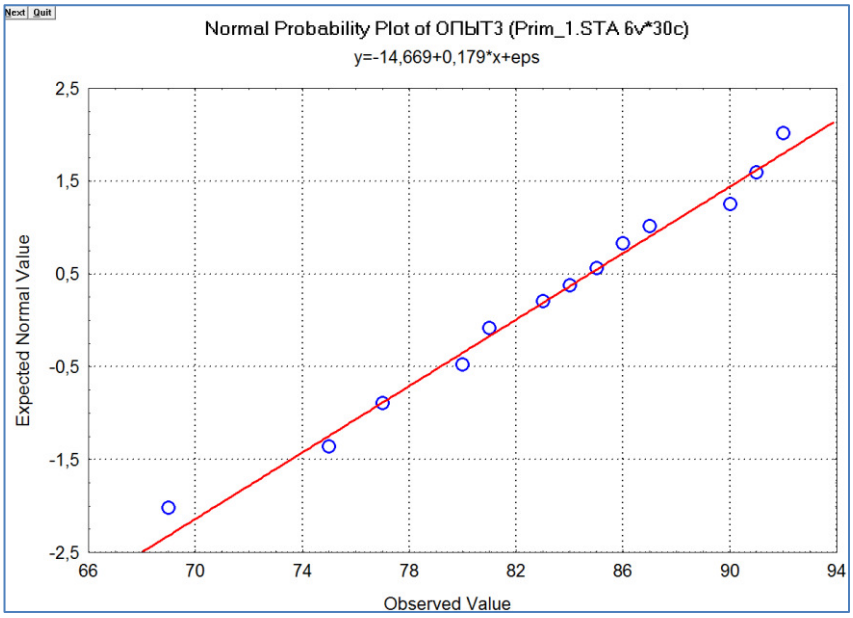


Рисунок 2.9 – График вероятностей для переменной Опыт3 на нормальной вероятностной бумаге

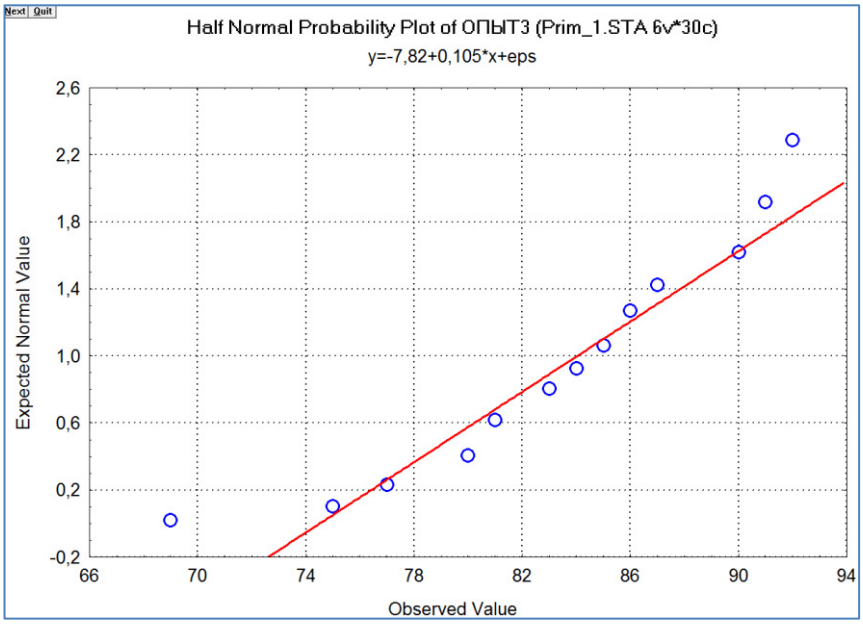


Рисунок 2.10 – График вероятностей для переменной Опыт3 на нормальной вероятностной бумаге

Упражнение 2

В этом упражнении вы научитесь строить диаграммы рассеяния.

1. В модуле Basic Statistics and Tables создайте новый файл с именем **Massa.sta**.
2. Уменьшите число переменных до трех.
3. Заполните файл в соответствии с таблицей 2.7. В таблице представлены две серии измерений массы 1000 семян при различных влажностях зерна.

Таблица 2.7 – Результаты изменения массы 1000 семян при разных влажностях зерна

φ, %	20	19	18	17	16	15	14	13	12	11
m1, г	46,2	45,4	43,9	42,7	42,1	41,3	43	40,7	39,5	38,2
m2, г	45,3	44,6	43,1	42,0	41,5	39,9	39,4	38,2	37,8	37,1

4. Постройте Диаграмму рассеяния (Scatterplots...) для переменной **m1**. По оси **X** откладывайте значения влажности, по оси **Y** – значения массы. Включите отображение сглаживающих линий (рисунок 2.11).

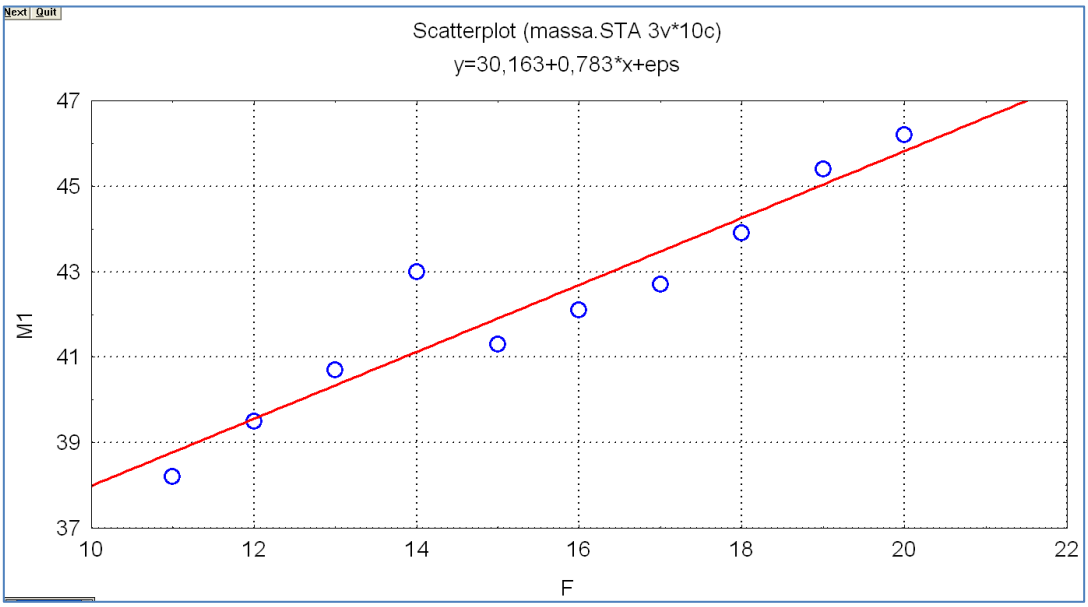


Рисунок 2.11 – Диаграмма рассеяния для переменной m1

5. Проанализируйте график. Обратите внимание, что значение массы семян при влажности 14% значительно удалено от сглаживающей прямой. В этом случае, следует скорректировать это значение, например, взять равным 41,3.
6. Постройте Диаграмму рассеяния (Scatterplots...) для переменных **m**1 и **m2**. Включите отображение сглаживающих линий. Сохраните полученный график в файле **ScatterMassa** (рисунок 2.12).
7. Сохраните файл данных. Закройте все файлы.

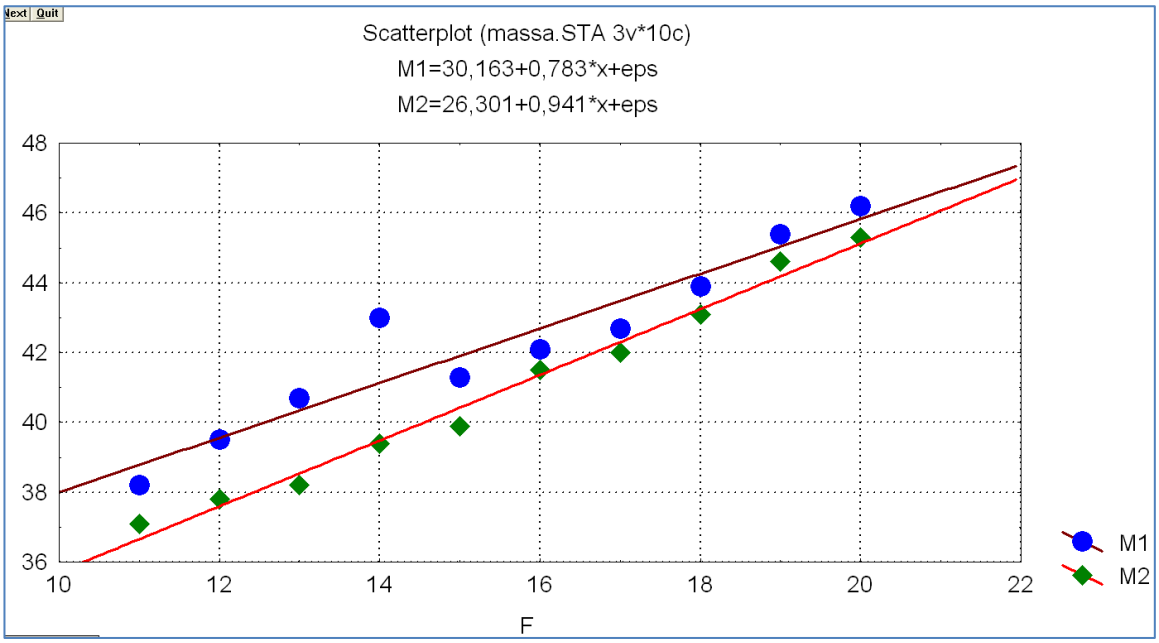


Рисунок 2.12 – Диаграмма рассеяния для переменных m1, m2

Упражнение 3

В этом упражнении вы научитесь выполнять настройку статистических графиков.

1. Откройте график **Histogramma2**, созданный на предыдущем занятии.
2. Для установки настроек статистического графика по умолчанию выполните команду Options – Global Defaults... На экране появится диалоговое окно Global Defaults (Общие установки по умолчанию) (рисунок 2.13).
3. При помощи диалогового окна Global Defaults настройте следующие параметры графика **Histogramma2**:
4. установите белый цвет внешней области графика. Для этого в разделе BACKGROUND в списке Outside: выберите самый светлый цвет. Примените выбранные настройки к текущему графику, для этого установите флажок в пункте Copy Options to Current Graph;
5. установите размер основных рисок на оси **Y** равный 6 пикселам, риски должны пересекать ось. Для этого в разделе AXES выберите Y left и нажмите кнопку Settings. В появившемся диалоговом окне (рисунок 2.14) в разделе TICKMARKS в списке Major: выберите Crossed, в списке Size: выберите **9** и нажмите OK;
6. увеличьте толщину линий осей **X** и **Y** до **1** пиксела. Для этого в диалоговом окне (рисунок 2.14) нажмите кнопку Scale Line. На экране появится окно для настройки линий осей, где в разделе Size в списке Point необходимо выбрать **1** и нажать OK (см. рисунок 2.15);

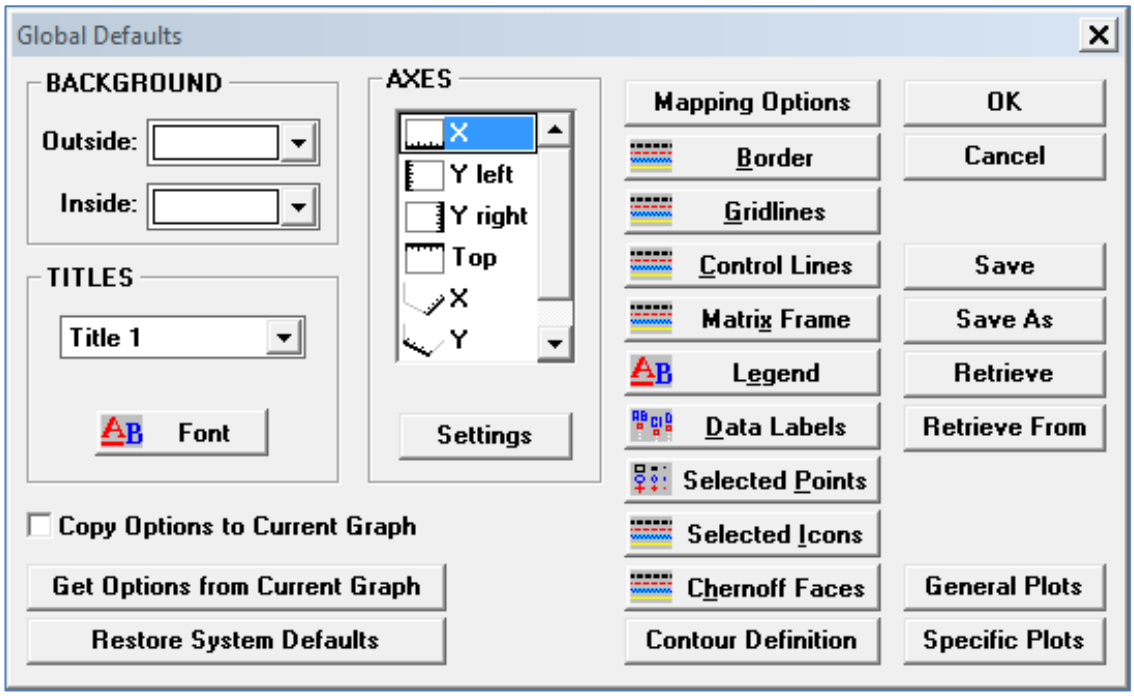


Рисунок 2.13 – Диалоговое окно Global Defaults

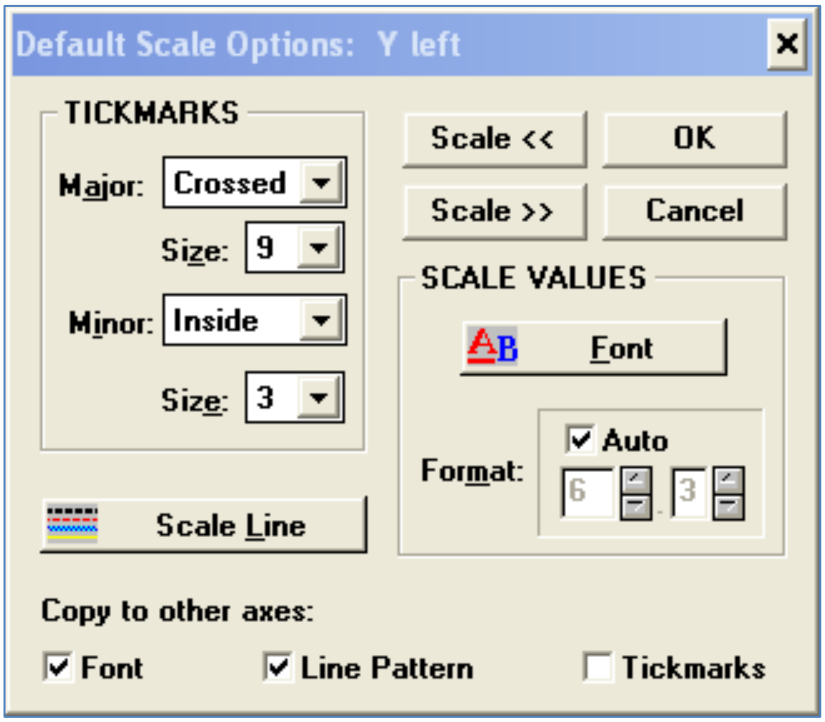


Рисунок 2.14 – Диалоговое окно настройки опций оси Y

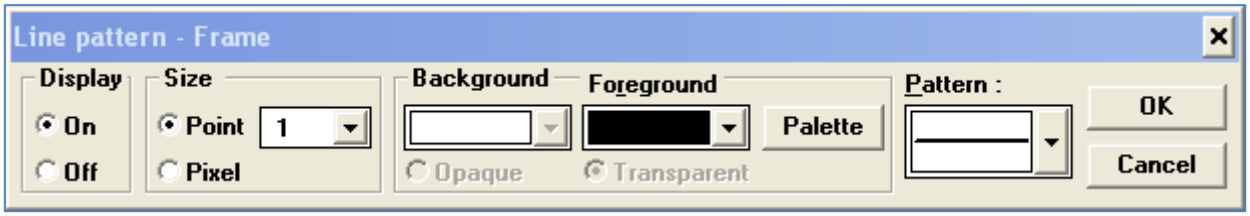


Рисунок 2.15 – Диалоговое окно настройки линий осей

– настройте отображение сетки в виде пунктирных линий черного цвета. Для этого нажмите на кнопку Gridlines и в появившемся диалоговом окне в списке Pattern: выберите пунктирную линию;

– замените цвета контуров и заполнения столбцов для варианта на малиновый, для контроля – на темно-зеленый. Для этого нажмите кнопку General Plots и в появившемся окне и в разделе Area щелкните по кнопке с цветом, соответствующим цвету варианта. На экране появится окно, в котором необходимо выбрать малиновый цвет и нажать OK. Повторите действия и выберите цвет для Контроля;

– сохраните произведенные настройки для графиков в файле настроек **Histogram.ini**.

Для того чтобы выбранные настройки применились к текущему графику, необходимо выбрать опцию Copy Options to Current Graph (Применить установки к текущему графику) до нажатия кнопки OK.

Результаты работы сравните с рисунком 2.16.

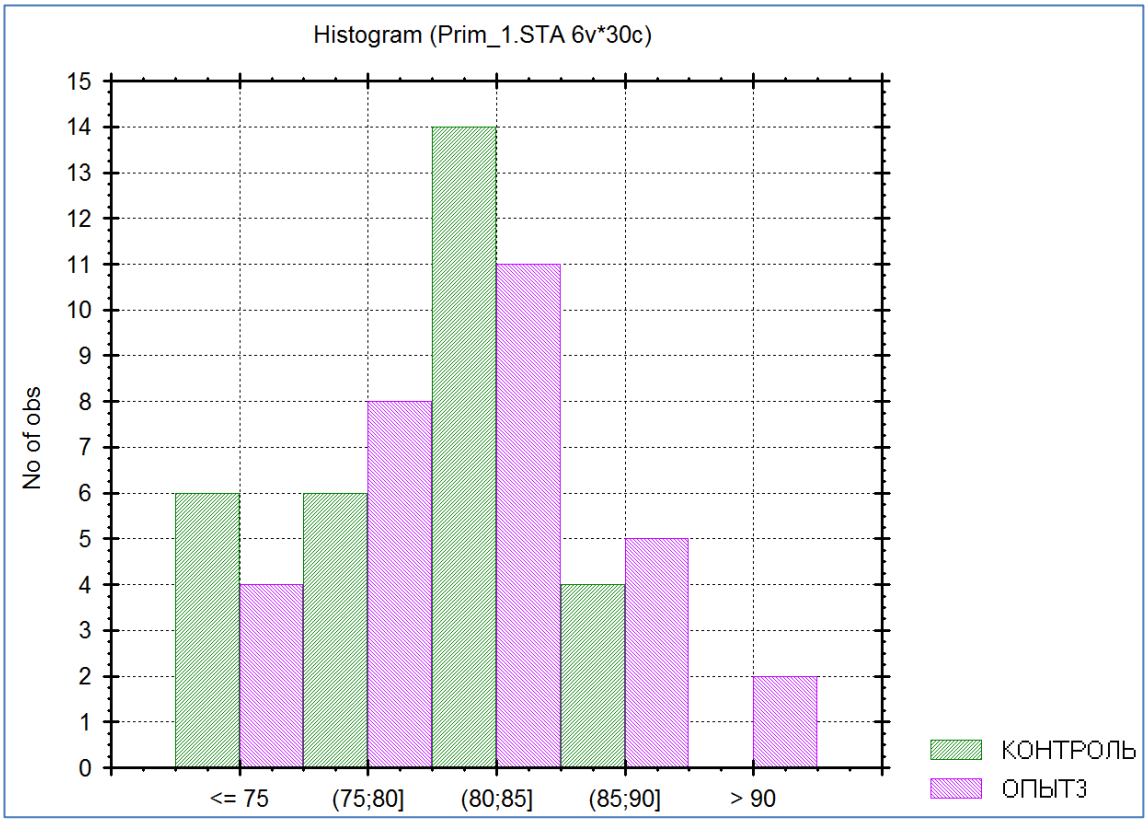


Рисунок 2.16 – Результат настройки гистограммы с помощью диалогового окна Global Defaults

7. При помощи диалогового окна General Layout (Общая разметка) (рисунок 2.17) настройте следующие параметры графика **Histogramma2**:

– установите максимальное значение шкалы **Y** равным **20** и ее шаг равным двум. Для этого в разделе SCALES выберите Y left, в списке Max: выберите **20**, в списке Step: выберите **2**;

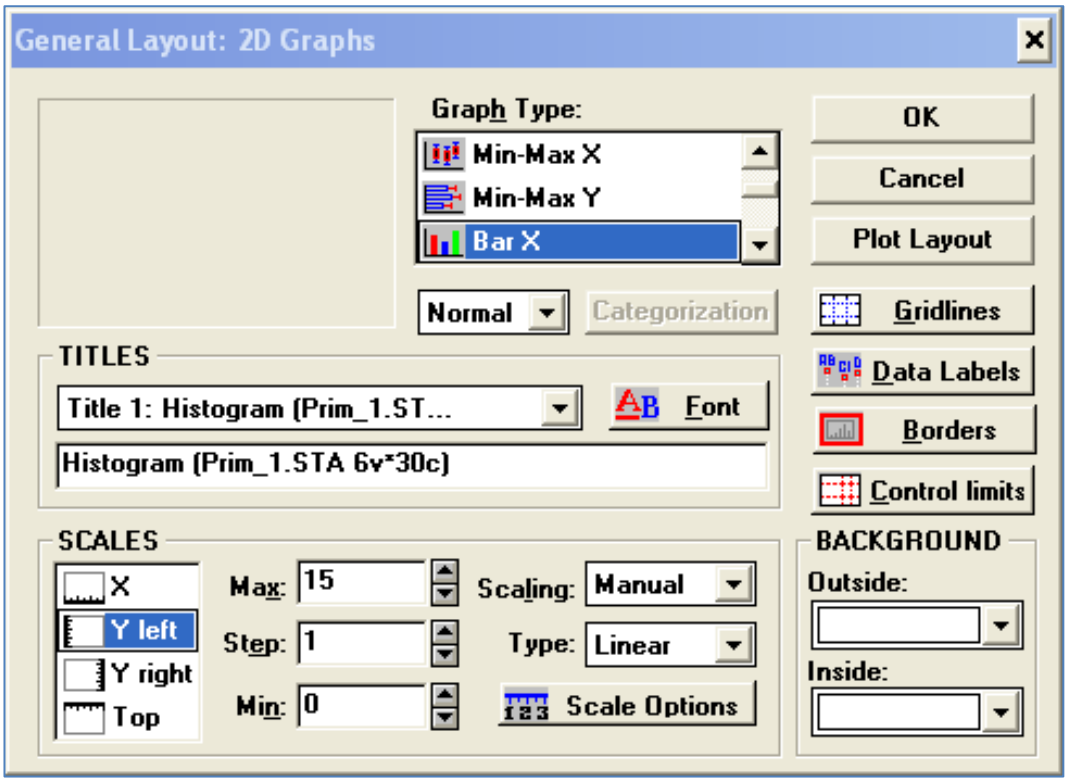


Рисунок 2.17 – Диалоговое окно настройки графика General Layout

– в качестве заголовка первого уровня введите текст: «Сравнительная гистограмма»; в качестве заголовка второго уровня – текст: «для вариантов…» вместо многоточия укажите свои варианты, укажите названия осей **Y** – **«Число наблюдений, шт.»**, **X** – **«Длина растения, см»**, выберите шрифт, корректно отображающий названия осей. Для этого в разделе TITLES в раскрывающемся списке выберите Title 1:, в текстовом поле ниже введите **Сравнительная гистограмма**. Далее нажмите кнопку Font и в появившемся диалоговом окне выберите шрифт **MS Sans Serif**, нажмите OK. Аналогично введите заголовок второго уровня и название осей;

– окрасьте внешнюю область графика в светло-голубой цвет. Для этого в разделе BACKGROUND в списке Outside: выберите нужный цвет;

– при помощи кнопки Gridlines (Линии сетки) выключите отображение вертикальных линий сетки.

Результаты работы сравните с рисунком 2.18.

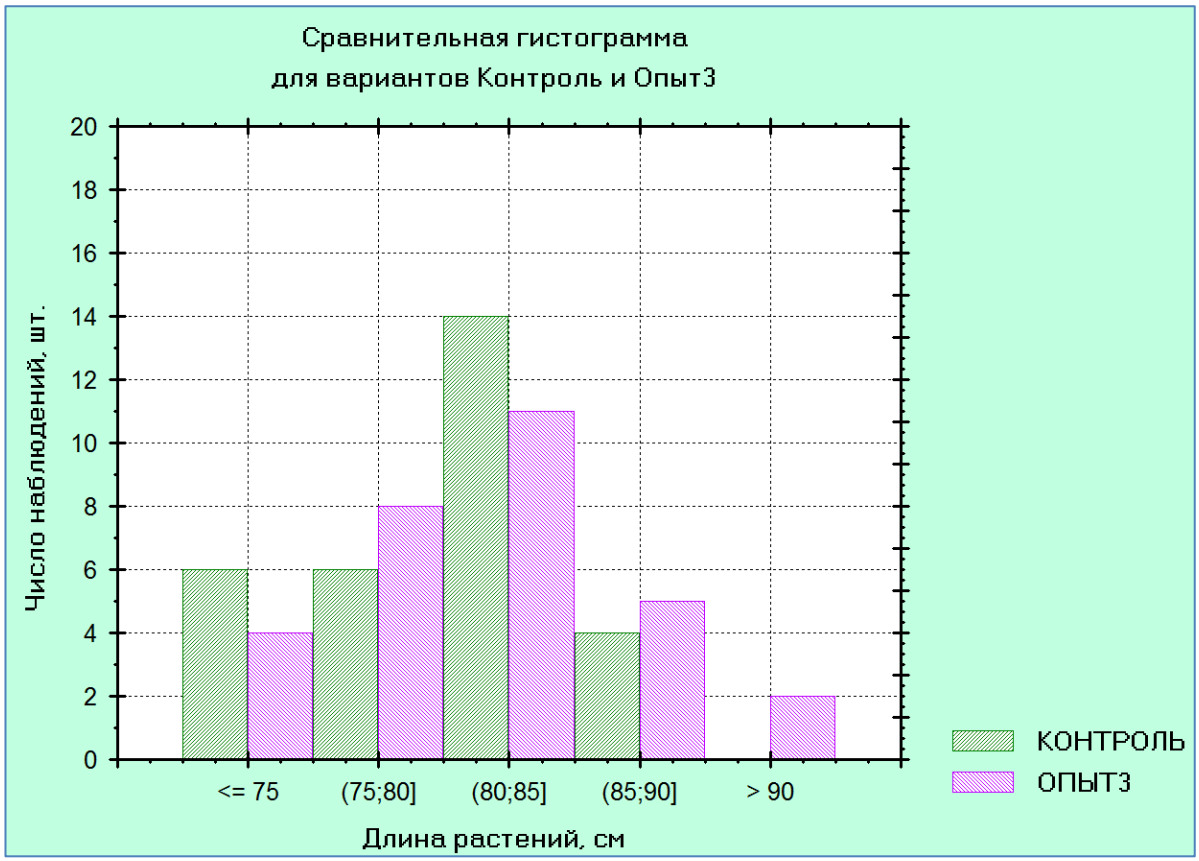


Рисунок 2.18 – Результат настройки гистограммы
с помощью диалогового окна General Layout

8. При помощи контекстных меню к элементам графиков или меню, вызываемых двойным щелчком, настройте следующие параметры графика **Histogramma2**:

– увеличьте на два пункта размер шрифта заголовков первого и второго уровней. Для этого щелкните правой кнопкой мыши по заголовку гистограммы и в появившемся контекстном меню выберите Edit Text... В появившемся диалоговом окне Edit Titles нажмите на кнопку  рядом с выделенным заголовком и в появившемся диалоговом окне в списке Size: увеличьте размер шрифта на два пункта. Повторите действия для второго заголовка;

– установите курсивное начертание названий осей;

– установите размер основных рисок на оси **Y** равный **9 пикселам**, риски должны находиться внутри области построения, установите толщину линий осей **X** и **Y** равной **0,75 пиксела**. Для этого щелкните правой кнопкой мыши по указанной оси и в появившемся списке выберите пункт Scaling/Style... На экране появится диалоговое окно, в котором в разделе MAJOR TICKMARKS установите **Inside**, **9**. Щелкните по кнопке Scale Line и установите толщину осей **0,75**. Нажмите OK;

– включите отображение значений над столбцами гистограммы. Для этого щелкните правой кнопкой мыши по любому столбцу гистограммы и выберите Change this Plot/series Layout... В появившемся диалоговом окне щелкните по кнопке Data Lables и выберите ось **Y**. Нажмите OK.

9. Сохраните график **Histogramma2**.

Результаты работы сравните с рисунком 2.19.

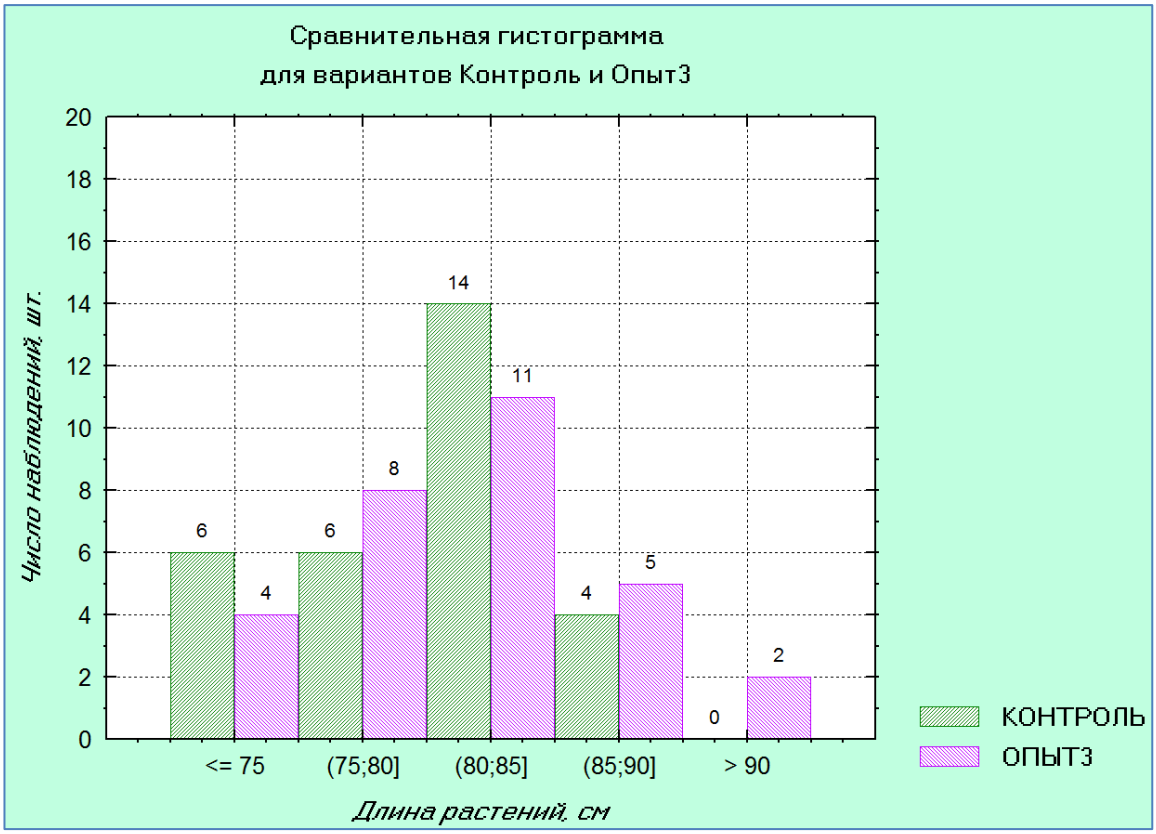


Рисунок 2.19 – Результат настройки гистограммы

10. При помощи диалогового окна Global Defaults (Общие установки по умолчанию) сбросьте настройки при помощи кнопки Restore System Defaults (Восстановить установки по умолчанию).

11. Закройте файл графика без сохранения. Закройте программу.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Запустите модуль Basic Statistics and Tables.
2. Откройте файл данных **Visota.sta**.
3. Постройте линейный график с двойной осью **Y** для заданной переменной (левая ось) и контроля (правая ось). Сохраните файл с именем **GraphDoubleY.stg**.
4. Постройте для этих двух переменных линейный график с одной осью **Y**. Сохраните файл с именем **GraphDoubleVar.stg**.
5. Постройте для выбранной переменной круговую диаграмму. Сохраните ее в файле **GraphPie.stg**.
6. При помощи быстрых статистических графиков для указанной преподавателем переменной постройте двумерный график размаха типа Прямоугольник-Отрезки (Box-Whisker). Сохраните график в файле **GraphBox.stg**.

7. Постройте поверхность в трехмерном пространстве (Stats 3D XYZ Graphs), заданную уравнением **$Z=2X^2+5,5Y$**. Сохраните ее в файле **MySurface**.
8. Откройте график **Histogramma1**.
9. Введите название графика.
10. Задайте заголовки осей.
11. Выключите отображение сетки в области построения графика.
12. Измените шрифтовое и цветовое оформление элементов графика по своему усмотрению.
13. Сохраните график с произведенными настройками в исходном файле.

5 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Какие элементы графиков Вам известны?
2. Какие категории графиков имеются в программе Statistica?
3. Что представляет собой графическая галерея Statistica? Как к ней обратиться?
4. Какие типы гистограмм имеются в Statistica?
5. В чем отличие простых, составных графиков и графиков с двойной осью?
6. Что такое подгонка или сглаживание?
7. В чем отличие двумерных диаграмм рассеяния (2D Scatterplots) от линейных графиков переменных (Line Plots(Variables)) и линейных графиков профилей наблюдений (Line Plots(Cases Profiles))?
8. Для чего нужна сглаживающая линия (Fit) на диаграммах рассеяния?
9. По каким правилам записывают выражения для графиков, заданных Функцией пользователя (Custom Function)?
10. Что представляют собой заголовки, легенда, метки данных?
11. Какие методы настройки графиков Вам известны?
12. Какие способы быстрой настройки элементов графиков Вам известны?
13. Какие шрифты корректно отображают Кириллицу на графиках?

Практическое занятие № 3

Тема: ЗАКОНЫ РАСПРЕДЕЛЕНИЯ ВЕРОЯТНОСТЕЙ. ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ В ПРАКТИКЕ УЧЕНОГО. ПОДБОР ТЕОРЕТИЧЕСКОГО РАСПРЕДЕЛЕНИЯ К ЭМПИРИЧЕСКИМ ДАННЫМ С ИСПОЛЬЗОВАНИЕМ КРИТЕРИЯ χ^2

1 ЦЕЛЬ РАБОТЫ

Изучить:

– основные законы распределения вероятностей, их свойства и графики;

– методику выдвижения и проверки статистических гипотез;

– виды возможных ошибок;

– методы подбора теоретического распределения и проверки нормальности распределения опытных данных.

Научиться:

– пользоваться вероятностным калькулятором;

– принимать решения о законе распределения вероятностей, руководствуясь статистическими критериями;

– проводить проверку статистических гипотез;

– подбирать теоретическое распределение для статистической выборки.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Законы распределения вероятностей

2.1.1 Нормальное распределение

Одним из наиболее часто встречающихся распределений является нормальное Гауссово распределение. Это связано с тем, что результаты экспериментов являются следствием воздействия на объект исследования множества факторов.

Непрерывная случайная величина X имеет нормальное распределение, если плотность распределения вероятности $f(X)$ имеет вид

$$f(X) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2e^2}}, \tag{3.1}$$

где μ – генеральная средняя;

σ – стандартное отклонение генеральной совокупности.

Из формулы видно, что *параметрами нормального распределения* являются *среднее значение* и *стандартное отклонение*. Нормальный закон с параметрами $\mu = 0$ и $\sigma = 1$ называется *стандартным*.

Для нормального распределения, независимо от значений среднего стандартного отклонения, характерны следующие особенности:

1. В области ограниченной одним стандартным отклонением влево и вправо от точки максимума $(\mu \pm \sigma)$, лежит 68,27% всех наблюдений.
2. В области, ограниченной двумя стандартными отклонениями $(\mu \pm 2\sigma)$, лежат 95,45% всех наблюдений.
3. В области, ограниченной тремя стандартными отклонениями, лежат 99,73% всех наблюдений.

Последняя особенность известна как правило трех сигм. Это правило часто применяется на практике. Например, если значение наблюдаемой величины отклоняется от среднего значения выборки более чем на 3σ, то следует обратить внимание на качество наблюдений. Возможно, это значение является ошибочным и его следует исключить из последующего анализа. Естественно это правило справедливо только при нормальном распределении вероятностей.

2.1.2 Распределения, связанные с нормальным

2.1.2.1 Хи-квадрат распределение (Распределение Пирсона)

Пусть X_i $(i = 1, 2, 3, \dots, k)$ независимые случайные величины со стандартным нормальным распределением (математическое ожидание каждой из них равно нолю, а среднее квадратическое отклонение равно единице).

Тогда сумма квадратов этих величин

$$\chi^2(k) = \sum_{i=1}^{n} X_i^2 \qquad (3.2)$$

распределена по закону χ^2 с k степенями свободы.

Число степеней свободы k является единственным параметром распределения χ^2. Квантили распределения χ^2 для различного числа степеней свободы k приведены в приложении Г.

Свойства распределения $\chi^2(k)$:

1. Случайная величина $\chi^2(k)$ имеет нулевую плотность распределения при $x \leq 0$, т.е. функция плотности вероятности $f(x)$ этого распределения находится в положительной области. Это следует из определения, т.к. данная случайная величина есть сумма квадратов и не может принимать отрицательные значения (вероятность попасть в отрицательный интервал равна нулю).
2. С увеличением числа степеней свободы $(k \to \infty)$ распределение χ^2 приближается к нормальному.

3. Математическое ожидание случайной величины, распределенной по закону $\chi^2(k)$, равно k.

$$M_{\chi^2(k)} = k. \tag{3.3}$$

4. Дисперсия случайной величины, распределенной по закону $\chi^2(k)$, равна $2k$.

$$D_{\chi^2(k)} = 2k. \tag{3.4}$$

2.1.2.2 Распределение Стьюдента (t-распределение)

Пусть X – стандартная нормально распределенная случайная величина ($M = 0,\ \sigma = 1$), а $\chi^2(k)$ – случайная величина, имеющая $\chi^2(k)$ – квадрат распределения с k степенями свободы, причем X и $\chi^2(k)$ – независимые величины.

Тогда, распределение случайной величины

$$t(k) = \frac{X}{\sqrt{\chi^2(k)/k}}. \tag{3.5}$$

называется распределением Стьюдента или t-распределением с k степенями свободы.

Число степеней свободы k является параметром t-распределения. Квантили t-распределения для различного числа степеней свободы k приведены в приложении Б.

Свойства t-распределения:

1. Функция плотности t-распределения симметрична относительно оси ординат, причем $M_{t(k)} = 0$.
2. С возрастанием числа степеней свободы k распределение Стьюдента быстро приближается к стандартному нормальному распределению.

$$D_{t(k)} = \frac{k}{k-2} \tag{3.6}$$

2.1.2.3 Распределение Фишера или F-распределение

Пусть X и Y – независимые случайные величины, распределенные по закону χ^2 со степенями свободы k_1 и k_2.Тогда величина

$$F(k_1,\ k_2) = \frac{X/k_1}{Y/k_2}. \tag{3.7}$$

имеет распределение Фишера (или F-распределение) со степенями свободы k_1 и k_2.

Функция плотности распределения Фишера задается двумя параметрами: числом степеней свободы k_1 и k_2. Квантили данного распределения приведены в приложении В.

Свойства F-распределения:

1. Функция плотности распределения случайной величины $F(k_1,\ k_2)$ равна нулю при $x \leq 0$, т.к. случайные величины X и Y имеют нулевые плотности распределения при $x \leq 0$.
2. Математическое ожидание случайной величины $F(k_1,\ k_2)$ равно:

$$M_{F(k_1,k_2)} = \frac{k_2}{k_2 - 2} \tag{3.8}$$

2.2 Методика проверки статистических гипотез

Под *статистической гипотезой* понимают всякое высказывание о генеральной совокупности (случайной величине), проверяемое по выборке (по результатам наблюдений).

Статистические гипотезы выдвигают относительно вида неизвестного распределения или параметров известных распределений.

Не располагая сведениями о всей генеральной совокупности, высказанную гипотезу сопоставляют с выборочными данными и делают вывод о правильности предположений. Такую процедуру называют проверкой гипотезы.

Процедура проверки статистических гипотез включает следующие этапы:

Этап 1. Формулировка нулевой H_0 и альтернативной H_1 гипотез. Проверяемую гипотезу называют нулевой или основной (H_0). Альтернативной или конкурирующей (H_1) называют гипотезу, противоположную основной, т.е. ту, которая будет принята в случае опровержения нулевой гипотезы.

Этап 2. Задание уровня значимости α. Под уровнем значимости понимают вероятность принять нулевую гипотезу, в то время как она неверна.

Этап 3. Выбор статистики К критерия для проверки нулевой гипотезы. Эта статистика должна удовлетворять следующим условиям:

– ее значение можно просчитать только на основании выборки;
– ее распределение известно в предположении, что нулевая гипотеза верна.

В качестве критериев могут использоваться и статистики t, F, χ^2 и т.д.

Используют критерии двух видов: параметрические и непараметрические. *Параметрическими* называют критерии, которые основаны на предположении, что распределение признака подчиняется нормальному закону распределения. К таким критериям относятся, в частности, критерии t и F.

Непараметрическими называют критерии, не требующие предварительного вычисления параметров и закона распределения. Они могут применяться, когда распределение отличается от нормального.

Непараметрические статистики менее эффективны (обладают меньшей мощностью) в сравнении с параметрическими.

Этап 4. Используя функцию плотности распределения выбранной статистики К в предположении, что нулевая гипотеза H_0 верна, определяют на числовой оси критическую область ω.

Критической областью ω называют такую подобласть таких значений из области допустимых значений критерия К, которые свидетельствуют о существенном расхождении выборки с гипотезой H_0.

Вероятность того, что критерий К примет значение из критической области ω должна быть равна заданному значению α.

$$P(K \in \omega) = \alpha. \tag{3.9}$$

Однако данная критическая область определяется неоднозначно. Возможны три вида расположения критической области (в зависимости от вида нулевой и альтернативной гипотез, вида и распределения критерия К):

1) Правосторонняя критическая область (рисунок 3.1 а), лежит в интервале $\left(X^{\text{кр}}_{\text{пр}\alpha}, +\infty\right)$, где точка $X^{\text{кр}}_{\text{пр}}$ определяется из условия

$$P\left(K > X^{\text{кр}}_{\text{пр}\alpha}\right) = \alpha. \tag{3.10}$$

Величина $X^{\text{кр}}_{\text{пр}\alpha}$ называется *правосторонней критической точкой*, отвечающей уровню значимости α.

2) Левосторонняя критическая область (рисунок 3.1 б) лежит в интервале $\left(-\infty,\ X^{\text{кр}}_{\text{лев}\alpha}\right)$, где точка $X^{\text{кр}}_{\text{лев}\alpha}$ определяется из условия

$$P\left(K < X^{\text{кр}}_{\text{лев}\alpha}\right) = \alpha. \tag{3.11}$$

Величина $X^{\text{кр}}_{\text{лев}\alpha}$ называется *левосторонней критической точкой*, отвечающей уровню значимости α.

3) Двусторонняя критическая область (рисунок 3.1 в) лежит в интервалах $\left(-\infty,\ X^{\text{кр}}_{\text{лев}\alpha/2}\right)$ и $\left(X^{\text{кр}}_{\text{пр}\alpha/2}, +\infty\right)$, где точки $X^{\text{кр}}_{\text{лев}\alpha/2}$ и $X^{\text{кр}}_{\text{пр}\alpha/2}$ определяются из условий:

$$\begin{gathered} P\left(K < X^{\text{кр}}_{\text{лев}\alpha/2}\right) = \frac{\alpha}{2}, \\ P\left(K > X^{\text{кр}}_{\text{пр}\alpha/2}\right) = \frac{\alpha}{2}. \end{gathered} \tag{3.12}$$

и называются *двусторонними критическими точками*.

Этап 5. По выборке вычислить значение статистики К.

Этап 6. Принять решение.

Если вычисленное значение статистики К лежит в критической области,

то нулевую гипотезу H_0 отвергают и принимают альтернативную гипотезу H_1.

Так как нулевая гипотеза H_0 проверяется по выборочным данным, то для генеральной совокупности принятое решение может быть ошибочным.

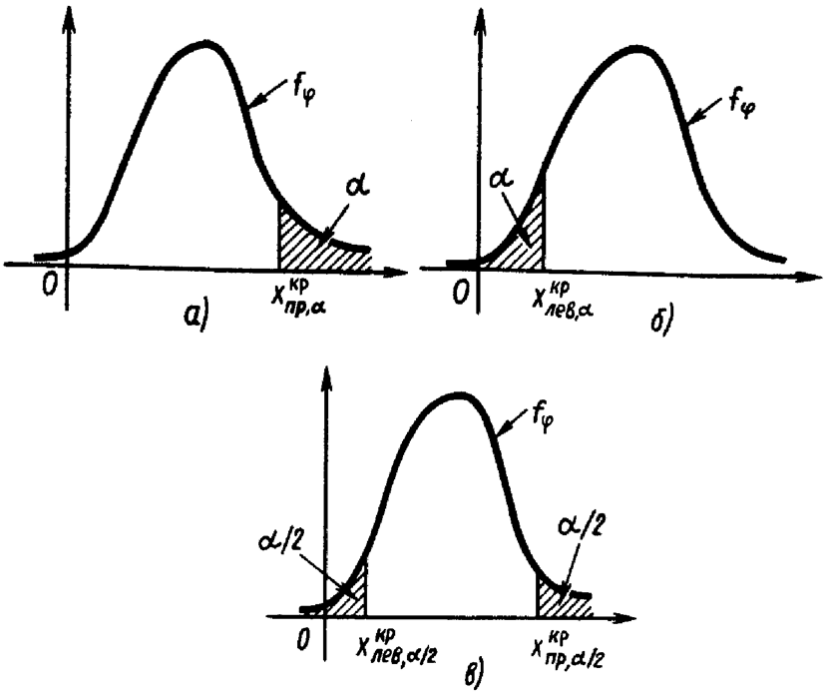


Рисунок 3.1 – Виды расположения критической области

Принимая решение на основе статистического критерия, возможны следующие ситуации:

– гипотеза H_0 верна и не отвергается, при этом отражается действительное положение;

– гипотеза H_0 верна, но она отвергается. В этом случае говорят, что допущена ошибка первого рода. Вероятность этого события равна уровню значимости α;

– гипотеза H_0 неверна и она отвергается. При этом принято правильное решение: отвергнута неверная гипотеза;

– гипотеза H_0 неверна, но она не отвергается. Говорят, что допущена ошибка второго рода. Вероятность ошибки второго рода обозначают через β.

Критическую область выбирают таким образом, чтобы значение β (вероятность принять неверную гипотезу) было наименьшим. В этом случае вероятность отвергнуть неверную гипотезу $1-\beta$ максимальна. Это число называют *мощностью критерия*.

Выше изложенное можно кратко сформулировать в виде таблицы 3.1.

Таблица 3.1 – Принятие решения на основе статистического критерия

Верна ли гипотеза H_0	Решение, принимаемое о гипотезе H_0 по выборке	
	Гипотеза H_0 отвергается, т.е. принимается гипотеза H_1	Гипотеза H_0принимается
Гипотеза H_0 верна	Ошибка первого рода, ее вероятность равна α	Правильное решение, его вероятность равна $1-\alpha$
Гипотеза H_0неверна, т.е. верна гипотеза H_1	Правильное решение, его вероятность равна $1-\beta$	Ошибка второго рода, ее вероятность равна β

2.2.1 Проверка гипотезы о виде распределения генеральной совокупности. Критерий согласия Хи-квадрат К. Пирсона

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Процедуру подбора теоретического распределения, характеризующего наблюдаемые данные, называют подгонкой (на англ. языке fitting).

Проверка гипотезы о типе распределения по критерию χ^2 не отличается от стандартной схемы проверки гипотез и заключается в следующем:

1) В качестве нулевой гипотезы H_0 принимают предположение, что распределение выборочных данных хорошо согласуется с выбранным теоретическим распределением. В качестве альтернативной гипотезы H_1 – предположение о значительном различии распределений.

2) Задаются уровнем значимости. Обычно принимают $\alpha = 0{,}05$.

3) Вычисляют статистику χ^2 Пирсона по формуле

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} \tag{3.13}$$

где n – объем выборки,

n_i – количество значений i-го интервала,

k – число интервалов,

p_i – теоретическая вероятность попадания в i-й интервал.

4) Определяют критическую область и область принятия нулевой гипотезы.

Область принятия нулевой гипотезы определяется выражением:

$$\chi^2_{\text{выч}} < \chi^2_{\text{кр}\alpha,\nu}, \tag{3.14}$$

где α – принятый уровень значимости,

ν – число степеней свободы.

Число степеней свободы ν определяется из выражения:

$$\nu = k - m - 1, \tag{3.15}$$

где m – число параметров вероятностной модели (теоретического распределения).

Например, у нормального распределения таких параметров два – это среднее значение μ и стандартное отклонение σ. Таким образом, для нормального распределения $\nu = k - 2 - 1 = k - 3$.

5) Принимают решение.

Примечание. Для применения критерия χ^2 Пирсона объем выборки должен быть достаточно велик (более 50 значений), каждый интервал должен содержать не менее 5 значений. Если количество в интервале менее 5, то его объединяют с соседним.

2.2.2 Критерии Колмогорова, Смирнова и типа Колмогорова-Смирнова

Часто при проверке гипотез о виде распределения недостаточно применить один критерий, особенно, когда результаты проверки не дают значительного отклонения от границы критической области. В таких случаях имеется возможность использовать другие критерии, основанные на других вероятностных идеях.

Критерий χ^2 Пирсона является весьма универсальным и может применяться в самом общем случае. Для непрерывных функций распределения применяют критерии Колмогорова и Смирнова. При проверке на нормальность, когда μ и σ оцениваются посредством $\bar{X}$ и S, эти критерии обладают большей мощностью, чем критерий Пирсона.

Основная идея критериев Колмогорова, Смирнова и аналогичных им состоит в измерении расстояния между функцией эмпирического распределения $F_n(X)$ и функцией теоретического распределения $F(X)$. Различаются эти критерии лишь видом расстояний в пространстве функций распределения.

В качестве нулевой гипотезы выдвигается предположение о соответствии эмпирического распределения теоретическому, т.е.:

$$H_0: F_n(X) = F(X). \tag{3.16}$$

В 1933 году А.Н. Колмогоров предложил использовать в качестве критерия согласия статистику D_n:

$$D_n = \sup_{-\infty<x<\infty} |F_n(X) - F(X)| \approx \max_{i=1,\dots,k} |F_n(i) - F(i)|, \tag{3.17}$$

где $F_n(X)$ – эмпирическая (полученная по опытным данным) функция распределения,

$F(X)$ – гипотетическая функция распределения,

$F_n(i) = \frac{n_1+n_2+\cdots+n_i}{n}$ – эмпирическая функция накопленных относительных частот,

$F(i) = p_1 + p_2 + \cdots + p_i$ – теоретическая функция накопленных относительных частот i-го интервала,

k – число интервалов.

2.2.3 Визуальные методы проверки нормальности распределения

Проверку соответствия эмпирического распределения нормальному закону распределения можно производить как с помощью критериев проверки гипотез, так и посредством визуальных (графических) средств.

Одним из методов оценки нормальности распределения является построение гистограммы исследуемой выборки с наложением кривой теоретического нормального распределения.

Другой метод визуальной оценки нормальности распределения частот предполагает построение графика распределения на так называемой *нормальной вероятностной бумаге*.

Нормальная вероятностная бумага представляет собой график, по оси абсцисс которого откладываются эмпирически полученные значения, а по оси ординат – преобразованные значения. Преобразование производится при помощи обратной функции нормального распределения (Лапласа) $\Phi^{-1}(u)$.

В результате в плоскости (x, z) получаем линейную зависимость

$$z = \frac{x-\mu}{\sigma}. \qquad (3.18)$$

Таким образом, график теоретического нормального вероятностного распределения на нормальной вероятностной бумаге представляет собой прямую линию (рисунок 3.2). И чем ближе точки эмпирических данных располагаются к теоретической прямой, тем ближе эмпирическое распределение вероятностей к нормальному.

Кроме графиков на нормальной вероятностной бумаге используют так же графики на *полунормальной вероятностной бумаге* (рисунок 3.3). Отличие заключается лишь в том, что преобразованная координата оси ординат не имеет отрицательных значений. Построения ведутся только в положительной части вероятностной плоскости.

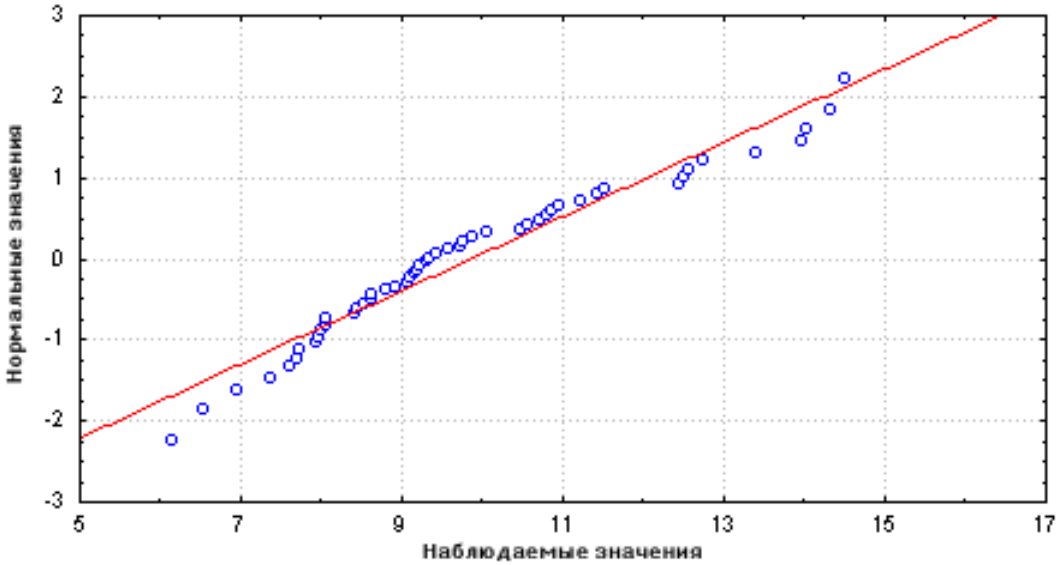


Рисунок 3.2 – График распределения частот
на нормальной вероятностной бумаге

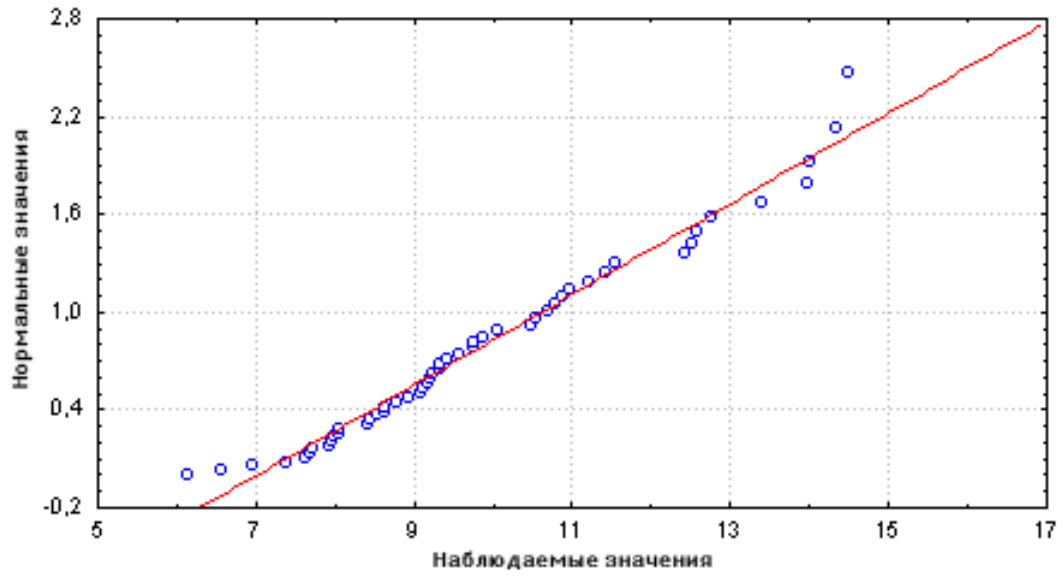


Рисунок 3.3 – График распределения частот
на полунормальной вероятностной бумаге

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

Описанные выше законы распределения вероятностей часто используются в различных процедурах анализа данных.

При использовании закона распределения необходимо по его параметрам, для выбранного уровня значимости определить значение функции плотности распределения. Для этих целей используются таблицы, приведенные в литературе.

Поиск необходимых значений по таблицам не всегда удобен. При работе в программе Statistica имеется возможность непосредственного вычисления и построения графика функции плотности заданного распределения посред-

ством задания необходимых параметров. Для этой цели служит вероятностный калькулятор.

3.1 Вероятностный калькулятор

Вероятностный калькулятор находится в модуле Basic Statistics and Tables. Запустить его можно командой Analysis – Probability calculator (Вероятностный калькулятор). После выбора данной команды появляется диалоговое окно Probability Distribution Calculator (Калькулятор вероятностных распределений), представленное на рисунке 3.4.

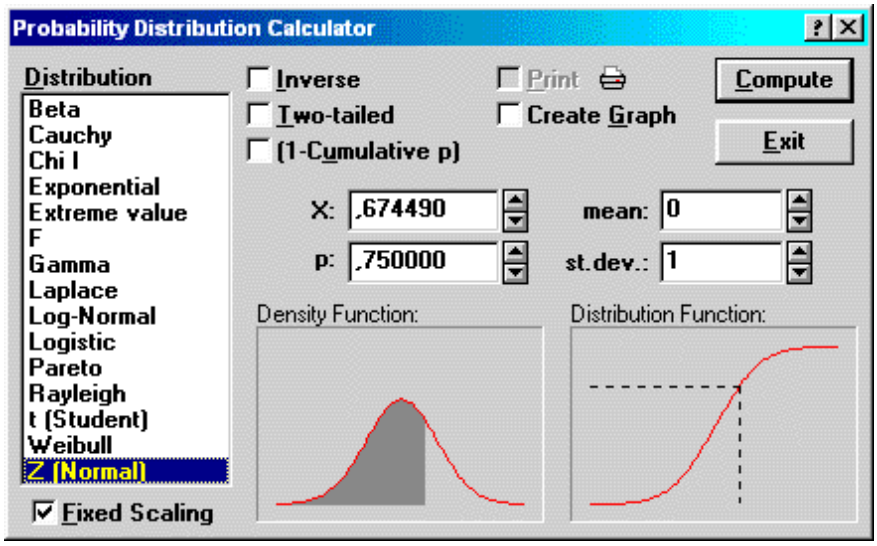


Рисунок 3.4 – Калькулятор вероятностных распределений

В поле Distribution (Распределение) представлен список доступных для вычисления распределений (таблица 3.2).

Справа от поля выбора распределения Distribution появляются поля, соответствующие выбранному распределению. В эти поля вводятся (заносятся) параметры распределения. Например, для нормального распределения – это среднее значение (mean) и стандартное отклонение (St.dev), для распределений Хи-квадрат, Стьюдента и Фишера – это степени свободы (df) (для распределения Фишера их две df1 и df2).

В поле p указывается вероятность. При этом вычисляется соответствующий ей *квантиль* (значение, которое заданная случайная величина не превышает с фиксированной вероятностью). Название поля квантиля зависит от выбранного распределения. Так, для нормального распределения это X, для распределения Хи-квадрат – это Chi I, для распределения Стьюдента – t, для распределения Фишера – F.

Если внести в поле ввода значение квантиля, то будет вычислена соответствующая ему вероятность.

Таблица 3.2 – Распределения, представленные в вероятностном калькуляторе

Наименования в списке распределений	Полное название распределения
Beta	Бета-распределение
Cauchy	Распределение Коши
Chi I	Хи-квадрат распределение (χ^2-распределение)
Exponential	Экспоненциальное распределение
Extreme value	Распределение экстремальных значений
F	Распределение Фишера (F-распределение)
Gamma	Гамма-распределение
Laplace	Распределение Лапласа
Log-Normal	Логнормальное распределение
Logistic	Логистическое распределение
Pareto	Распределения Парето
Rayleigh	Распределение Релея
t(student)	Распределение Стьюдента (t-распределение)
weibull	Распределение Вейбулла
Z(Normal)	Нормальное распределение

В нижней части окна Probability Distribution Calculator отражаются графики функции плотности (Density Function) и функции распределения (Distribution Function).

Если выбрать опцию Create graph (Создать график), то будет создано окно, в котором будут построены данные графики. Причем, если выбрать опцию Fixed Scaling (Фиксированная шкала), будет использоваться фиксированная шкала для отображения графиков.

Опция Invers (Инвертировать) позволяет вычислить обратную функцию распределения.

Если опция Invers активна, то вычисляется квантиль заданной функции распределения по ее параметрам и указанной вероятности. Если же сбросить установку этой опции, то вычисляется вероятность по заданным параметрам и значению квантиля.

Опция Two-tailed (Двухсторонний) доступна для симметричных относительно оси Y распределений и позволяет отсекать заданный квантиль симметрично слева и справа от оси ординат.

Опция 1-Cumulative p позволяет вычислить вероятность и построить графики для вероятности равной **1-p**.

Упражнение 1

В этом упражнении вы научитесь работать с вероятностным калькулятором.

1 Откройте программу Statistica.

2 В модуле Basic Statistics and Tables запустите вероятностный калькулятор с помощью команды Analysis – Probability calculator.

В появившемся диалоговом окне Probability Distribution Calculator (см. рисунок 3.4) в поле Distribution выберите нормальное распределение – Z (Normal). Настройте диалоговое окно как показано на рисунке 3.5 и нажмите Compute.

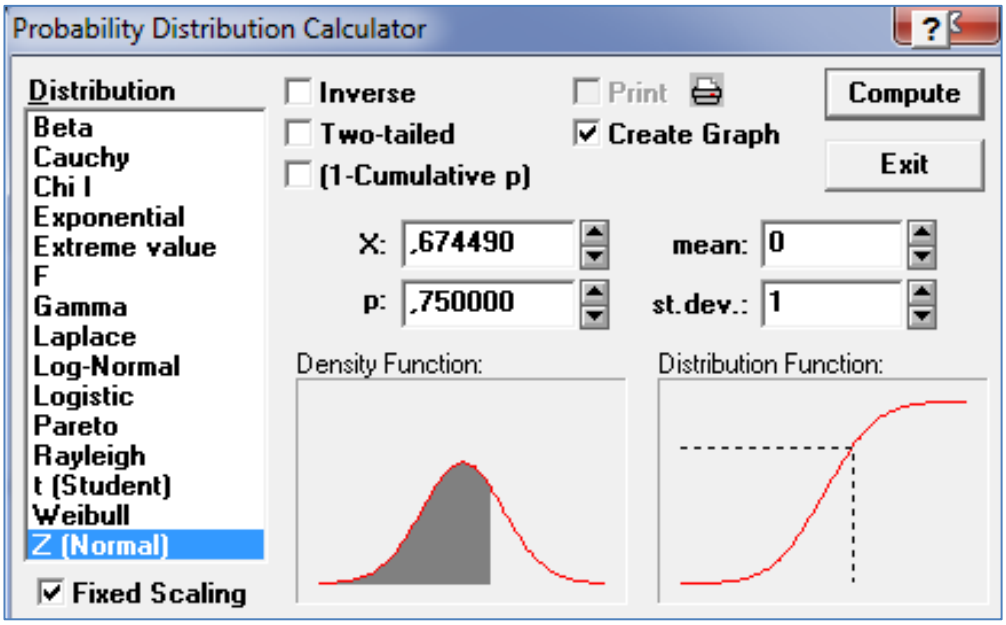


Рисунок 3.5 – Настройки диалогового окна Probability Distribution Calculator для стандартного нормального распределения с параметрами (0, 1)

На экране появится окно с двумя графиками. Закройте диалоговое окно, щелкнув по кнопке Exit, и разверните окно с графиками на весь экран (см. рисунок 3.6).

Обратите внимание, что в окне два графика: слева – график функции плотности распределения, а справа – график самой функции распределения. Для того чтобы развернуть один из графиков на весь экран, щелкните по нему дважды.

Вертикальная пунктирная линия на графике функции плотности распределения проходит через точку x, равной значению квантиля (в данном случае **0,674**).

На графике функции распределения пунктирной линией показано значение вычисленного квантиля и соответствующая ему вероятность.

3 При помощи вероятностного калькулятора постройте графики функции плотности нормального распределения при изменении среднего значения от **–3** до **3**. При этом должна быть выбрана опция Fixed Scaling. Обратите внимание на изменения графика функции.

4 При помощи вероятностного калькулятора постройте графики функции плотности нормального распределения при изменении стандартного от-

клонения от **0,5** до **3**. При этом должна быть выбрана опция Fixed Scaling. Обратите внимание на изменения графика функции.

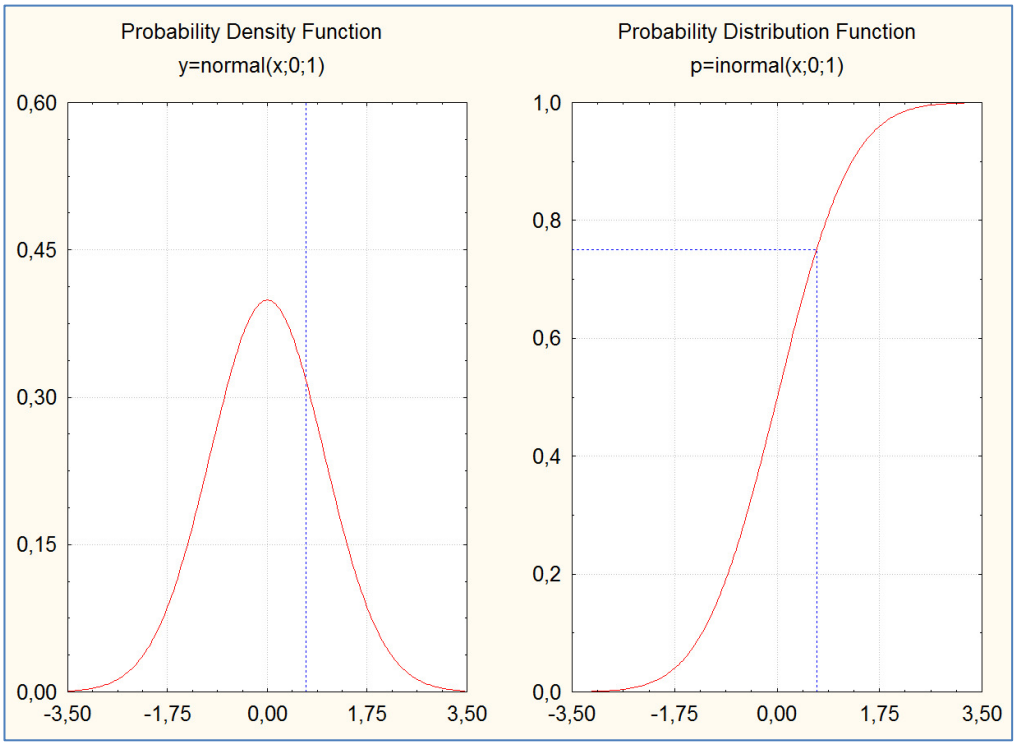


Рисунок 3.6 – Графики плотности и функции нормального распределения для параметров (0, 1)

5 При помощи вероятностного калькулятора постройте графики функции плотности распределения Хи-квадрат при изменении числа степеней свободы от **2** до **100**. При этом опцию Fixed Scaling отключите.

6 Обратите внимание на изменения графика функции.

7 Закройте вероятностный калькулятор.

3.2 Подбор теоретического распределения вероятностей

Проверку гипотезы о виде распределения совокупности при помощи параметрического критерия согласия Хи-квадрат Пирсона и непараметрических критериев типа Колмогорова-Смирнова производят в модуле Nonparametric Statistics / Distributions (Непараметрические статистики / распределения).

Стартовая панель модуля имеет вид, представленный на рисунке 3.7.

На данной панели имеется переключатель, позволяющий выбрать опции Непараметрических тестов (Nonparametric stats) или опции Подгонки вероятностных распределений (Distribution fitting).

Среди Непараметрических тестов (Nonparametric stats) имеется опция Observed versus expected XI (Наблюдаемые частоты против ожидаемых), которая позволяет сравнить по критерию χ^2 два распределения вероятностей, заданных в виде вариационных рядов.

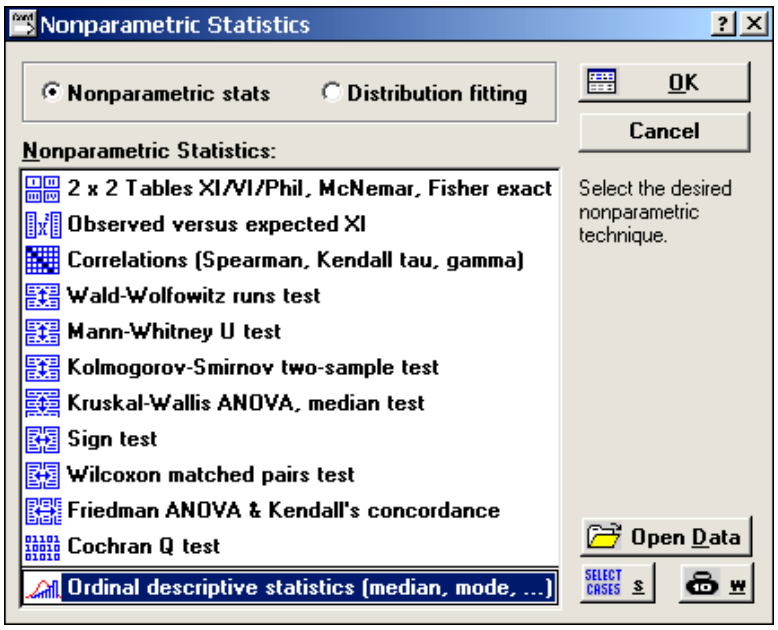


Рисунок 3.7 – Стартовая панель Nonparametric Statistics

При переключении в положение Подгонки вероятностных распределений (Distribution fitting) стартовая панель принимает вид, представленный на рисунке 3.8. На панели имеются списки теоретических Непрерывных (Continuous Distributions) и Дискретных (Discrete Distributions) распределений (см. таблицу 3.3), которые сравнивают с эмпирическим (опытным) распределением.

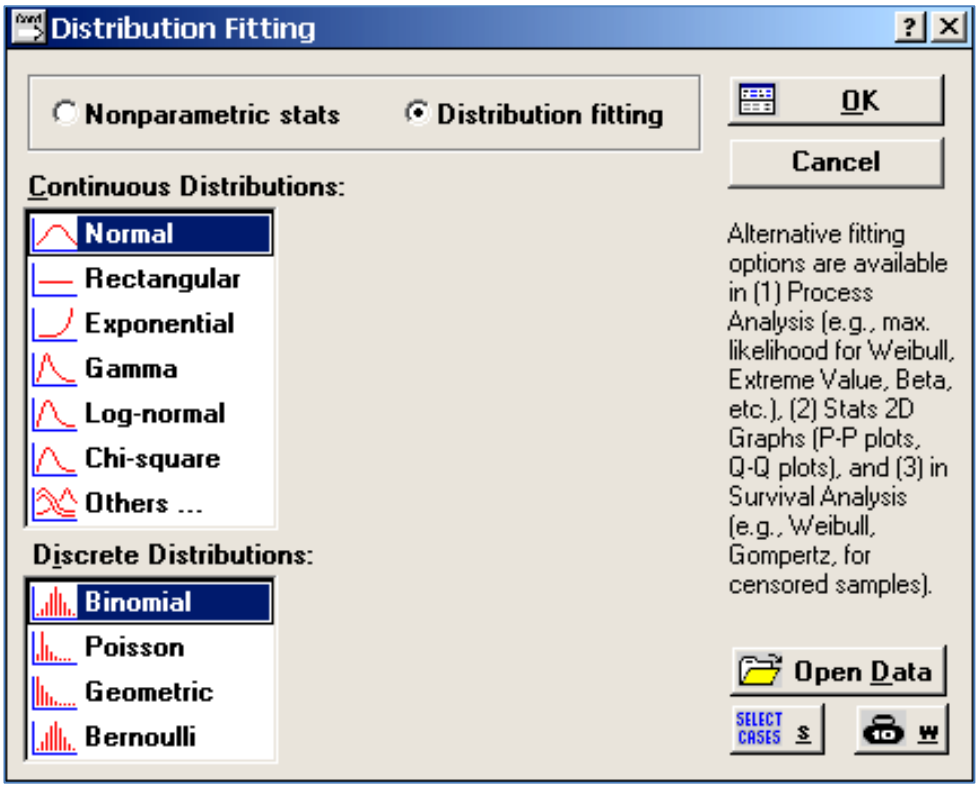


Рисунок 3.8 – Стартовая панель Подгонка вероятностных распределений

Таблица 3.3 – Теоретические вероятностные распределения

Пиктограмма	Распределения	
Непрерывные (Continuous)		
	Normal	Нормальное распределение
	Rectangular	Равномерное распределение
	Exponential	Экспоненциальное распределение
	Gamma	Гамма-распределение
	Log-normal	Логнормальное распределение
	Chi-square	Хи-квадрат распределение
Дискретные (Discrete)		
	Binomial	Биномиальное распределение
	Poisson	Распределение Пуассона
	Geometric	Геометрическое распределение
	Bernoulli	Распределение Бернулли

При выборе одного из непрерывных распределений появляется диалоговое окно Fitting Continuous Distributions (Подгонка непрерывных вероятностных распределений) (рисунок 3.9). Причем часть опций этого окна неактивна до тех пор, пока не выбрана переменная (Variable), эмпирическое распределение которой сравнивается с теоретическим. Назначение опций данного диалога приведено в таблице 3.4.

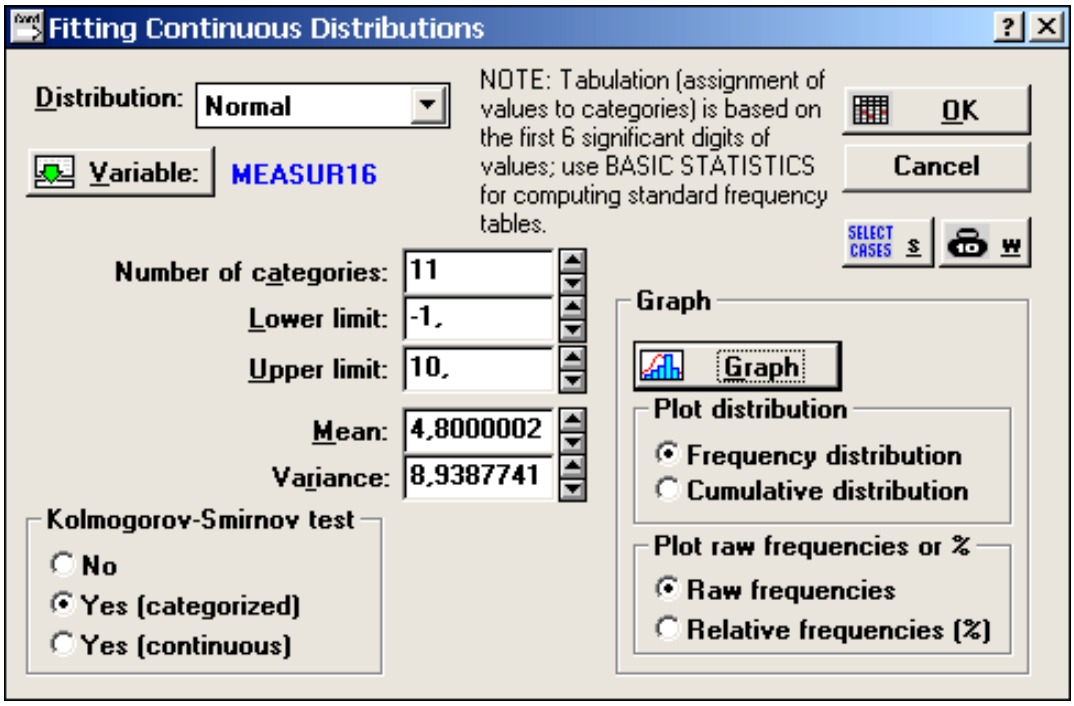


Рисунок 3.9 – Диалоговое окно Fitting Continuous Distributions

Таблица 3.4 – Назначение опций диалогового окна Fitting Continuous Distributions (Подгонка непрерывных распределений)

Элемент окна	Назначение элемента
Distribution (Распределение)	Выбор теоретического непрерывного распределения
Variable (Переменная)	Задание переменной, распределение которой исследуется
Number of categories (Число интервалов)	Задание количества интервалов, на которые разбивается выборка
Lower limit (Нижний предел)	Минимальное значение переменной
Upper limit (Верхний предел)	Максимальное значение переменной
Mean (Среднее значение)	Среднее значение выборки*
Variance (Дисперсия)	Дисперсия выборки*
Kolmogorov-Smirnov test (Тест Колмогорова-Смирнова)	Группа опций, позволяющая отобразить статистики Колмогорова-Смирнова на экране. Имеет значения: No (нет) – не отображать; Yes (categorized) (Да (категоризированный)) – отобразить статистики для категоризированной функции; Yes (continuous) (Да (непрерывный)) – отобразить статистики для непрерывной функции
Graph (График)	Кнопка для вывода графика на экран
Plot distribution (График распределения)	Группа опций для выбора вида отображения распределения вероятностей. Имеет значения: Frequency distribution (Распределение частот) или Cumulative distribution (Распределение накопленных частот)
Plot raw frequencies or % (Столбцы графика в виде частот или %)	Группа опций для выбора формата отображения данных в столбцах. Имеет значения: Raw frequencies (Столбцы частот) или Relative frequencies (%) (Относительные частоты (%))
OK	Отображение подробной таблицы категоризированных данных с эмпирическим и теоретическим распределениями

* параметры нормального распределения. При выборе другого теоретического распределения отображаются параметры этого распределения

Упражнение 2

В этом упражнении вы научитесь проверять соответствие опытных данных теоретическому распределению.

Проверить соответствие результатов опыта Менделя по скрещиванию двух сортов гороха теоретическому ожидаемому отношению желтых семян к

зеленым как **3:1**. В результате опыта во втором поколении получено **355** желтых семян и **123** зеленых.

Так как сумма полученных в опыте семян равна **478**, то теоретически должны получить **3/4·478=358,5** желтых семян и **1/4·478=119,5** зеленых семян.

1 Запустите модуль Nonparametric Statistics с помощью переключателя модулей.

2 Создайте новую таблицу, имеющую две переменные. Первая из них будет иметь два значения: **355** и **123**, вторая: **358,5** и **119,5** (рисунок 3.10). Сохраните файл с именем **Mendel.sta**.

3 Активируйте стартовую панель модуля и выберите среди Непараметрических статистик (Nonparametric stats) опцию Observed versus expected XI (Наблюдаемые частоты против ожидаемых).

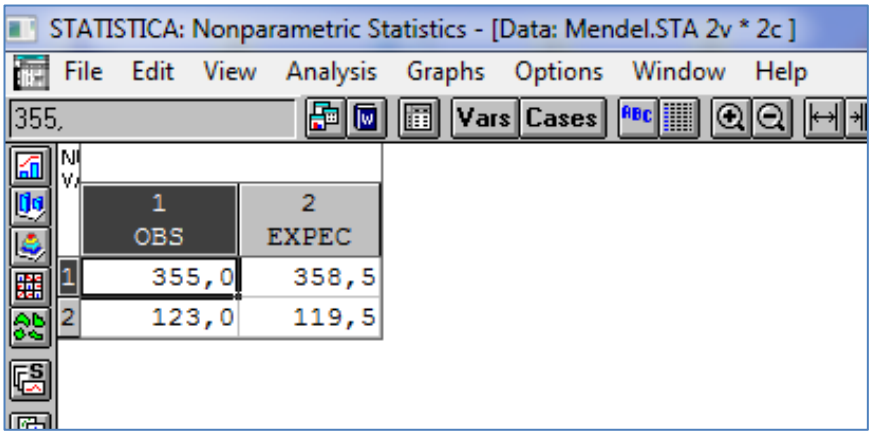


Рисунок 3.10 – Таблица с исходными данными

4 В появившемся диалоговом окне укажите переменные для наблюдаемых (Observed) и ожидаемых (Expected) частот (рисунок 3.11). Нажмите кнопку OK.

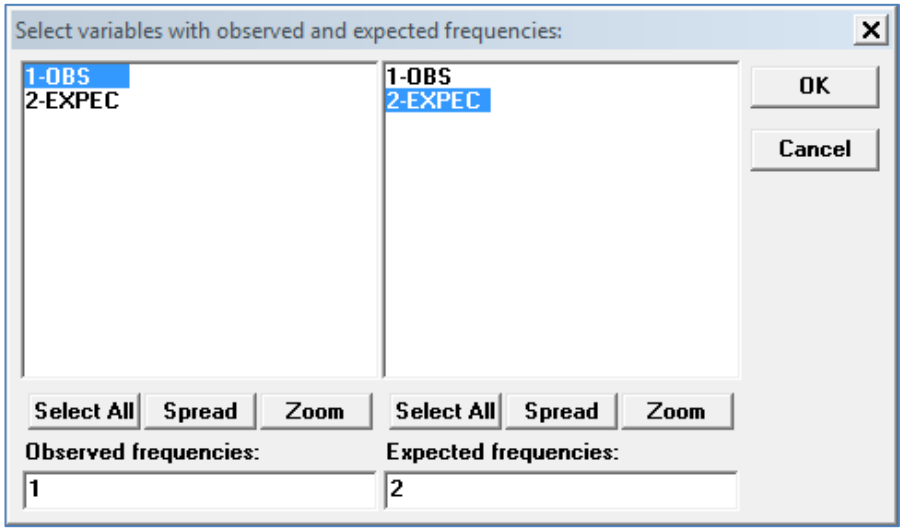


Рисунок 3.11 – Диалоговое окно для выбора переменных

В таблице результатов (рисунок 3.12) видим значение критерия Хи-квадрат (Chi-Square), равное **0,137**, число степеней свободы (df), равное единице и вероятность, соответствующая квантилю распределения Хи-квадрат, равная **0,71**.

Observed vs. Expected Frequencies (mendel.sta)

Continue... Chi-Square = ,1366806 df = 1 p < ,711606

Case	observed OBS	expected EXPEC	O - E	(O-E)**2 /E
C: 1	355,0000	358,5000	-3,50000	,034170
C: 2	123,0000	119,5000	3,50000	,102510
Sum	478,0000	478,0000	0,00000	,136681

Рисунок 3.12 – Результаты расчета критерия χ^2

Нулевая гипотеза H_0 (различия между наблюдаемыми и ожидаемыми частотами несущественны) не отвергается, когда значение фактического критерия $\chi^2_{\text{факт}}$ меньше значения теоретического распределения χ^2_{α} при данном числе степеней свободы df и уровне значимости α=0,05 (p=0,95)

$$\chi^2_{\text{факт}} < \chi^2_{05}.$$

Это условие выполняется, когда вычисленная вероятность, соответствующая квантилю распределения Хи-квадрат не превышает 0,95.

В нашем случае $\chi^2_{\text{факт}} = 0{,}137 < \chi^2_{05} = 3{,}84$ и p=0,71 < 0,95, следовательно, нулевая гипотеза не отвергается, т.е. различия между наблюдаемыми и ожидаемыми частотами несущественны.

Вывод. **Результаты опыта Менделя по скрещиванию двух сортов гороха соответствуют теоретическому ожидаемому отношению желтых семян к зеленым как 3:1.**

Упражнение 3

В этом упражнении вы научитесь проверять соответствие опытных данных теоретическому распределению при гибридном расщеплении во втором поколении (опыт Г. Менделя).

Установите, соответствует ли эмпирическое распределение частот (см. таблица 3.5) теоретически ожидаемому наследованию в соотношении **9:3:3:1** при гибридном расщеплении во втором поколении (опыт Г. Менделя).

1 Создайте новую таблицу, имеющую 2 столбца и 4 строки, и заполните ее исходными данными (см. таблица 3.5). Сохраните файл с именем **Mendel2.sta**.

2 Проверьте соответствие опытных данных теоретическому распределению при гибридном расщеплении во втором поколении с использованием модуля Nonparametric Statistics. Результат сравните с рисунком 3.13. Сделайте выводы.

Observed vs. Expected Frequencies (mendel2.sta)

Continue...

Chi-Square = ,5103070 df = 3 p < ,916620

Case	observed OBS	expected EXPEC	O - E	(O-E)**2 /E
C: 1	315,0000	313,0000	2,00000	,012780
C: 2	101,0000	104,0000	-3,00000	,086538
C: 3	108,0000	104,0000	4,00000	,153846
C: 4	32,0000	35,0000	-3,00000	,257143
Sum	556,0000	556,0000	0,00000	,510307

Рисунок 3.13 – Результаты расчета критерия χ^2

Таблица 3.5 – Результаты опытов Г. Менделя по изучению наследования признаков у гороха

Показатели	Семена			
	гладкие		морщинистые	
	желтые	зеленые	желтые	зеленые
Наблюдаемые частоты	315	101	108	32
Ожидаемое расщепление	9	3	3	1
Ожидаемые частоты	313	104	104	35

Упражнение 4

В этом упражнении вы научитесь проверять соответствие опытных данных теоретическому распределению с помощью визуальных методов.

С помощью визуальных методов определите, соответствуют ли нормальному закону распределения частоты переменных, представленных в таблице 3.6.

3 Создайте новую таблицу, имеющую 3 столбца и 50 строк, и заполните ее исходными данными (таблица 3.6). Сохраните файл с именем **V123.sta**.

Таблица 3.6 – Исходные данные для подбора теоретических распределений

V1	**V2**	**V3**
1	2	3
12,6	9,9	0,5
9,2	4,1	0,2
12,4	2,0	4,9
12,5	7,1	5,4
11,0	18,3	10,8
6,9	5,6	0,4
10,8	11,7	5,8
8,8	11,9	7,3
6,1	4,0	3,8
7,7	8,3	8,9
10,7	11,8	3,2

Окончание таблицы 3.6

1	2	3
9,9	15,2	1,4
14,5	8,4	19,9
9,7	9,5	2,6
8,9	9,1	0,5
13,4	3,4	2,1
8,5	3,9	6,4
10,1	8,3	3,7
8,6	4,7	0,4
8,0	9,8	6,9
7,4	8,2	17,3
11,5	13,1	16,1
8,0	8,2	2,0
9,3	4,2	3,8
8,0	7,3	3,4
10,5	9,9	0,1
7,6	8,5	25,1
9,8	9,6	0,8
8,4	4,3	12,1
6,5	3,9	2,7
9,3	13,3	0,7
14,0	2,9	7,8
14,3	6,9	9,7
8,0	6,1	27,2
9,6	11,5	6,0
11,4	23,9	9,1
10,5	10,1	6,2
9,1	6,4	11,1
9,2	5,9	0,7
9,4	8,2	1,4
7,9	8,1	3,3
7,7	14,1	11,2
8,6	2,0	11,8
11,2	9,3	3,9
9,1	5,2	4,5
8,4	1,4	6,3
10,9	4,9	1,4
14,0	15,7	14,1
9,2	11,5	1,8
12,7	12,6	1,8

4 Выполните команду Graphs – Stats 2D Graphs – Normal Probability Plots ... Настройте появившееся диалоговое окно как показано на рисунке 3.14. Нажмите OK.

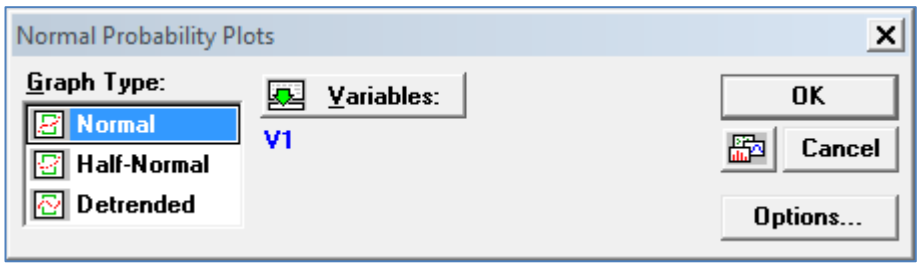


Рисунок 3.14 – Диалоговое окно Normal Probability Plots

5 На экране появится график распределения частот на нормальной вероятностной бумаге (рисунок 3.15).

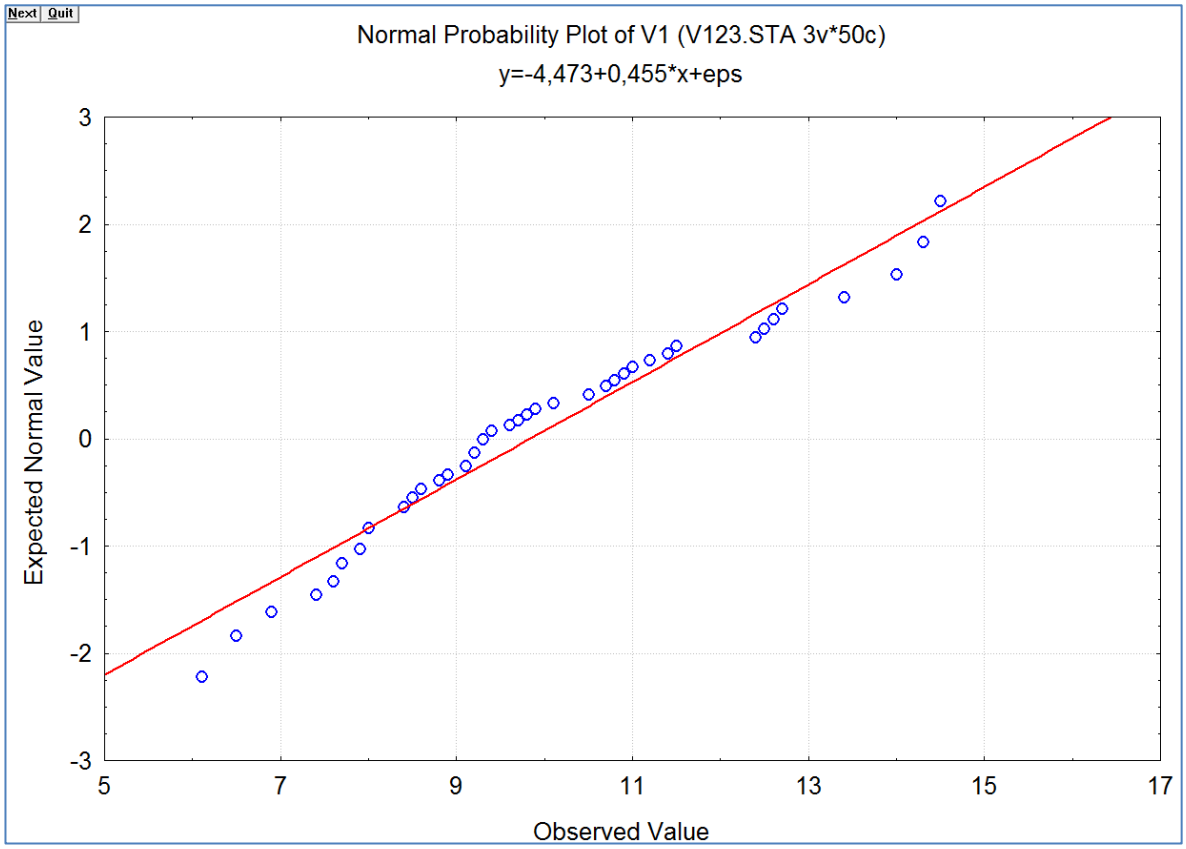


Рисунок 3.15 – График распределения частот
на нормальной вероятностной бумаге

Вывод. **Точки эмпирических данных близко располагаются к теоретической прямой, поэтому эмпирическое распределение вероятностей близко к нормальному.**

6 Определите, соответствуют ли нормальному закону распределения частоты переменных **V2** и **V3**. Сделайте выводы.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Проверьте правила двух и трех сигм при помощи вероятностного калькулятора. Для этого задайте параметры стандартного нормального распределения и в поле ввода квантиля X задайте 2 и 3 соответственно. Опция Two-tailed должна быть активна.

2. Вычислите значение функции плотности t-распределения для 100 степеней свободы и вероятности 95 %. Сравните это значение с табличным (Приложение Б).

3. Вычислите значение функции плотности распределения Фишера для 12 и 40 степеней свободы и вероятности 95 %. Сравните это значение с табличным (Приложение В).

1. Вычислите значение функции плотности χ^2-распределения для 25 степеней свободы и вероятности 95 %. Сравните это значение с табличным (Приложение Г).

2. В опытах Т.С. Фадеевой по изучению наследования признаков у земляники получено следующее количество растений, отличающихся окраской ягоды и формой чашечки (таблица 3.7). Определите по критерию χ^2 соответствие эмпирического расщепления ожидаемому теоретическому **(1:2:1)^2** или **1:2:2:4:1:2:1:2:1** (для второго поколения дигибридного скрещивания при неполном доминировании двух параллелей).

Таблица 3.7 – Результаты опытов Т.С. Фадеевой по изучению наследования признаков у земляники

Показатели	Форма чашечки								
	нормальная	промежуточная	листовидная	нормальная	промежуточная	листовидная	нормальная	промежуточная	листовидная
	красная ягода			розовая ягода			белая ягода		
Наблюдаемые частоты	25	33	17	45	85	30	20	42	10
Ожидаемое расщепление	1	2	1	2	4	2	1	2	1
Ожидаемые частоты	19,19	38,37	19,19	38,37	76,76	38,37	19,19	38,37	19,19

5 СОДЕРЖАНИЕ ОТЧЕТА

Отчет оформляется в соответствии с требованиями, принятыми в АЧИИ. Отчет должен содержать:

- ответы на контрольные вопросы;
- результаты выполнения самостоятельной работы.

6 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Каковы параметры нормального распределения?
2. Какое распределение называют стандартным нормальным?
3. Сформулируйте правила двух и трех сигм.
4. Каковы параметры распределения Хи-квадрат?
5. Каковы параметры распределения Стьюдента?
6. Каковы параметры распределения Фишера?
7. Как запустить вероятностный калькулятор?
8. Назовите свойства распределения Хи-квадрат?
9. Назовите свойства распределения Стьюдента?
10. Назовите свойства распределения Фишера?
11. Как изменяется график плотности нормального распределения при изменении среднего значения?
12. Как изменяется график плотности нормального распределения при изменении стандартного отклонения?
13. Как изменяются графики функций плотности распределений Стьюдента, Фишера, Хи-квадрат при изменении числа степеней свободы?
14. Что такое статистическая гипотеза?
15. Назовите этапы проверки статистических гипотез.
16. Что такое нулевая (основная) и альтернативная (конкурирующая) гипотезы?
17. Что такое уровень значимости? Какие уровни значимости Вы знаете?
18. Что такое критерий проверки гипотезы? Какие критерии Вы знаете?
19. Какие статистики могут использоваться в качестве критериев проверки гипотез?
20. Какие критерии называют параметрическими, непараметрическими?
21. Что такое критическая область? Какие виды расположения критической области Вы знаете?
22. Какие виды ошибок могут быть допущены при принятии решения на основе статистического критерия?
23. Что такое критерий согласия? Что такое мощность критерия?
24. Что представляет собой критерий согласия Хи-квадрат Пирсона?
25. Как подобрать теоретическое распределение, соответствующее эмпирическому (опытному) распределению частот?
26. Что такое нормальная и полунормальная вероятностная бумага?
27. Как визуально определить соответствие опытных данных нормальному закону распределения?

Практическое занятие № 4

Тема: **ОЦЕНКА РАЗЛИЧИЙ МЕЖДУ ДИСПЕРСИЯМИ ПО КРИТЕРИЮ ФИШЕРА. ОДНОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ. МНОГОФАКТОРНЫЙ ДИСПЕРСИОННЫЙ АНАЛИЗ**

1 ЦЕЛЬ РАБОТЫ

Изучить:

– назначение и области применения критерия Фишера.

Научиться:

– проводить однофакторный дисперсионный анализ в модулях Basic Statistics and Tables (Основные статистики и таблицы), ANOVA / MANOVA (Однофакторный / Многофакторный дисперсионный анализ);

– проводить многофакторный дисперсионный анализ в модуле ANOVA / MANOVA (Однофакторный / Многофакторный дисперсионный анализ).

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Оценка различий между дисперсиями по критерию Фишера. Дисперсионный анализ

Дисперсионный анализ – совокупность статистических методов обработки данных, позволяющих анализировать изменчивость одного или нескольких результативных признаков (откликов) под влиянием контролируемых факторов (независимых переменных), качественных или количественных.

Дисперсионный анализ используется для оценки влияния на отклик Y контролируемых переменных факторов $(X_1, X_2, \dots X_q)$. То есть, в результате этого анализа можно выяснить: какие из факторов $(X_1, X_2, \dots X_q)$ вносят существенные изменения в отклик Y на заданном уровне значимости, а какие существенно не влияют.

В *однофакторном дисперсионном анализе* фактор всего один, поэтому в результате такого анализа делают вывод, имеются ли значимые различия выходной величины Y для различных значений фактора X. Если изменения отклика Y обусловлены изменениями фактора X, то часто представляет интерес, какие значения фактора X вызывают значимые изменения отклика Y.

Например, если исследуется зависимость урожайности (отклик) от сорта (фактор), то необходимо выяснить какой из сортов (уровень фактора) значимо влияет на урожайность (отклик).

Для проведения дисперсионного анализа необходимо, чтобы для каждого уровня (значения) фактора X было произведено несколько измерений значений отклика Y (минимум три измерения).

Таким образом, массив данных для анализа имеет m наборов значений Y, соответствующих m уровням факторов, с k повторениями в каждом варианте (таблица 4.1). Общее количество элементов в массиве составляет $N = k \cdot m$.

Таблица 4.1 – Массив данных для дисперсионного анализа

y_{11}	y_{12}	…	y_{1m}
y_{21}	y_{22}	…	y_{2m}
…	…	…	…
y_{k1}	y_{k2}	…	y_{km}
$\bar{y}_1$	$\bar{y}_2$	…	$\bar{y}_m$

Дисперсионный анализ осуществляют в два этапа:

1. *Оценивают различия между вариантами по критерию Фишера*. На этом этапе делают заключение о наличии значимых различий средних значений выходной величины $\bar{y}_i$ для каждого из значений фактора X.

2. *Оценивают значимости разностей между средними*. Если на первом этапе выяснилось, что имеются значимые различия между средними $\bar{y}_i$, необходимо выяснить какие из вариантов значимо отличаются от других, то есть какие значения фактора X вызывают значимые изменения отклика Y. Для этого используют критерий *наименьшей существенной разницы (НСР)*.

Замечание. *Если значимые различия между средними значениями вариантов отсутствуют, то выполнять второй этап анализа не имеет смысла.*

2.2 Оценка различий между вариантами

Сущность дисперсионного анализа заключается в разделении общей дисперсии на части, соответствующие структуре эксперимента. При этом изменения выходной величины Y связаны с варьированием из-за различных значений фактора X и наличием ошибки опыта

$$SS_y = SS_v + SS_o, \tag{4.1}$$

где SS_y – сумма квадратов отклонений выходной величины;

SS_v – сумма квадратов отклонений вариантов при различных значениях фактора X;

SS_o – сумма квадратов отклонений, связанных с ошибкой опыта.

Сумма квадратов отклонений выходной величины SS_y вычисляется по формуле

$$SS_y = \sum_{i=1}^{N} (y_i - \bar{y})^2, \tag{4.2}$$

где y_i – i-й элемент массива данных;

$\bar{y}$ – среднее значение выходной величины (для всего массива данных).

Сумма квадратов отклонений вариантов SS_v определяется в соответствии с формулой

$$SS_v = \sum_{i=1}^{L} (\overline{y_i} - \bar{y})^2, \tag{4.3}$$

где $\overline{y_i}$ – среднее значение по i-му варианту, соответствующему i-му значению (уровню) фактора X.

Сумма квадратов отклонений, связанных с ошибкой опыта SS_o может быть найдена из выражения (3.1)

$$SS_o = SS_y - SS_v. \tag{4.4}$$

Дисперсия для каждого вида варьирования определяется из выражения

$$S^2 = \frac{SS}{df} \tag{4.5}$$

где df – число степеней свободы.

Общее число степеней свободы выходной величины df_y составляет $N - 1$ ($df_y = N - 1$), число степеней свободы вариантов $df_v = L - 1$, число степеней свободы ошибки $df_o = N - L$.

Дисперсионный анализ разработан и введен в практику сельскохозяйственных и биологических исследований английским ученым Р.А. Фишером, который открыл закон распределения отношения дисперсий

$$F = \frac{S_v^2}{S_o^2}, \tag{4.6}$$

где S_v^2 – дисперсия выборочных средних;

S_o^2 – дисперсия ошибки.

Критерий Фишера используется для поверки гипотезы о равенстве выборочных средних вариантов. Проверка гипотезы производится в соответствии с общепринятой схемой.

2.2.1 Этапы дисперсионного анализа по критерию Фишера

Этап 1. Формулируются нулевая H_0 и альтернативная H_1 гипотезы.

В качестве нулевой гипотезы H_0 выдвигается предположение о том, что все выборочные средние являются оценками одной генеральной средней и между вариантами нет существенных различий, т.е.

$$\bar{y}_1 = \bar{y}_2 = \cdots = \bar{y}_i. \tag{4.7}$$

Альтернативной гипотезой H_1 является предположение о том, что хотя бы одно среднее значение варианта отличается от других.

Этап 2. Задается уровень значимости α. В большинстве случаев при биологических исследованиях принимают 5% уровень значимости ($\alpha = 0{,}05$). При более строгом подходе принимают 1% или 0,1% ($\alpha = 0{,}01$ или $\alpha = 0{,}001$).

Этап 3. Выбирается статистика критерия для проверки нулевой гипотезы. Как уже отмечалось, для поверки гипотезы H_0 используется статистика F критерия Фишера.

Этап 4. Используя теоретическую функцию плотности распределения выбранной статистики F в предположении, что нулевая гипотеза H_0 верна, определяется граница критической области $F_{\text{кр}}$. Так как эта граница рассчитывается для теоретической функции F, то ее чаще обозначают F_α. Это значение либо вычисляется при помощи вероятностного калькулятора, либо принимается из таблицы (приложение В). Поэтому статистику F_α иногда в литературе называют табличным значением.

Значение F_α определяется исходя из количества степеней свободы вариантов df_v и ошибки df_o для принятого уровня значимости α.

Этап 5. По выборке вычисляется значение критерия Фишера F с использованием формулы (4.6). Предварительно для исследуемых данных рассчитываются дисперсия выборочных средних и дисперсия ошибки.

Этап 6. Полученное значение критерия Фишера F сравнивают с теоретическим значением F_α.

Если $F \geq F_\alpha$, то нулевая гипотеза отвергается, то есть хотя бы один из вариантов отличается от других вариантов на заданном уровне значимости.

При этом переходят ко второму этапу дисперсионного анализа – определяют между какими вариантами имеются существенные разности.

В противном случае нулевую гипотезу не отвергают и считают, что различия между любыми парами вариантов находятся в пределах ошибки опыта. Оценку частных различий не проводят.

2.2.2. Оценка значимости разностей между средними

Критерий Фишера устанавливает только факт присутствия значимых различий между вариантами, но не указывает между средними каких вариантов имеются различия. Чтобы установить, какие варианты значимо отличаются, проводят дополнительные тесты, которые называют *апостериорным анализом*.

Одним из способов оценки значимости разности между средними является определение *наименьшей существенной разницы (НСР)*. Часто рядом с НСР пишут уровень значимости для которого она рассчитана, например $НСР_{05}$ или $НСР_{01}$ (т.е. 5% или 1% уровень значимости соответственно).

Критерий НСР указывает предельную ошибку для разности двух выборочных средних δ. Если фактическая разность $\delta \geq$ НСР, то она существенная для выбранного уровня значимости.

Пример. Средние по вариантам имеют следующие значения $\bar{y}_1 = 10{,}1,\ \bar{y}_2 = 9{,}4,\ \bar{y}_3 = 9{,}8,$ наименьшая существенная разница на 5% уровне значимости $НСР_{05}$=0,6. Определить, между какими вариантами имеются значимые различия на уровне значимости 5%.

Значения $\delta_{12} = \bar{y}_1 - \bar{y}_2 = 10{,}1 - 9{,}4 = 0{,}7$ больше $НСР_{05}$ = 0,6, следовательно, $\bar{y}_1$ отличается от $\bar{y}_2$ на 5%-м уровне значимости. А $\delta_{13} = 0{,}3$ и $\delta_{23} = 0{,}4$ меньше $НСР_{05}$ = 0,6, следовательно, различия между средними $\bar{y}_1$ и $\bar{y}_3$, а также $\bar{y}_2$ и $\bar{y}_3$ не существенны для указанного уровня значимости, то есть находятся в пределах ошибки опыта.

Наименьшая существенная разница определяется как произведение значения критерия Стьюдента на ошибку разности средних

$$НСР = t \cdot S_d\ , \tag{4.8}$$

где t – значение критерия Стьюдента;
S_d – ошибка разности средних.

Значение t-критерия выбирают из таблицы (приложение А) для заданного уровня значимости α и количества степеней свободы $df_v = N - 1$, где N – количество элементов в массиве данных.

Ошибка разности средних вычисляется по формуле

$$S_d = \sqrt{\frac{2S_o^2}{n}}, \tag{4.9}$$

где S_o^2 – дисперсия ошибки (остатка);
n – число наблюдений (случаев) в каждом варианте.

2.3 Многофакторный дисперсионный анализ

Принципиальной разницы между многофакторным и однофакторным дисперсионным анализом нет. Многофакторный анализ не меняет общую логику дисперсионного анализа, а лишь несколько усложняет ее, поскольку, кроме учета влияния на зависимую переменную каждого из факторов по отдельности, следует оценивать и их совместное действие.

Таким образом, то новое, что вносит в анализ данных многофакторный дисперсионный анализ, касается в основном возможности оценить межфакторное взаимодействие.

Рассмотрим многофакторный анализ на примере двухфакторного. Двухфакторный дисперсионный анализ позволяет проверить эффекты влияния обоих факторов на зависимую переменную одновременно, а не по отдельности. Кроме этого, можно проверить гипотезу об эффекте взаимодействия между двумя независимыми переменными (рисунок 4.1).

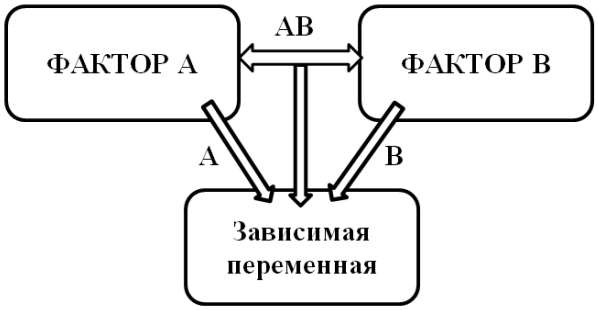


Рисунок 4.1 – Двухфакторный дисперсионный анализ

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

3.1 Подготовка исходных данных

Для проведения однофакторного дисперсионного анализа в программе Statistica исходные данные представляют в виде двух переменных.

Первая переменная содержит случаи с названиями вариантов – уровнями (значениями) фактора X. Например, **Опыт1**, **Опыт2**, **Контроль** и т.д. Эта переменная называется группирующей (Grouping) или независимой (Independent).

Вторая переменная содержит случаи со значениями исследуемой величины (отклика Y) в нескольких повторениях для каждого варианта. Эта переменная называется зависимой (Dependent). Пример исходных данных для однофакторного дисперсионного анализа двух вариантов с тремя повторениями приведен в таблице 4.2.

При двухфакторном дисперсионном анализе группирующих (Grouping) или независимых (Independent) переменных две. Третья переменная является зависимой (Dependent). Пример исходных данных для двухфакторного дисперсионного анализа приведен в таблице 4.3.

Дальнейшее увеличение факторов ведет к увеличению группирующих переменных. Таким образом, *таблица исходных данных для многофакторного дисперсионного анализа содержит группирующие переменные* $(X_1, X_2, \dots)$, *количество которых равно числу факторов опыта, и одну независимую переменную – отклик* (Y).

Таблица 4.2 – Пример таблицы исходных данных для однофакторного дисперсионного анализа

X	Y
Опыт1	10
Опыт1	12
Опыт1	9
Опыт2	15
Опыт2	18
Опыт2	14

Таблица 4.3 – Пример таблицы исходных данных для двухфакторного дисперсионного анализа

X_1	X_2	Y
1	2	3
Вариант1	Опыт1	10
Вариант1	Опыт1	12
Вариант1	Опыт1	9
Вариант1	Опыт2	15
Вариант1	Опыт2	18
Вариант1	Опыт2	16
Вариант2	Опыт1	12
Вариант2	Опыт1	8
Вариант2	Опыт1	11
Вариант2	Опыт2	14
Вариант2	Опыт2	12
Вариант2	Опыт2	11

3.2 Однофакторный дисперсионный анализ в модуле Basic Statistics and Tables

Для начала дисперсионного анализа в модуле Basic Statistics and Tables выбирают опцию Breakdown & one-way ANOVA на стартовой панели модуля или в меню Analysis. При этом открывается окно Descriptive Statistics and Correlations by Groups (рисунок 4.2), в котором необходимо задать группирующую (Grouping) и зависимую (Dependent) переменные. Выбор группирующей (Grouping) и зависимой (Dependent) переменных производится при помощи кнопки Variables.

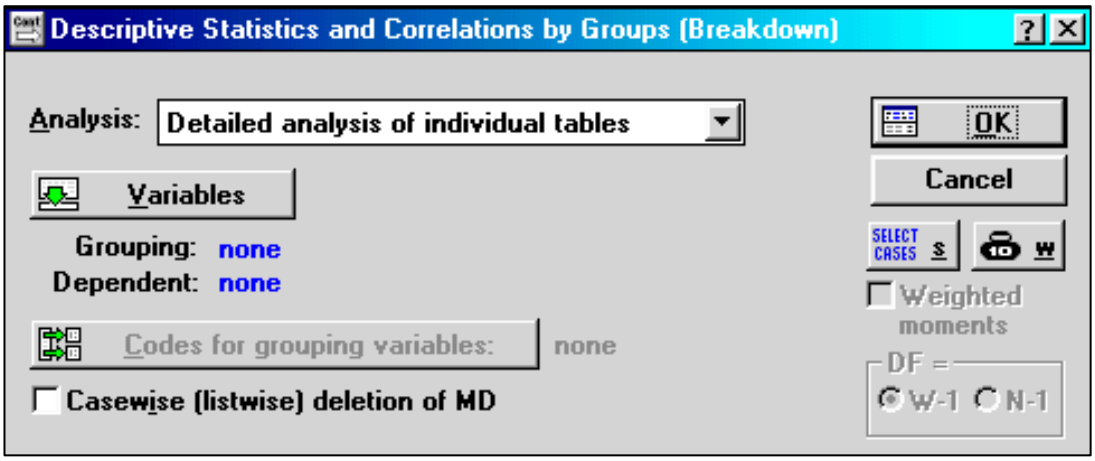


Рисунок 4.2 – Окно Descriptive Statistics and Correlations by Groups

После выбора переменных и нажатия кнопки OK появляется окно Descriptive Statistics and Correlations by Groups – Results (рисунок 4.3), в верхней части которого отображаются основные результаты анализа. Более подробную информацию можно получить, выбирая опции в нижней части окна. Назначение опций представлено в таблице 4.4.

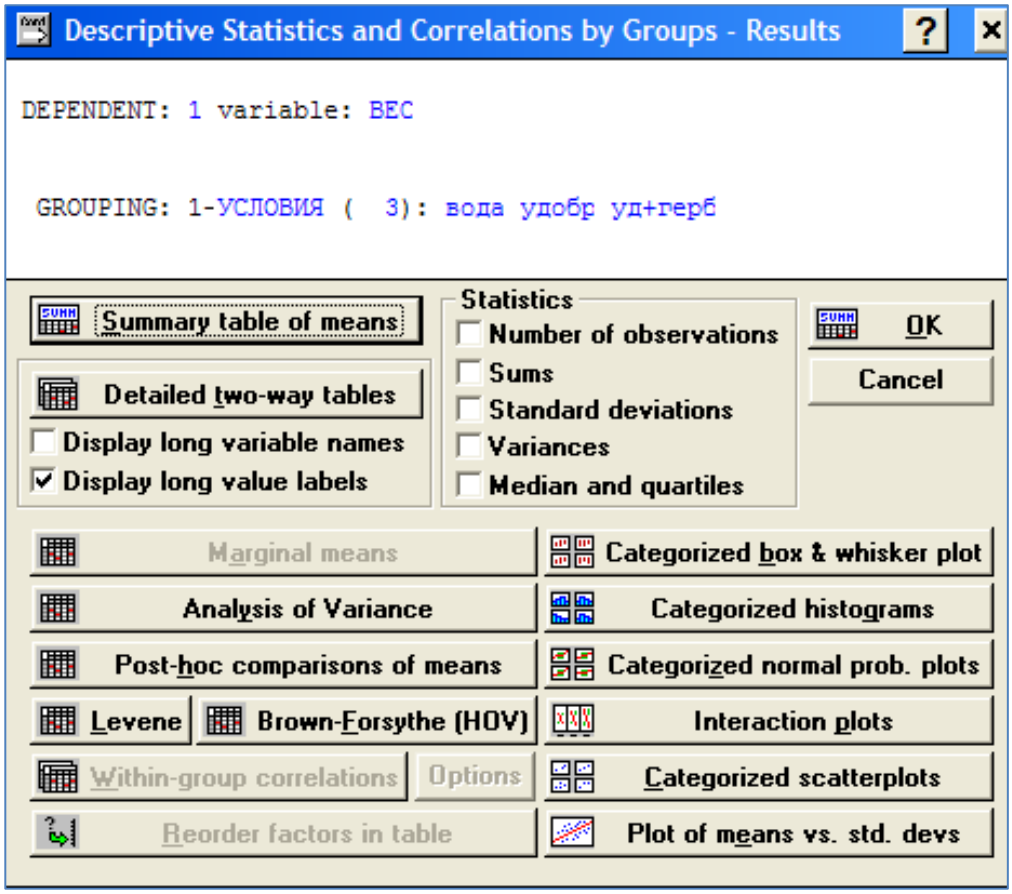


Рисунок 4.3 – Окно Descriptive Statistics and Correlations by Groups - Results

Таблица 4.4 – Назначение опций диалогового окна Descriptive Statistics and Correlations by Groups - Results

Элемент окна	Назначение элемента окна
Summary table of means (Итоговая таблица средних)	Расчет средних для вариантов и всей выборки. Дополнительно можно рассчитать статистики, отмеченные в группе опций Statistics (Статистики)
Detailed two-way tables (Детальные таблицы)	Расчет средних для вариантов и всей выборки
Display long variable names (Отображать длинные имена переменных)	Отображение в таблицах длинных имен переменных
Display long value labels (Отображать длинные метки значений)	Отображение в таблицах длинных меток значений
Statistics (Статистики)	Расчет выбранных статистик, отображаемых при нажатии кнопки Summary table of means (Суммарная таблица средних)
Number of observations (Число наблюдений)	Расчет числа наблюдений
Sums (Суммы)	Расчет суммы
Standard deviations (Стандартные отклонения)	Расчет стандартных отклонений
Variances (Дисперсии)	Расчет дисперсии
Median and quartiles (Медиана и квартили)	Расчет медианы и квартилей
Analysis of Variance (Дисперсионный анализ)	Вывод результатов дисперсионного анализа
Post-hoc comparisons of means (Апостериорные сравнения средних)	Проведение апостериорного анализа

После нажатия на кнопку Analysis of Variance появляется таблица результатов дисперсионного анализа (рисунок 4.4).

Analysis of Variance (tomat.sta)

BASIC STATS — Marked effects are significant at p < ,05000

Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
BEC	,626944	2	,313472	3,912083	15	,260806	1,201938	,327974

Рисунок 4.4 – Окно результатов дисперсионного анализа Analysis of Variance

В таблице результатов приводятся следующие статистические характеристики (таблица 4.5).

Таблица 4.5 – Обозначения, принятые в таблице результатов дисперсионного анализа

Обозначение	Пояснения
SS (Sum of Square) Effect	Сумма квадратов отклонений эффекта (вариантов)
df (Degree of freedom) Effect	Степени свободы эффекта
MS (Mean Square) Effect	Средний квадрат (дисперсия) эффекта
SS Error	Сумма квадратов отклонений ошибки (остатков)
df Error	Степени свободы ошибки
MS Error	Средний квадрат отклонений (дисперсия) ошибки
F	Критерий Фишера
p	Вероятность, соответствующая вычисленной квантили F распределения Фишера

Рассчитанное значение критерия Фишера F_α сравнивают с табличным значением F для уровня значимости $\alpha = 0{,}05$ и степеней свободы эффекта (df Effect) и ошибки (df Error). ***Если $F \geq F_\alpha$, то нулевая гипотеза отвергается, то есть хотя бы один из вариантов отличается от других вариантов на заданном уровне значимости.***

Сделать вывод о значимости различий средних можно без определения F_α. В этом случае используют рассчитанную вероятность p для квантили критерия Фишера F. Эту вероятность сравнивают с принятым уровнем значимости α. Если $p < \alpha$, то нулевая гипотеза H_0 отвергается и принимаются альтернативная гипотеза H_1. При этом строка с вероятностью $p < 0{,}05$ выделяется красным цветом, о чем свидетельствует сообщение в заголовке таблицы Marked effects are significant at p < ,05 (Выделенные эффекты соответствуют вероятности ниже 0,05).

Если в результате проверки по критерию Фишера выяснилось, что между средними значениями вариантов имеются значимые различия, то проводят апостериорный анализ. Выяснить, какие из вариантов имеют значимые различия средних, позволяет тест НСР.

Для проведения теста НСР нажимают кнопку Post-hoc comparisons of means окна Окно Descriptive Statistics and Correlations by Groups – Results (см. рисунок 4.3). При этом появляется окно апостериорного анализа (см. рисунок 4.5). В этом окне выбирают опцию LSD test planned comparison. Результаты теста получают в виде таблицы (см. рисунок 4.6).

В данной таблице именами строк являются названия и номера всех вариантов выборочной совокупности, а именами столбцов – только номера вариантов. В ячейках приводятся значения вероятности p.

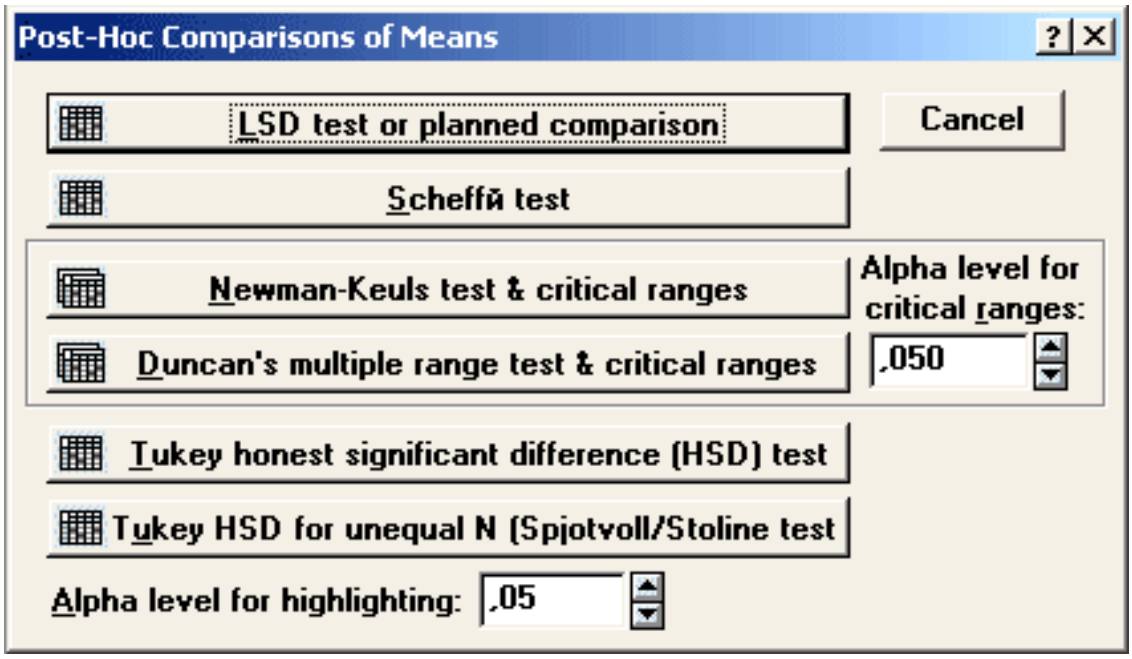


Рисунок 4.5 – Окно Post-hoc comparisons of means

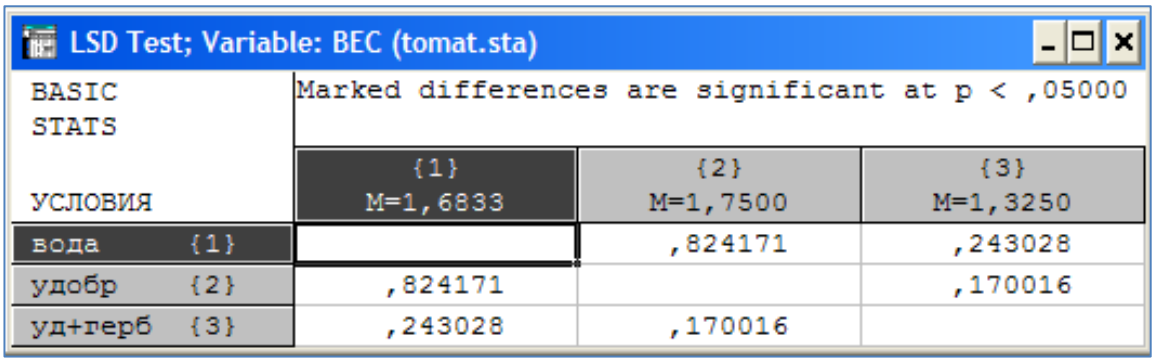


Рисунок 4.6 – Результаты теста НСР

Если значение вероятности p в ячейке менее 0,05, то имеются значимые различия между вариантом, название которого записано в заголовке строки, и вариантом, номер которого записан в заголовке столбца таблицы.

При этом ячейки, имеющие вероятность $p < 0{,}05$, выделяются красным цветом. Об этом свидетельствует надпись в заголовке таблицы: Marked differences are significant at p < ,05000 – Выделенные различия имеют значимость на уровне p < 0,05000.

При помощи теста НСР (LSD test) программы Statistica определяется только уровень вероятности различий между средними значениями. Значение же наименьшей существенной разницы (НСР) не отображается. Его можно рассчитать, используя формулы 4.8 и 4.9. Значение S_0^2 (дисперсия ошибки) берется из таблицы (см. рисунок 4.4) – MS Error.

Упражнение 1

В этом упражнении вы научитесь выполнять дисперсионный анализ в модуле Basic Statistics and Tables.

Задача. Выяснить, имеются ли значимые различия на уровне значимости 0,05 между вариантами в опыте по определению энергии прорастания семян яровой пшеницы после физического воздействия. Рассчитать значения $НСР_{05}$.

Опыт проводился в трех повторениях. Данные, полученные в опыте, приведены в таблице 4.6. Вариант 1 – необработанные семена, вариант 2 – тепловое воздействие, вариант 3 – воздействие полем сверхвысокой частоты (СВЧ).

Таблица 4.6 – Данные опыта по определению энергии прорастания семян яровой пшеницы после физического воздействия

Вариант 1	Вариант 2	Вариант 3
91	92	94
89	94	95
90	91	94

1. *В качестве нулевой гипотезы H_0 сделаем предположение о том, что все выборочные средние являются оценками одной генеральной средней и между вариантами нет существенных различий. Альтернативной гипотезой H_1 является предположение о том, что хотя бы одно среднее значение варианта отличается от других,* **т.е., в данном примере необходимо проверить нулевую гипотезу утверждающую, что фактор – физическая обработка семян ярового ячменя – не влияет на вариацию зависимой переменной – энергию прорастания семян ярового ячменя, т.е. вся вариация сводится к случайной.**
2. Задайте уровень значимости α. В большинстве случаев принимают 5% уровень значимости ($\alpha = 0,05$).
3. Откройте программу Statistica.
4. В модуле Basic Statistics and Tables создайте новую таблицу, содержащую 2 столбца и 9 строк. Заполните таблицу данными как показано на рисунке 4.7. Сохраните файл с именем **Energia_Prorastania.sta**.

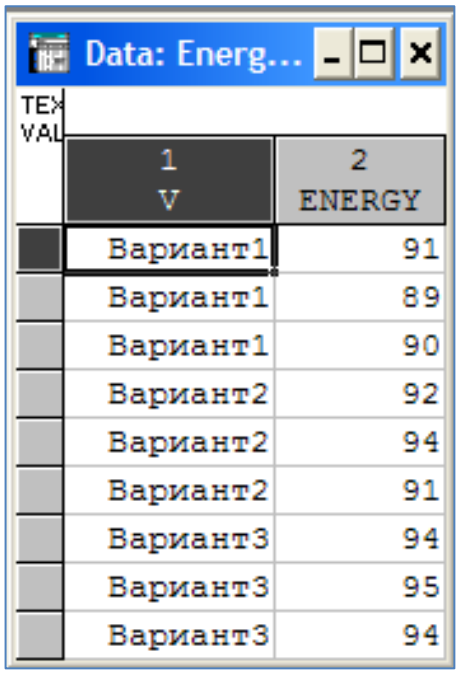


Рисунок 4.7 – Таблица исходных данных для дисперсионного анализа данных опыта по определению энергии прорастания семян яровой пшеницы

5. Выполните команду Analysis – Breakdown & one-way ANOVA. В появившемся диалоговом окне щелкните по кнопке Variables и выберите группирующую (Grouping) и зависимую (Dependent) переменные (рисунок 4.8). Нажмите OK. Нажмите OK еще раз.

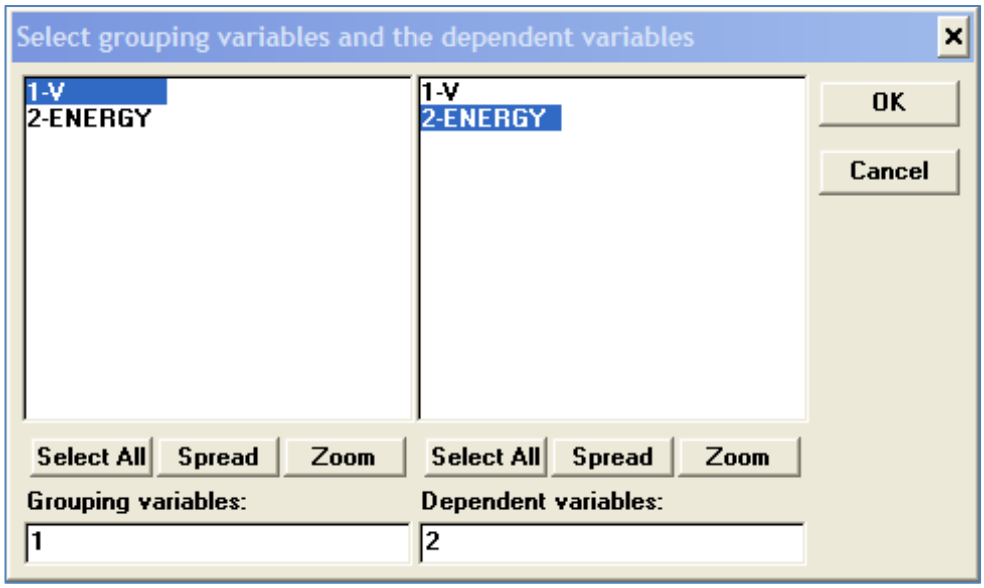
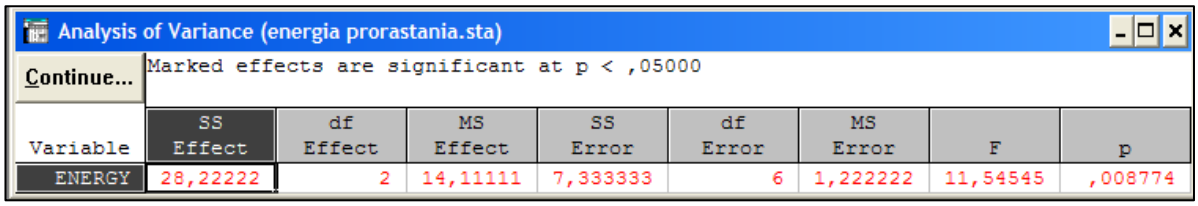


Рисунок 4.8 – Выбор переменных для дисперсионного анализа

6. В появившемся диалоговом окне Descriptive Statistics and Correlations by Groups – Results нажмите кнопку Analysis of Variance для проведения дисперсионного анализа. На экране появится таблица с результатом (рисунок 4.9).


Analysis of Variance (energia prorastania.sta)

Continue...

Marked effects are significant at p < ,05000

Variable	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
ENERGY	28,22222	2	14,11111	7,333333	6	1,222222	11,54545	,008774

Рисунок 4.9 – Окно с результатами дисперсионного анализа

По результатам дисперсионного анализа (рисунок 4.9) сделаем следующие **выводы:**

1) Критерий Фишера, полученный в результате дисперсионного анализа, $F = 11{,}54545$ (рисунок 4.9).

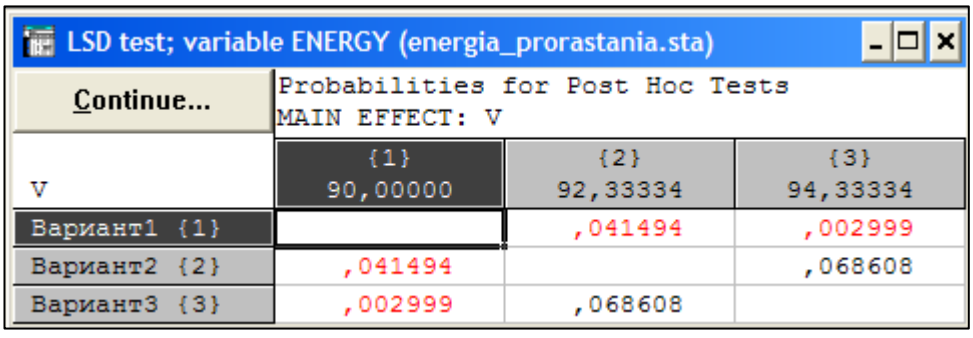
2) Теоретическую функцию F_α для степеней свободы $df_v = 2$ (df Effect, рисунок 4.9) и $df_0 = 6$ (df Error, рисунок 4.9) определяем по таблице при $\alpha = 0{,}05$ (приложение В). Получим: $F_\alpha(2{,}6) = 5{,}14$.

3) Сравниваем критерии Фишера: $F > F_\alpha$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: хотя бы один из вариантов отличается от других вариантов на заданном уровне значимости.***

4) Поскольку принята альтернативная гипотеза, то необходимо сделать апостериорный анализ (тест HCP), чтобы узнать, между какими вариантами есть существенные различия.

7. Сохраните результаты дисперсионного анализа в файл с именем **Disp_1.scr**.

8. В окне с результатами дисперсионного анализа нажмите кнопку Continue. В появившемся окне нажмите кнопку Post-hoc comparisons of means. В новом окне нажмите кнопку LSD test planned comparison. На экране появится окно с результатами апостериорного анализа (рисунок 4.10).


LSD test; variable ENERGY (energia_prorastania.sta)

Continue...

Probabilities for Post Hoc Tests
MAIN EFFECT: V

V	{1} 90,00000	{2} 92,33334	{3} 94,33334
Вариант1 {1}		,041494	,002999
Вариант2 {2}	,041494		,068608
Вариант3 {3}	,002999	,068608	

Рисунок 4.10 – Результаты теста HCP

По результатам апостериорного анализа (рисунок 4.10) сделаем следующие **выводы:**

1) Из таблицы результатов теста HCP видно, что значимые различия на уровне значимости 0,05 имеются между Вариантом 1 и Вариантом 2 (p=0,041<0,05), а также между Вариантом 1 и Вариантом 3 (p=0,0029<0,05). Различия между Вариантом 2 и Вариантом 3 не значимы (p=0,0686>0,05).

Таким образом, результаты опыта показывают, что физические воздействия как тепловое, так и СВЧ повышают энергию прорастания семян, хотя значимых различий между обработкой тепловым полем по сравнению с полем СВЧ не наблюдается.

2) Рассчитаем значение $НСР_{05}$ по формуле 4.8, предварительно вычислив значение ошибки разности средних по формуле 4.9, где $n = 3$ (количество повторений опыта), $S_o^2 = 1{,}22$ (значение MS Error в окне на рисунке 4.9) и критерий Стьюдента для степеней свободы $df = 6$ (это значение df Error в окне на рисунке 4.9) t(6) = 2,45.

$$S_d = \sqrt{\frac{2 \cdot 1{,}22}{3}} = 0{,}9,$$

$$НСР = 2{,}45 \cdot 0{,}9 = 2{,}205.$$

3) Сравним разницы средних значений по вариантам со значением НСР:

$$\delta_{21} = \bar{y}_2 - \bar{y}_1 = 92{,}33 - 90 = 2{,}33 > 2{,}205,$$
$$\delta_{31} = \bar{y}_3 - \bar{y}_1 = 94{,}33 - 90 = 4{,}33 > 2{,}205,$$
$$\delta_{32} = \bar{y}_3 - \bar{y}_2 = 94{,}33 - 92{,}33 = 2 < 2{,}205.$$

Видим, что средние значения вариантов 1 и 2, 1 и 3 имеют значимые различия, а разница средних значений вариантов 2 и 3 находится в пределах ошибки опыта для принятого уровня значимости 0,05.

9. Сохраните результаты апостериорного анализа в файл с именем **LSD.scr**. Закройте все окна дисперсионного анализа.

3.3 Однофакторный дисперсионный анализ в модуле ANOVA/MANOVA

Кроме модуля Basic Statistics and Tables дисперсионный анализ можно проводить в специальном модуле, который называется ANOVA/MANOVA.

Упражнение 2

В этом упражнении вы научитесь выполнять дисперсионный анализ в модуле ANOVA/MANOVA.

Для проведения дисперсионного анализа воспользуемся данными из файла **Energia_Prorastania.sta** (см. рисунок 4.7).

1. Переключите модуль Basic Statistics and Tables на модуль ANOVA/MANOVA, используя переключатель модулей.

2. В диалоговом окне General ANOVA/MANOVA укажите группирующую (или независимую) Independent и зависимую (Dependent) переменные (Variables) (рисунок 4.11). Нажмите OK.

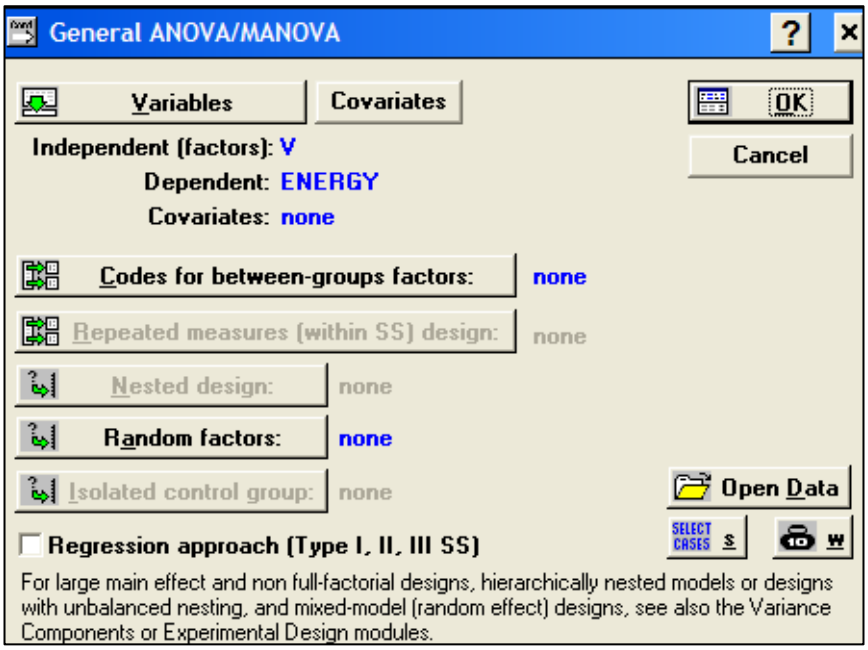


Рисунок 4.11 – Основное окно модуля ANOVA/MANOVA

3. На экране появится окно с результатами дисперсионного анализа (рисунок 4.12). Для просмотра результатов нажмите кнопку All effects.

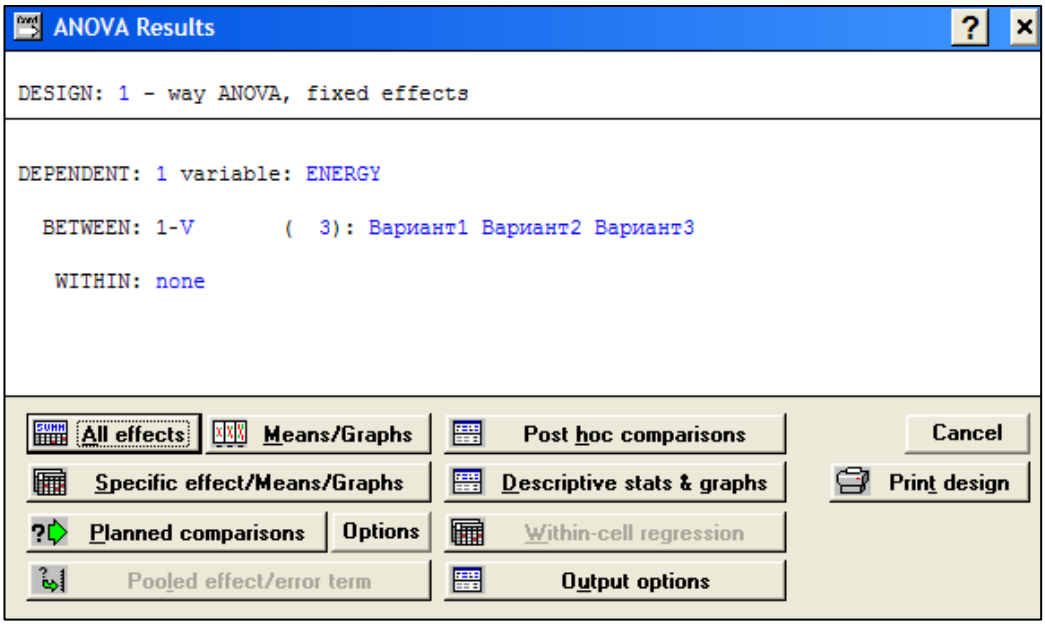


Рисунок 4.12 – Окно с результатами дисперсионного анализа модуля ANOVA/MANOVA

4. На экране появится окно с таблицей результатов (см. рисунок 4.13).

5. Сохраните результаты дисперсионного анализа в файл с именем **Disp_2.scr**.

6. Для вычисления средних значений вариантов, в окне с таблицей результатов нажмите кнопку Continue. В появившемся окне нажмите кнопку Specific effect/Means/Graphs, либо кнопку Mean/Graphs и установите переключатель ScrollSheet в разделе Display.

Summary of all Effects; design: (energia_prorastania.sta)

Continue... 1-V

Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	2	14,11111	6	1,222222	11,54545	,008774

Рисунок 4.13 – Таблица результатов дисперсионного анализа

Summary of all effects

7. Для проведения теста НСР в окне ANOVA Results нажмите кнопку Post hoc comparisons (рисунок 4.12), при этом необходимо выделить переменную в поле Factors (рисунок 4.14). Нажмите OK. В появившемся диалоговом окне нажмите кнопку LSD test on planned comparison.

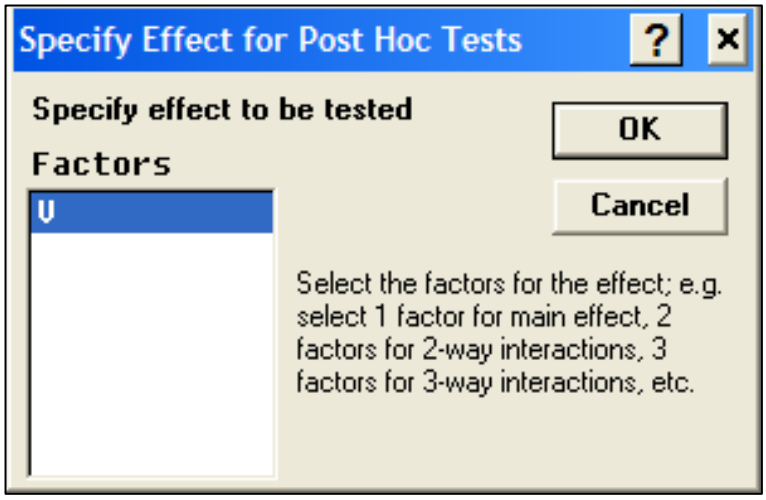


Рисунок 4.14 – Окно выбора переменной для теста НСР

LSD test; variable ENERGY (energia_prorastania.sta)

Continue... Probabilities for Post Hoc Tests
MAIN EFFECT: V

V	{1} 90,00000	{2} 92,33334	{3} 94,33334
Вариант1 {1}		,041494	,002999
Вариант2 {2}	,041494		,068608
Вариант3 {3}	,002999	,068608	

Рисунок 4.15 – Результаты теста НСР

По результатам дисперсионного анализа (рисунок 4.13) сделаем следующие **выводы:**

1) Критерий Фишера, полученный в результате дисперсионного анализа будет равен: $F = 11{,}54545$ (рисунок 4.13).

2) Теоретическую функцию F_α для степеней свободы $df_v = 2$ (df Effect, рисунок 4.13) и $df_0 = 6$ (df Error, рисунок 4.13) определяем по таблице при $\alpha = 0{,}05$ (приложение В). Получим: $F_\alpha(2{,}6) = 5{,}14$.

3) Сравниваем критерии Фишера: $F > F_\alpha$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: хотя бы один из вариантов отличается от других вариантов на заданном уровне значимости.***

4) Поскольку принята альтернативная гипотеза, то необходимо сделать апостериорный анализ (тест НСР), чтобы узнать, между какими вариантами есть существенные различия.

По результатам апостериорного анализа (см. рисунок 4.15) сделаем следующие **выводы:**

1) Из таблицы результатов теста НСР видно, что значимые различия на уровне значимости 0,05 имеются между Вариантом 1 и Вариантом 2 (p=0,041<0,05), а также между Вариантом 1 и Вариантом 3 (p=0,0029<0,05). Различия между Вариантом 2 и Вариантом 3 не значимы (p=0,0686>0,05). Таким образом, результаты опыта показывают, что физические воздействия как тепловое, так и СВЧ повышают энергию прорастания семян, хотя значимых различий между обработкой тепловым полем по сравнению с полем СВЧ не наблюдается.

2) Рассчитаем значение $НСР_{05}$ по формуле 4.8, предварительно вычислив значение ошибки разности средних по формуле 4.9, где $n = 3$ (количество повторений опыта), $S_o^2 = 1{,}22$ (значение MS Error в окне на рисунке 4.13) и критерий Стьюдента для степеней свободы $df = 6$ (это значение df Error в окне на рисунке 4.13) t(6) = 2,45.

$$S_d = \sqrt{\frac{2 \cdot 1{,}22}{3}} = 0{,}9,$$

$$НСР = 2{,}45 \cdot 0{,}9 = 2{,}205.$$

3) Сравним разницы средних значений по вариантам со значением НСР:

$$\delta_{21} = \bar{y}_2 - \bar{y}_1 = 92{,}33 - 90 = 2{,}33 > 2{,}205,$$
$$\delta_{31} = \bar{y}_3 - \bar{y}_1 = 94{,}33 - 90 = 4{,}33 > 2{,}205,$$
$$\delta_{32} = \bar{y}_3 - \bar{y}_2 = 94{,}33 - 92{,}33 = 2 < 2{,}205.$$

Видим, что средние значения вариантов 1 и 2, 1 и 3 имеют значимые различия, а разница средних значений вариантов 2 и 3 находится в пределах ошибки опыта для принятого уровня значимости 0,05.

Замечание. *Обратите внимание, что независимо от выбора модуля для проведения дисперсионного анализа, результаты дисперсионного анализа совпадают!*

3.4 Подготовка исходных данных для многофакторного анализа

При двухфакторном дисперсионном анализе группирующих (Grouping) или независимых (Independent) переменных две. Третья переменная является зависимой (Dependent). Пример исходных данных для двухфакторного дисперсионного анализа приведен в таблице 4.7.

Дальнейшее увеличение факторов ведет к увеличению группирующих переменных. Таким образом, *таблица исходных данных для многофакторного дисперсионного анализа содержит группирующие переменные* $(X_1, X_2, \dots)$, *количество которых равно числу факторов опыта, и одну независимую переменную – отклик* (Y).

Таблица 4.7 – Пример таблицы исходных данных для двухфакторного дисперсионного анализа

X_1	X_2	Y
Вариант1	Опыт1	10
Вариант1	Опыт1	12
Вариант1	Опыт1	9
Вариант1	Опыт2	15
Вариант1	Опыт2	18
Вариант1	Опыт2	16
Вариант2	Опыт1	12
Вариант2	Опыт1	8
Вариант2	Опыт1	11
Вариант2	Опыт2	14
Вариант2	Опыт2	12
Вариант2	Опыт2	11

3.5 Многофакторный дисперсионный анализ в модуле ANOVA /MANOVA

Многофакторный дисперсионный анализ в программе Statistica выполняется только в модуле ANOVA/MANOVA.

Упражнение 3

В этом упражнении вы научитесь выполнять двухфакторный дисперсионный анализ в модуле ANOVA/ MANOVA.

Выполнить двухфакторный дисперсионный анализ данных, приведенных в таблице 4.7.

1. Откройте программу Statistica.
2. Переключите модуль Basic Statistics and Tables на модуль ANOVA/MANOVA, используя переключатель модулей.
3. Создайте новый файл с именем **Disp_3.sta**. Подготовьте данные в соответствии с рисунком 4.16.

Data: Primer1.STA 3v * 1...

	1 X1	2 X2	3 Y
1	Вариант1	Опыт1	10
2	Вариант1	Опыт1	12
3	Вариант1	Опыт1	9
4	Вариант1	Опыт2	15
5	Вариант1	Опыт2	18
6	Вариант1	Опыт2	16
7	Вариант2	Опыт1	12
8	Вариант2	Опыт1	8
9	Вариант2	Опыт1	11
10	Вариант2	Опыт2	14
11	Вариант2	Опыт2	12
12	Вариант2	Опыт2	11

Рисунок 4.16 – Таблица исходных данных для двухфакторного дисперсионного анализа

4. Вызовите стартовую панель модуля, где в качестве независимых (Independent) переменных укажите **X1**, **X2**, а в качестве зависимой (Dependent) – **Y** (рисунок 4.17). Нажмите OK.

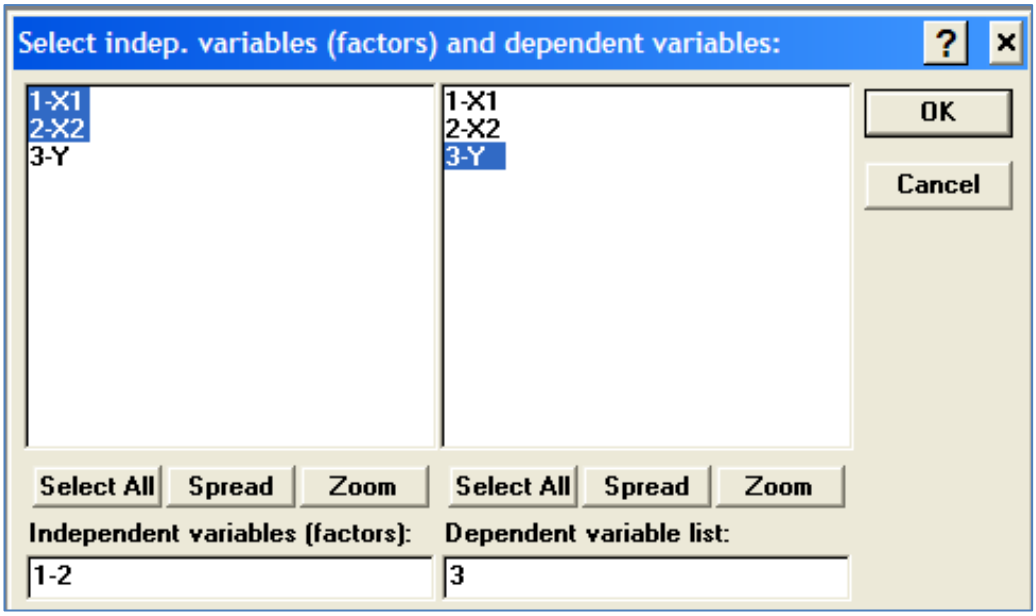


Рисунок 4.17 – Окно для выбора зависимой и независимой переменных

5. На экране появится окно с результатами дисперсионного анализа (см. рисунок 4.18). Для просмотра результатов нажмите кнопку All effects.

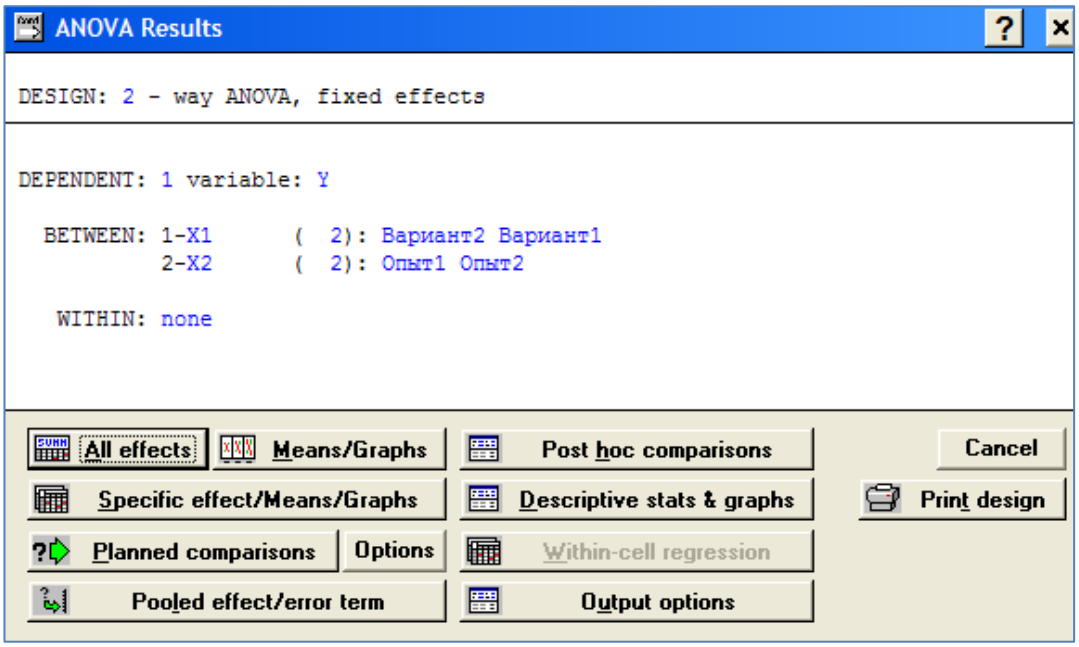
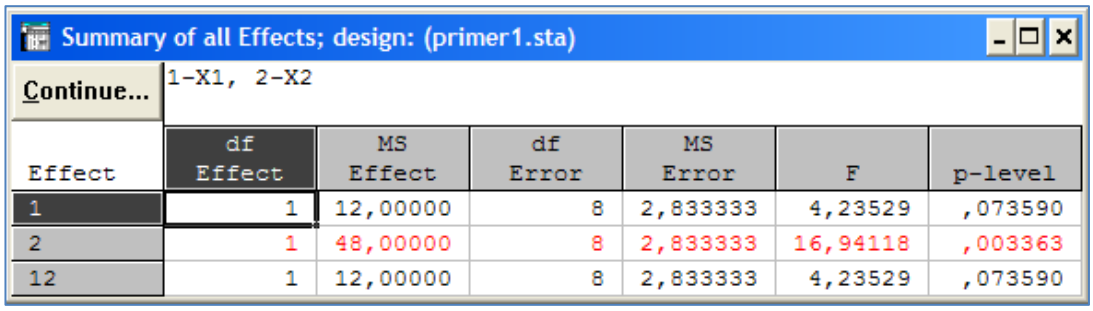


Рисунок 4.18 – Окно с результатами дисперсионного анализа модуля ANOVA/MANOVA

6. На экране появится окно с таблицей результатов (рисунок 4.19).



Effect	df Effect	MS Effect	df Error	MS Error	F	p-level
1	1	12,00000	8	2,833333	4,23529	,073590
2	1	48,00000	8	2,833333	16,94118	,003363
12	1	12,00000	8	2,833333	4,23529	,073590

Рисунок 4.19 – Таблица результатов двухфакторного дисперсионного анализа Summary of all effects

7. Сохраните результаты дисперсионного анализа в файл с именем **Disp_3.scr**.

8. Для проведения теста НСР в окне ANOVA Results нажмите кнопку Post hoc comparisons (рисунок 4.18), при этом необходимо выделить переменные в поле Factors (см. рисунок 4.20). Нажмите OK. В появившемся диалоговом окне нажмите кнопку LSD test on planned comparison.

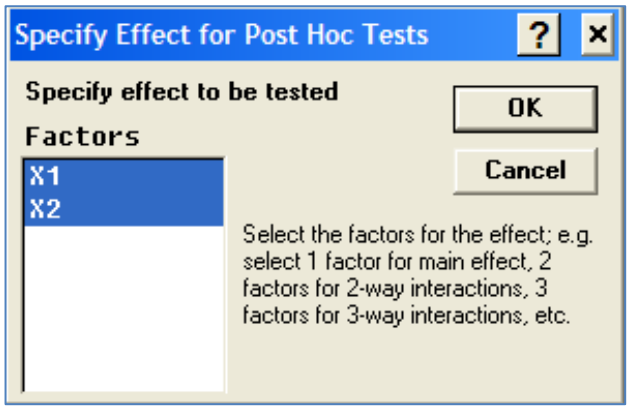


Рисунок 4.20 – Окно выбора переменной для теста HCP

Многофакторный дисперсионный анализ выполняется по тем же пунктам и правилам, которые указаны для однофакторного дисперсионного анализа.

По результатам дисперсионного анализа (см. рисунок 4.19) сделаем следующие **выводы**:

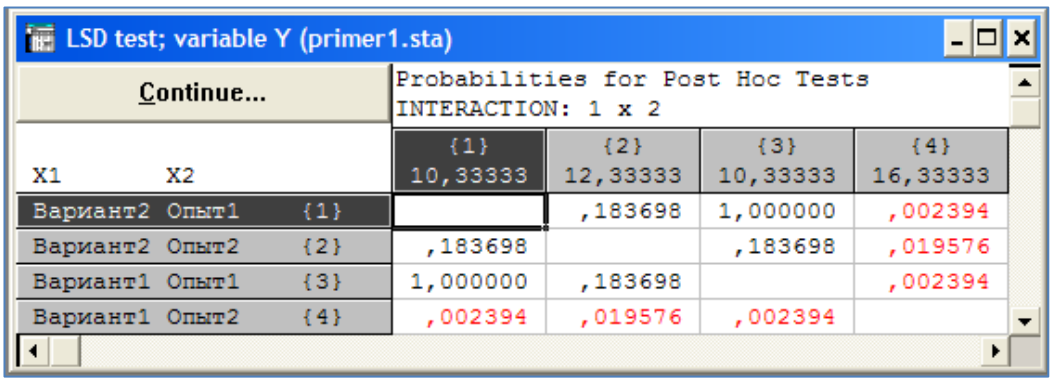
1) Из таблицы результатов видно, что в изменения значений наблюдаемой величины **Y** (отклика) вносит эффект только второй фактор – **X2**, так как p = 0,003363 < 0,05. Первый фактор **X1** значимого влияния (на уровне значимости 0,05) на отклик не оказывает.

2) Критерий Фишера для значимого фактора, полученный в результате дисперсионного анализа, $F = 16{,}94118$ (рисунок 4.19).

3) Теоретическую функцию F_α для степеней свободы $df_v = 1$ (df Effect, рисунок 4.19) и $df_o = 8$ (df Error, рисунок 4.19) определяем по таблице при $\alpha = 0{,}05$ (приложение В). Получим: $F_\alpha(1, 8) = 5{,}32$.

4) Сравниваем критерии Фишера: $F > F_\alpha$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: хотя бы один из вариантов отличается от других вариантов на заданном уровне значимости.***

5) Поскольку принята альтернативная гипотеза, то необходимо сделать апостериорный анализ (тест HCP), чтобы узнать, между какими вариантами есть существенные различия.



X1	X2		{1} 10,33333	{2} 12,33333	{3} 10,33333	{4} 16,33333
Вариант2	Опыт1	{1}		,183698	1,000000	,002394
Вариант2	Опыт2	{2}	,183698		,183698	,019576
Вариант1	Опыт1	{3}	1,000000	,183698		,002394
Вариант1	Опыт2	{4}	,002394	,019576	,002394	

Рисунок 4.21 – Результаты теста HCP

По результатам апостериорного анализа (см. рисунок 4.21) сделаем следующие **выводы**:

Из таблицы результатов теста НСР видно, что значимые различия на уровне значимости 0,05 имеются между случаем Опыт 2 – Вариант1 и всеми остальными случаями. Таким образом, результаты опыта показывают, что случай Опыт 2 – Вариант1 значимо влияет на функцию отклика Y по сравнению с остальными случаями.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Откройте файл **Visota.sta** в модуле Basic Statistics and Tables. Сохраните его с именем **DispVisota.sta**.

Data: DispVis...

	1 N_ОПЫТА	2 ДАННЫЕ
1	Опыт1	80
2	Опыт1	85
3	Опыт1	89
4	Опыт1	92
5	Опыт1	84
6	Опыт1	89
7	Опыт1	76
8	Опыт1	93
9	Опыт1	81
10	Опыт1	87
11	Опыт1	91
12	Опыт1	89
13	Опыт1	87
14	Опыт1	87
15	Опыт1	92
16	Опыт1	89
17	Опыт1	92
18	Опыт1	80
19	Опыт1	92
20	Опыт1	92
21	Опыт1	79

Рисунок 4.22 – Таблица, подготовленная для дисперсионного анализа

2. Подготовьте таблицу для дисперсионного анализа (рисунок 4.22):

– удалите столбец **Повтор**;

– добавьте два столбца с именами **N_Опыта** и **Данные**, добавьте **120** случаев;

– заполните столбец переменной **N_Опыта** значениями **Опыт 1**, **Опыт 2**, **Контроль**, **Опыт 3**, **Опыт 4** по **30** случаев для каждого варианта (используйте автозаполнение);

– перенесите данные из столбцов **Опыт 1**, **Опыт 2**, **Контроль**, **Опыт 3** и **Опыт 4** в соответствующие ячейки переменной **Данные**;

– удалите переменные **Опыт 1**, **Опыт 2**, **Контроль**, **Опыт 3** и **Опыт 4**. Сохраните файл.

3. Проведите дисперсионный анализ в модуле Basic Statistics and Tables. Сохраните таблицу результатов в файл с именем **DispResult.scr**.

4. В группе опций Statistics отметьте все опции.

5. Просмотрите итоговую таблицу средних и сохраните ее с именем **DispMean.scr**.

6. Выполните апостериорное сравнение средних значений вариантов с помощью критерия НСР. Сохраните результаты в файле с именем **LSDTest.scr**.

7. Рассчитайте значение наименьшей существенной разности на 5% уровне значимости $НСР_{05}$.

8. Рассчитайте разности выборочных средних δ для каждой пары вариантов. Сравните полученные δ с $НСР_{05}$ и сделайте вывод о существенности различий. Сравните полученные результаты с результатами апостериорных сравнений средних (LSD Test).

9. Запишите выводы о проведенном анализе в файл отчета с именем **DispReport.docx**. В отчете должно быть отражено, какие варианты имеют значимые различия, вероятность ошибки первого рода, значение наименьшей средней разницы.

10. Перейдите в модуль ANOVA/MANOVA. Если файл **DispVisota.sta** не открыт автоматически, откройте его.

11. Укажите независимые и зависимые переменные.

12. Просмотрите результаты дисперсионного анализа.

13. Просмотрите средние значения вариантов.

14. Проведите НСР-тест для вариантов.

15. Сравните результаты анализа в модулях Basic Statistics and Tables и ANOVA/MANOVA.

16. Проведите дисперсионный анализ в модуле ANOVA/MANOVA следующих данных (см. таблицу 4.8). Файл исходных данных сохраните с именем **UR.sta**. Результаты анализа сохраните с именем **DispUR.scr**. Запишите результаты и выводы в файл **DispUR.docx**.

17. Подготовьте таблицу исходных данных для двухфакторного дисперсионного анализа, результаты которого представлены в таблице 4.9. Для этого: создайте таблицу 3×36 с тремя переменными (**Сорт**, **Норма** и **Урожайность**) и **36** случаями; заполните таблицу данными, как показано на рисунке 4.23, сохраните файл с именем **UR_2.sta**.

18. Выполните двухфакторный дисперсионный анализ данных, представленных в таблице 4.9.

19. Сохраните результаты анализа в файл **DispUR_2.scr**.

Таблица 4.8 – Урожайность различных сортов ячменя

Задонский 8	Приазовский 9	Зерноградец 770
2,01	2,53	2,40
2,77	2,15	2,52
2,70	2,79	2,33
2,20	1,67	1,46
1,97	1,59	1,60
2,04	1,51	1,41
1,79	1,37	1,35
1,69	1,42	1,39
1,67	1,26	1,35
2,04	1,51	1,74
1,88	1,46	1,84
1,79	1,66	1,74

20. Анализируя значения F-критерия и уровня значимости p, сделайте вывод об эффектах, вносимых каждым фактором и взаимодействием факторов.

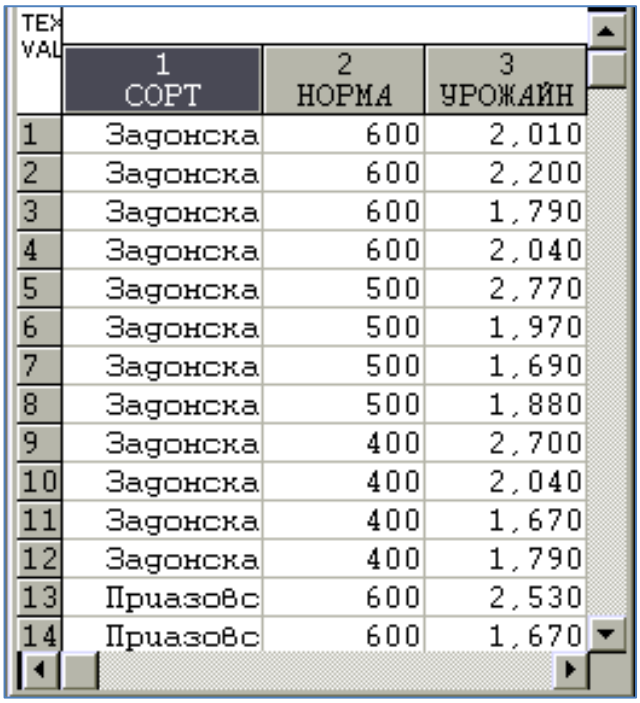
21. Выводы запишите в файле **DispUR_2.docx**.

22. Выполните дисперсионный анализ трехфакторного полевого эксперимента, результаты которого представлены в таблице 4.10.

23. Сохраните файл исходных данных с именем **UR_3.sta**.

Таблица 4.9 – Урожайность ячменя, т/га

Сорт	Норма	Повторения			
		I	II	III	IV
Задонский 8	600	2,01	2,20	1,79	2,04
Задонский 8	500	2,77	1,97	1,69	1,88
Задонский 8	400	2,70	2,04	1,67	1,79
Приазовский 9	600	2,53	1,67	1,37	1,51
Приазовский 9	500	2,15	1,59	1,42	1,46
Приазовский 9	400	2,79	1,51	1,26	1,66
Зерноградец 770	600	2,40	1,46	1,35	1,74
Зерноградец 770	500	2,52	1,60	1,39	1,84
Зерноградец 770	400	2,33	1,41	1,35	1,74



	1 СОРТ	2 НОРМА	3 УРОЖАЙН
1	Задонска	600	2,010
2	Задонска	600	2,200
3	Задонска	600	1,790
4	Задонска	600	2,040
5	Задонска	500	2,770
6	Задонска	500	1,970
7	Задонска	500	1,690
8	Задонска	500	1,880
9	Задонска	400	2,700
10	Задонска	400	2,040
11	Задонска	400	1,670
12	Задонска	400	1,790
13	Приазовс	600	2,530
14	Приазовс	600	1,670

Рисунок 4.23 – Таблица исходных данных для двухфакторного дисперсионного анализа

24. Сохраните результаты дисперсионного анализа в файл **DispUR_3.scr**.

25. Проведите апостериорные сравнения средних для каждого фактора в отдельности.

26. Выводы запишите в файле **DispUR_3.docx**.

Таблица 4.10 – Урожай 2002 года

Предше-ственник	Сорт	Нор-ма высе-ва, шт./м^2	Повторения			
			I	II	III	IV
1	2	3	4	5	6	7
Озимая пшеница	Задонский 8	600	20,1	22,0	1,79	2,04
	Задонский 8	500	2,77	1,97	1,69	1,88
	Задонский 8	400	2,70	2,04	1,67	1,79
	Приазовский 9	600	2,53	1,67	1,37	1,51
	Приазовский 9	500	2,15	1,59	1,42	1,46
	Приазовский 9	400	2,79	1,51	1,26	1,66
	Зерноградец 770	600	2,40	1,46	1,35	1,74
	Зерноградец 770	500	2,52	1,60	1,39	1,84
	Зерноградец 770	400	2,33	1,41	1,35	1,74

Окончание таблицы 4.10

1	2	3	4	5	6	7
Подсолнечник	Задонский 8	600	3,08	2,60	2,90	2,88
	Задонский 8	500	2,96	3,40	2,93	3,19
	Задонский 8	400	2,91	3,17	2,73	2,97
	Приазовский 9	600	2,66	2,97	2,78	2,81
	Приазовский 9	500	2,12	2,55	2,85	2,60
	Приазовский 9	400	2,58	2,56	2,63	2,61
	Зерноградец 770	600	2,84	2,63	2,74	2,75
	Зерноградец 770	500	3,10	2,76	2,57	2,92
	Зерноградец 770	400	2,83	2,55	2,43	2,64
Кукуруза на зерно	Задонский 8	600	4,58	4,07	3,77	4,18
	Задонский 8	500	4,32	4,09	4,07	4,30
	Задонский 8	400	3,99	4,04	3,89	4,02
	Приазовский 9	600	4,67	3,63	4,27	4,22
	Приазовский 9	500	4,42	3,52	3,81	4,05
	Приазовский 9	400	4,38	4,00	3,80	3,84
	Зерноградец 770	600	4,32	3,55	3,56	3,83
	Зерноградец 770	500	4,27	3,76	3,37	3,94
	Зерноградец 770	400	4,12	3,71	3,09	3,66

5 СОДЕРЖАНИЕ ОТЧЕТА

Отчет оформляется в соответствии с требованиями, принятыми в АЧИИ. Отчет должен содержать:

- ответы на контрольные вопросы;
- результаты дисперсионного и апостериорного анализа переменных файлов **Visota.sta**, **UR.sta**;
- результаты двухфакторного и трехфакторного дисперсионного анализа урожайности ячменя;
- выводы о различиях выборочных средних вариантов по каждому проведенному статистическому исследованию.

6 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Для чего проводят дисперсионный анализ? Какие выводы необходимо сделать в результате анализа?
2. В чем заключается сущность дисперсионного анализа?
3. Назовите основные этапы проверки гипотезы о равенстве выборочных средних вариантов по критерию Фишера.
4. Сформулируйте нулевую и альтернативную гипотезы, которые выдвигается при дисперсионном анализе.
5. Как определяется значение критерия Фишера F?
6. Как определяется критическое (табличное) значение критерия F?

7. Какие уровни значимости могут быть приняты при дисперсионном анализе?

8. Как определяют число степеней свободы вариантов, ошибки?

9. Для чего проводят апостериорный анализ?

10. Что такое наименьшая существенная разность (НСР)? Для чего применяют НСР?

11. Как рассчитывают значение НСР?

12. Как вычисляется ошибка разности средних?

13. В каких модулях производится дисперсионный анализ?

14. Как представляют исходные данные для дисперсионного анализа в Statistica?

15. Какую переменную называют группирующей (Grouping) или независимой (Independent)? Какую зависимой (Dependent)?

16. Сколько переменных должна содержать таблица исходных данных для однофакторного, двухфакторного и трехфакторного дисперсионного анализа?

17. Как вывести на экран результаты дисперсионного анализа?

18. Как просмотреть средние значения вариантов?

19. В чем отличие НСР-теста в модуле ANOVA /MANOVA?

20. Как определить, что некоторый вариант отличается от других вариантов?

Практическое занятие № 5

Тема: **ЛИНЕЙНАЯ И НЕЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИИ. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ**

1 ЦЕЛЬ РАБОТЫ

Изучить:

– основы регрессионного анализа.

Научиться:

– проводить линейный и нелинейный регрессионный анализ в модуле Multiple Regression (Множественная регрессия);

– рассчитывать коэффициенты уравнения множественной регрессии в модуле Multiple Regression (Множественная регрессия).

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Основы регрессионного анализа

Корреляционный анализ дает представление о направлении и степени сопряженности признаков, но не позволяет судить о количественном изменении результирующего признака при изменении факторов.

Для получения зависимости результирующего признака (отклика) Y от количественных факториальных признаков $X_1, X_2, \dots X_q$ в виде математического выражения $Y = f(X_1, X_2, \dots X_q)$ применяют *регрессионный анализ*. Полученное в результате регрессионного анализа выражение называют *уравнением регрессии*.

В зависимости от количества факторов X различают множественную и парную регрессии. Если исследуется зависимость выходной величины Y от одного фактора X, то такая *регрессия* называется *парной*, так как она вычисляется по парам точек $(x_i,\ y_i)$. Когда число факторов составляет два и более *регрессию* называют *множественной*.

В соответствии с видом зависимости $Y = f(X)$ различают линейную и нелинейную (криволинейную) регрессии. Функция линейной регрессии принадлежит классу параметрического семейства $f(x) = b_0 + b_1 \cdot x$.

В регрессионном анализе обрабатывается массив опытных данных, в котором каждому набору уровней факторов $X_1, X_2, \dots, X_q$ соответствует значение отклика Y. Массив имеет n опытов (таблицы 5.1, 5.2).

Эксперимент может быть реализован в нескольких повторениях. Это позволяет повысить точность эксперимента, оценить воспроизводимость эксперимента и оценить дисперсию ошибки (шума). В таком случае каждому набору факторов $X_1, X_2, \dots, X_q$ соответствует m измерений отклика Y. Функ-

цию регрессии получают для зависимости среднего значения $\bar{Y}$ от факторов $X_1, X_2, \dots, X_q$ (таблица 5.2).

Таблица 5.1 – Массив исходных данных для получения регрессионной модели без повторений опытов

Номер опыта	X_1	X_2	…	X_q	Y
1	x_{11}	x_{21}	…	x_{q1}	y_1
2	x_{12}	x_{22}	…	x_{q2}	y_2
…	…	…	…	…	…
i	x_{1i}	x_{2i}	…	x_{qi}	y_i
…	…	…	…	…	…
n	x_{1n}	x_{2n}	…	x_{qn}	y_n

Таблица 5.2 – Массив исходных данных для получения регрессионной модели при нескольких повторениях опытов

Номер опыта	X_1	X_2	…	X_q	Y_1	…	Y_m	$\bar{Y}$
1	x_{11}	x_{21}	…	x_{q1}	y_{11}	…	y_{m1}	$\bar{y}_1$
2	x_{12}	x_{22}	…	x_{q2}	y_{12}	…	y_{m2}	$\bar{y}_2$
…	…	…	…	…	…	…	…	…
i	x_{1i}	x_{2i}	…	x_{qi}	y_{1i}	…	y_{mi}	$\bar{y}_i$
…	…	…	…	…	…	…	…	…
n	x_{1n}	x_{2n}	…	x_{qn}	y_{1n}	…	y_{mn}	$\bar{y}_n$

Обозначение данных в таблицах 5.1, 5.2:

- $X_1, X_2, \dots, X_q$ – факторы эксперимента;
- $Y_1, Y_2, \dots, Y_m$ – значения функции отклика при m повторениях;
- q – количество факторов;
- n – количество опытов;
- m – количество повторений опытов.

Общая схема регрессионного анализа включает в себя несколько этапов:

1. Определение общего вида функции $Y = f(X)$ – линейная, нелинейная (полиномиальная, экспоненциальная, гиперболическая и т.д.).

2. Анализ структуры функции – отбор наиболее информативных факторов (производится на основе корреляционного анализа).

3. Вычисление коэффициентов уравнения регрессии с использованием метода наименьших квадратов.

4. Оценка значимости коэффициентов уравнения регрессии по критерию Стьюдента t, исключение незначимых коэффициентов.

5. Оценка адекватности, качества регрессионной модели, визуальный анализ остатков.

При проведении регрессионного анализа перечисленные этапы могут повторяться.

Этап 1 – Определение общего вида функции $\mathbf{Y = f(X)}$

Регрессионный анализ начинают с визуального изучения зависимостей факторов и отклика при помощи двумерных и трехмерных диаграмм рассеяния.

Диаграмма рассеяния используется для изучения возможной связи между двумя переменными величинами. По виду диаграммы рассеяния нельзя определить зависимость одной переменной от другой, однако диаграмма показывает, существует ли связь между ними и какова сила этой связи.

В случае парной регрессии эти графики оказывают неоценимую помощь и позволяют оценить характер и степень связи. В случае множественной регрессии можно также строить двумерные графики рассеяния для отклика и каждого фактора в отдельности. Данные графики могут дать предварительное представление о степени влияния каждого из факторов на отклик.

Графическое представление данных в начале регрессионного анализа позволяет обратить внимание на так называемые «выбросы», то есть точки, значения которых резко отличаются от соседних. Возможно, выбросы связаны с ошибками или «промахами» в измерениях и эти значения следует исключить из массива данных. К исключению данных следует подходить осторожно, особенно, если выбросы находятся близко к краям графика.

Произведя визуальную оценку, определяются с общим видом уравнения регрессии. В случае парной регрессии при помощи двумерной диаграммы можно установить необходимое семейство функций (например, парабола или гипербола и т.д.).

В общем случае вначале задаются функцией линейного вида $f(x) = b_0 + b_1 \cdot x$ и переходят к другим этапам регрессионного анализа. Если в результате выбранная регрессионная модель по каким-либо причинам не устраивает исследователя, то модель усложняется.

Этап 2 – Отбор наиболее информативных факторов

Регрессионному анализу всегда предшествует корреляционный, в котором определяют степень корреляционной связи всех факторов между собой и каждого фактора с откликом. В результате корреляционного анализа можно получить первое представление о степени взаимосвязи факторов и отклика.

При большом количестве факторов (трех и более) производится отбор наиболее информативных факторов. Если два фактора имеют сильную корреляцию друг с другом и с откликом, то один из факторов исключают. Обычно исключают фактор, у которого коэффициент корреляции с откликом меньше.

Если не исключить сильно коррелированные факторы, то соответствующие оценки коэффициентов уравнения регрессии могут оказаться незначимыми. В итоге, качество модели ухудшается.

Этап 3 – Вычисление коэффициентов уравнения регрессии с использованием метода наименьших квадратов

После выбора вида функции регрессии, т.е. вида рассматриваемой модели зависимости Y от X, например, линейной модели $Y(x) = b_0 + b_1 \cdot x$, необходимо определить конкретные значения коэффициентов модели.

При различных значениях b_0 и b_1 можно построить бесконечное число зависимостей вида $Y(x) = b_0 + b_1 \cdot x$, т.е. на координатной плоскости имеется бесконечное количество прямых. Необходимо получить такую зависимость $Y(x) = b_0 + b_1 \cdot x$, которая соответствует наблюдаемым значениям наилучшим образом. Таким образом, задача сводится к подбору наилучших коэффициентов уравнения регрессии.

Линейная функция $b_0 + b_1 \cdot x$ подбирается исходя из некоторого количества имеющихся опытов. Для нахождения функции с наилучшим соответствием наблюдаемым значениям используется метод наименьших квадратов.

Обозначим:

– Y_i – значение, вычисленное по уравнению $Y(x) = b_0 + b_1 \cdot x$,

– y_i – измеренное значение;

– $\varepsilon_i = y_i - Y_i$ – разность между измеренными и вычисленными по уравнению значениям, $\varepsilon_i = y_i - b_0 - b_1 \cdot x$.

В методе наименьших квадратов требуется, чтобы ε_i – разность между измеренными y_i и вычисленными по уравнению значениям Y_i была минимальной. Следовательно, коэффициенты b_0 и b_1 вычисляются так, чтобы сумма квадратов отклонений наблюдаемых значений от значений на прямой линии регрессии оказалась наименьшей:

$$SS = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - Y_i)^2 = \sum_{i=1}^{n} (y_i - b_0 - b_1 \cdot x)^2 \to min \qquad (5.1)$$

Функция SS принимает минимальное значение, если коэффициенты b_0 и b_1 являются решениями системы:

$$\begin{cases} b_0 \cdot \sum_{i=1}^{n} x_i + b_1 \cdot \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} (x_i \cdot y_i) \\ n \cdot b_0 + b_1 \cdot \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \end{cases} \tag{5.2}$$

Если разделить обе части уравнений (13.2) на n, то получится:

$$\begin{cases} \frac{b_0}{n} \cdot \sum_{i=1}^{n} x_i + \frac{b_1}{n} \cdot \sum_{i=1}^{n} x_i^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i \cdot y_i) \\ b_0 + \frac{b_1}{n} \cdot \sum_{i=1}^{n} x_i = \frac{1}{n} \cdot \sum_{i=1}^{n} y_i \end{cases} \tag{5.3}$$

Учитывая, что

$$\begin{gathered} \frac{1}{n} \cdot \sum_{i=1}^{n} x_i = \bar{x}, \qquad \frac{1}{n} \cdot \sum_{i=1}^{n} y_i = \bar{y}, \\ \frac{1}{n} \cdot \sum_{i=1}^{n} x_i^2 = \overline{x^2}, \qquad \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i \cdot y_i) = \overline{x \cdot y}, \end{gathered} \tag{5.4}$$

получим:

$$\begin{cases} b_0 \cdot \bar{x} + b_1 \cdot \overline{x^2} = \overline{x \cdot y} \\ b_0 + b_1 \cdot \bar{x} = \bar{y}. \end{cases} \tag{5.5}$$

Если $b_0 = \bar{y} - b_1 \cdot \bar{x}$, то подставляя значение b_0 в первое уравнение системы (4.5), получим:

$$b_1 = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2}, \tag{5.6}$$

$$b_0 = \frac{\overline{x^2} \cdot \bar{y} - \bar{x} \cdot \overline{x \cdot y}}{\overline{x^2} - \bar{x}^2}. \tag{5.7}$$

При этом коэффициент b_1 называют коэффициентом уравнения регрессии; b_0 называют свободным членом уравнения регрессии.

Этап 4 – Оценка значимости коэффициентов уравнения регрессии по критерию Стьюдента t, исключение незначимых коэффициентов

Критерий существенности (значимости) коэффициента регрессии определяют по формуле

$$t = \frac{b_i}{s_{b_i}}, \tag{5.8}$$

где s_{b_i} – ошибка i-го коэффициента регрессии, $s_{b_i} = s_r \cdot \sqrt{\frac{\sum(Y-\bar{y})^2}{\sum(X-\bar{x})^2}}$.

Вычисленный критерий t_b сравнивают с табличным значением t_α для принятого уровня значимости α и степени свободы $df_v = n - 2$.

Если $t > t_\alpha$, то нулевую гипотезу H_0 отвергают. В качестве нулевой гипотезы выдвигается предположение, что коэффициент b_i равен нулю.

Этап 5 – Оценка адекватности, качества регрессионной модели, визуальный анализ остатков

Проверка адекватности модели производится на основе анализа составляющих общей дисперсии отклика и выясняется, связан ли разброс значений отклика с варьированием уровней фактора или он носит случайный характер.

Изменения величины отклика Y вызваны изменениями, обусловленными регрессионной зависимостью и ошибкой опыта (остатками). То есть общая сумма квадратов отклонений SS_y включает в себя сумму квадратов отклонений, обусловленных регрессией, SS_r и сумму квадратов остатков SS_o

$$SS_y = SS_r + SS_o, \tag{5.9}$$

где SS_y – общая сумма квадратов отклонений отклика Y;

SS_r – сумма квадратов отклонений Y, обусловленных регрессионной зависимостью;

SS_o – сумма квадратов остатков.

В свою очередь остатки связаны с ошибкой опыта (действие неучтенных факторов и погрешности измерений), а также с неудачным выбором параметрического семейства функций для аппроксимации опытных данных (неадекватностью модели). Сумма квадратов остатков SS_o в таком случае будет определяться суммой квадратов отклонений, обусловленных неадекватностью SS_n, и суммой квадратов ошибки (шума) SS_e

$$SS_o = SS_n + SS_e, \tag{5.10}$$

где SS_n – сумма квадратов отклонений отклика Y, обусловленная неадекватностью модели;

SS_e – сумма квадратов ошибки (шума).

Проверка адекватности регрессионной модели на основе анализа общей вариации отклика

В качестве нулевой гипотезы H_0 выдвигается предположение о том, что связь между откликом Y и факторами $X_1, X_2, \dots, X_q$ отсутствует.

Когда модель строится на основе массива данных без повторений (для каждого набора факторов $X_i = (X_1, X_2, \dots, X_q)$ производилось только одно измерение значения отклика Y_i) нет возможности оценить дисперсию ошибки S_e^2. Поэтому для проверки адекватности используют отношение дисперсии, обусловленной регрессионной зависимостью S_r^2, к дисперсии остатков S_o^2.

Для сравнения этих дисперсий применяют критерий Фишера

$$F = \frac{S_r^2}{S_o^2}. \tag{5.11}$$

Дисперсия отклика, связанная с регрессионной зависимостью S_r^2, вычисляется по формуле

$$S_r^2 = \frac{SS_r}{df_r}, \tag{5.12}$$

где SS_r – сумма квадратов отклонений отклика, обусловленных регрессионной зависимостью;

df_r – число степеней свободы отклика.

Сумма квадратов отклонений отклика Y, обусловленных регрессионной зависимостью определяется из выражения

$$SS_r = \sum_{i=1}^{n} (Y_i - \bar{y})^2, \tag{5.13}$$

где Y_i – значение отклика, вычисленное по уравнению регрессии для i-го сочетания факторов;

$\bar{y}$ – среднее значение отклика.

Число степеней свободы df_r определяется как разность числа опытов n и количества значимых коэффициентов уравнения регрессии d

$$df_r = n - d. \tag{5.14}$$

Дисперсия остатков S_o^2 вычисляется по формуле

$$S_o^2 = \frac{SS_o}{df_o}, \tag{5.15}$$

где SS_o – сумма квадратов остатков;

df_o – число степеней свободы остатков.

Сумма квадратов остатков SS_o определяется как сумма квадратов разности значения отклика y_i и вычисленного по уравнению регрессии значения Y_i

$$SS_o = \sum_{i=1}^{n} (y_i - Y_i)^2, \tag{5.16}$$

где y_i – значение отклика для i-го сочетания факторов.

Число степеней свободы остатков df_o на единицу меньше числа значимых коэффициентов уравнения регрессии d

$$df_o = d - 1. \tag{5.17}$$

Рассчитанное значение критерия Фишера F сравнивают с табличным значением F_α для df_r и df_o степеней свободы и принятого уровня значимости α.

Если $F > F_\alpha$, то нулевая гипотеза отвергается и принимается альтернативная: между откликом Y и факторами $X_1, X_2, \dots, X_q$ существует зависимость. Полученная регрессионная модель адекватно отображает экспериментальные данные.

Оценка качества регрессионного уравнения

Оценку качества построенной модели дает *коэффициент ции* R^2.

Коэффициент детерминации численно выражает долю вариации зависимой переменной, объясненную с помощью регрессионного уравнения. Чем больше R^2, тем большую долю вариации объясняют переменные, включенные в модель. Таким образом, коэффициент детерминации показывает, какой процент значений отклика Y зависит от фактора (или факторов) X. Если $R^2 < 0{,}75$, то от модели следует отказаться; если $R^2 \geq 0{,}75$ то модель может использоваться для прогноза. Из двух моделей лучше та, у которой коэффициент детерминации ближе к 1.

Визуальный анализ остатков регрессионного уравнения

Для оценки качества уравнения регрессии дополнительно проводят анализ остатков. Он позволяет получить представление, о том насколько хорошо подобрана регрессионная модель.

Остатки – это отклонение наблюдаемых значений от значений, вычисленных по уравнению регрессии.

Чем качественнее регрессионная модель, тем ближе распределение остатков к нормальному, остатки независимы от величин X и Y, имеют среднее значение равное нулю и постоянную дисперсию.

Чем лучше регрессионная модель согласуется с данными, тем меньше величина остатков. Каждый i-ый остаток (ε_i) вычисляется так:

$$\varepsilon_i = y_i - Y_i. \qquad (6.18)$$

где Y_i – значение отклика, вычисленное по уравнению регрессии для i-го сочетания факторов;

y_i – наблюдаемое значение отклика.

Для оценки *пригодности эмпирической модели* регрессии остатки откладываются по вертикальной оси, а значения Y_i – по горизонтальной. Если эмпирическая модель пригодна, график не должен иметь ярко выраженной закономерности. Если же модель регрессии не пригодна, на рисунке проявится зависимость между значениями Y_i и остатками ε_i.

2.2 Нелинейные регрессионные модели

В соответствии с видом регрессионного уравнения $Y = f(X)$ различают линейную и нелинейную (криволинейную) регрессии. Функция линейной регрессии рассмотрена в лабораторной работе № 4.

В тех случаях, когда линейная модель неадекватно описывает опытные данные, следует использовать нелинейные модели (полиномиальную, логарифмическую, экспоненциальную и т.д.).

Широкое распространение получили следующие регрессионные модели:

- полиномиальная модель $y = b_0 + b_1 \cdot x + b_2 \cdot x^2 + \cdots + b_n \cdot x^n$;
- степенная модель $y = b_0 \cdot x^{b_1}$;
- показательная модель $y = b_0 \cdot {b_1}^x$;
- экспоненциальная модель $y = e^{(b_0 + b_1 \cdot x)}$;
- гиперболическая модель $y = b_0 + \frac{b_1}{x}$;
- логарифмическая модель $\ln y = \ln b_0 + b_1 \ln x$.

Как показывает опыт большинства исследователей, среди нелинейной полиномиальной регрессии, чаще всего используется полином второй степени; в отдельных случаях – полином третьего порядка. Ограничения в использовании полиномов более высоких степеней связаны с требованием однородности исследуемой совокупности: чем выше порядок полинома, тем больше изгибов имеет кривая и соответственно менее однородна совокупность по результативному признаку.

2.3 Уравнение множественной регрессии

До сих пор рассматривались вопросы, связанные с исследованием функции одной переменной, т.е. изучали совместное изменение двух переменных, одна из которых зависела от другой. Значением независимой переменной полностью определялось значение зависимой переменной, или функции.

На практике значения зависимых переменных в агробиологических опытах определяются обычно влиянием не одного, а многих объясняющих факторов. Таким образом, независимых переменных часто оказывается несколько, и для определения значения функции необходимо установить значения,

совместно принимаемые всеми этими независимыми переменными. В таком случае говорят, что существует зависимость $y = f(x_1,\ x_2, \dots, x_m)$.

При этом следует помнить, что все факторы учесть в модели невозможно по целому ряду причин: часть факторов просто неизвестна современной науке, по части известных факторов нет достоверной информации или количество включаемых в модель факторов может быть ограничено объемом выборки (количество факторных признаков должно быть на порядок меньше численности изучаемой совокупности).

Множественная регрессия описывает форму связи в виде уравнения множественной регрессии (таблица 5.3).

Таблица 5.3 – Основные виды множественной регрессии

Наименование формы регрессии	Вид уравнения регрессии
Линейная	$y = b_0 + b_1 \cdot x_1 + \dots + b_m \cdot x_m$
Гиперболическая	$y = b_0 + b_1 \cdot \frac{1}{x_1} + \dots + b_m \cdot \frac{1}{x_m}$
Параболическая	$y = b_0 + b_1 \cdot {x_1}^2 + \dots + b_m \cdot {x_m}^2$
Степенная	$y = b_0 \cdot {x_1}^{b_1} \cdot {x_2}^{b_2} \cdot \dots \cdot {x_m}^{b_m}$

Параметры уравнения множественной регрессии $b_0,\ b_1, \dots, b_m$ называют коэффициентами множественной регрессии и определяют с помощью метода наименьших квадратов (МНК) путем решения системы нормальных уравнений МНК. При этом число нормальных уравнений в общем случае будет равно числу параметров. Если связь отдельного фактора с результатом не является линейной, то производят линеаризацию уравнения.

Полученные коэффициенты множественной регрессии являются именованными числами и показывают: насколько изменится результативный признак (по отношению к своей средней величине) при отклонении факторного признака от своей средней на единицу и при постоянстве (фиксированном уровне) других факторов.

Коэффициенты регрессии можно преобразовать в сравнимые относительные показатели – стандартизованные коэффициенты регрессии или β-коэффициенты.

Оценить меру влияния вариации факторного признака на вариацию результата при фиксированном уровне других факторов позволяет β-коэффициент:

$$\beta_{x_i} = b_i \cdot \frac{\sigma_{x_i}}{\sigma_y}. \qquad (5.19)$$

где σ_{x_i} – среднее квадратическое отклонение факторного признака;

σ_y – среднее квадратическое отклонение результативного признака;

b_i – коэффициент регрессии при соответствующем факторном признаке x_i.

При интерпретации результатов корреляционно-регрессионного анализа часто используют частные коэффициенты эластичности (E_{x_i}). Коэффициент эластичности показывает: насколько процентов в среднем изменится значение результативного признака при изменении факторного на 1% и при постоянстве (фиксированном уровне) других факторов

$$E_{x_i} = b_i \cdot \frac{\bar{x}_i}{\bar{y}}. \qquad (5.20)$$

где $\bar{x}_i$ – среднее значение факторного признака;

$\bar{y}$ – среднее значение результативного признака.

Этапы проведения регрессионного анализа для получения уравнения множественной регрессии такие же, как и для получения уравнения линейной или нелинейной парной регрессии.

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

3.1 Линейный регрессионный анализ в модуле Multiple Regression

Линейный регрессионный анализ выполняется в модуле Multiple Regression, который открывается с помощью переключателя модулей Module Switcher (рисунок 5.1).

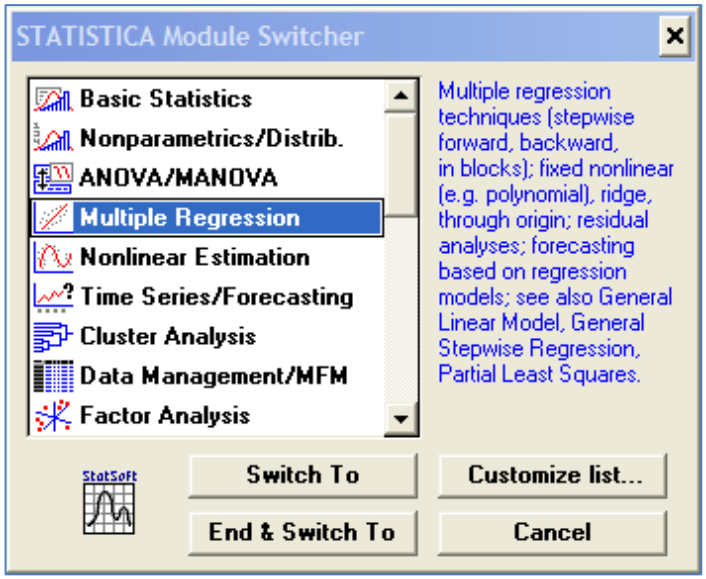


Рисунок 5.1 – Окно переключателя модулей Module Switcher

В стартовом диалоговом окне модуля Multiple Regression при помощи кнопки Variables указываются зависимая (Dependent) и независимые (Independent) переменные (рисунок 5.2).

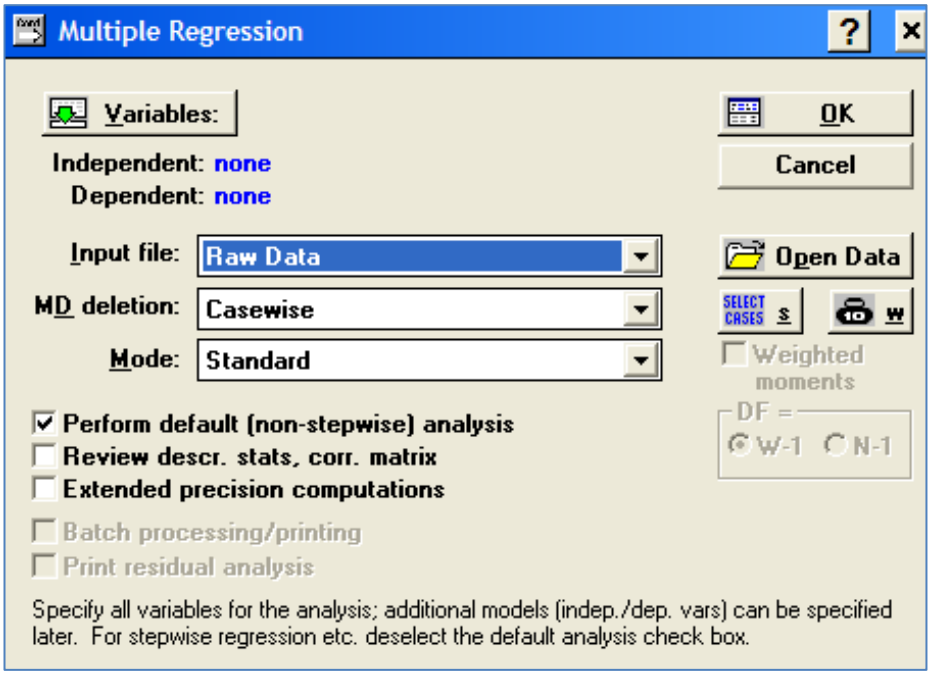


Рисунок 5.2 – Стартовое диалоговое окно модуля Multiple Regression

Рассмотрим основные компоненты диалогового окна модуля Multiple Regression.

В поле Input file указывается тип файла с данными:

– Raw Date – данные в виде строчной таблицы;

– Correlation Matrix – данные в виде корреляционной матрицы.

В поле MD deletion указывается способ исключения из обработки недостающих данных:

– Casewise – игнорируется вся строка, в которой есть хотя бы одно пропущенное значение;

– Mean Substitution – взамен пропущенных данных подставляются средние значения переменных;

– Pairwise – попарное исключение данных с пропусками из тех переменных, корреляция которых вычисляется.

В поле Mode указывается тип регрессионной модели:

– Standard – стандартная линейная модель вида:

$$Y = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \cdots + b_n \cdot x_n.$$

– Fixed non-linear – фиксированная нелинейная, т.е. нелинейная модель, но которая может быть приведена к линейному виду путем преобразования переменных.

После установки опций стартовой панели нажатие на кнопку OK приведет к появлению окна Multiple Regressions Results (результаты регрессионного анализа), с помощью которого можно просмотреть результаты анализа в деталях (рисунок 5.3).

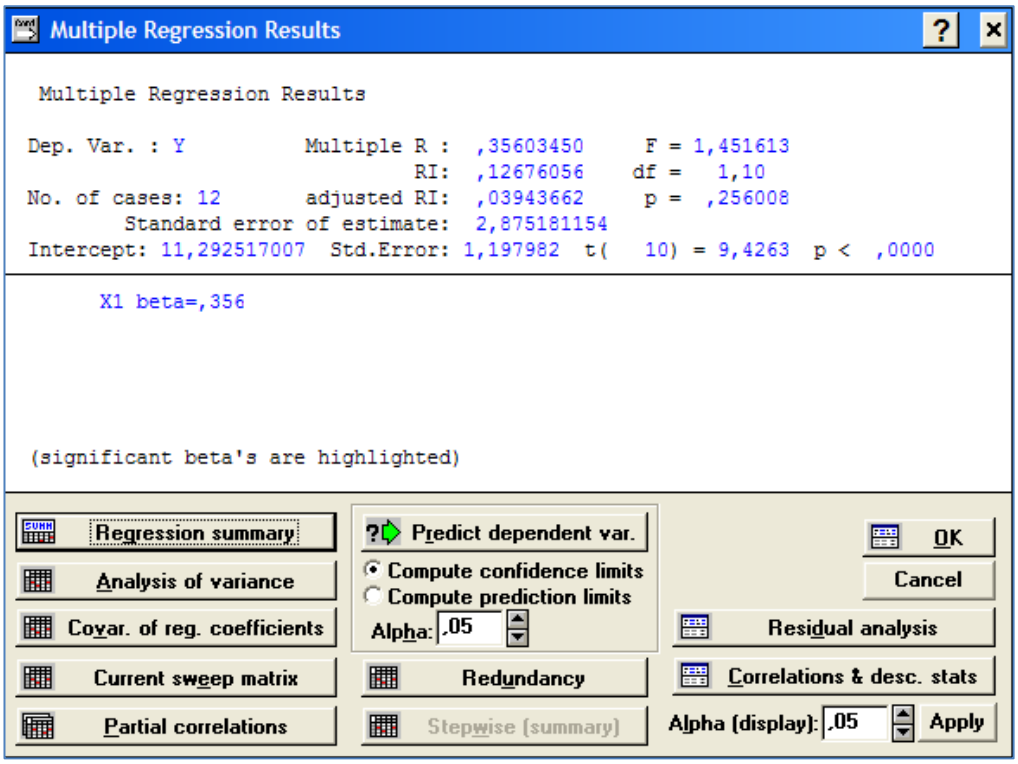


Рисунок 5.3 – Окно просмотра результатов регрессионного анализа

В верхней части окна приводятся наиболее важные параметры полученной регрессионной модели:

– Multiple R – коэффициент множественной корреляции – характеризует тесноту линейной связи между зависимой и всеми независимыми переменными, может принимать значения от 0 до 1;

– RI (R^2) – коэффициент детерминации – численно выражает долю вариации зависимой переменной, объясненную с помощью регрессионного уравнения; чем больше R^2, тем большую долю вариации объясняют переменные, включенные в модель; если $R^2 < 0{,}75$, то от модели следует отказаться; если $R^2 \geq 0{,}75$, то модель может использоваться для прогнозов; из двух моделей лучше та, у которой коэффициент детерминации ближе к 1;

– adjusted R – скорректированный коэффициент множественной корреляции – этот коэффициент лишен недостатков коэффициента множественной корреляции;

– adjusted RI (adjusted R^2) – скорректированный коэффициент детерминации, скорректированный коэффициент R^2 можно с большим успехом (по сравнению с обычным R^2) применять для выбора наилучшего подмножества независимых переменных в регрессионном уравнении; не всегда включение новой переменной в регрессионное уравнение увеличивает RI, а только в том случае, когда частный t-критерий при проверке гипотезы о значимости включаемой переменной больше или равен 1; в противном случае включение новой переменной уменьшает значение RI и adjusted R^2;

– F – критерий Фишера;

– df – число степеней свободы для критерия Фишера;

– p – вероятность нулевой гипотезы для критерия Фишера;
– Standard error of estimate – стандартная ошибка оценки (уравнения);
– Intercept – свободный член уравнения;
– Std.Error – стандартная ошибка свободного члена уравнения;
– t – критерий Стьюдента для свободного члена уравнения;
– p – вероятность нулевой гипотезы для свободного члена уравнения;
– Beta – β-коэффициенты уравнения регрессии – это стандартизированные регрессионные коэффициенты, рассчитанные по стандартизированным значениям переменных; по их величине можно сравнить и оценить значимость зависимых переменных, так как β -коэффициент показывает на сколько единиц стандартного отклонения изменится зависимая переменная при изменении на одно стандартное отклонение независимой переменной при условии постоянства остальных независимых переменных, свободный член в таком уравнении равен 0.

Кнопка Regression summary – позволяет просмотреть основные результаты регрессионного анализа (см. рисунок 5.4):
– BETA – β-коэффициенты уравнения;
– St. Err. of BETA – стандартные ошибки *β*-коэффициентов;
– B – коэффициенты уравнения регрессии;
– St. Err. of B – стандартные ошибки коэффициентов уравнения регрессии;
– t(10) – значения критерия Стьюдента для коэффициентов уравнения регрессии;
– p-level – вероятность нулевой гипотезы для коэффициентов уравнения регрессии.

Regression Summary for Dependent Variable: Y

Continue...

R= ,35603450 RI= ,12676056 Adjusted RI= ,03943662
F(1,10)=1,4516 p<,25601 Std.Error of estimate: 2,8752

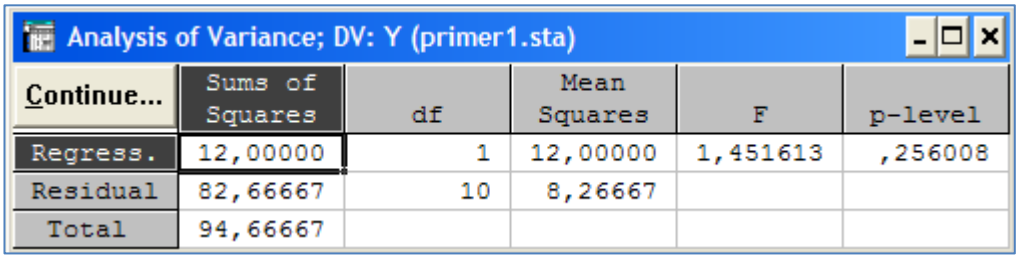
N=12	BETA	St. Err. of BETA	B	St. Err. of B	t(10)	p-level
Intercpt			11,29252	1,197982	9,426281	,000003
X1	,356034	,295506	,02041	,016939	1,204829	,256008

Рисунок 5.4 – Окно основных результатов регрессионного анализа

Кнопка Analysis of variance – позволяет ознакомиться с результатами дисперсионного анализа уравнения регрессии. В строках таблицы дисперсионного анализа уравнения регрессии – источники вариации (см. рисунок 5.5):
– Regress – обусловленная регрессией;
– Residual – остаточная;
– Total – общая.

В столбцах таблицы:

– Sums of Squares – сумма квадратов;
– df – число степеней свободы;
– Mean Squares – средний квадрат;
– F – значение критерия Фишера;
– p-level – вероятность нулевой гипотезы для критерия Фишера.


Analysis of Variance; DV: Y (primer1.sta)

Continue...	Sums of Squares	df	Mean Squares	F	p-level
Regress.	12,00000	1	12,00000	1,451613	,256008
Residual	82,66667	10	8,26667		
Total	94,66667				

Рисунок 5.5 – Окно результатов дисперсионного анализа уравнения регрессии

Критерий Фишера полученного уравнения регрессии значим на 5% уровне. Вероятность нулевой гипотезы (p-level) значительно меньше 0,05, что говорит об общей значимости уравнения регрессии.

Кнопка Partial correlations – позволяет просмотреть частные коэффициенты корреляции (Partial Cor.) между переменными. Частная корреляция – это корреляция между двумя переменными, когда одна или больше из оставшихся переменных удерживаются на постоянном уровне (т.е. имеют постоянное значение). Частные коэффициенты корреляции, как и парные, могут принимать значения от (–1) до (+1).

В идеальной регрессионной модели независимые переменные вообще не коррелируют друг с другом. Однако в моделях, разрабатываемых для природных объектов, сильная коррелированность переменных является довольно частым явлением. Это приводит к увеличению ошибок уравнения, уменьшению точности оценивания, снижает эффективность использования регрессионной модели. Поэтому выбор независимых переменных, включаемых в регрессионную модель, должен быть очень тщательным.

Кнопка Predict dependent var. – позволяет рассчитать по полученному регрессионному уравнению значение зависимой переменной по значениям независимых переменных.

Кнопка Correlations and desc. stats позволяет просмотреть описательные статистики и корреляционную матрицу с парными коэффициентами корреляции переменных, участвующих в регрессионной модели.

Кнопка Residual analysis запускает процедуру всестороннего анализа остатков регрессионного уравнения (см. рисунок 5.6).

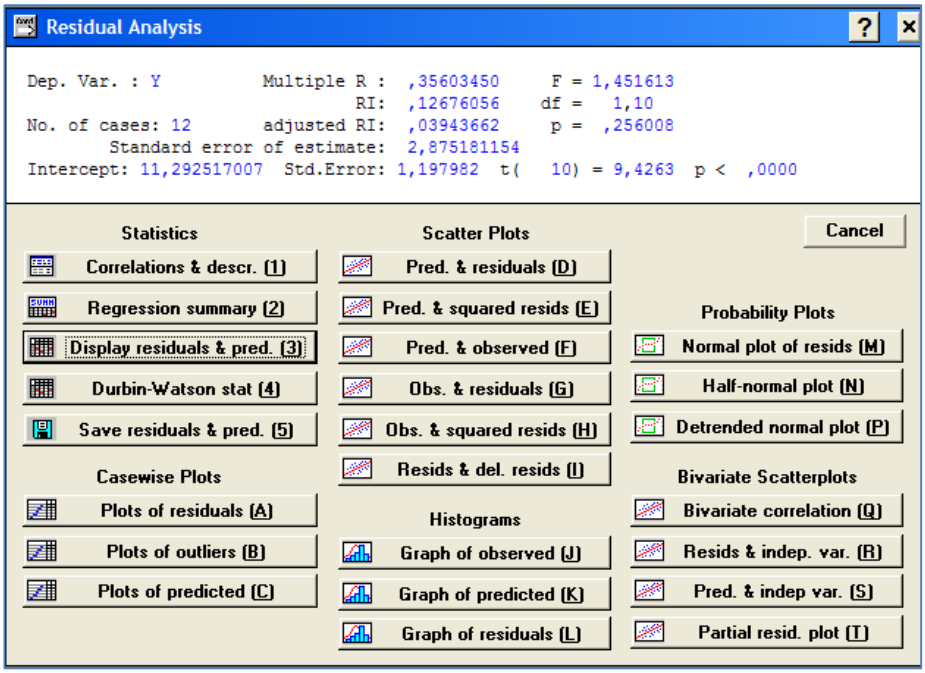


Рисунок 5.6 – Диалоговое окно Residual analysis

Кнопка Redundancy (см. рисунок 5.3) предназначена для поиска выбросов. *Выбросы* – это остатки, которые значительно превосходят по абсолютной величине остальных. Выбросы показывают опытные данные, которые являются не типичными по отношению к остальным данным, и требуют выяснения причин их возникновения. Выбросы должны исключаться из обработки, если они вызваны ошибками измерения.

Для выделения имеющихся в регрессионных остатках выбросов предложен ряд показателей:

– показатель Кука (Cook's Distance) – принимает только положительное значение и показывает расстояние между коэффициентами уравнения регрессии после исключения из обработки i-ой точки данных; большое значение показателя Кука указывает на сильно влияющий случай;

– расстояние Махаланобиса (Mahalns. Distance) – показывает насколько каждый случай или точка в p-мерном пространстве независимых переменных отклоняется от центра статистической совокупности.

Просмотр величин остатков и специальных критериев, их оценивающих, осуществляется после нажатия на кнопку Display residuals & pred. окна Residual analysis. Вполне достаточно бывает одного графического анализа остатков.

О нормальности остатков можно судить по графику остатков на нормальной вероятностной бумаге. Чем ближе распределение к нормальному виду, тем лучше значения остатков выстраиваются на прямой линии. График строится после нажатия на кнопку Normal plot of resids. окна Residual analysis.

Важно просмотреть графики зависимости остатков от каждой из независимых переменных. Их легко просмотреть при помощи кнопки Resids & indep. var. окна Residual analysis. Остатки должны быть нормаль-

но распределены, т.е. на графике они должны представлять приблизительно горизонтальную полосу одинаковой ширины на всем ее протяжении. Коэффициент корреляции (r) между регрессионными остатками и переменными должен равняться нулю.

Для выявления нестабильности дисперсии ошибки уравнения после нажатия на кнопку Pred. & residuals окна Residual analysis можно создать график зависимости регрессионных остатков от предсказанного значения зависимой переменной.

Очень удобным визуальным способом оценки адекватности регрессионной модели является анализ графического изображения опытных и полученных по регрессионному уравнению значений отклика. Оно строится при помощи кнопки Pred. & observed окна Residual analysis.

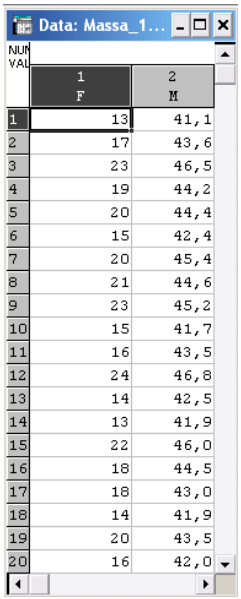
Упражнение 1

В этом упражнении вы научитесь строить уравнение линейной регрессии.

1. Откройте программу Statistica.
2. В модуле Basic Statistics and Tables создайте новую таблицу, содержащую две переменные F и M по 20 случаев.
3. Заполните новую таблицу данными в соответствии с таблицей 5.4. Сохраните файл с именем **Massa_1.sta** (рисунок 5.7).

Таблица 5.4 – Результаты измерений массы 1000 семян пшеницы при различной влажности зерна

φ, %	13	17	23	19	20	15	20	21	23	15
m, г	41,1	43,6	46,5	44,2	44,4	42,4	45,4	44,6	45,2	41,7
φ, %	16	24	14	13	22	18	18	14	20	16
m, г	43,5	46,8	42,5	41,9	46	44,5	43	41,9	43,5	42



	1 F	2 M
1	13	41,1
2	17	43,6
3	23	46,5
4	19	44,2
5	20	44,4
6	15	42,4
7	20	45,4
8	21	44,6
9	23	45,2
10	15	41,7
11	16	43,5
12	24	46,8
13	14	42,5
14	13	41,9
15	22	46,0
16	18	44,5
17	18	43,0
18	14	41,9
19	20	43,5
20	16	42,0

Рисунок 5.7 – Таблица исходных данных (файл Massa.sta)

4. Переключите модуль Basic Statistics and Tables на модуль Multiple Regression, используя переключатель модулей.

5. В диалоговом окне Multiple Regression укажите группирующую (или независимую) – Independent и зависимую (Dependent) переменные (Variables) (рисунок 5.8). Нажмите OK.

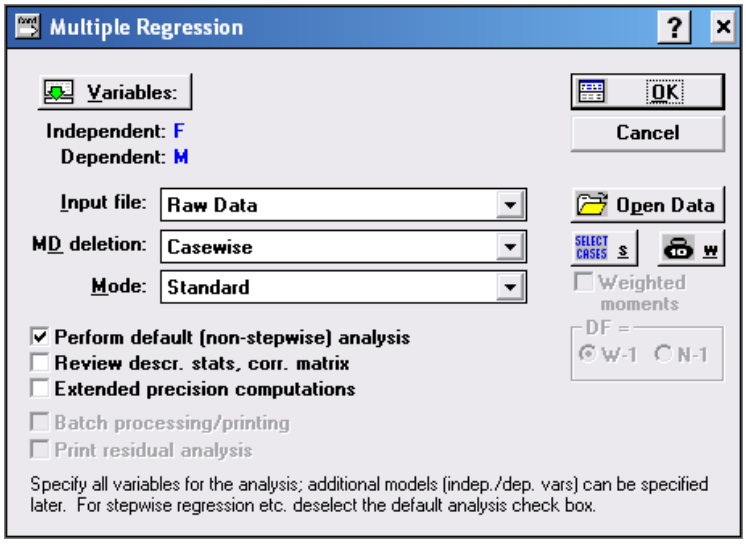


Рисунок 5.8 – Основное окно модуля Multiple Regression

6. На экране появится окно с результатами регрессионного анализа (рисунок 5.9), в котором в верхней части указаны все статистики регрессионного анализа. Для определения коэффициентов уравнения регрессии нажмите кнопку Regression Summary.

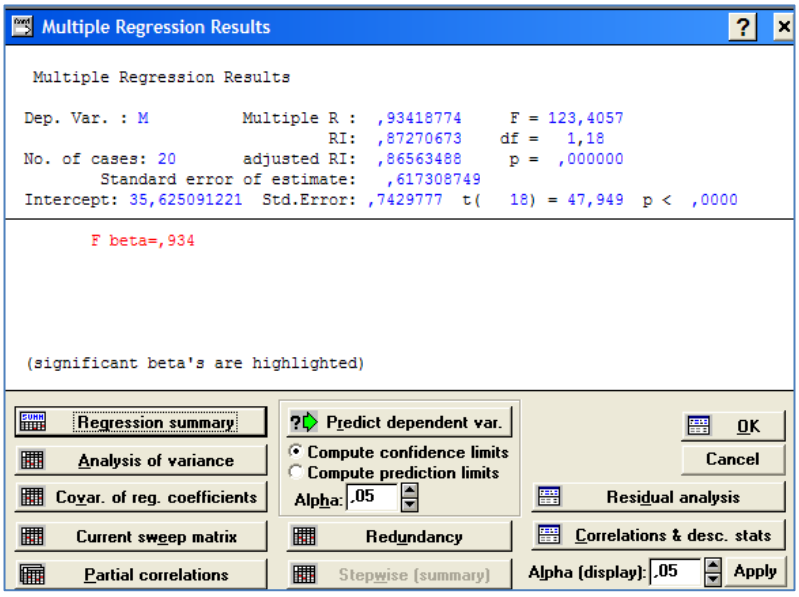
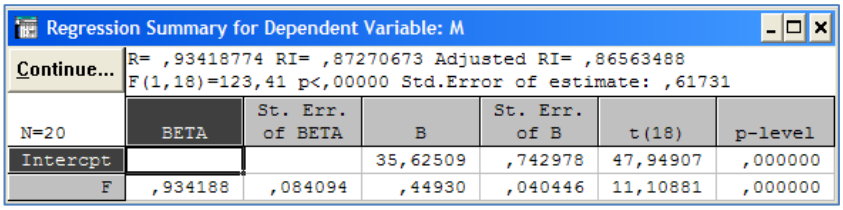


Рисунок 5.9 – Окно с результатами регрессионного анализа Multiple Regression Results

7. На экране появится окно Regression Summary for Dependent Variable: M (см. рисунок 5.10).



Regression Summary for Dependent Variable: M

Continue...

R= ,93418774 RI= ,87270673 Adjusted RI= ,86563488
F(1,18)=123,41 p<,00000 Std.Error of estimate: ,61731

N=20	BETA	St. Err. of BETA	B	St. Err. of B	t(18)	p-level
Intercpt			35,62509	,742978	47,94907	,000000
F	,934188	,084094	,44930	,040446	11,10881	,000000

Рисунок 5.10 – Таблица результатов регрессионного анализа для определения коэффициентов уравнения регрессии

Regression Summary for Dependent Variable: M

По результатам регрессионного анализа (рисунок 5.10) сделаем следующие **выводы:**

1) Определим коэффициенты уравнения линейной регрессии. Коэффициенты уравнения регрессии вычислены в столбце **B** (рисунок 5.10). Таким образом, $b_0 = 35{,}62509,\ \ b_1 = 0{,}44930$.

2) Оценим значимость коэффициентов уравнения регрессии по критерию Стьюдента. ***В качестве нулевой гипотезы выдвигается предположение, что коэффициент b_i равен нулю.*** Для каждого коэффициента уравнения регрессии необходимо сравнить вычисленный критерий Стьюдента с табличным значением.

Вычисленные значения критерия Стьюдента для коэффициентов уравнения находятся в столбце $t(18)$: для b_0 критерий Стьюдента $t_{b_0} = 47{,}94907$, для b_1 критерий Стьюдента $t_{b_1} = 11{,}10881$. Теоретическое значение критерия Стьюдента для степени свободы $df_v = 18$ (значение степени свободы df_v для критерия Стьюдента указано в столбце таблицы t(18) на рисунке 5.10) и уровня значимости $\alpha = 5\%$ равно $t_{05}(18) = 2{,}1$. Для всех коэффициентов уравнения регрессии выполняется условие $|t_{b_i}| > t_{05}(18)$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: коэффициенты уравнения регрессии отличны от нуля***. Все коэффициенты полученного уравнения значимы. Таким образом, уравнение регрессии будет иметь следующий вид

$$M = 35{,}62509 + 0{,}4493 \cdot F.$$

3) Проверим адекватность полученной модели по критерию Фишера. ***В качестве нулевой гипотезы H_0 выдвигаем предположение о том, что линейная связь между откликом M и фактором F отсутствует.*** Теоретическая функция $F_{05}(1, 18) = 4{,}41$ (значения степеней свободы выбираем из таблицы результатов регрессионного анализа – рисунок 5.10). Вычисленное значение критерия Фишера $F(1, 18) = 123{,}41$. Сравниваем критерии Фишера: $F > F_{05}$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: между откликом M и фактором F существует линейная зависимость.***

4) Оценим качество полученной регрессионной модели. Коэффициент детерминации $R^2 = 0{,}873$ (см. коэффициент RI в таблице результатов ре-

грессионного анализа – рисунок 5.10), следовательно, 87% значений отклика объяснимы значениями фактора. Вычисленный $R^2 > 0{,}75$, значит, полученная регрессионная модель может использоваться для прогноза.

8. Выполните анализ остатков (постройте график зависимости остатков регрессионной модели от значений зависимой переменной, вычисленных по регрессионному уравнению; просмотрите таблицу наблюдаемых, вычисленных значений и остатков; постройте график остатков на нормальной вероятностной бумаге). Для этого в окне Multiple Regression Results нажмите кнопку Residual Analysis. На экране появится диалоговое окно Residual Analysis (см. рисунок 5.11).

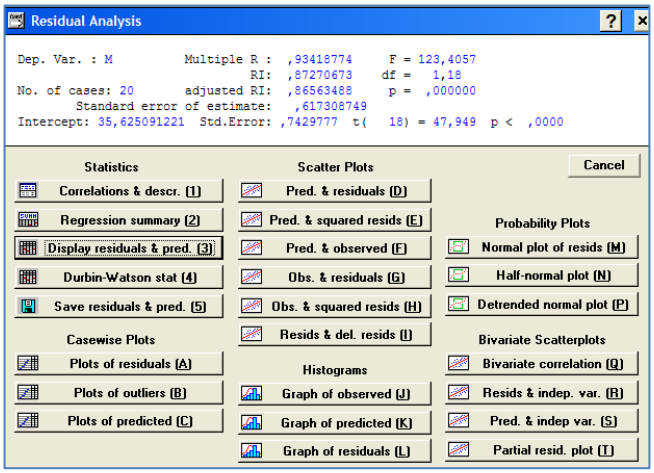


Рисунок 5.11 – Окно для анализа остатков регрессионного уравнения Residual Analysis

9. В окне Residual Analysis нажмите кнопку Pred. & residuals (D). В результате на экране появится график зависимости остатков регрессионной модели от значений зависимой переменной (функции отклика), вычисленных по регрессионному уравнению (см. рисунок 5.12).

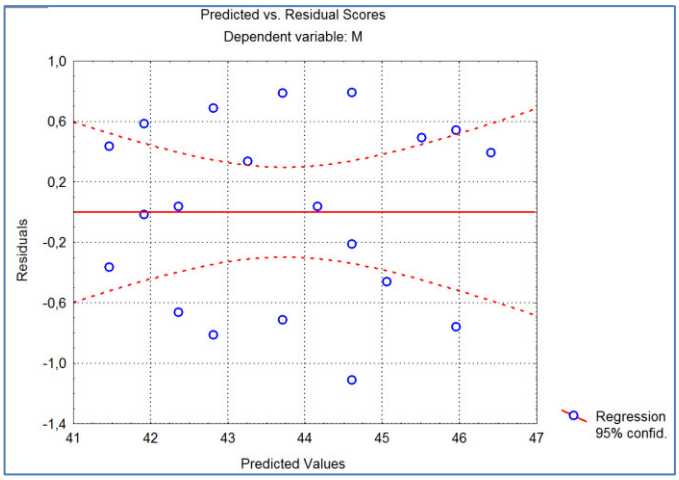


Рисунок 5.12 – График зависимости остатков регрессионной модели от значений зависимой переменной, вычисленных по регрессионному уравнению

По графику зависимости остатков регрессионной модели от значений зависимой переменной сделаем **вывод:** на рисунке 5.12 точки облака рассеяния разбросаны на координатной плоскости в хаотичном порядке (т.е. не выстраиваются в виде какой-либо линейной или нелинейной зависимости), поэтому связь между остатками регрессионного уравнения и значениями зависимой переменной, вычисленных по регрессионному уравнению, отсутствует. Это говорит о том, что полученная регрессионная модель пригодна для использования.

10. Для просмотра таблицы наблюдаемых, вычисленных значений и остатков в окне Residual Analysis нажмите кнопку Display residuals & pred.(см. рисунок 5.11). В результате на экране появится таблица (рисунок 5.13).

По данным таблицы на рисунке 5.13 **сделаем вывод:** среднее значение остатков равно нулю (третий столбец таблицы на рисунке 5.13). Это говорит о том, что полученная регрессионная модель адекватно описывает экспериментальные данные.

Continue... Dependent variable: M

Case No.	Observed Value	Predictd Value	Residual	Standard Pred. v.	Standard Residual	Std.Err. Pred.Val	Mahalns. Distance	Deleted Residual	Cook's Distance
1	41,10000	41,46602	-,36602	-1,44224	-,59294	,246519	2,080049	-,43547	,039681
2	43,60000	43,26323	,33677	-,29987	,54554	,144420	,089923	,35626	,009115
3	46,50000	45,95905	,54095	1,41368	,87631	,243179	1,998487	,64032	,083485
4	44,20000	44,16184	,03816	,27131	,06182	,143282	,073610	,04034	,000115
5	44,40000	44,61114	-,21114	,55690	-,34203	,158977	,310142	-,22614	,004450
6	42,40000	42,36463	,03537	-,87105	,05730	,185124	,758736	,03887	,000178
7	45,40000	44,61114	,78886	,55690	1,27790	,158977	,310142	,84490	,062121
8	44,60000	45,06044	-,46045	,84250	-,74589	,182454	,709798	-,50452	,029176
9	45,20000	45,95905	-,75904	1,41368	-1,22960	,243179	1,998487	-,89847	,164369
10	41,70000	42,36463	-,66463	-,87105	-1,07665	,185124	,758736	-,73031	,062936
11	43,50000	42,81393	,68607	-,58546	1,11139	,161022	,342767	,73616	,048381
12	46,80000	46,40835	,39165	1,69927	,63445	,277428	2,887519	,49078	,063830
13	42,50000	41,91533	,58467	-1,15665	,94714	,214209	1,337830	,66471	,069808
14	41,90000	41,46602	,43398	-1,44224	,70302	,246519	2,080049	,51632	,055783
15	46,00000	45,50974	,49026	1,12809	,79418	,211132	1,272580	,55520	,047312
16	44,50000	43,71254	,78746	-,01428	1,27564	,138049	,000204	,82892	,045087
17	43,00000	43,71254	-,71254	-,01428	-1,15426	,138049	,000204	-,75005	,036915
18	41,90000	41,91533	-,01532	-1,15665	-,02482	,214209	1,337830	-,01742	,000048
19	43,50000	44,61114	-1,11114	,55690	-1,79998	,158977	,310142	-1,19007	,123247
20	42,00000	42,81393	-,81393	-,58546	-1,31851	,161022	,342767	-,87335	,068095
Minimum	41,10000	41,46602	-1,11114	-1,44224	-1,79998	,138049	,000204	-1,19007	,000048
Maximum	46,80000	46,40835	,78886	1,69927	1,27790	,277428	2,887519	,84490	,164369
Mean	43,73500	43,73500	,00000	-,00000	-,00000	,190593	,950000	,00435	,050707
Median	43,55000	43,71254	,03677	-,01428	,05956	,183789	,734267	,03960	,047847

Рисунок 5.13 – Таблица наблюдаемых, вычисленных значений и остатков

О нормальности остатков можно судить по графику остатков на нормальной вероятностной бумаге. Чем ближе распределение к нормальному виду, тем лучше значения остатков выстраиваются на прямой линии.

11. Для построения графика остатков на нормальной вероятностной бумаге нажмите кнопку Normal plot of resids.(см. рисунок 5.11). На экране появится график остатков (рисунок 5.14).

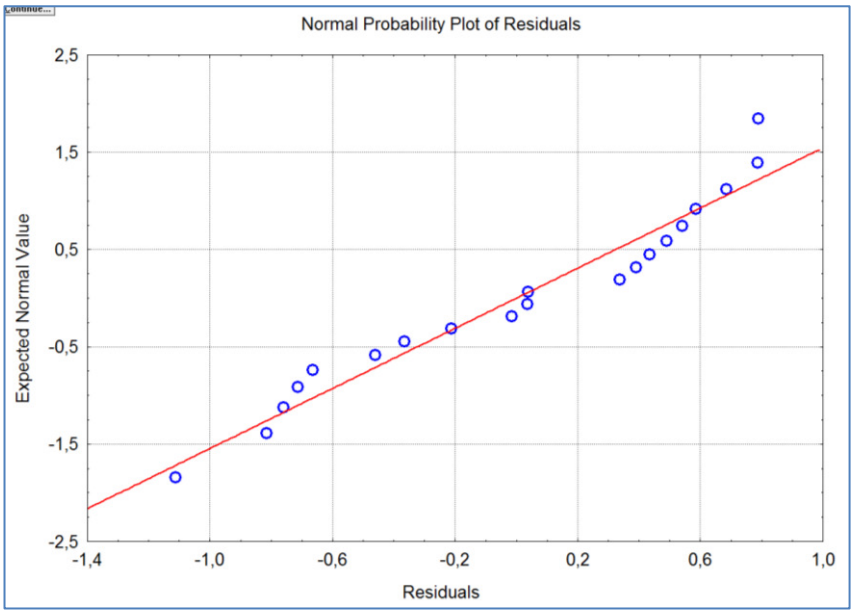


Рисунок 5.14 – График остатков построенный на нормальной вероятностной бумаге

По данным графика остатков на рисунке 5.14 сделаем **вывод:** чем ближе остатки расположены к прямой линии, тем качественнее регрессионная модель отражает исходные данные.

12. Рассчитайте значение массы 1000 семян при влажности равной 23%. Для этого в окне Multiple Regression Results нажмите кнопку Predict dependent var. На экране появится диалоговое окно Specify values for indep. vars (рисунок 5.15).

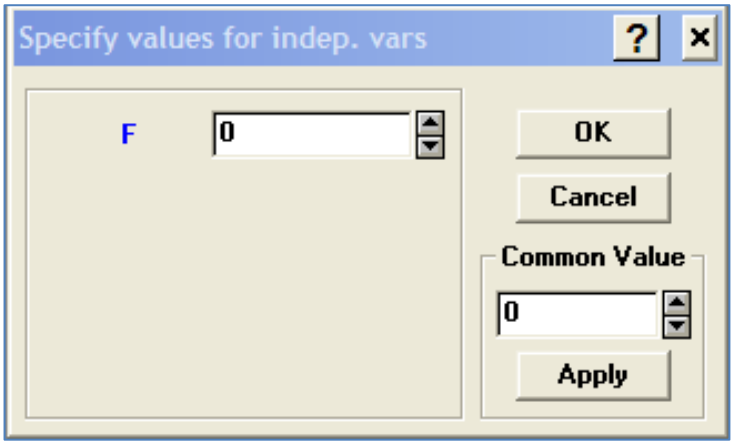


Рисунок 5.15 – Окно для вычисления значения отклика от произвольного значения независимой переменной по полученному уравнению регрессии

13. В появившемся окне в поле ввода значения переменной F введите 23 и нажмите OK. На экране появится окно с результатом вычислений, в котором по полученному уравнению регрессии вычислено значение отклика для F = 23 (рисунок 5.16). Масса 1000 семян при влажности 23% равна 45,95905 грамм.

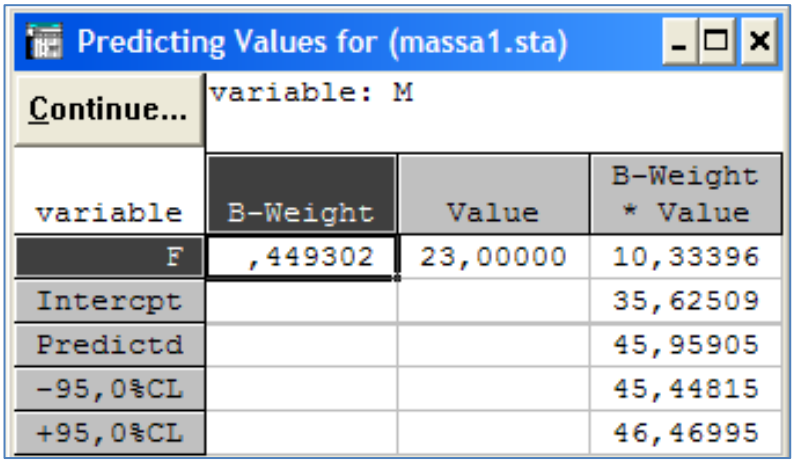


Рисунок 5.16 – Окно с результатом вычислений отклика от заданного значения независимой переменной

3.2 Нелинейный регрессионный анализ в модуле Multiple Regression

Рассмотрим выполнение нелинейного регрессионного анализа в модуле Multiple Regression на примере данных файла **RU_1.sta**.

Для выполнения нелинейного регрессионного анализа в модуле Multiple Regression необходимо в стартовом окне модуля в раскрывающемся списке Mode: выбрать Fixed non-linear (рисунок 5.17).

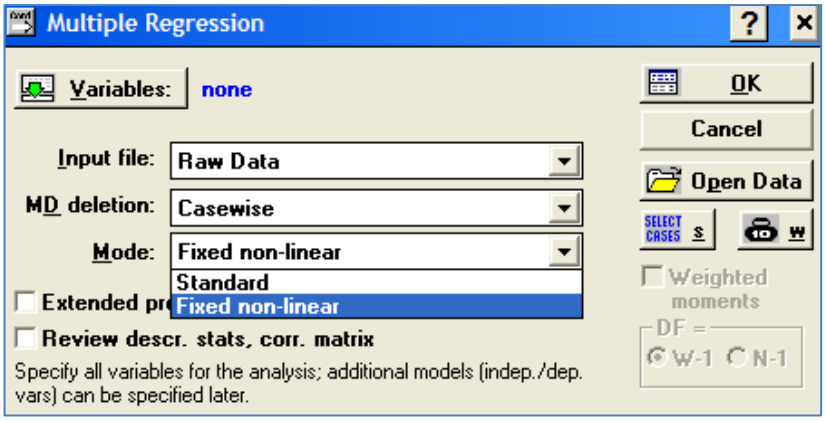


Рисунок 5.17 – Стартовая панель модуля Multiple Regression

Далее необходимо выбрать переменные, щелкнув по кнопке Variables. После нажатия кнопки OK в диалоговом окне Multiple Regressions, появляется окно Non-linear Components Regression, в котором можно выбрать виды регрессионных моделей и их совокупности. Например, выберем в качестве регрессионной модели полином второй степени (см. рисунок 5.18).

После того, как тип нелинейной регрессионной модели определен, необходимо нажать OK. На экране появится диалоговое окно Model Definition, в котором нужно уточнить зависимую и независимую переменные (см. рисунок 5.19).

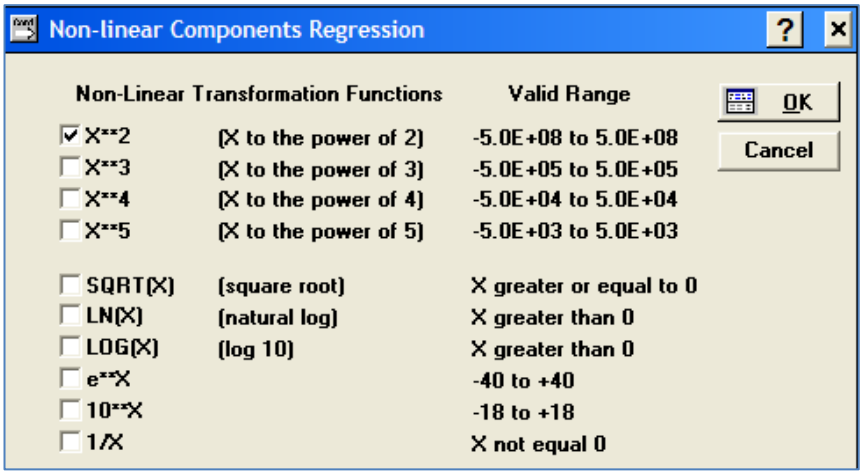


Рисунок 5.18 – Окно для выбора вида нелинейной регрессионной модели

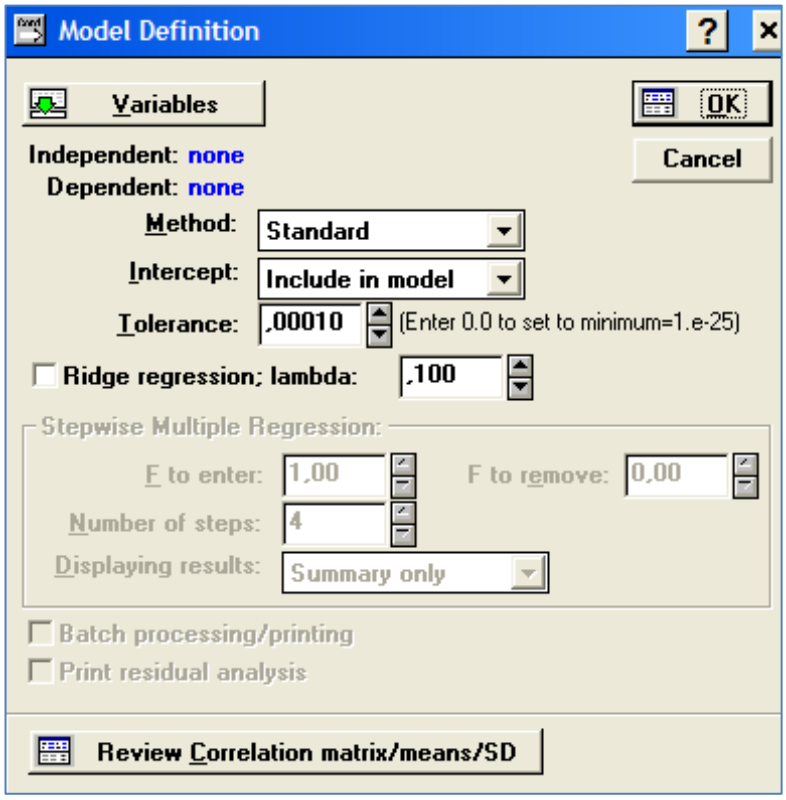


Рисунок 5.19 – Диалоговое окно Model Definition

Для этого нужно щелкнуть по кнопке Variables и выбрать в качестве зависимой переменной функцию отклика – ***Y***, а в качестве независимой переменной – ***X*** и переменную, соответствующую типу выбранной регрессионной модели – **V1**2** (полином второй степени) (см. рисунок 5.20). Затем щелкнуть по кнопке OK.

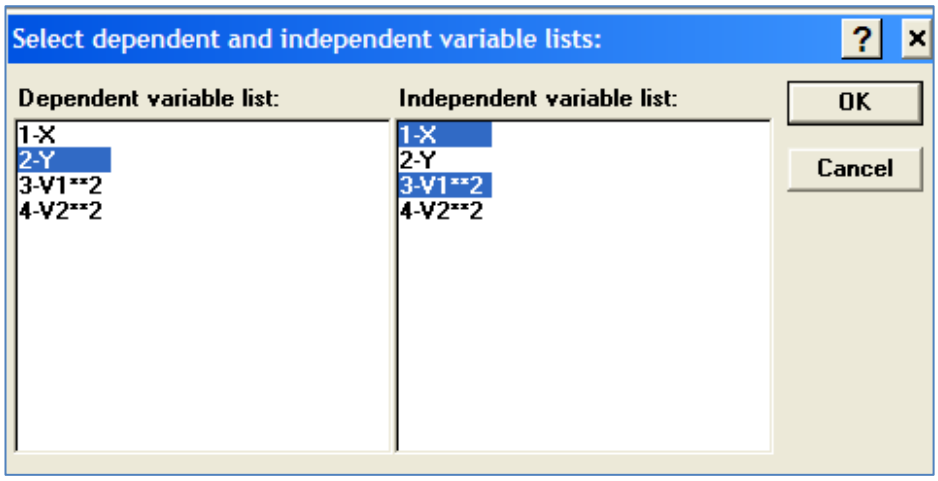


Рисунок 5.20 – Выбор зависимой и независимой переменных

Для просмотра результатов регрессионного анализа (рисунок 5.21) необходимо нажать кнопку Regression Summary.

Regression Summary for Dependent Variable: Y (ru_1.sta)

Continue...

R= ,95380125 RI= ,90973682 Adjusted RI= ,88394734
F(2,7)=35,276 p<,00022 Std.Error of estimate: ,62300

N=10	BETA	St. Err. of BETA	B	St. Err. of B	t(7)	p-level
Intercpt			3,989676	,640906	6,225056	,000435
X	,196620	,421339	,109584	,234830	,466655	,654922
V1**2	,762984	,421339	,032253	,017811	1,810854	,113073

Рисунок 5.21 – Результаты регрессионного анализа

Вывод о качестве полученной регрессионной модели можно сделать по значению коэффициента детерминации или по графику соответствия вычисленных и опытных данных.

Вывод по коэффициенту детерминации: *чем ближе к единице коэффициент детерминации, тем качественнее модель.*

Вывод по графику: *на графике опытные и вычисленные значения должны уложиться на линию регрессии. Чем больше точек лежит в стороне от линии регрессии и чем дальше от этой линии они находятся, тем хуже соответствие модели опытным данным.*

Поиск наилучшей регрессионной модели представляет собой довольно трудоемкий процесс, который заключается в последовательном построении различных видов моделей (полиномиальной, степенной, гиперболической и т.д.) и их сравнении.

Модель определенного вида может с различной точностью описывать опытные данные. Это зависит от количества коэффициентов, включаемых в регрессионную модель. Например, регрессионное уравнение вида $y = b_0 + b_1 \cdot x + b_3 \cdot x^3$ может лучше описывать данные, чем уравнение вида $y = b_0 + b_1 \cdot x + b_2 \cdot x^2 + b_3 \cdot x^3$. Для улучшения качества регрессионной

модели необходимо пошагово включать или исключать коэффициенты уравнения регрессии.

При помощи раскрывающегося списка Method (см. рисунок 5.19) можно отказаться от стандартного проведения регрессионного анализа (Standard) и воспользоваться методами пошагового включения переменных в регрессионную модель (Forward stepwise) или пошагового исключения переменных (Backward stepwise) из регрессионной модели. Опция Displaying results позволяет просматривать или же только итоговые результаты регрессионного анализа (Summary only) или после каждого шага включения или исключения переменных (At each step). Если необходимо получить регрессионную модель без свободного члена уравнения, тогда в списке поля Intercept нужно выбрать – Set to zero.

Рассмотрим построение уравнения регрессии методом пошагового включения коэффициентов.

После выбора переменных (см. рисунок 5.20) в раскрывающемся списке Method диалогового окна Model Definition нужно выбрать метод последовательного включения или исключения коэффициентов уравнения регрессии, например, Forward stepwise. В раскрывающемся списке Displaying results выбрать просмотр результатов пересчета коэффициентов уравнения регрессии после каждого шага включения или исключения коэффициентов (At each step) (рисунок 5.22).

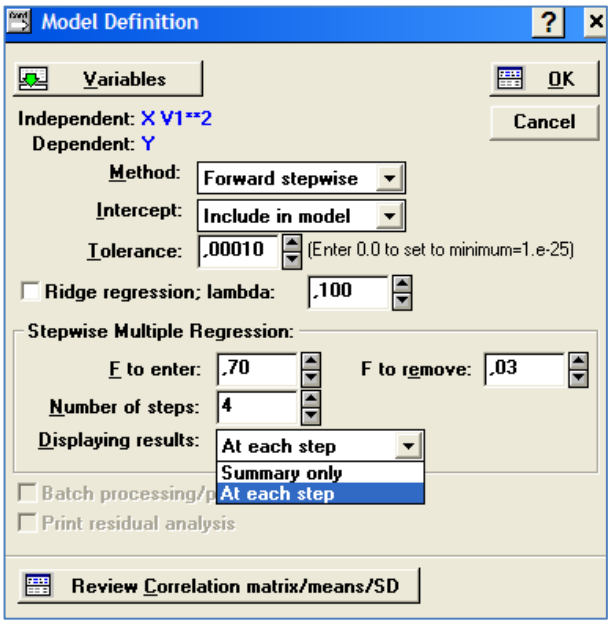


Рисунок 5.22 – Выбор метода включения / исключения коэффициентов уравнения регрессии и просмотра результатов

После нажатия кнопки OK на экране появится диалоговое окно результатов регрессионного анализа с нулевым шагом обработки (рисунок 5.23). На данном шаге (step 0) все статистики регрессионного уравнения равны нулю.

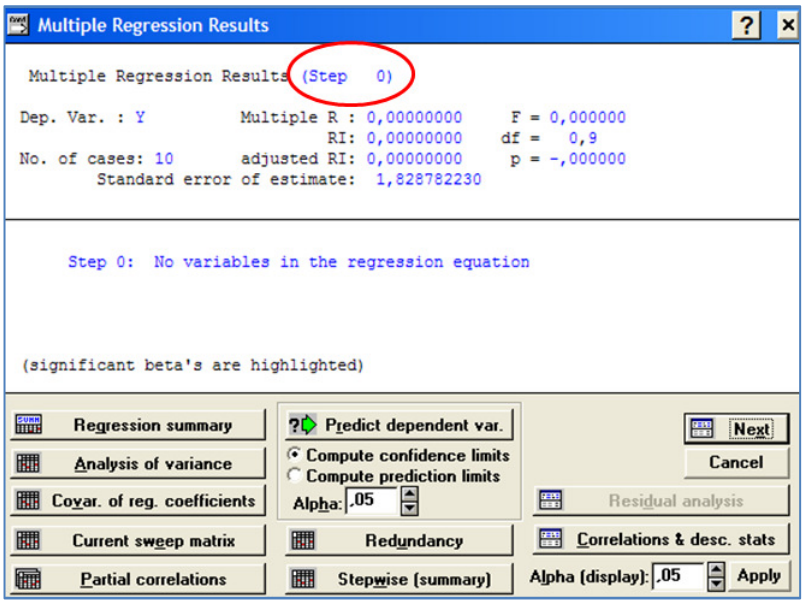


Рисунок 5.23 – Диалоговое окно результатов регрессионного анализа с нулевым шагом обработки

Для дальнейшего пересчета коэффициентов уравнения регрессии необходимо нажать кнопку Next. На экране появится диалоговое окно результатов регрессионного анализа с первым (следующим) шагом обработки (рисунок 5.24). На рисунке 5.24 указано, что это первый шаг обработки и, в данном случае, финальный, т.е. все статистики и коэффициенты для уравнения регрессии рассчитаны.

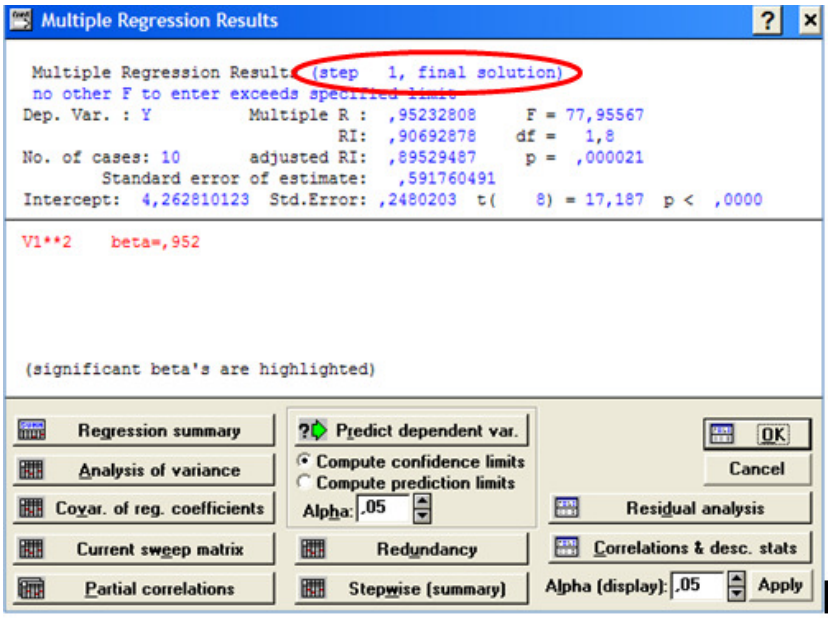


Рисунок 5.24 – Диалоговое окно результатов регрессионного анализа с первым шагом обработки

После нажатия кнопки OK на экране появятся результаты регрессионного анализа после пошагового включения коэффициентов уравнения регрессии (см. рисунок 5.25).

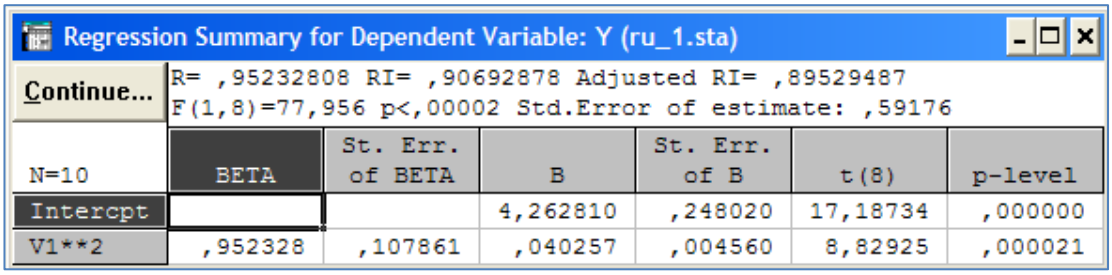


Рисунок 5.25 – Результаты регрессионного анализа после пошагового включения коэффициентов

Пересчет коэффициентов уравнения регрессии происходит автоматически. Количество шагов пересчета коэффициентов уравнения может быть разным. Это зависит от количества коэффициентов уравнения регрессии.

Сравните результаты регрессионного анализа стандартным методом и методом включения коэффициентов регрессионного уравнения (см. рисунки 5.21, 5.25). Выпишите полученные уравнения. Обратите внимание, что во втором случае в уравнении отсутствует член уравнения регрессии x, а оставшиеся члены уравнения регрессии имеют уточненные коэффициенты.

Для пошаговых методов регрессионного анализа важно установить величину Tolerance (толерантность) и величины частного F-критерия для включения в модель (F to enter) и исключения из нее (F to remove). Установив величину толерантности, создается барьер для включения в модель переменных, толерантность которых меньше установленной. Если величина толерантности переменной мала, то переменная несет малую дополнительную информацию и включение ее в модель нецелесообразно. Какая-либо новая независимая переменная, включаемая в модель, может сильно влиять на зависимую переменную, но если она включается в модель после других переменных, она может уже мало влиять на переменную отклика (например, из-за сильной коррелированности с переменными, уже включенными в модель). По умолчанию в пакете Statistica переменная включается в модель, если частный F-критерий больше или равен 1. Численное значение F-критерия для включения никогда не выбирается меньше, чем численное значение F-критерия для исключения.

Необходимо помнить, что при поиске лучшей регрессионной модели следует руководствоваться следующими наиболее общими требованиями:

1. Регрессионная модель должна объяснять не менее 75% вариации зависимой переменной, т.е. $R^2 > 0{,}75$.
2. Коэффициенты уравнения регрессии и его свободный член должны быть значимы на 5%-ом уровне.
3. Остатки от регрессии должны быть без заметной автокорреляции ($r < 0{,}3$), нормально распределены и без систематической составляющей. Чем меньше сумма квадратов остатков, чем меньше стандартная ошибка оценки и чем больше R^2, тем лучше уравнение регрессии.

Одним из недостатков классического регрессионного анализа, в основе которого лежит метод наименьших квадратов, является недостаточная устойчивость к изменениям входной информации.

Сейчас довольно широко стали применяться альтернативные регрессионные модели, одной из которых является *гребневая регрессия*, которая отличается устойчивостью для случаев сильной коррелированности зависимых переменных друг с другом. В отличие от метода наименьших квадратов, дающего несмещенные оценки коэффициентов уравнения, в методе гребневой регрессии оценки смещенные, но при этом они имеют меньшую дисперсию. Поэтому такие оценки могут давать более точные и приемлемые для практического использования модели.

Для расчета гребневой регрессии следует установить флажок в опции Ridge regression диалогового окна Model Definition (см. рисунок 5.22).

При практическом использовании метода гребневой регрессии одним из основных вопросов является выбор параметра λ lambda). Существует несколько численных методов расчета параметра, но чаще используют простой эмпирический подход: выбирают такой параметр, при котором коэффициенты стабилизируются и при дальнейшем увеличении параметра изменяются мало. Значение принятого параметра является мерой смещения оценок от истинного значения, поэтому стараются не придавать λ слишком больших значений. Обычно λ выбирают меньше 0,5, а шаг при подборе выбирают небольшим, например, 0,02.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Создайте новую таблицу, содержащую две переменные X и Y по 10 случаев.

2. Заполните новую таблицу данными в соответствии с таблицей 5.5. Сохраните файл с именем **RU_1.sta**.

3. Выполните регрессионный анализ исходных данных и постройте уравнение линейной регрессии. Сделайте анализ остатков. Результаты регрессионного анализа сохраните в отчете с именем **RU_1.docx**.

Таблица 5.5 – Исходные данные

x_i	1	2	3	3	4	5	6	7	8	12
y_i	4	4	4	6	5	5	6	6	7	10

4. В модуле Multiple Regression создайте файл исходных данных в соответствии с таблицей 5.6. В таблице представлена зависимость урожайности ячменя от дозы внесения минеральных удобрений.

Таблица 5.6 – Зависимость урожайности ячменя от дозы внесения минеральных удобрений

D, кг/га	280	500	150	130	250	450	550	320	180	100	200	470	350	230	420	600
У, ц/га	58,2	63,1	50,6	45,4	58,5	62,4	65,6	62,3	50,8	43,1	55	66,7	60,2	55,8	64,2	66,9

5. Сохраните файл с именем **NlinRegr.sta**.
6. Подберите модель, описывающую наилучшим образом опытные данные.
7. При построении различных моделей используйте метод пошагового включения коэффициентов.
8. Для оценки качества и адекватности модели используйте коэффициент детерминации и анализ остатков.
9. В модуле Multiple Regression создайте файл исходных данных в соответствии с таблицей 5.7. В таблице представлена зависимость урожайности ячменя Y (ц/га) от дозы внесения азотных X_1 и фосфорных X_2 удобрений (кг/га).

Таблица 5.7 – Зависимость урожайности ячменя $\boldsymbol{Y}$ (ц/га) от дозы внесения азотных $\boldsymbol{X_1}$ и фосфорных $\boldsymbol{X_2}$ удобрений (кг/га)

X_1	X_2	Y
200	240	33,264
350	100	32,85
450	220	37,884
150	140	29,64
550	120	40,128
300	280	38,64
500	260	43,824
600	160	43,89
100	180	29,016
400	300	44,1
250	200	32,64

10. Сохраните файл с именем **MultRegr.sta**.
11. Подберите регрессионную модель, описывающую наилучшим образом опытные данные.
12. При построении различных моделей используйте метод пошагового включения коэффициентов. Для оценки качества модели используйте коэффициент детерминации и анализ остатков.
13. Постройте поверхность отклика полученной регрессионной модели.

5 СОДЕРЖАНИЕ ОТЧЕТА

Отчет оформляется в соответствии с требованиями, принятыми в АЧИИ. Отчет должен содержать:

- ответы на контрольные вопросы;
- результаты регрессионного анализа.

6 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Для чего предназначен регрессионный анализ?
1. Какая регрессия называется парной, множественной?
2. Какие этапы включает в себя общая схема регрессионного анализа?
3. Какая диаграмма называется диаграммой рассеяния?
4. По какому критерию оценивается значимость коэффициентов уравнения регрессии? Сформулируйте нулевую гипотезу.
5. По какому критерию оценивается адекватность уравнения регрессии? Сформулируйте нулевую гипотезу.
6. Как оценивается качество регрессионной модели?
7. В каком модуле программы Statistica проводится линейный регрессионный анализ?
8. Чем отличается построение линейного регрессионного уравнения от нелинейного в программе Statistica?
9. Перечислите нелинейные регрессионные модели.
10. В каком модуле программы Statistica выполняется построение нелинейной регрессионной модели?
11. Опишите метод пошагового включения коэффициентов.
12. Перечислите основные модели множественной регрессии.

Практическое занятие № 6

КОНТРОЛЬНАЯ РАБОТА № 1

«Применение методов и способов решения исследовательских задач. Дисперсионный и регрессионный анализ»

Задача 1. Выполните дисперсионный анализ данных в программе Statistica.
1. Подготовьте таблицу 1 и проведите дисперсионный анализ для переменных Опыт1, Опыт3 и Контроль.
2. Результаты сохраните в файле отчета с именем Kreport.doc. Отчет должен содержать средние значения вариантов, таблицу результатов дисперсионного анализа, расчет значения НСР, выводы.

Таблица 1 – Высота растений ячменя

Повтор	Опыт1	Опыт2	Контроль	Опыт3	Опыт4
Повтор1	80	83	78	83	80
Повтор1	88	77	85	81	86
Повтор1	89	82	86	83	87
Повтор1	92	87	75	77	85
Повтор1	84	77	85	77	79
Повтор1	89	84	80	75	91
Повтор1	76	77	86	69	85
Повтор1	93	81	81	86	84
Повтор1	81	91	79	77	82
Повтор1	87	78	78	85	83
Повтор2	91	80	82	81	90
Повтор2	89	80	81	75	81
Повтор2	87	84	87	80	86
Повтор2	87	72	85	80	83
Повтор2	92	80	85	81	94
Повтор2	89	81	70	80	82
Повтор2	92	75	81	83	87
Повтор2	80	80	77	75	86
Повтор2	92	81	84	81	91
Повтор2	92	81	81	80	86
Повтор3	79	80	81	90	82
Повтор3	90	89	70	91	90
Повтор3	82	82	74	85	70
Повтор3	79	93	84	87	85
Повтор3	81	89	81	85	82
Повтор3	85	94	75	86	86
Повтор3	84	76	86	92	78
Повтор3	78	87	80	84	82
Повтор3	91	80	83	90	81
Повтор3	77	90	75	80	81

Задача 2. Подберите уравнение регрессии по данным, представленным в таблице, с использованием программы Statistica.

1. Подготовьте таблицу 2 и проведите регрессионный анализ данных.
2. Результаты регрессионного анализа сохраните в файле отчета. Отчет должен содержать:

– коэффициенты уравнения регрессии и коэффициент детерминации полученной модели;

– оценку значимости коэффициентов регрессии полученной модели;

– выводы об адекватности полученной модели;

– графики остатков на нормальной вероятностной бумаге полученной модели.

Таблица 2 – Данные для регрессионного анализа

x	0,32	0,08	1,02	0,19	0	1,06	0,26	0,16	0,45	1,58	1,32	1,47	0,55	0,63	1,89
y	2,62	1,05	7,21	1,9	0,7	7,55	2,34	1,58	3,5	10,74	9,24	10,15	4,08	4,77	12,88

Практическое занятие № 7

Тема: **ПЛАНИРОВАНИЕ НАУЧНОГО ЭКСПЕРИМЕНТА. ПЛАНЫ ПЕРВОГО И ВТОРОГО ПОРЯДКОВ**

1 ЦЕЛЬ РАБОТЫ

Изучить:

- основные понятия планирования эксперимента;
- планы эксперимента первого порядка;
- планы эксперимента второго порядка.

Научиться генерировать планы первого и второго порядков в программе Statistica.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Планирование эксперимента. Основные понятия и определения

Под *экспериментом* понимают совокупность операций совершаемых над объектом исследования с целью получения информации о свойствах исследуемого объекта.

Эксперимент, в котором исследователь по своему усмотрению может изменять условия его проведения, называется *активным экспериментом*. Если исследователь не может самостоятельно изменять условия его проведения, а лишь регистрирует их, то это *пассивный эксперимент*.

Важнейшей задачей методов обработки полученной в ходе эксперимента информации является задача построения математической модели изучаемого явления, процесса, объекта.

Полученную математическую модель можно использовать при анализе процессов и при проектировании объектов. Можно получить хорошо аппроксимирующую математическую модель, если целенаправленно применяется активный эксперимент.

Другой задачей обработки полученной в ходе эксперимента информации, является задача оптимизации, т.е. нахождения такой комбинации влияющих независимых переменных, при которой выбранный показатель оптимальности принимает экстремальное значение.

Опыт – метод исследования, направленный на решение конкретной задачи. Цель – подтверждение имеющийся гипотезы, а также сбор данных. Опыт проводится в естественных условиях.

План эксперимента – совокупность данных определяющих число, условия и порядок проведения опытов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям, совокупность действий направленных на разработку стратегии экспериментирования (от получения априорной ин-

формации до получения работоспособной математической модели или определения оптимальных условий).

Планирование эксперимента заключается в выборе такой стратегии экспериментирования, которая позволяет принимать обоснованные решения после каждой серии опытов. Оно позволяет определить заранее схему пошагового процесса проведения эксперимента, включить в него минимальное число опытов при одновременном варьировании всеми факторами без снижения количества и качества полученной информации.

В процессе измерений, последующей обработки данных, а также формализации результатов в виде математической модели, возникают погрешности и теряется часть информации, содержащейся в исходных данных. Применение методов планирования эксперимента позволяет определить погрешность математической модели и судить о ее адекватности. Если точность модели оказывается недостаточной, то применение методов планирования эксперимента позволяет модернизировать математическую модель с проведением дополнительных опытов без потери предыдущей информации и с минимальными затратами.

Цель планирования эксперимента – нахождение таких условий и правил проведения опытов, при которых удается получить надежную и достоверную информацию об объекте с наименьшими затратами труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности.

Пусть интересующее нас свойство Y объекта зависит от n независимых переменных $X_1, X_2, \dots, X_n$ и мы хотим выяснить характер этой зависимости – $Y = F(X_1, X_2, \dots, X_n)$, о которой мы имеем лишь общее представление.

Величина Y называется *откликом*, а сама зависимость $Y = F(X_1, X_2, \dots, X_n)$ – *функцией отклика (целевой функцией)*.

Независимые переменные $X_1, X_2, \dots, X_n$, иначе *факторы*, также должны иметь количественную оценку. Для построения эффективной математической модели целесообразно провести предварительный анализ значимости факторов (степени влияния на функцию), их ранжирование и исключить малозначащие факторы.

Факторы должны быть управляемыми и непосредственно воздействовать на объект.

Под *управляемостью* подразумевается возможность установки и поддержания выбранного нужного уровня факторов постоянным в течение всего опыта или его изменения по заданной программе.

Требования к факторам:

– *совместимость* – осуществимость и безопасность всех запланированных комбинаций факторов;

– *независимость* – возможность установления факторов на любом уровне вне зависимости от уровней других.

Геометрическое представление функции отклика в факторном пространстве $X_1, X_2, \dots, X_n$ называется *поверхностью отклика*.

Диапазоны изменения факторов задают область определения Y. Если принять, что каждому фактору соответствует координатная ось, то полученное пространство называется *факторным пространством*. При $n = 2$ область определения Y представляет собой прямоугольник, при $n = 3$ – куб, при $n > 3$ – гиперкуб.

При выборе диапазонов изменения факторов нужно учитывать их совместимость, т.е. контролировать, чтобы в этих диапазонах любые сочетания факторов были бы реализуемы в опытах и не приводили бы к абсурду. Для каждого из факторов указывают граничные значения

$$X_{i\,min} \le X_i \le X_{i\,max}, \qquad i = 1, \dots, n.$$

Регрессионный анализ функции отклика предназначен для получения её математической модели в виде *уравнения регрессии*

$$Y = B_0 + \sum_{i=1}^{n} B_i X_i + \sum_{1 \le i < j}^{n} B_{ij} X_i X_j + \sum_{i=1}^{n} B_{ii} X_i^2 + \cdots,$$

где $B_0, B_1, \dots, B_m$ – некоторые коэффициенты.

Таким образом, формальной целью эксперимента является определение численных значений коэффициентов уравнения регрессии, а для определения оптимальных условий протекания процессов находят значения факторов $X_1, X_2, \dots, X_n$, соответствующих экстремуму целевой функции.

Нахождение уравнения регрессии методом планирования экспериментов состоит из следующих этапов:

- *выбор основных факторов эксперимента;*
- *планирование и проведение эксперимента;*
- *определение коэффициентов уравнения регрессии;*
- *статистический анализ результатов эксперимента.*

Среди основных методов планирования, применяемых на разных этапах исследования, используют:

- планирование отсеивающего эксперимента, основное значение которого, выделение из всей совокупности факторов группы существенных факторов, подлежащих дальнейшему детальному изучению;
- планирование эксперимента для дисперсионного анализа, т.е. составление планов для объектов с качественными факторами;
- планирование регрессионного эксперимента, позволяющего получать регрессионные модели (полиномиальные и иные);
- планирование экстремального эксперимента, в котором главная задача – экспериментальная оптимизация объекта исследования;
- планирование при изучении динамических процессов и т.д.

При использовании методов планирования эксперимента необходимо найти ответы на 4 вопроса:

1. Какие сочетания факторов и сколько таких сочетаний необходимо взять для определения функции отклика?
2. Как найти коэффициенты $B_0, B_1, \dots, B_m$?
3. Как оценить точность представления функции отклика?
4. Как использовать полученное представление для поиска оптимальных значений целевой функции Y?

2.2 Разложение функции отклика в степенной ряд. Кодирование факторов

Если заранее неизвестно аналитическое выражение функции отклика, то можно рассматривать не саму функцию, а её разложение, например, в степенной ряд в виде полинома

$$Y = B_0 + B_1X_1 + \cdots + B_nX_n + B_{12}X_1X_2 + \cdots + B_{n\,n-1}X_nX_{n-1} + B_{11}X_1^2 + \cdots + B_{n\,n}X_n^2$$

Разложение в степенной ряд функции возможно в том случае, если сама функция является непрерывной и гладкой. На практике обычно ограничиваются числом членов степенного ряда и аппроксимируют функцию полиномом некоторой степени (чаще второй).

Функция отклика может зависеть от факторов, имеющих разные размерности и резко отличающихся количественно. Поэтому в теории планирования эксперимента часто используют кодирование факторов.

Эта операция заключается в выборе нового масштаба для кодированных факторов, причем такого, чтобы минимальное значение кодированных факторов соответствовало (–1), а максимальное значение – (+1), а также в переносе начала координат в точку с координатами $X_{1\text{ср}}, X_{2\text{ср}}, \dots, X_{n\text{ср}}$.

Используя формулу

$$X_{i\text{ср}} = \frac{X_{i\,max} - X_{i\,min}}{2},$$

получим текущее значение кодированного фактора

$$x_i = \frac{X_i - X_{i\text{ср}}}{X_{i\text{ср}} - X_{i\,min}} = \frac{X_i - X_{i\text{ср}}}{X_{i\,max} - X_{i\text{ср}}} = \frac{2X_i - X_{i\,max} - X_{i\,min}}{X_{i\,max} - X_{i\,min}},$$

где X_i – именованное (абсолютное) значение фактора;

x_i – кодированное значение фактора;

$X_{i\text{ср}} - X_{i\,min} = X_{i\,max} - X_{i\text{ср}}$ – интервал варьирования фактора.

Если фактор изменяется дискретно, то каждому уровню этого кодированного фактора присваиваются числа в диапазоне от +1 до –1. Например, при двух уровнях это +1 и –1, при трех уровнях +1, 0, –1 и т.д.

Функция отклика может быть выражена через кодированные факторы $Y = f(x_1, x_2, \dots, x_n)$ и записана в полиномиальном виде:

$$Y = b_0 + b_1x_1 + b_2x_2 + \cdots + b_nx_n + b_{12}x_1x_2 + \cdots + b_{n\,n-1}x_nx_{n-1} + b_{11}x_1^2 + \cdots + b_{n\,n}x_n^2.$$

Очевидно, что $B_i \neq b_i$, но ***регрессионные уравнения в именованных и кодированных факторах должны быть одного вида и значения функции отклика, вычисленные по обоим уравнениям, должны быть равны***:

$$Y = F(X_1, X_2, \dots, X_n) = f(x_1, x_2, \dots, x_n).$$

Для полинома, записанного в кодированных факторах, степень влияния факторов или их сочетаний на функцию отклика определяется величиной их коэффициента b_i.

Для полинома в именованных факторах величина коэффициента B_i еще не говорит однозначно о степени влияния этого фактора или их сочетаний на функцию отклика.

Полный полином второй степени для двух факторов выглядит следующим образом:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_{12}X_1X_2 + B_{11}X_1^2 + B_{22}X_2^2$$

2.3 Планы первого порядка. Планы полного факторного эксперимента 2^n (планы ПФЭ 2^n)

Под *планом первого порядка* понимают такие планы эксперимента, которые позволяют найти уравнение регрессии, содержащее только первые степени факторов и их произведения:

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{u,i=1(u\neq i)}^{k} b_{ui}x_u x_i + \sum_{g,u,i=1(g\neq u\neq i)}^{k} b_{gui}x_g x_u x_i + \cdots$$

Планы второго порядка позволяют найти уравнение регрессии, содержащее и вторые степени факторов:

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{u,i=1(u\neq i)}^{k} b_{ui}x_u x_i + \sum_{g,u,i=1(g\neq u\neq i)}^{k} b_{gui}x_g x_u x_i + \sum_{i=1}^{k} b_{ii}x_i^2 + \cdots$$

Планы ПФЭ 2^n являются простейшими планами первого порядка. Основание 2 означает, что принято два уровня варьирования, на которых фиксируются факторы (–1, +1), n – число факторов.

Для плана ПФЭ 2^2 число факторов равно двум ($n = 2$) и число уровней фиксирования факторов также равно 2. Значения кодированных факторов выбираются в виде +1 и –1. Полное число возможных сочетаний значений n факторов (число опытов, а значит и число строк плана) $N = 2^2 = 4$. В матри-

це планирования эксперимента будет 4 строки и, соответственно, необходимо провести 4 опыта в эксперименте.

Для плана ПФЭ 2^3 число факторов $n = 3$ выполняется $N = 2^3 = 8$ опытов.

Планы ПФЭ используются только в тех случаях, когда:

a) количество проводимых опытов достаточно небольшое, т.к. опыты необходимо воспроизводить минимум дважды;

b) все факторы контролируемы на уровнях варьирования.

2.4 Матричное представление факторов и результатов эксперимента

Матричная запись результатов N опытов для полиномиального представления результата может быть представлена в виде

$$Y_U = \sum_{i=0}^{m} B_i X_{iU}$$

где $Y = X \cdot B$; Y – матрица-столбец результатов опыта; B – матрица-столбец коэффициентов полинома; X – матрица сочетаний факторов.

$$X = \begin{vmatrix} X_{01} & X_{11} \dots & X_{i1} \dots & X_{m1} \\ X_{02} & X_{12} \dots & X_{i2} \dots & X_{m2} \\ \dots & \dots & \dots & \dots \\ X_{0U} & X_{1U} \dots & X_{1U} \dots & X_{mU} \\ \dots & \dots & \dots & \dots \\ X_{0N} & X_{1N} \dots & X_{iN} \dots & X_{mN} \end{vmatrix}, Y = \begin{vmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_U \\ \dots \\ Y_N \end{vmatrix}, B = \begin{vmatrix} B_0 \\ B_1 \\ \dots \\ B_i \\ \dots \\ B_m \end{vmatrix}.$$

Здесь $0, 1, \dots, i, \dots, m$ – номера членов уравнения; $1, \dots, U, \dots, N$ ... – номера опытов. Матрица X – прямоугольная, содержащая $(m + 1)$ столбцов и N строк.

Если учесть, что в матрице X элементы $X_{0U} = 1, U = 1, \dots, N$, то матрицу X можно записать таким образом:

$$X = \begin{vmatrix} 1 & X_{11} \dots & X_{i1} \dots & X_{m1} \\ 1 & X_{12} \dots & X_{i2} \dots & X_{m2} \\ 1 & \dots & \dots & \dots \\ 1 & X_{1U} \dots & X_{1U} \dots & x_{mU} \\ \dots & \dots & \dots & \dots \\ 1 & X_{1N} \dots & X_{iN} \dots & X_{mN} \end{vmatrix}.$$

Полученные коэффициенты B позволяют сформировать уравнение функции отклика при $(m + 1)$ членах уравнения. Если точность этого уравнения оказалась недостаточной, то требуется взять уравнение с большим числом членов и начать все заново, так как все коэффициенты B оказываются

зависимыми друг от друга. Это возникает при использовании пассивного эксперимента. Однако если целенаправленно использовать активный эксперимент и особым образом построить матрицу сочетаний факторов в опытах X, использовать планирование эксперимента, то коэффициенты полинома определяются независимо друг от друга.

Стратегия применения планов заключается в принципе постепенного планирования – постепенного усложнения модели. Начинают с простейшей модели, находятся для нее коэффициенты, определяется ее точность. Если точность не удовлетворяет, то планирование и модель постепенно усложняются.

Задача планирования заключается в том, как нужно строить матрицу X, чтобы коэффициенты B определялись независимо друг от друга. Эти требования выполняются, если матрица для вычисления коэффициентов уравнения является диагональной, то есть все элементы, расположенные не на главной диагонали матрицы, равны нулю

$$C_{ij} = 0;\ \ i \neq j;\ \ i = 0, 1, 2, \dots, m;\ \ j = 0, 1, 2, \dots, m$$

или

$$C = \begin{vmatrix} C_{00} & 0 \dots & 0 \dots & 0 \\ 0 & C_{11} \dots & 0 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 \dots & C_{ii} \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 \dots & 0 \dots & C_{mm} \end{vmatrix}.$$

Так как любое скалярное произведение двух различных столбцов в матрице X должно быть равно нулю, то это условие называется *условием ортогональности матрицы* X, а соответствующее планирование эксперимента (определение сочетаний факторов) называется *ортогональным планированием*.

Для ортогонального планирования при учете того, что $X_{0U} = 1$, $U = 1, \dots, N$

$$\sum_{U=1}^{N} X_{0U} \cdot X_{iU} = \sum_{U=1}^{N} X_{iU} = 0.$$

Таким образом, при ортогональном планировании сумма элементов любого столбца матрицы X, кроме первого столбца должна быть равна нулю. Это правило показывает, что в ортогональном планировании при четном числе уровней, на которых фиксируется каждый фактор, эти уровни должны быть симметрично расположены относительно центральной точки $x = 0$, при нечетном числе уровней должна использоваться и центральная точка (см. рисунок 7.1).

Кроме свойства ортогональности план может обладать свойством насыщенности, ротатабельности и др.

План является насыщенным, если общее число опытов N равняется числу неизвестных коэффициентов полинома $(m+1)$.

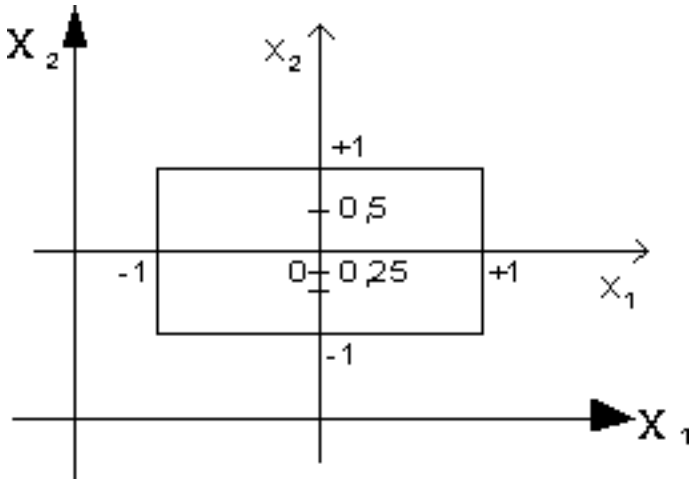


Рисунок 7.1 – Выбор уровней варьирования
при ортогональном планировании

План называется *ротатабельным*, если дисперсия отклика одинакова на одном расстоянии от центра плана при любом направлении в факторном пространстве. В упрощенном виде это означает, что все точки плана лежат на окружности (сфере, гиперсфере).

2.5 Планы дробного факторного эксперимента (планы ДФЭ)

При многофакторном эксперименте, особенно когда число факторов больше шести ($n > 6$), число опытов планов ПФЭ 2^n ($N = 2^n$) становится чрезмерным. Если не требуется определение всех коэффициентов неполного квадратичного полинома, то переходят к дробному факторному эксперименту (ДФЭ) – части полного факторного эксперимента.

Так, например, если требуется определить лишь коэффициенты при самих факторах

$$Y = B_0 + B_1X_1 + B_2X_2 + \cdots + B_nX_n,$$

то план ПФЭ 2^n дает избыточную информацию. Например, при $n = 6$, требуется определить $n + 1 = 7$ коэффициентов, тогда как по плану ПФЭ необходимо провести $N = 2^6 = 64$ опыта.

Хотя эта избыточная информация не является бесполезной, она позволяет более точно определить коэффициенты, но все же часто используют планы ДФЭ 2^{n-k}, где k – показатель дробности плана ПФЭ (число эффектов факторов, которые смешаны с эффектами взаимодействия).

При $k = 1$ число опытов в плане ДФЭ в два раза меньше, чем в плане ПФЭ, поэтому такие планы называют *полуреплика плана ПФЭ*.

При $k = 1$ для плана ДФЭ 2^{6-1} $N = 2^{6-1} = 32$.

При $k = 2$ для плана ДФЭ 2^{6-2} $N = 2^{6-2} = 16$ и такой план называют *четвертьрепликой.*

При $k = 3$ для плана ДФЭ 2^{6-3} $N = 2^{6-3} = 8$.

При выборе дробности плана k необходимо учитывать, что число опытов должно быть больше числа членов уравнения.

Достоинство планов ДФЭ заключается и в том, что если построенный на его основе неполный полином не удовлетворяет требованиям по точности, то план ДФЭ легко достраивается до плана ПФЭ, без потери информации прежних опытов, с формированием более точного полинома.

Как показывает опыт экспериментальной работы, полуреплики от ПФЭ целесообразно получать лишь при числе факторов более трех.

2.6 Планы второго порядка. Основные понятия и определения

Планы второго порядка позволяют сформировать функцию отклика в виде полного полинома второй степени, который содержит большее число членов, чем неполный полином второй степени, сформированный по планам первого порядка, и поэтому требуют большего числа выполняемых опытов.

Полный квадратичный полином при n=2 содержит 6 членов:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_{12}X_1X_2 + B_{11}X_1^2 + B_{22}X_2^2$$

При n=3 квадратичный полином содержит 11 членов:

$$Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_{12}X_1X_2 + B_{13}X_1X_3$$
$$+ B_{23}X_2X_3 + B_{123}X_1X_2X_3 + B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2$$

Известно, что для получения квадратичной зависимости каждый фактор должен фиксироваться как минимум на трех уровнях.

Наиболее распространенными планами второго порядка являются планы типа 3^{n-k}*, планы Бокса-Бенкина и центральные композиционные планы.*

Наиболее полную информацию об объекте исследования можно получить, используя планы 3^{n-k}. Эти планы используют все сочетания трех уровней каждого фактора. Они довольно громоздки (имеют большое количество опытов) и зачастую содержат избыточную информацию.

Регрессионные модели второго порядка можно получить, если проделать меньшее число опытов при помощи трехуровневых планов Бокса-Бенкина.

Главной отличительной чертой этих планов является тот факт, что все факторы, кроме двух (для 3–5 факторов) или трех (для 6–9 факторов), фиксируются на нулевом уровне.

Планы Бокса-Бенкина строятся на гиперкубе, т.е. все факторы имеют уровни –1, 0, 1. Это бывает полезно при изучении некоторых явлений.

Если уже был ранее сформирован план ПФЭ 2^n, но точность его функции отклика не удовлетворяет, то мы можем достроить этот план до плана второго порядка (композиционный план) и сформировать функцию отклика в

виде полного полинома второй степени, без потери информации о ранее сделанных опытах.

Для описания поверхности отклика полиномом второй степени «ядро» плана достраивается некоторым количеством «звездных» точек, расположенных от центра эксперимента на расстоянии «звездного» плеча α. *Такие планы называются композиционными или последовательными*.

Кроме того, для оценки кривизны поверхности отклика добавляют еще одну точку или несколько точек в центре плана (точка ноль), поэтому *план будет центральным*, т.к. он симметричен относительно центра.

Центрально-композиционные планы могут быть ротатабельными и ортогональными.

2.7 Ортогональные центрально-композиционные планы второго порядка (ОЦКП)

Ортогональным планом называется такой план, у которого матрица планирования X строится так, чтобы матрица коэффициентов оказалась диагональной.

Этот подход используется при построении планов второго порядка.

План называется центральным, если все точки расположены симметрично относительно центра плана.

ОЦКП – ортогональный центрально-композиционный план (симметричный). В ОЦКП входят: ядро-план ПФЭ с $N_0 = 2^n$ точками плана, $n_0 = 1$ (одна для этого плана) центральная точка плана $(x_i = 0,\ \ i = 1, 2, 3, \dots, n)$ и по две «звездные» точки для каждого фактора

$$x_i = \pm\alpha, \qquad x_j = 0, \qquad i = 1, \dots, n, \qquad j = 1, \dots, n, \qquad i \neq j,$$

α – плечо «звездных» точек (расстояние от центральной точки до звёздной).

Общее количество опытов в плане ОЦКП определяется выражением

$$N = 2^n + 2 \cdot n + 1,$$

где 2^n – число опытов, составляющих ПФЭ при двух уровнях варьирования;

$2 \cdot n$ – число «звездных» точек.

В ОЦКП каждый фактор фиксируется, в общем случае, на пяти уровнях $(-\alpha, -1, 0, +1, +\alpha)$.

Значение плеча α для $n = 2, 3, 4, 5$ приведено в таблице 7.1.

Таблица 7.1 – Значение плеча $\boldsymbol{\alpha}$ для ОЦКП

n	2	3	4	5
ПФЭ (или ДФЭ)	2^2	2^3	2^4	2^5
α	1,000	1,215	1,414	1,547

Если пренебрегать эффектами взаимодействия, то можно реализовать дробный факторный эксперимент и тем самым сократить число опытов.

При $n > 2$ в ОЦКП оказывается меньшее количество точек, чем в плане ПФЭ 3^n.

В таблице 7.2 для сравнения приведено число опытов в ОЦКП и ПФЭ 3^n.

Таблица 7.2 – Число опытов в ОЦКП и ПФЭ $\mathbf{3^n}$

n	2	3	4	5	6
ОЦКП	9	15	25	43	77
ПФЭ 3^n	9	27	81	243	729

По результатам опытов плана ОЦКП формируется полином вида:

$$Y = B_0' + B_1X_1 + B_2X_2 + B_3X_3 + B_{12}X_1X_2 + B_{13}X_1X_3$$
$$+ B_{23}X_2X_3 + B_{123}X_1X_2X_3 + B_4X_1^2 + B_5X_2^2 + B_6X_3^2$$

2.8 Ротатабельные центрально-композиционные планы второго порядка (РЦКП)

Ротатабельным называется план, в котором полученная информация о поверхности отклика является одинаковой для всех направлений (факторов) в точках, удаленных на одинаковое расстояние от центра эксперимента.

Ротатабельный план может быть симметричным, когда точки плана располагаются симметрично друг друга. Рассмотренный ранее план ПФЭ 2^n – ротатабельный симметричный план первого порядка.

Ротатабельный план может быть *ортогональным* при выполнении условия ортогональности матрицы коэффициентов.

Ротатабельный центрально-композиционный план (*РЦКП*) строится аналогично рассмотренному ранее плану ОЦКП. К использованному в качестве ядра плану ПФЭ 2^n добавляются «звездные» точки – по две на каждый фактор и несколько точек в центре плана.

Схема ротатабельного планирования экспериментов на трех уровнях отличается от ортогонального планирования плечом «звездных» точек α и числом центральных точек.

Число опытов ротатабельного планирования (таблица 7.3) вычисляется по формуле

$$N = 2^n + 2n + n_0,$$

где n_0 – число центральных точек.

Таблица 7.3 – Число опытов в РЦКП

Число факторов n	Число точек ядра 2^n, N_0	Число «звездных точек» $2n$	Число нулевых точек n_0	Общее число точек N (число опытов)	Примечание
2	4	4	3	11	
3	8	6	2	16	
4	16	8	2	26	
5	32	10	3	45	
5	16	10	1	27	Полуреплика
6	64	12	3	79	
6	32	12	2	46	Полуреплика
7	128	14	1	143	
7	64	14	3	81	Полуреплика
8	256	16	0	272	

Величина «звездной» точки α в ротатабельных планах отличается от величины α при ортогональном планировании. Для $n = 2,3,4,5,6,7$ значения α приведены в таблице 7.4.

Таблица 7.4 – Величина плеча $\boldsymbol{\alpha}$ в ротатабельных планах

	Число факторов, n								
	2	3	4	5	5	6	6	7	7
Тип плана ПФЭ или ДФЭ	2^2	2^3	2^4	2^5	$\frac{1}{2}\cdot 2^5$	2^6	$\frac{1}{2}\cdot 2^6$	2^7	$\frac{1}{2}\cdot 2^7$
Величина плеча, α	1,414	1,682	2,000	2,378	2,000	2,828	2,378	3,333	2,828
Число нулевых точек, n_0	5	6	7	10	6	15	9	21	14

Ротатабельные планы позволяют хорошо оценить коэффициенты регрессии полинома, но они очень трудоемки из-за большого числа опытов и не всегда ортогональны, а ортогональное планирование позволяет получить независимые оценки коэффициентов регрессии.

2.9 Ротатабельные ортогональные центрально-композиционные планы (РОЦКП)

План эксперимента называется ортогональным, если скалярное произведение всех векторов-столбцов матрицы равно нулю.

Ротатабельный ортогональный центрально-композиционный план (РОЦКП) строится аналогично рассмотренному ранее плану ОЦКП. К использованному в качестве ядра плану ПФЭ 2^n добавляются «звездные» точки – по две на каждый фактор и несколько точек в центре плана.

При построении РОЦКП с ядром из плана ПФЭ 2^n плечо «звездных» точек определяется числом факторов $\alpha = \sqrt{n}$.

Для РОЦКП число наблюдений в центре плана вычисляется по формуле

$$n_0 = \frac{4n^2}{2^n} + 2n.$$

Если n_0 не целое, то при практическом построении плана его округляют до целого, при этом свойство ортогональности плана нарушается.

Параметры РОЦКП в зависимости от числа факторов указаны в таблице 7.5.

Таблица 7.5 – Параметры РОЦКП

	Число факторов, n							
	1	2	3	4	5	6	7	8
Величина плеча, $\alpha = \sqrt{n}$	1	1,414	1,732	2	2,236	2,45	2,646	2,83
Число нулевых точек, $n_0 = \frac{4n^2}{2^n} + 2n$	4	8	10,5	12	13,13	14,25	15,53	17
Число точек ядра, $N_0 = 2^n$	2	4	8	16	32	64	128	256
Количество опытов, N	8	16	24,5	36	55,13	90	157,55	289

2.10 Композиционные планы Бокса B_n

Когда нежелателен выход «звездных» точек за пределы области планирования исходного полного факторного эксперимента, используют планы типа B_n, которые отличаются от ОЦКП и РКЦП лишь значением α ($\pm\alpha$).

Данные планы содержат ядро ПФЭ 2^n или ДФЭ 2^{n-k} и с этой точки зрения полностью аналогичны ОЦКП и РЦКП. Кроме ядра планы B_n включают $2n$ «звездных» точек. Планы B_n отличаются простотой построения: не содержат центральных точек; величина α получается равной 1.

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

3.1 Построение планов полных и дробных факторных экспериментов в программе Statistica

Планирование и последующий анализ результатов активных экспериментов производят в модуле Experimental Design.

Упражнение 1. В этом упражнении вы научитесь составлять план первого порядка для проведения эксперимента и заполнять таблицу плана эксперимента факторами и уровнями их варьирования в программе Statistica.

Составьте матрицу планирования для проведения эксперимента по определению влияния параметров агента сушки (T_a, V_a), начальной влажности зерна W_n, времени сушки τ на изменение влажности зерна при сушке.

Итак, в эксперименте участвуют 4 фактора на двух уровнях варьирования (таблица 7.6).

Таблица 7.6 – Факторы и уровни варьирования (два уровня варьирования факторов)

Факторы	Уровни варьирования	
	-1	1
Продолжительность сушки, τ, ч.	1	2
Влажность зерна, W_n, %	16	24
Температура агента сушки, T_a,℃	20	34
Скорость агента сушки, V_a, м/с	0,6	1

Модуль Experimental Design программы Statistica можно запустить двумя способами:

Способ 1. С помощью приложения **STA_EXP**.

Способ 2. С помощью переключателя модулей Module Switcher, выбрав в списке модуль Experimental Design.

1. Запустите файл **STA_EXP** с помощью команды Пуск – Мой компьютер – C:\STAT\STA_EXP. На экране появится окно модуля Experimental Design.

2. Для вызова на экран стартового окна модуля выполните команду Analysis – Startup Panel. На экране появится стартовая панель Design & Analysis of Experiments модуля Experimental Design (рисунок 7.2).

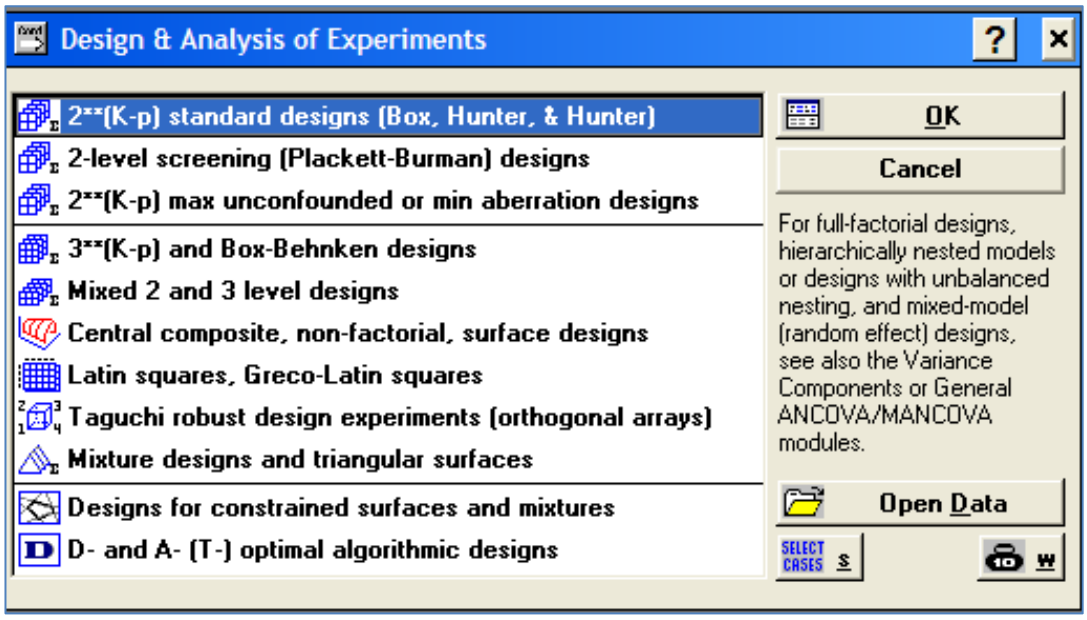


Рисунок 7.2 – Стартовая панель модуля Experimental Design

В окне Design & Analysis of Experiments находится список для выбора плана эксперимента. В первых трех строках расположены планы первого порядка:

– 2**(K-p) standard designs (Box, Hunter, & Hunter) – планы первого порядка полных и дробных факторных экспериментов.

– 2-level screening (Plackett-Burman) designs – планы Плакетта-Бермана (матрица Адамара) для отсеивания. При планировании эксперимента часто встает проблема определения факторов и их количества, влияющих на функцию отклика. В этих условиях неопределенности выбирают те факторы, которые могут оказывать такое влияние. При этом, если использовать полный факторный план, то необходимо выполнить большое количество опытов. Для проведения таких опытов потребуется много времени и затрат средств. Например, для 16 факторов, варьируя их на двух уровнях необходимо провести $2^{16} = 65536$ опытов. Плакетт и Берман предложили планы эксперимента, при которых оказалось возможным сократить число опытов до числа переменных плюс один. Характерной особенностью планов Плакетта-Бермана является то, что их результаты легко поддаются обработке, а линейные эффекты рассчитываются независимо друг от друга. Данные планы являются ортогональными и нормированными.

– 2**(K-p) max unconfounded or min aberration designs – планы максимально несмешанные или с минимальной аберрацией. В дополнение к стандартным 2**(K-p) планам модуль планирования включает в себя общую опцию поиска плана для создания дробных факторных планов с минимальной аберрацией (наименее смешанных) с блоками или без блоков с более чем 100 факторами и более чем 2000 опытами. Эти типы эффективных планов позволят оценить большее число (частных) факторных взаимодействий, чем в стандартных планах Бокса-Хантера.

3. Для построения планов первого порядка, полных и дробных фактор-

ных экспериментов, в списке планов эксперимента выберите план 2**(K-p) standard designs (Box, Hunter, & Hunter). Нажмите кнопку OK. На экране появится окно Design Experiments with Two-Level Factors (Планирование экспериментов с двухуровневыми факторами) (рисунок 7.3).

4. Для генерирования плана эксперимента установите переключатель в позицию Generate design.

Для задания числа факторов Number of factors служит специальное поле ввода. Это число может составлять от двух до одиннадцати факторов. Вид плана выбирается в списке No of factors/blocks/runs (Факторы/блоки/опыты). Использование блоков подразумевает определенную последовательность реализации (выполнения) частей плана. По количеству блоков судят о дробности плана. Максимальное число опытов соответствует полному факторному эксперименту. Если число опытов меньше максимального в два раза, то это дробная полуреплика, если в четыре раза, то – четвертьреплика.

Согласно условию задачи, количество выбранных факторов – 4, следовательно, в поле Number of factors (min=2, max=11) установите значение **4**, в списке No of factors/blocks/runs: выберите **4/1/16** – 4 фактора, 1 блок, 16 опытов в эксперименте (рисунок 7.3). Нажмите кнопку OK.

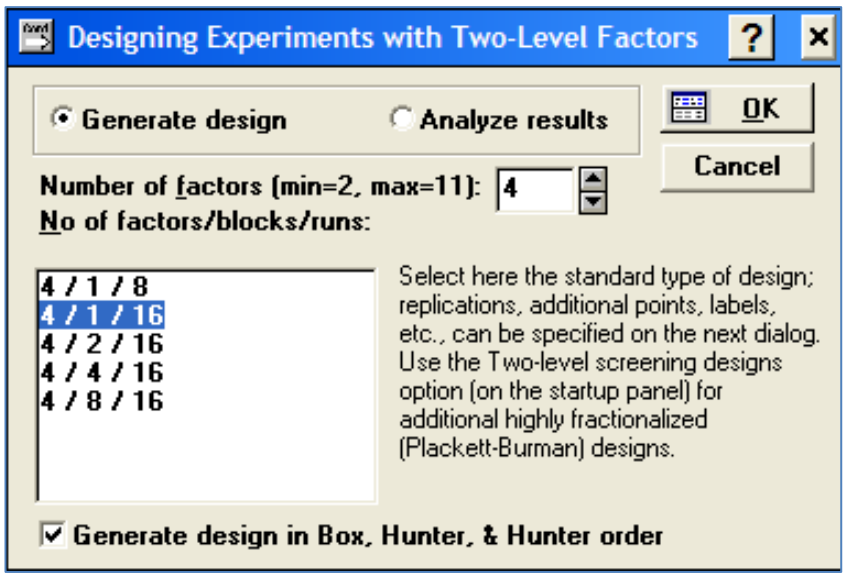


Рисунок 7.3 – Окно Design Experiments with Two-Level Factors

5. После выбора вида (типа) плана на экране появится окно Design of an Experiment with Two-Level Factors (Планирование эксперимента с двухуровневыми факторами) (рисунок 7.4).

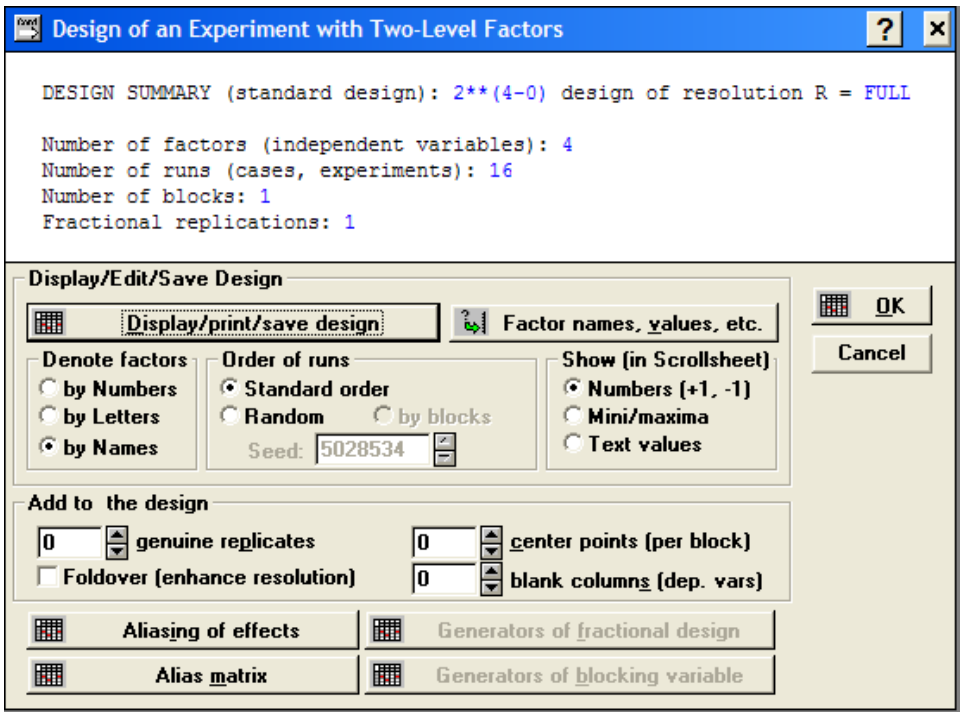


Рисунок 7.4 – Окно Design of an Experiment with Two-Level Factors

В верхней части окна представлена информация о параметрах выбранного плана: число факторов, опытов, блоков, часть реплики.

Ниже расположена группа опций для просмотра/правки/сохранения плана Display/Edit/save design.

Факторы могут обозначаться Denote Factors номерами by Numbers, буквами by Letters или именами by Numes.

Порядок опытов Order of runs может быть стандартным Standard order, псевдослучайным Random, блоковым by blocks. Для задания псевдослучайной последовательности служит опция Seed (Случайное число).

Группа опций Show (Отобразить) задает вид уровня фактора в таблице, который может быть числовым Numbers, минимумом/максимумом Mini/maxima, текстовым Text values.

Просмотреть план эксперимента можно с помощью кнопки Display/print/save design (Просмотр/правка/сохранение).

Факторы в плане можно задавать как в кодированном, так и в именованном виде. Для задания имен и значений факторов служит кнопка Factor names,values, ets.

В план могут быть добавлены: полные реплики genuine replicates, центральные точки на блок center points (per block), пустые столбцы для зависимой переменной blank columns (dep. vars).

Кнопки Aliasing of effects и Alias matrix предназначены для просмотра таблицы и матрицы взаимного влияния коэффициентов уравнения регрессии.

С помощью кнопки Generators of fractional design можно просмотреть генерирующее соотношение для дробного факторного эксперимента.

6. Установите опции для просмотра плана эксперимента, как показано на рисунке 7.4. Нажмите кнопку Display/print/save design. На экране появится матрица планирования эксперимента для факторов в кодированном виде (рисунок 7.5).

Design: 2**(4-0) design (new.sta)

Continue...	A	B	C	D
1	-1	-1	-1	-1
2	1	-1	-1	-1
3	-1	1	-1	-1
4	1	1	-1	-1
5	-1	-1	1	-1
6	1	-1	1	-1
7	-1	1	1	-1
8	1	1	1	-1
9	-1	-1	-1	1
10	1	-1	-1	1
11	-1	1	-1	1
12	1	1	-1	1
13	-1	-1	1	1
14	1	-1	1	1
15	-1	1	1	1
16	1	1	1	1

Рисунок 7.5 – Матрица планирования эксперимента для четырех факторов на двух уровнях в кодированном виде

Планируемый эксперимент проводится по плану, изображенному на рисунке 7.5. Число (–1) обозначает нижний уровень варьирования фактора, число (1) – верхний. Согласно таблице на рисунке 7.5, факторы на всех уровнях варьирования управляемы.

7. Для продолжения работы нажмите кнопку Continue.

8. Для получения матрицы планирования эксперимента с именами и значениями факторов в окне Design of an Experiment with Two-Level Factors нажмите кнопку Factor names,values, ets. На экране появится окно с таблицей для ввода имен факторов и их значений (см. рисунок 7.6).

9. Отредактируйте таблицу в окне на рисунке 7.6 согласно данным задачи (см. таблицу 7.1). Сравните полученную таблицу с рисунком 7.7.

10. Для продолжения работы нажмите кнопку Continue.

11. В окне Design of an Experiment with Two-Level Factors в разделе **Show** установите переключатель в положение **Mini/maxima** и нажмите кнопку **Display/print/save design**. На экране появится матрица планирования эксперимента в именованном виде (рисунок 7.8).

Summary for Variables (Factors) (new.sta)

Continue... To change labels, values, etc., type in the desired changes, then click Continue.

Factor	Factor Name	Low Value	Low Label	High Value	High Label	C/Q Cont or Qual.
A (1)	A	-1,00000	Low	1,000000	High	C
B (2)	B	-1,00000	Low	1,000000	High	C
C (3)	C	-1,00000	Low	1,000000	High	C
D (4)	D	-1,00000	Low	1,000000	High	C

Рисунок 7.6 – Окно для ввода имен факторов и их значений

Summary for Variables (Factors) (new.sta)

Continue... To change labels, values, etc., type in the desired changes, then click Continue.

Factor	Factor Name	Low Value	Low Label	High Value	High Label	C/Q Cont or Qual.
A (1)	t	1,00000	Low	2,000000	High	C
B (2)	W	16,00000	Low	24,00000	High	C
C (3)	Ta	20,00000	Low	34,00000	High	C
D (4)	Va	,60000	Low	1,00000	High	C

Рисунок 7.7 – Окно с отредактированной таблицей имен факторов и их значений

Design: 2**(4-0) design (new.sta)

Continue...

	t	W	Ta	Va
1	1,000000	16,00000	20,00000	,600000
2	2,000000	16,00000	20,00000	,600000
3	1,000000	24,00000	20,00000	,600000
4	2,000000	24,00000	20,00000	,600000
5	1,000000	16,00000	34,00000	,600000
6	2,000000	16,00000	34,00000	,600000
7	1,000000	24,00000	34,00000	,600000
8	2,000000	24,00000	34,00000	,600000
9	1,000000	16,00000	20,00000	1,000000
10	2,000000	16,00000	20,00000	1,000000
11	1,000000	24,00000	20,00000	1,000000
12	2,000000	24,00000	20,00000	1,000000
13	1,000000	16,00000	34,00000	1,000000
14	2,000000	16,00000	34,00000	1,000000
15	1,000000	24,00000	34,00000	1,000000
16	2,000000	24,00000	34,00000	1,000000

Рисунок 7.8 – Окно с матрицей планирования эксперимента для факторов в именованном виде

12. Сохраните матрицу планирования эксперимента для факторов в именованном виде в файл с именем **Plan_1.sta**.

13. Для продолжения работы нажмите кнопку Continue.

14. Для просмотра матрицы взаимного влияния коэффициентов уравнения регрессии в окне Design of an Experiment with Two-Level Factors нажмите кнопку Alias matrix. На экране появится матрица взаимного влияния коэффициентов уравнения регрессии (рисунок 7.9). Матрица диагональная, элементы главной диагонали равны 1. Такой вид матрицы говорит о том, что коэффициенты уравнения регрессии определяются однозначно, независимо друг от друга.

Alias Matrix (new.sta)

EXPERIM. DESIGN: 2**(4-0) design

Factor	t	W	Ta	Va	1*2	1*3	1*4	2*3	2*4	3*4
t	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
W	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Ta	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Va	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00
1*2	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
1*3	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
1*4	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
2*3	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00
2*4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00
3*4	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Рисунок 7.9 – Матрица взаимного влияния коэффициентов уравнения регрессии

15. Сгенерируйте полуреплику для рассматриваемого эксперимента. Для этого в списке **No of factors/blocks/runs:** окна Design Experiments with Two-Level Factors выберите **4/1/8** (см. рисунок 7.3).

16. Выполните анализ матрицы взаимного влияния коэффициентов уравнения регрессии для полуреплики: согласно матрице взаимного влияния коэффициентов уравнения регрессии (рисунок 7.9), все коэффициенты уравнения регрессии определяются независимо друг от друга.

Вывод: *факторы на всех уровнях варьирования управляемы, коэффициенты уравнения регрессии определяются независимо друг от друга, следовательно,* ***полученный план эксперимента применим для проведения эксперимента и построения регрессионной модели***.

17. Сохраните результаты работы.

Упражнение 2. В этом упражнении вы научитесь составлять план второго порядка – план Бокса-Бенкина и заполнять таблицу эксперимента факторами и уровнями их варьирования в программе Statistica.

Составьте матрицу планирования для проведения эксперимента по определению влияния параметров агента сушки (T_a, V_a), начальной влажности зерна W_n, времени сушки τ на изменение влажности зерна при сушке.

В эксперименте участвуют четыре фактора на трёх уровнях варьирования (таблица 7.7).

Таблица 7.7 – Факторы и уровни варьирования (три уровня варьирования факторов)

Факторы	Уровни варьирования		
	-1	0	1
Продолжительность сушки, τ, ч.	1	1,5	2
Влажность зерна, W_n, %	16	20	24
Температура агента сушки, T_a,℃	20	27	34
Скорость агента сушки, V_a, м/с	0,6	0,8	1

1. Запустите модуль Experimental Design, если открыт другой модуль.
2. Для вызова на экран стартового окна модуля выполните команду Analysis – Startup Panel. На экране появится стартовая панель Design & Analysis of Experiments модуля Experimental Design (рисунок 7.10).

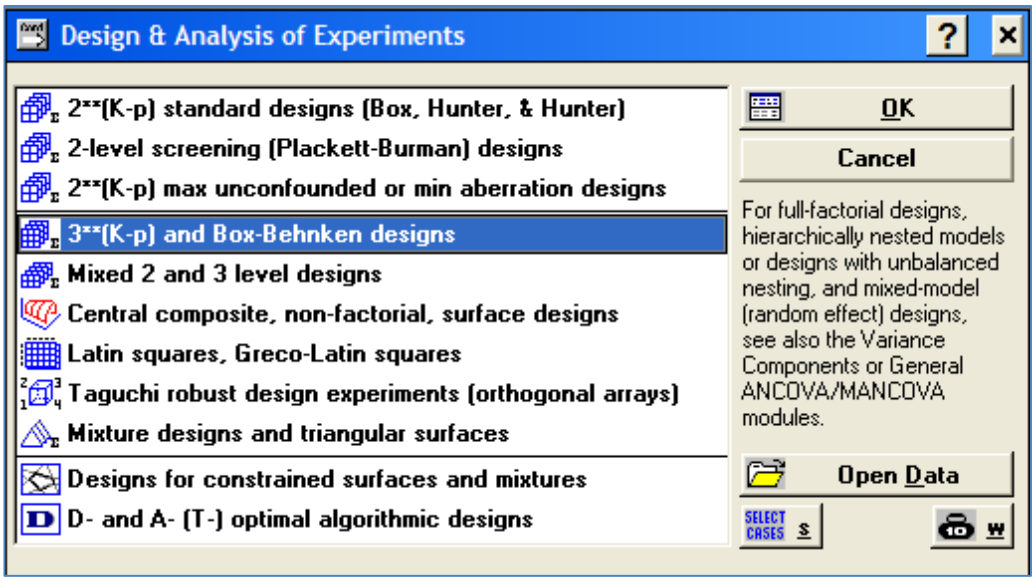


Рисунок 7.10 – Стартовая панель модуля Experimental Design

В окне Design & Analysis of Experiments находится список для выбора плана эксперимента. Планы второго порядка расположены в 3–10 строках:

– 3**(K-p) and Box-Behnken designs – планы второго порядка полных и дробных факторных экспериментов и планы Бокса-Бенкина. В некомпозиционных планах Бокса-Бенкина каждый фактор может изменяться на трех уровнях. Указанные планы состоят из основной части, представляющей собой определенную выборку из строк матрицы ПФЭ 3^n и некоторого количества центральных точек. Использование некомпозиционных планов, предусматривающих всего три уровня варьирования факторов, упрощает и удешевляет проведение эксперимента. Некомпозиционные планы характеризуются наличием в строках матрицы планирования большого числа нулей, в результате чего существенно упрощается вычисление коэффициентов моде-

ли. Кроме этого, некомпозиционные планы для 3, 4, 6 и 7 факторов требуют постановки меньшего числа опытов по сравнению с соответствующими ротатабельными центральными композиционными планами второго порядка.

– Mixed 2 and 3 level designs – смешанные 2-х и 3-х уровневые планы. В некотором смысле, способ построения указанных планов является комбинацией процедур, описанных для планов 2**(K-p) и 3**(K-p). Следует отметить, что хотя все эти планы очень эффективны, но они необязательно ортогональны. Указанные планы уменьшают количество опытов в эксперименте по сравнению с соответствующими планами ПФЭ.

– Central composite, non-factorial, surface design – центральные композиционные планы. В дополнение к стандартным опциям, доступным для всех планов (добавление опытов, рандомизация, реплицирование, задание верхних и нижних значений факторов, и т.п.) можно задать звездные точки как центрированные, вычислить их для ротатабельности, ортогональности, или для того и другого.

– Latin squares, Greco-Latin squares – планы на «Латинских квадратах». Упрощением полного плана с тремя независимыми переменными является планирование по методу «латинского квадрата». «Латинский квадрат» применяют тогда, когда нужно исследовать одновременное влияние трех переменных, имеющих два уровня или более. Принцип «латинского квадрата» состоит в том, что два уровня разных переменных встречаются в экспериментальном плане только один раз. Тем самым процедура значительно упрощается, не говоря о том, что экспериментатор избавляется от необходимости работать с огромными выборками. В программе Statistica пользователь может выбирать между различными планами на «Латинских квадратах», имеющих до девяти уровней. Если это возможно, программа также сгенерирует доступные Греко-Латинские и Гипер-Греко-Латинские квадраты. В случаях, когда имеется несколько альтернативных «Латинских квадратов», программа выберет один из них случайным образом, либо предложит пользователю выбрать желаемый «Латинский квадрат». Пользователь может просмотреть планы в таблице результатов, рандомизировать порядок опытов, добавить пустые столбцы для удобства ввода данных. План можно также сохранить в стандартном файле данных Statistica. После добавления наблюдаемых значений в этот файл, эксперимент можно легко проанализировать. В дополнение к полной таблице дисперсионного анализа модуль «Планирование эксперимента» вычислит средние значения для всех факторов. Эти средние могут быть представлены в графическом виде на итоговом графике.

– Taguchi robust design experiments (orthogonal arrays) – робастные планы Тагучи. Программа Statistica позволяет сгенерировать ортогональные массивы с числом факторов до 31; могут быть проанализированы планы с числом факторов до 65. Как и во всех других типах планов, опыты эксперимента могут быть рандомизированы, и пользователь может добавлять пустые столбцы в таблице результатов для удобства ввода данных.

– Mixture designs and triangular surface – планы для смесей и тер-

нарные поверхности. Эта процедура содержит опции для построения симплекс-вершинных и симплекс-центроидных планов для переменных смеси. Эти планы могут быть расширены путем добавления внутренних и центральных точек. Пользователь программы Statistica может ввести ограничения снизу для каждого фактора, тогда программа автоматически построит соответствующий план на субсимплексе, определенном этими ограничениями. Множественные верхние и нижние ограничения могут быть заданы с помощью общих средств построения планов в ограниченных экспериментальных областях. Пользователь может добавить отдельные опыты или реплики, отобразить и сохранить планы в стандартном или случайном порядке.

– Designs for constrained surfaces and mixtures – планы для поверхностей и смесей с ограничениями. Модуль «Планирование эксперимента» содержит процедуры для вычисления вершин и центральных точек для поверхностей с ограничениями и смесей, заданных с помощью линейных ограничений.

– D- and A- (T-) optimal algorithmic designs – D- и A-оптимальные планы. Программа Statistica включает несколько алгоритмов для построения оптимальных планов. Пользователь может задать критерий D оптимальности (по определителю матрицы плана) или A оптимальности (по следу), а также модели для поверхностей и смесей. Список точек-кандидатов для плана может быть введен вручную либо образован из файла данных (например, из существующего плана, созданного с помощью вычисления вершин и центральных точек для поверхностей с ограничениями и смесей, смотрите выше). Пользователь может пометить точки в списке кандидатов, чтобы принудительно включить их в окончательный план, и, таким образом, «расширить» или «подправить» существующие эксперименты. Программа содержит все основные алгоритмы поиска для построения D- и A-оптимальных планов.

3. Для построения планов второго порядка полных и дробных факторных экспериментов в списке планов эксперимента выберите план 3**(K-p) and Box-Behnken designs. Нажмите кнопку OK. На экране появится окно Designing Experiments with Three-Level Factors (Планирование экспериментов с трехуровневыми факторами) (рисунок 7.11). В этом окне выбирается либо полный факторный план (окно выбора параметров плана слева), либо дробная полуреплика (окно выбора параметров плана справа).

Согласно условию задачи, количество выбранных факторов – 4, уровней варьирования – 3. Следовательно, план ПФЭ будет содержать 81 опыт.

В этом случае в списке Factors/blocks/runs: необходимо выбрать 4/1/81 – **4** фактора, **1** блок, **81** опыт в эксперименте (см. рисунок 7.11). Однако эксперимент нужно повторять не менее трёх раз, значит, общее количество опытов вырастет до 343. Поэтому, для уменьшения количества опытов в эксперименте выберем план Бокса-Бенкина.

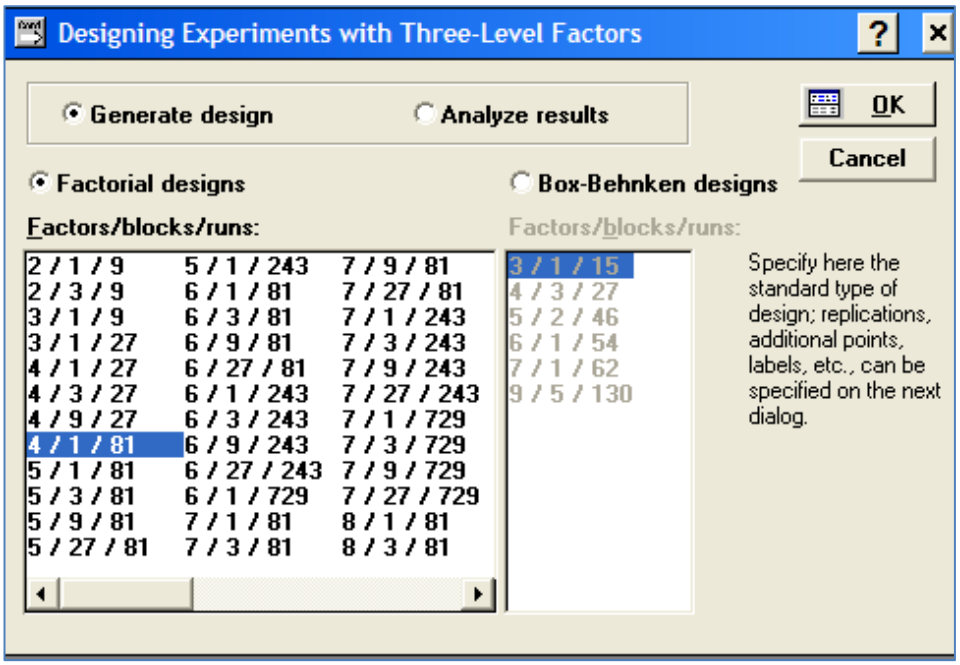


Рисунок 7.11 – Окно Designing Experiments with Three-Level Factors

4. Для построения плана Бокса-Бенкина в окне Design Experiments with Three-Level Factors установите переключатель в позицию Box-Behnken designs. Далее в списке Factors/blocks/runs: необходимо выбрать **4/3/27** – **4** фактора, **3** блока, **27** опытов в эксперименте. Нажмите кнопку OK.

5. После выбора вида (типа) плана на экране появится окно Design of an Experiment with Three-Level Factors (Планирование эксперимента с трёхуровневыми факторами) (рисунок 7.12).

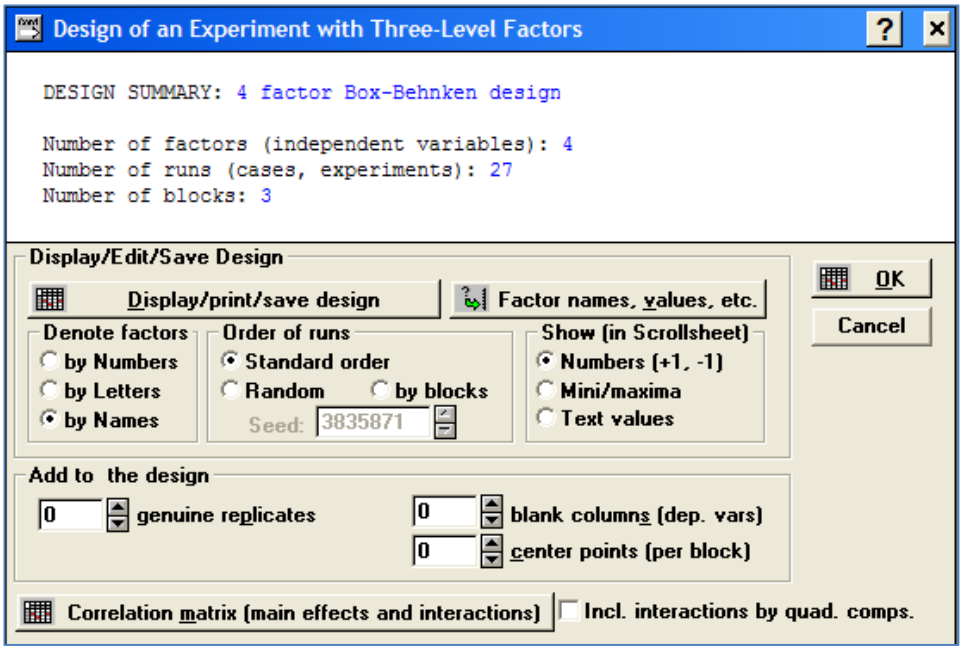


Рисунок 7.12 – Окно Design of an Experiment with Three-Level Factors

6. Установите опции для просмотра плана эксперимента, как показано на рисунке 7.12. Нажмите кнопку Display/print/save design. На экране по-

явится матрица планирования эксперимента для факторов в кодированном виде (рисунок 7.13).

4 factor Box-Behnken design, 3 blocks, 27 runs (new.sta)

Continue...

	Block	A	B	C	D
1	1	-1	-1	0	0
2	1	1	-1	0	0
3	1	-1	1	0	0
4	1	1	1	0	0
5	1	0	0	-1	-1
6	1	0	0	1	-1
7	1	0	0	-1	1
8	1	0	0	1	1
9	1	0	0	0	0
10	2	-1	0	0	-1
11	2	1	0	0	-1
12	2	-1	0	0	1
13	2	1	0	0	1
14	2	0	-1	-1	0
15	2	0	1	-1	0
16	2	0	-1	1	0
17	2	0	1	1	0
18	2	0	0	0	0
19	3	-1	0	-1	0
20	3	1	0	-1	0
21	3	-1	0	1	0
22	3	1	0	1	0
23	3	0	-1	0	-1
24	3	0	1	0	-1
25	3	0	-1	0	1
26	3	0	1	0	1
27	3	0	0	0	0

Рисунок 7.13 – Матрица планирования эксперимента для четырех факторов на трех уровнях в кодированном виде (план Бокса-Бенкина)

Планируемый эксперимент проводится по плану, изображенному на рисунке 7.13. Число (–1) обозначает нижний уровень варьирования фактора, число (0) – центр, число (1) – верхний.

7. Для продолжения работы нажмите кнопку Continue.

8. Выведите на экран таблицу для ввода имен факторов и уровней варьирования и заполните её. Для этого нажмите кнопку Factor names, values, etc. Результат сравните с рисунком 7.14.

Summary for Variables (Factors) (new.sta)

Continue...

To change labels, values, etc., type in the desired changes, then click Continue.

Factor	Factor Name	Low Value	Low Label	Center Value	Center Label	High Value	High Label
A (1)	t	1,00000	Low	1,500000	CenterPt	2,000000	High
B (2)	W	16,00000	Low	20,00000	CenterPt	24,00000	High
C (3)	Ta	20,00000	Low	27,00000	CenterPt	34,00000	High
D (4)	Va	,60000	Low	,80000	CenterPt	1,00000	High

Рисунок 7.14 – Окно с отредактированной таблицей имен факторов и их значений

9. Для продолжения работы нажмите кнопку Continue.

10. В окне Design of an Experiment with Three-Level Factors в разделе Show (см. рисунок 7.12) установите переключатель в положение Mini/maxima и нажмите кнопку Display/print/save design. На экране появится матрица планирования эксперимента в именованном виде (рисунок 7.15).

11. Сохраните матрицу планирования эксперимента для факторов в именованном виде в файл с именем **Plan_2.sta**.

12. Для продолжения работы нажмите кнопку Continue.

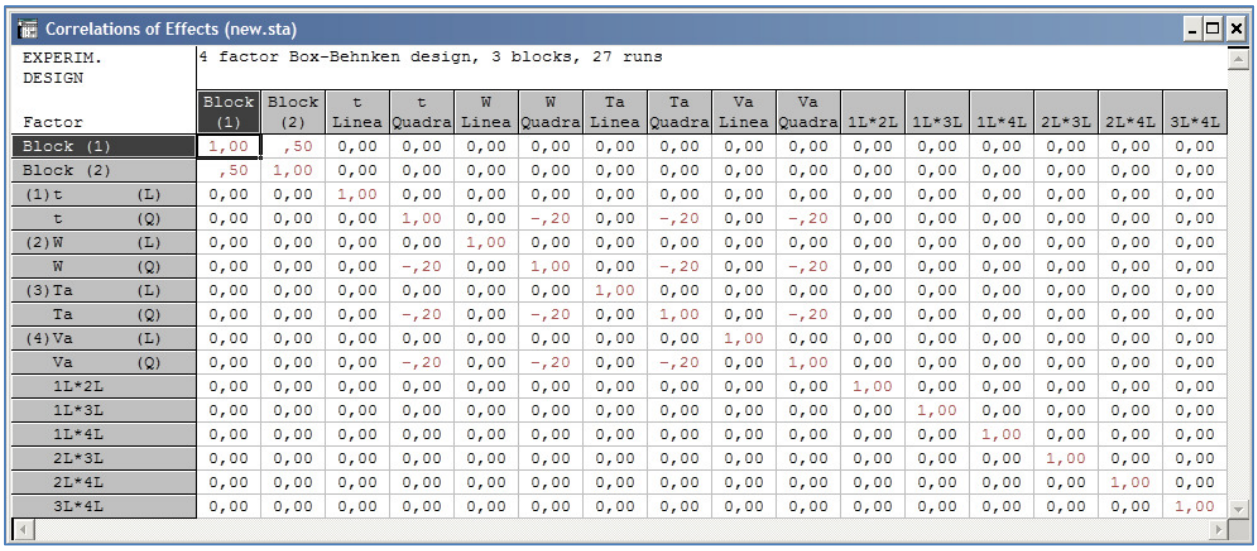
4 factor Box-Behnken design, 3 blocks, 27 runs (new.sta)

Continue...

	Block	t	W	Ta	Va
1	1	1,000000	16,00000	27,00000	,800000
2	1	2,000000	16,00000	27,00000	,800000
3	1	1,000000	24,00000	27,00000	,800000
4	1	2,000000	24,00000	27,00000	,800000
5	1	1,500000	20,00000	20,00000	,600000
6	1	1,500000	20,00000	34,00000	,600000
7	1	1,500000	20,00000	20,00000	1,000000
8	1	1,500000	20,00000	34,00000	1,000000
9	1	1,500000	20,00000	27,00000	,800000
10	2	1,000000	20,00000	27,00000	,600000
11	2	2,000000	20,00000	27,00000	,600000
12	2	1,000000	20,00000	27,00000	1,000000
13	2	2,000000	20,00000	27,00000	1,000000
14	2	1,500000	16,00000	20,00000	,800000
15	2	1,500000	24,00000	20,00000	,800000
16	2	1,500000	16,00000	34,00000	,800000
17	2	1,500000	24,00000	34,00000	,800000
18	2	1,500000	20,00000	27,00000	,800000
19	3	1,000000	20,00000	20,00000	,800000
20	3	2,000000	20,00000	20,00000	,800000
21	3	1,000000	20,00000	34,00000	,800000
22	3	2,000000	20,00000	34,00000	,800000
23	3	1,500000	16,00000	27,00000	,600000
24	3	1,500000	24,00000	27,00000	,600000
25	3	1,500000	16,00000	27,00000	1,000000
26	3	1,500000	24,00000	27,00000	1,000000
27	3	1,500000	20,00000	27,00000	,800000

Рисунок 7.15 – Окно с матрицей планирования эксперимента в именованном виде

13. Для просмотра матрицы взаимного влияния коэффициентов уравнения регрессии в окне Design of an Experiment with Three-Level Factors (см. рисунок 7.12) нажмите кнопку Correlation matrix (main effects and interactions). На экране появится матрица взаимного влияния коэффициентов уравнения регрессии (рисунок 7.16). Согласно рисунку 7.16 матрица взаимного влияния коэффициентов уравнения регрессии не является диагональной. Следовательно, существует зависимость между шестью коэффициентами уравнения регрессии, но поскольку, значения элементов матрицы, стоящие не на главной диагонали, по модулю меньше 0,5, то согласно корреляционному анализу, связь слабая.



Correlations of Effects (new.sta)

EXPERIM. DESIGN — 4 factor Box-Behnken design, 3 blocks, 27 runs

Factor	Block (1)	Block (2)	t Linea	t Quadra	W Linea	W Quadra	Ta Linea	Ta Quadra	Va Linea	Va Quadra	1L*2L	1L*3L	1L*4L	2L*3L	2L*4L	3L*4L
Block (1)	1,00	,50	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Block (2)	,50	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
(1)t (L)	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t (Q)	0,00	0,00	0,00	1,00	0,00	-,20	0,00	-,20	0,00	-,20	0,00	0,00	0,00	0,00	0,00	0,00
(2)W (L)	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
W (Q)	0,00	0,00	0,00	-,20	0,00	1,00	0,00	-,20	0,00	-,20	0,00	0,00	0,00	0,00	0,00	0,00
(3)Ta (L)	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Ta (Q)	0,00	0,00	0,00	-,20	0,00	-,20	0,00	1,00	0,00	-,20	0,00	0,00	0,00	0,00	0,00	0,00
(4)Va (L)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Va (Q)	0,00	0,00	0,00	-,20	0,00	-,20	0,00	-,20	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00
1L*2L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
1L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
1L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
2L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00
2L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00
3L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Рисунок 7.16 – Матрица взаимного влияния коэффициентов уравнения регрессии

14.Сохраните данные и закройте все открытые окна.

Упражнение 3. В этом упражнении вы научитесь составлять композиционные планы (РЦКП, ОЦКП, типа Бокса B_n) в программе Statistica.

Составьте матрицу планирования для проведения эксперимента по определению влияния параметров агента сушки (T_a, V_a), начальной влажности зерна W_n, времени сушки τ на изменение влажности зерна при сушке. В эксперименте четыре фактора на трех уровнях варьирования (таблицу 7.8).

Таблица 7.8 – Факторы и их уровни варьирования (три уровня варьирования факторов)

Факторы	Уровни варьирования		
	-1	0	1
Продолжительность сушки, τ, ч.	1	1,5	2
Влажность зерна, W_n, %	16	20	24
Температура агента сушки, T_a, ℃	20	27	34
Скорость агента сушки, V_a, м/с	0,6	0,8	1

При выполнении задачи составьте несколько планов эксперимента (РЦКП, ОЦКП, типа Бокса B_n) и обоснуйте выбор наилучшего плана.

Построение плана РЦКП

1. В стартовой панели Design & Analysis of Experiments модуля Experimental Design (см. рисунок 7.10) в списке планов эксперимента выберите план Central composite, non-factorial, surface design и нажмите кнопку OK. На экране появится диалоговое окно Central Composite (Response Surface) Designs (рисунок 7.17).

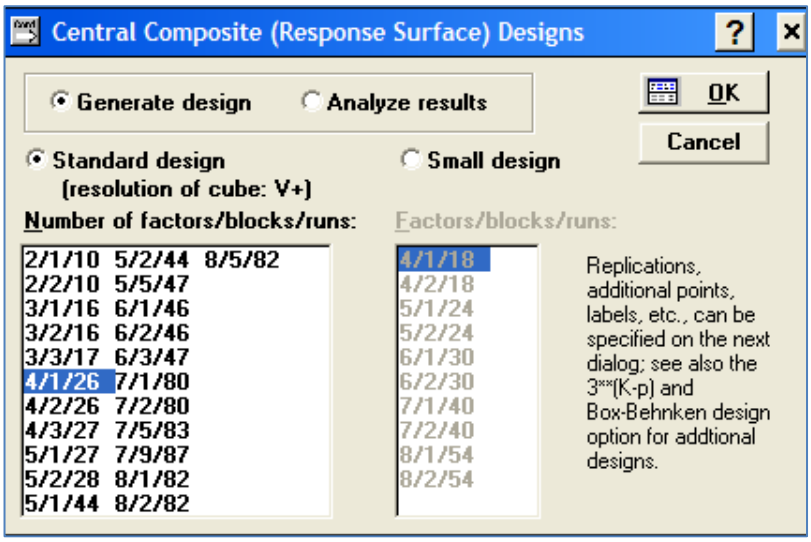


Рисунок 7.17 – Окно выбора центральных композиционных планов

2. В окне Central Composite (Response Surface) Designs в списке Standard Design выберите **4/1/26** – **4** фактора, **1** блок, **26** опытов. Нажмите кнопку OK.

3. На экране появится окно Designs of a Response Surface Experiment (Central Composite Designs) для настройки параметров плана (рисунок 7.18). В этом окне выбирают значения α для ортогональности, рота-табельности.

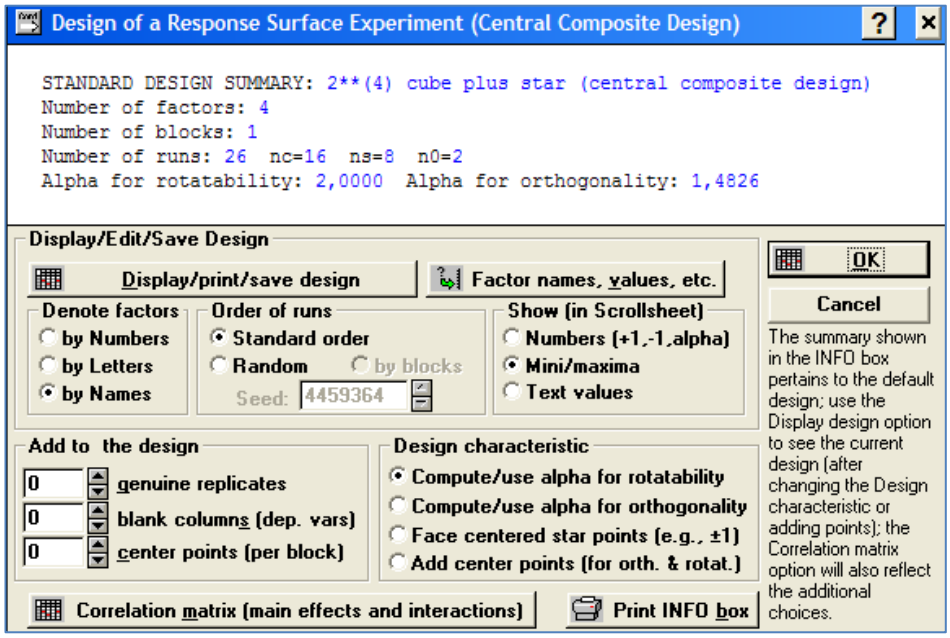


Рисунок 7.18 – Окно для настройки параметров центральных композиционных планов

Для выбора служит группа опций Design Charecteristic, в которой содержатся следующие опции:

– Compute/use alpha for rotatability – Вычислить/использовать альфа для ротатабельности;

– Compute/use alpha for ortohonality – Вычислить/использовать альфа для ортогональности;

– Face center star points (±1) – Звездные точки в центре плоскости (±1);

– Add center points (for ort. and rotat.) – Добавить звездные точки (для ортогональности и ротатабельности).

4. В разделе Design characteristic (см. рисунок 7.18) установите пере-

5. ключатель в позицию Compute/use alpha for rotatability. Нажмите кнопку Display/print/save design. На экране появится матрица плана РЦКП в кодированном виде (рисунок 7.19).

6. Сохраните матрицу плана РЦКП в кодированном виде в файле с именем **Plan_RCKP_kod.sta**.

Continue...	A	B	C	D
1	-1,00000	-1,00000	-1,00000	-1,00000
2	-1,00000	-1,00000	-1,00000	1,00000
3	-1,00000	-1,00000	1,00000	-1,00000
4	-1,00000	-1,00000	1,00000	1,00000
5	-1,00000	1,00000	-1,00000	-1,00000
6	-1,00000	1,00000	-1,00000	1,00000
7	-1,00000	1,00000	1,00000	-1,00000
8	-1,00000	1,00000	1,00000	1,00000
9	1,00000	-1,00000	-1,00000	-1,00000
10	1,00000	-1,00000	-1,00000	1,00000
11	1,00000	-1,00000	1,00000	-1,00000
12	1,00000	-1,00000	1,00000	1,00000
13	1,00000	1,00000	-1,00000	-1,00000
14	1,00000	1,00000	-1,00000	1,00000
15	1,00000	1,00000	1,00000	-1,00000
16	1,00000	1,00000	1,00000	1,00000
17	-2,00000	0,00000	0,00000	0,00000
18	2,00000	0,00000	0,00000	0,00000
19	0,00000	-2,00000	0,00000	0,00000
20	0,00000	2,00000	0,00000	0,00000
21	0,00000	0,00000	-2,00000	0,00000
22	0,00000	0,00000	2,00000	0,00000
23	0,00000	0,00000	0,00000	-2,00000
24	0,00000	0,00000	0,00000	2,00000
25 (C)	0,00000	0,00000	0,00000	0,00000
26 (C)	0,00000	0,00000	0,00000	0,00000

Рисунок 7.19 – Матрица плана РЦКП для четырех факторов в кодированном виде

В полученной матрице плана РЦКП (рисунок 7.19) видим пять уровней варьирования факторов $(-\alpha, -1, 0, +1, +\alpha)$, т.е. согласно полученному плану $(-2, -1, 0, +1, +2)$. Следовательно, значения уровней варьирования факторов надо задать таким образом, чтобы эти значения не выходили за пределы факторного пространства, т.е. значения факторов в «звездных» точках $(-\alpha)$, $(+\alpha)$ соответствуют нижнему и верхнему уровням варьирования факторов.

Плечо «звездной» точки для соответствующего плана указаны в таблицах 7.1, 7.4, 7.5.

В нашем примере таким пространством являются уровни варьирования факторов в таблице 7.8. Поэтому для плана РЦКП таблица 7.8 будет иметь другой вид – таблица 7.9.

Обратите внимание, что полученные значения факторов в точках (–1) и (+1) возможно реализовать в ходе данного эксперимента.

Таблица 7.9 – Факторы и их уровни варьирования (пять уровней варьирования факторов)

Факторы	Уровни варьирования				
	–2	–1	0	+1	+2
Продолжительность сушки, τ, ч	1	1,25	1,5	1,75	2
Влажность зерна, W_n, %	16	18	20	22	24
Температура агента сушки, T_a,℃	20	23,5	27	30,5	34
Скорость агента сушки, V_a, м/с	0,6	0,7	0,8	0,9	1

7. Для продолжения работы нажмите кнопку Continue.

8. Выведите на экран таблицу для ввода имен факторов и уровней варьирования и заполните её. Результат сравните с рисунком 7.20.

Continue... | To change labels, values, etc., type in the desired changes, then click Continue.

Factor	Factor Name	Low Value	Low Label	Center Value	Center Label	High Value	High Label	Star Low Label	Star Hi. Label
A (1)	t	1,25000	Low	1,500000	CenterPt	1,750000	High	StarLow	StarHigh
B (2)	W	18,00000	Low	20,00000	CenterPt	22,00000	High	StarLow	StarHigh
C (3)	Ta	23,50000	Low	27,00000	CenterPt	30,50000	High	StarLow	StarHigh
D (4)	Va	,70000	Low	,80000	CenterPt	,90000	High	StarLow	StarHigh

Рисунок 7.20 – Окно с отредактированной таблицей имен факторов и их значений

9. Для продолжения работы нажмите кнопку Continue.

10. В окне Design of an Experiment with Three-Level Factors в разделе Show установите переключатель в положение Mini/maxima и нажмите кнопку Display/print/save design. На экране появится матрица планирования эксперимента для четырех факторов в именованном виде (рисунок 7.21).

11. Сохраните матрицу планирования эксперимента в именованном виде в файл с именем **Plan_RCKP_f.sta**.

12. Для продолжения работы нажмите кнопку Continue.

13. Для просмотра матрицы взаимного влияния коэффициентов уравнения регрессии в окне Design of an Experiment with Three-Level Factors нажмите кнопку Correlation matrix (main effects and interactions). На экране появится матрица взаимного влияния коэффициентов уравнения регрессии (рисунок 7.22). Сделайте вывод о взаимном влиянии коэффициентов уравнения регрессии.

Continue...	t	W	Ta	Va
1	1,250000	18,00000	23,50000	,700000
2	1,250000	18,00000	23,50000	,900000
3	1,250000	18,00000	30,50000	,700000
4	1,250000	18,00000	30,50000	,900000
5	1,250000	22,00000	23,50000	,700000
6	1,250000	22,00000	23,50000	,900000
7	1,250000	22,00000	30,50000	,700000
8	1,250000	22,00000	30,50000	,900000
9	1,750000	18,00000	23,50000	,700000
10	1,750000	18,00000	23,50000	,900000
11	1,750000	18,00000	30,50000	,700000
12	1,750000	18,00000	30,50000	,900000
13	1,750000	22,00000	23,50000	,700000
14	1,750000	22,00000	23,50000	,900000
15	1,750000	22,00000	30,50000	,700000
16	1,750000	22,00000	30,50000	,900000
17	1,000000	20,00000	27,00000	,800000
18	2,000000	20,00000	27,00000	,800000
19	1,500000	16,00000	27,00000	,800000
20	1,500000	24,00000	27,00000	,800000
21	1,500000	20,00000	20,00000	,800000
22	1,500000	20,00000	34,00000	,800000
23	1,500000	20,00000	27,00000	,600000
24	1,500000	20,00000	27,00000	1,000000
25 (C)	1,500000	20,00000	27,00000	,800000
26 (C)	1,500000	20,00000	27,00000	,800000

Рисунок 7.21 – Окно с матрицей планирования эксперимента для четырех факторов в именованном виде

Continue...	2**(4) central composite, nc=16 ns=8 n0=2 Runs=26													
Effect	A Linea	A Quadra	B Linea	B Quadra	C Linea	C Quadra	D Linea	D Quadra	1L*2L	1L*3L	1L*4L	2L*3L	2L*4L	3L*4L
(1)A (L)	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
A (Q)	0,00	1,00	0,00	-,24	0,00	-,24	0,00	-,24	0,00	0,00	0,00	0,00	0,00	0,00
(2)B (L)	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
B (Q)	0,00	-,24	0,00	1,00	0,00	-,24	0,00	-,24	0,00	0,00	0,00	0,00	0,00	0,00
(3)C (L)	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C (Q)	0,00	-,24	0,00	-,24	0,00	1,00	0,00	-,24	0,00	0,00	0,00	0,00	0,00	0,00
(4)D (L)	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
D (Q)	0,00	-,24	0,00	-,24	0,00	-,24	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00
1L*2L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
1L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
1L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
2L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00
2L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00
3L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Рисунок 7.22 – Матрица взаимного влияния коэффициентов уравнения регрессии

14. Сохраните данные и закройте окна с матрицами плана РЦКП.

Построение плана ОЦКП

1. В окне Designs of a Response Surface Experiment (Central Composite Designs) для настройки параметров плана (см. рисунок 7.18). В разделе Design characteristic установите переключатель в позицию

Compute/use alpha for orthogonality. Нажмите кнопку Display/print/save design. На экране появится матрица плана ОЦКП в кодированном виде (см. рисунок 7.23).

2. Сохраните матрицу плана ОЦКП в кодированном виде в файле с именем **Plan_OCKP_kod.sta**.

В нашем примере для плана ОЦКП таблица 7.8 будет иметь другой вид – таблица 7.10.

Обратите внимание, что полученные значения факторов в точках (–1) и (+1) затруднительно реализовать в ходе данного эксперимента.

Таблица 7.10 – Факторы и их уровни варьирования (пять уровней варьирования факторов)

Факторы	Уровни варьирования				
	-1,48258	-1	0	1	1,48258
Продолжительность сушки, τ, ч	1	1,17	1,5	1,83	2
Влажность зерна, W_n, %	16	17,35	20	22,65	24
Температура агента сушки, T_a, °С	20	22,36	27	31,64	34
Скорость агента сушки, V_a, м/с	0,6	0,67	0,8	0,93	1

3. Для продолжения работы нажмите кнопку Continue.

4. Выведите на экран таблицу для ввода имен факторов и уровней варьирования и заполните её. Результат сравните с рисунком 7.24.

Continue...	A	B	C	D
1	-1,00000	-1,00000	-1,00000	-1,00000
2	-1,00000	-1,00000	-1,00000	1,00000
3	-1,00000	-1,00000	1,00000	-1,00000
4	-1,00000	-1,00000	1,00000	1,00000
5	-1,00000	1,00000	-1,00000	-1,00000
6	-1,00000	1,00000	-1,00000	1,00000
7	-1,00000	1,00000	1,00000	-1,00000
8	-1,00000	1,00000	1,00000	1,00000
9	1,00000	-1,00000	-1,00000	-1,00000
10	1,00000	-1,00000	-1,00000	1,00000
11	1,00000	-1,00000	1,00000	-1,00000
12	1,00000	-1,00000	1,00000	1,00000
13	1,00000	1,00000	-1,00000	-1,00000
14	1,00000	1,00000	-1,00000	1,00000
15	1,00000	1,00000	1,00000	-1,00000
16	1,00000	1,00000	1,00000	1,00000
17	-1,48258	0,00000	0,00000	0,00000
18	1,48258	0,00000	0,00000	0,00000
19	0,00000	-1,48258	0,00000	0,00000
20	0,00000	1,48258	0,00000	0,00000
21	0,00000	0,00000	-1,48258	0,00000
22	0,00000	0,00000	1,48258	0,00000
23	0,00000	0,00000	0,00000	-1,48258
24	0,00000	0,00000	0,00000	1,48258
25 (C)	0,00000	0,00000	0,00000	0,00000
26 (C)	0,00000	0,00000	0,00000	0,00000

Рисунок 7.23 – Матрица плана ОЦКП для четырех факторов в кодированном виде

Continue...	To change labels, values, etc., type in the desired changes, then click Continue.								
Factor	Factor Name	Low Value	Low Label	Center Value	Center Label	High Value	High Label	Star Low Label	Star Hi. Label
A (1)	A	1,17000	Low	1,50000	CenterPt	1,83000	High	StarLow	StarHigh
B (2)	B	17,35000	Low	20,00000	CenterPt	22,65000	High	StarLow	StarHigh
C (3)	C	22,36000	Low	27,00000	CenterPt	31,64000	High	StarLow	StarHigh
D (4)	D	,67000	Low	,80000	CenterPt	,93000	High	StarLow	StarHigh

Рисунок 7.24 – Окно с отредактированной таблицей имен факторов и их значений

5. Для продолжения работы нажмите кнопку Continue.

6. В окне Design of an Experiment with Three-Level Factors в разделе Show установите переключатель в положение Mini/maxima и нажмите кнопку Display/print/save design. На экране появится матрица планирования эксперимента в именованном виде (рисунок 7.25).

Continue...	A	B	C	D
1	1,170000	17,35000	22,36000	,670000
2	1,170000	17,35000	22,36000	,930000
3	1,170000	17,35000	31,64000	,670000
4	1,170000	17,35000	31,64000	,930000
5	1,170000	22,65000	22,36000	,670000
6	1,170000	22,65000	22,36000	,930000
7	1,170000	22,65000	31,64000	,670000
8	1,170000	22,65000	31,64000	,930000
9	1,830000	17,35000	22,36000	,670000
10	1,830000	17,35000	22,36000	,930000
11	1,830000	17,35000	31,64000	,670000
12	1,830000	17,35000	31,64000	,930000
13	1,830000	22,65000	22,36000	,670000
14	1,830000	22,65000	22,36000	,930000
15	1,830000	22,65000	31,64000	,670000
16	1,830000	22,65000	31,64000	,930000
17	1,010749	20,00000	27,00000	,800000
18	1,989251	20,00000	27,00000	,800000
19	1,500000	16,07117	27,00000	,800000
20	1,500000	23,92883	27,00000	,800000
21	1,500000	20,00000	20,12084	,800000
22	1,500000	20,00000	33,87916	,800000
23	1,500000	20,00000	27,00000	,607265
24	1,500000	20,00000	27,00000	,992735
25 (C)	1,500000	20,00000	27,00000	,800000
26 (C)	1,500000	20,00000	27,00000	,800000

Рисунок 7.25 – Окно с матрицей планирования эксперимента в именованном виде

7. Сохраните матрицу планирования эксперимента в именованном виде в файл с именем **Plan_OCKP_f.sta**.

8. Для продолжения работы нажмите кнопку Continue.

9. Для просмотра матрицы взаимного влияния коэффициентов уравнения регрессии в окне Design of an Experiment with Three-Level Factors нажмите кнопку **Correlation matrix (main effects and interactions)**. На экране появится матрица взаимного влияния коэффициентов уравнения регрессии (см. рисунок 7.26). Сделайте вывод о взаимном влиянии коэффициентов уравнения регрессии.

10. Закройте окна с матрицами плана ОЦКП.

Continue...	2**(4) central composite, nc=16 ns=8 n0=2 Runs=26													
Effect	A Linea	A Quadra	B Linea	B Quadra	C Linea	C Quadra	D Linea	D Quadra	1L*2L	1L*3L	1L*4L	2L*3L	2L*4L	3L*4L
(1)A (L)	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
A (Q)	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
(2)B (L)	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
B (Q)	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
(3)C (L)	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
C (Q)	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
(4)D (L)	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
D (Q)	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00
1L*2L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
1L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
1L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
2L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00
2L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00
3L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Рисунок 7.26 – Матрица взаимного влияния коэффициентов уравнения регрессии

Построение плана типа Бокса B_n

1. В окне Designs of a Response Surface Experiment (Central Composite Designs) для настройки параметров плана (см. рисунок 7.18) в разделе Design characteristic установите переключатель в позицию Face centered star points (e.g., ±1). Нажмите кнопку Display/print/save design. На экране появится матрица плана ОЦКП в кодированном виде (рисунок 7.27).

Continue...	A	B	C	D
1	-1,00000	-1,00000	-1,00000	-1,00000
2	-1,00000	-1,00000	-1,00000	1,00000
3	-1,00000	-1,00000	1,00000	-1,00000
4	-1,00000	-1,00000	1,00000	1,00000
5	-1,00000	1,00000	-1,00000	-1,00000
6	-1,00000	1,00000	-1,00000	1,00000
7	-1,00000	1,00000	1,00000	-1,00000
8	-1,00000	1,00000	1,00000	1,00000
9	1,00000	-1,00000	-1,00000	-1,00000
10	1,00000	-1,00000	-1,00000	1,00000
11	1,00000	-1,00000	1,00000	-1,00000
12	1,00000	-1,00000	1,00000	1,00000
13	1,00000	1,00000	-1,00000	-1,00000
14	1,00000	1,00000	-1,00000	1,00000
15	1,00000	1,00000	1,00000	-1,00000
16	1,00000	1,00000	1,00000	1,00000
17	-1,00000	0,00000	0,00000	0,00000
18	1,00000	0,00000	0,00000	0,00000
19	0,00000	-1,00000	0,00000	0,00000
20	0,00000	1,00000	0,00000	0,00000
21	0,00000	0,00000	-1,00000	0,00000
22	0,00000	0,00000	1,00000	0,00000
23	0,00000	0,00000	0,00000	-1,00000
24	0,00000	0,00000	0,00000	1,00000
25 (C)	0,00000	0,00000	0,00000	0,00000
26 (C)	0,00000	0,00000	0,00000	0,00000

Рисунок 7.27 – Матрица плана типа Бокса B_n

2. Сохраните матрицу плана в кодированном виде в файле с именем **Plan_Bn_kod.sta**.

3. Для продолжения работы нажмите кнопку Continue.

4. Выведите на экран таблицу для ввода имен факторов и уровней варьирования и заполните её. Результат сравните с рисунком 7.28.

Continue... | To change labels, values, etc., type in the desired changes, then click Continue.

Factor	Factor Name	Low Value	Low Label	Center Value	Center Label	High Value	High Label	Star Low Label	Star Hi. Label
A (1)	t	1,00000	Low	1,500000	CenterPt	2,000000	High	StarLow	StarHigh
B (2)	W	16,00000	Low	20,00000	CenterPt	24,00000	High	StarLow	StarHigh
C (3)	Ta	20,00000	Low	27,00000	CenterPt	34,00000	High	StarLow	StarHigh
D (4)	Va	,60000	Low	,80000	CenterPt	1,00000	High	StarLow	StarHigh

Рисунок 7.28 – Окно с отредактированной таблицей имен факторов и их значений

5. Для продолжения работы нажмите кнопку Continue.

6. В окне Design of an Experiment with Three-Level Factors в разделе Show установите переключатель в положение **Mini/maxima** и нажмите кнопку Display/print/save design. На экране появится матрица планирования эксперимента в именованном виде (рисунок 7.29).

Continue...	t	W	Ta	Va
1	1,000000	16,00000	20,00000	,600000
2	1,000000	16,00000	20,00000	1,000000
3	1,000000	16,00000	34,00000	,600000
4	1,000000	16,00000	34,00000	1,000000
5	1,000000	24,00000	20,00000	,600000
6	1,000000	24,00000	20,00000	1,000000
7	1,000000	24,00000	34,00000	,600000
8	1,000000	24,00000	34,00000	1,000000
9	2,000000	16,00000	20,00000	,600000
10	2,000000	16,00000	20,00000	1,000000
11	2,000000	16,00000	34,00000	,600000
12	2,000000	16,00000	34,00000	1,000000
13	2,000000	24,00000	20,00000	,600000
14	2,000000	24,00000	20,00000	1,000000
15	2,000000	24,00000	34,00000	,600000
16	2,000000	24,00000	34,00000	1,000000
17	1,000000	20,00000	27,00000	,800000
18	2,000000	20,00000	27,00000	,800000
19	1,500000	16,00000	27,00000	,800000
20	1,500000	24,00000	27,00000	,800000
21	1,500000	20,00000	20,00000	,800000
22	1,500000	20,00000	34,00000	,800000
23	1,500000	20,00000	27,00000	,600000
24	1,500000	20,00000	27,00000	1,000000
25 (C)	1,500000	20,00000	27,00000	,800000
26 (C)	1,500000	20,00000	27,00000	,800000

Рисунок 7.29 – Окно с матрицей планирования эксперимента в именованном виде

7. Сохраните матрицу планирования эксперимента в именованном виде в файл с именем **Plan_Bn_f.sta**.

8. Для продолжения работы нажмите кнопку Continue.

9. Для просмотра матрицы взаимного влияния коэффициентов уравнения регрессии в окне Design of an Experiment with Three-Level Factors нажмите кнопку Correlation matrix (main effects and interactions). На экране появится матрица взаимного влияния коэффициентов уравнения регрессии (рисунок 7.30). Сделайте вывод о взаимном влиянии коэффициентов уравнения регрессии.

10. Закройте окна с матрицами плана. Закройте программу.

Continue...	2**(4) central composite, nc=16 ns=8 n0=2 Runs=26													
Effect	t Linea	t Quadra	W Linea	W Quadra	Ta Linea	Ta Quadra	Va Linea	Va Quadra	1L*2L	1L*3L	1L*4L	2L*3L	2L*4L	3L*4L
(1)t (L)	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
t (Q)	0,00	1,00	0,00	,64	0,00	,64	0,00	,64	0,00	0,00	0,00	0,00	0,00	0,00
(2)W (L)	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
W (Q)	0,00	,64	0,00	1,00	0,00	,64	0,00	,64	0,00	0,00	0,00	0,00	0,00	0,00
(3)Ta (L)	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Ta (Q)	0,00	,64	0,00	,64	0,00	1,00	0,00	,64	0,00	0,00	0,00	0,00	0,00	0,00
(4)Va (L)	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
Va (Q)	0,00	,64	0,00	,64	0,00	,64	0,00	1,00	0,00	0,00	0,00	0,00	0,00	0,00
1L*2L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00	0,00
1L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00	0,00
1L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00	0,00
2L*3L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00	0,00
2L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00	0,00
3L*4L	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	1,00

Рисунок 7.30 – Матрица взаимного влияния коэффициентов уравнения регрессии

По результатам построения планов эксперимента для четырех факторов на 3 уровнях сделаем следующие выводы:

– План ПФЭ 3^n, при $n = 4$ содержит 81 опыт. Коэффициенты уравнения регрессии определяются независимо друг от друга. Факторы на всех уровнях варьирования управляемы. Но, опыты требуют трехкратного повторения, поэтому общее число опытов в эксперименте – 343.

– План Бокса-Бенкина, при $n = 4$ содержит 27 опытов. Три коэффициента уравнения регрессии имеют слабое взаимное влияние. Факторы на всех уровнях варьирования управляемы. При трехкратном повторении – 81 опыт.

– План РЦКП, при $n = 4$ содержит 26 опытов. Три коэффициента уравнения регрессии имеют слабое взаимное влияние. Факторы на всех уровнях варьирования управляемы, но необходимо правильно вычислить плечо «звездных» точек. При трехкратном повторении – 78 опытов.

– План ОЦКП, при $n = 4$ содержит 26 опытов. Коэффициенты уравнения регрессии определяются независимо друг от друга. Но необходимо правильно вычислить плечо «звездных» точек и для рассмотренной выше задачи, при таком виде планирования факторами сложно управлять на уровнях (-1), (+1). При трехкратном повторении – 78 опытов.

– План типа Бокса B_n $n = 4$ содержит 26 опытов. Три коэффициента уравнения регрессии имеют умеренное взаимное влияние. Факторы на всех

уровнях варьирования управляемы. При трехкратном повторении – 78 опытов.

В результате, выберем план Бокса-Бенкина, т.к. он содержит 27 опытов (практически столько же, что и планы РЦКП, ОЦКП и B_n), но коэффициента уравнения регрессии имеют слабое взаимное влияние (отличие от B_n), факторы на всех уровнях варьирования управляемы и нет звездных точек.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Сгенерируйте план для полного пятифакторного эксперимента на двух уровнях. Оцените взаимное влияние коэффициентов уравнения регрессии.
2. Сгенерируйте полуреплику и четвертьреплику для пятифакторного эксперимента на двух уровнях. Оцените взаимное влияние коэффициентов уравнения регрессии. Сделайте выводы о применимости полученных планов.
3. Выберите план для эксперимента (таблица 7.11) из трех перечисленных: полный факторный, полуреплика или четвертьреплика. Обоснуйте выбор.
4. Задайте имена и уровни варьирования факторов для модели, представленной в таблице 7.11. В таблице представлены факторы и уровни и их варьирования для проведения многофакторного эксперимента по исследованию влияния количества внесенных удобрений U, температуры почвы T, количества вносимой суточной влаги W, освещенности O и густоты стояния растений G на скорость созревания редиса в защищенном грунте V (сут.).

Таблица 7.11 – Факторы и уровни варьирования (на двух уровнях варьирования факторов)

№ фактора	Имя фактора	Нижний уровень	Верхний уровень
1	Количество внесенных удобрений U, г/м^2	50	200
2	Температура почвы T, ℃	10	25
3	Количество вносимой суточной влаги W, л/м^2	0,2	0,8
4	Освещенность, O лк/м^2	100	500
5	Густота растений G, шт./м^2	100	300

5. Сгенерируйте план данной модели, добавив столбец для зависимой переменной (отклика).
6. Сохраните файл с именем **Model1.sta**.
7. Сгенерируйте план для трехуровневого полного пятифакторного эксперимента для данных, представленных в таблице 7.12. Оцените взаимодействие эффектов.

Таблица 7.12 – Факторы и уровни варьирования

№ фактора	Имя фактора	–1	0	+1
1	Количество внесенных удобрений U, г/м^2	50	125	200
2	Температура почвы T, ℃	10	18	26
3	Количество вносимой суточной влаги W, л/м^2	0,2	0,5	0,8
4	Освещенность, O лк/м^2	100	300	500
5	Густота растений G, шт./м^2	100	200	300

8. Сгенерируйте четвертьреплику для трехуровневого пятифакторного эксперимента.

9. Сгенерируйте план Бокса-Бенкина для пяти факторов.

10. Получите центральные композиционные планы с α для ортогональности, ротатабельности и $\alpha=\pm1$.

11. Сравните число опытов и степень взаимодействия эффектов для всех полученных планов.

12. Выберите наилучший план эксперимента для поставленной задачи. Обоснование выбора наилучшего плана запишите в отчет.

13. Задайте имена и уровни варьирования факторов для выбранного плана эксперимента.

14. Сгенерируйте план эксперимента в именованных факторах. Сохраните файл как данные с именем **Plan_3.sta**.

5 СОДЕРЖАНИЕ ОТЧЕТА

Отчет оформляется в соответствии с требованиями, принятыми в АЧИИ. Отчет должен содержать:

- ответы на контрольные вопросы;
- результаты планирования эксперимента.

6 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что понимается под экспериментом?
2. Что называется опытом, планом эксперимента, планированием эксперимента?
3. Сформулируйте цель планирования эксперимента.
4. Перечислите этапы нахождения уравнения регрессии методом планирования эксперимента.
5. Объясните суть кодирования факторов.
6. Какие планы называются планами первого порядка?
7. Какой план называется ортогональным? Ротатабельным?
8. В каком модуле программы Statistica строятся планы эксперимента?

Практическое занятие № 8

Тема: **РЕАЛИЗАЦИЯ ПЛАНА МНОГОФАКТОРНОГО ЭКСПЕРИМЕНТА. ОБРАБОТКА И АНАЛИЗ РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА. АНАЛИЗ РЕГРЕССИОННОЙ МОДЕЛИ**

1 ЦЕЛЬ РАБОТЫ

Научиться:

- обрабатывать результаты эксперимента в программе Statistica;
- анализировать результаты обработки экспериментальных данных.

2 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

В данной лабораторной работе вы рассмотрите на примере процесс обработки результатов эксперимента от составления плана эксперимента до анализа полученного регрессионного уравнения.

Задача. Провести планирование эксперимента по определению влияния количества внесенных в почву удобрений, температуры почвы, количества вносимой суточной влаги, освещенности и густоты растений на время созревания растений. Построить регрессионное уравнение зависимости времени созревания растений от перечисленных факторов. Выполнить анализ полученного регрессионного уравнения. Факторы и их уровни варьирования указаны в таблице 8.1.

Таблица 8.1 – Факторы и их уровни варьирования

№ фактора	Имя фактора	–1	0	+1
1	Количество внесенных удобрений, U, г/м^2	50	125	200
2	Температура почвы, Tp, ℃	10	17,5	25
3	Количество вносимой суточной влаги, W, л/м^2	0,2	0,5	0,8
4	Освещенность, L, лк/м^2	100	300	500
5	Густота растений, G, шт./м^2	100	200	300

1. Выбор плана эксперимента для пяти факторов на трех уровнях.

Необходимо выбрать такой план эксперимента, чтобы выполнялись следующие условия:

- количество опытов должно позволять выполнить эксперимент в трёхкратной повторности;
- коэффициенты уравнения регрессии должны определяться независимо друг от друга;
- факторы на всех уровнях должны быть управляемыми.

Сначала сгенерируйте планы эксперимента для пяти факторов на трех уровнях:

– План ПФЭ 3^n, при $n = 5$ содержит 243 опыта. Коэффициенты уравнения регрессии определяются независимо друг от друга. Факторы на всех уровнях варьирования управляемы.

– План Бокса-Бенкина, при $n = 5$ содержит 46 опытов. Четыре коэффициента уравнения регрессии имеют слабое взаимное влияние. Факторы на всех уровнях варьирования управляемы.

– План РЦКП, при $n = 5$ содержит 27 опытов. Четыре коэффициента уравнения регрессии имеют слабое взаимное влияние. Факторы на всех уровнях варьирования управляемы. Но необходимо правильно вычислить плечо «звездных» точек.

– План ОЦКП, при $n = 5$ содержит 27 опытов. Коэффициенты уравнения регрессии определяются независимо друг от друга. Но необходимо правильно вычислить плечо «звездных» точек и для рассмотренной выше задачи, при таком виде планирования, факторами сложно управлять на уровнях (-1), (+1).

– Композиционный план типа B_n при $n = 5$ содержит 27 опытов. Но четыре коэффициента уравнения регрессии имеют умеренное взаимное влияние.

Исходя из выше изложенного, выберите для планирования план типа B_n, т.к. количество опытов – 27, следовательно, в трёхкратной повторности будет 81 опыт и нет необходимости вычислять «звёздные» точки.

После выбора плана эксперимента и проведения опытов, будут получены значения зависимой переменной, согласно экспериментальному плану (таблица 8.2).

Таблица 8.2 – Значения зависимой переменной

Номер опыта	Время созревания, *t*, сутки
1	*2*
1	57
2	44
3	46
4	44
5	38
6	36
7	38
8	29
9	50
10	47
11	50
12	38
13	41
14	31

Продолжение таблицы 8.2

1	*2*
15	33
16	31
17	42
18	41
19	43
20	32
21	42
22	38
23	43
24	37
25	39
26	40
27	36

1.1. В стартовой панели Design & Analysis of Experiments модуля Experimental Design (рисунок 8.1) в списке планов эксперимента выберите план Central composite, non-factorial, surface design и нажмите кнопку OK. На экране появится диалоговое окно Central Composite (Response Surface) Designs (см. рисунок 8.2).

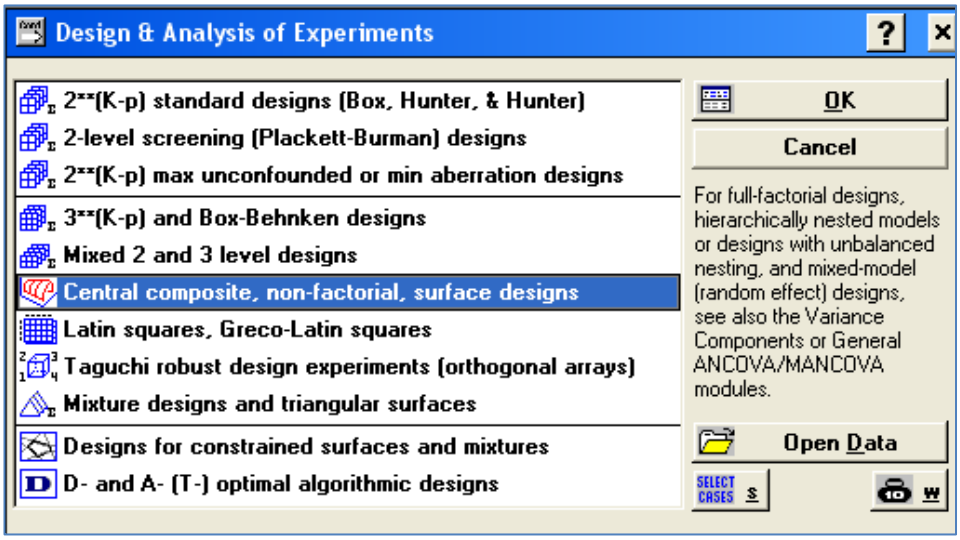


Рисунок 8.1 – Стартовая панель модуля Experimental Design

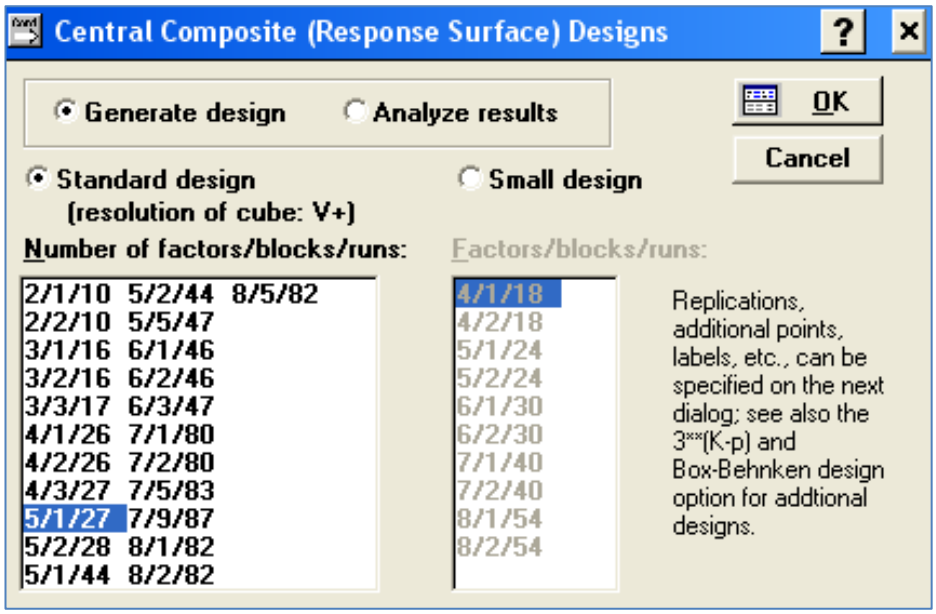


Рисунок 8.2 – Окно выбора центральных композиционных планов

1.2. В окне Central Composite (Response Surface) Designs в списке Standard Design выберите **5/1/27** – **5** факторов, **1** блок, **27** опытов. Нажмите кнопку OK.

1.3. На экране появится окно Designs of a Response Surface Experiment (Central Composite Designs) для настройки параметров плана (рисунок 8.3). В этом окне выбирают значения α для ортогональности, рота-табельности.

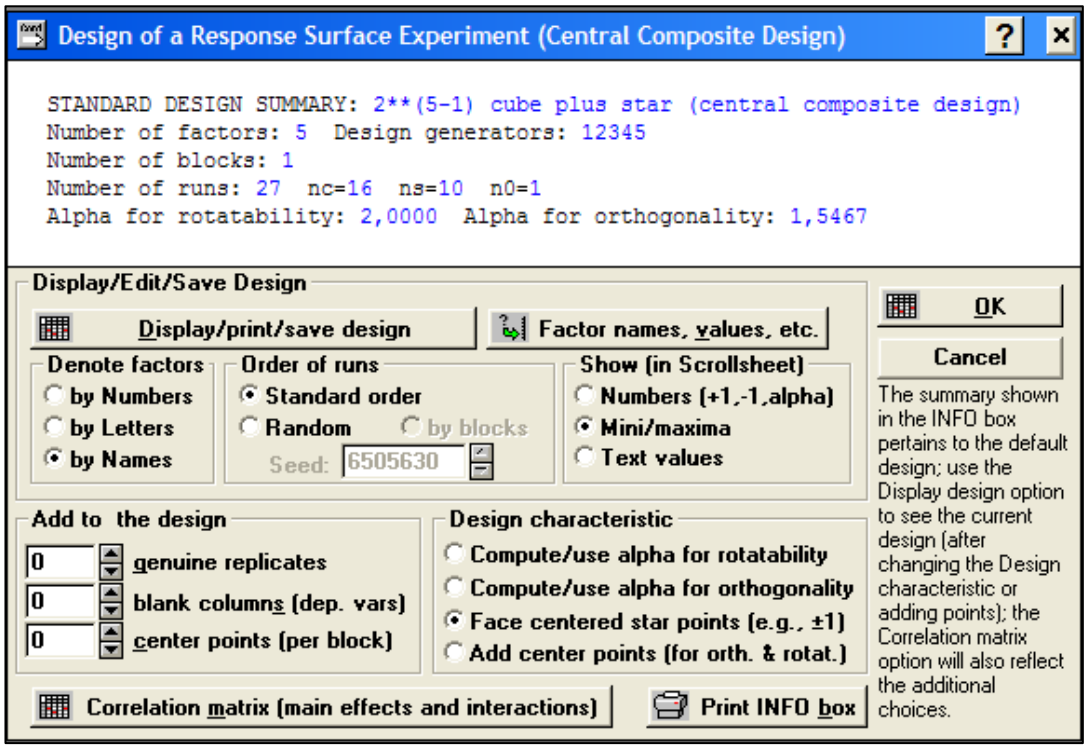


Рисунок 8.3 – Окно для настройки параметров
центральных композиционных планов

1.4. В разделе Design characteristic установите переключатель в позицию Face centered star points (e. g., ± 1). Нажмите кнопку Display/print/save design. На экране появится матрица плана РЦКП в кодированном виде (рисунок 8.4).

1.5. Сохраните матрицу плана типа B_n для факторов в кодированном виде в файле с именем **Plan_kod.sta**.

1.6. Закройте диалоговое окно выбора плана эксперимента. Закройте сохраненные файлы.

1.7. Выполните команду File – Open Data – Plan_kod.sta. На экране появится таблица плана эксперимента для факторов в кодированном виде. Переименуйте столбцы таблицы в X_1, X_2, X_3, X_4, X_5. Отформатируйте данные. Добавьте столбец для зависимой переменной t и заполните его значениями (см. рисунок 8.5). Сохраните файл.

Continue...	A	B	C	D	E
1	-1,00000	-1,00000	-1,00000	-1,00000	1,00000
2	-1,00000	-1,00000	-1,00000	1,00000	-1,00000
3	-1,00000	-1,00000	1,00000	-1,00000	-1,00000
4	-1,00000	-1,00000	1,00000	1,00000	1,00000
5	-1,00000	1,00000	-1,00000	-1,00000	-1,00000
6	-1,00000	1,00000	-1,00000	1,00000	1,00000
7	-1,00000	1,00000	1,00000	-1,00000	1,00000
8	-1,00000	1,00000	1,00000	1,00000	-1,00000
9	1,00000	-1,00000	-1,00000	-1,00000	-1,00000
10	1,00000	-1,00000	-1,00000	1,00000	1,00000
11	1,00000	-1,00000	1,00000	-1,00000	1,00000
12	1,00000	-1,00000	1,00000	1,00000	-1,00000
13	1,00000	1,00000	-1,00000	-1,00000	1,00000
14	1,00000	1,00000	-1,00000	1,00000	-1,00000
15	1,00000	1,00000	1,00000	-1,00000	-1,00000
16	1,00000	1,00000	1,00000	1,00000	1,00000
17	-1,00000	0,00000	0,00000	0,00000	0,00000
18	1,00000	0,00000	0,00000	0,00000	0,00000
19	0,00000	-1,00000	0,00000	0,00000	0,00000
20	0,00000	1,00000	0,00000	0,00000	0,00000
21	0,00000	0,00000	-1,00000	0,00000	0,00000
22	0,00000	0,00000	1,00000	0,00000	0,00000
23	0,00000	0,00000	0,00000	-1,00000	0,00000
24	0,00000	0,00000	0,00000	1,00000	0,00000
25	0,00000	0,00000	0,00000	0,00000	-1,00000
26	0,00000	0,00000	0,00000	0,00000	1,00000
27 (C)	0,00000	0,00000	0,00000	0,00000	0,00000

Рисунок 8.4 – Матрица плана типа Bn для пяти факторов в кодированном виде

NUMERIC VALUES	2**(5-1) central composite, nc=16 ns=10 n0=1					
	1 X1	2 X2	3 X3	4 X4	5 X5	6 T
1	-1	-1	-1	-1	1	57
2	-1	-1	-1	1	-1	44
3	-1	-1	1	-1	-1	46
4	-1	-1	1	1	1	44
5	-1	1	-1	-1	-1	38
6	-1	1	-1	1	1	36
7	-1	1	1	-1	1	38
8	-1	1	1	1	-1	29
9	1	-1	-1	-1	-1	50
10	1	-1	-1	1	1	47
11	1	-1	1	-1	1	50
12	1	-1	1	1	-1	38
13	1	1	-1	-1	1	41
14	1	1	-1	1	-1	31
15	1	1	1	-1	-1	33
16	1	1	1	1	1	31
17	-1	0	0	0	0	42
18	1	0	0	0	0	41
19	0	-1	0	0	0	43
20	0	1	0	0	0	32
21	0	0	-1	0	0	42
22	0	0	1	0	0	38
23	0	0	0	-1	0	43
24	0	0	0	1	0	37
25	0	0	0	0	-1	39
26	0	0	0	0	1	40
27 (C)	0	0	0	0	0	36

Рисунок 8.5 – Матрица плана типа Bn для пяти факторов в кодированном виде с зависимой переменной

2. Построение уравнения регрессии зависимости функции отклика от факторов в кодированном виде.

2.1. Выполните команду Analysis – Resume Analysis. В появившемся окне Design & Analysis of Experiments (см. рисунок 8.1) в списке планов эксперимента выберите план Central composite, non-factorial, surface design и нажмите кнопку OK.

2.2. На экране появится диалоговое окно Central Composite (Response Surface) Designs (см. рисунок 8.2). Установите переключатель в позицию Analyze results (рисунок 8.6).

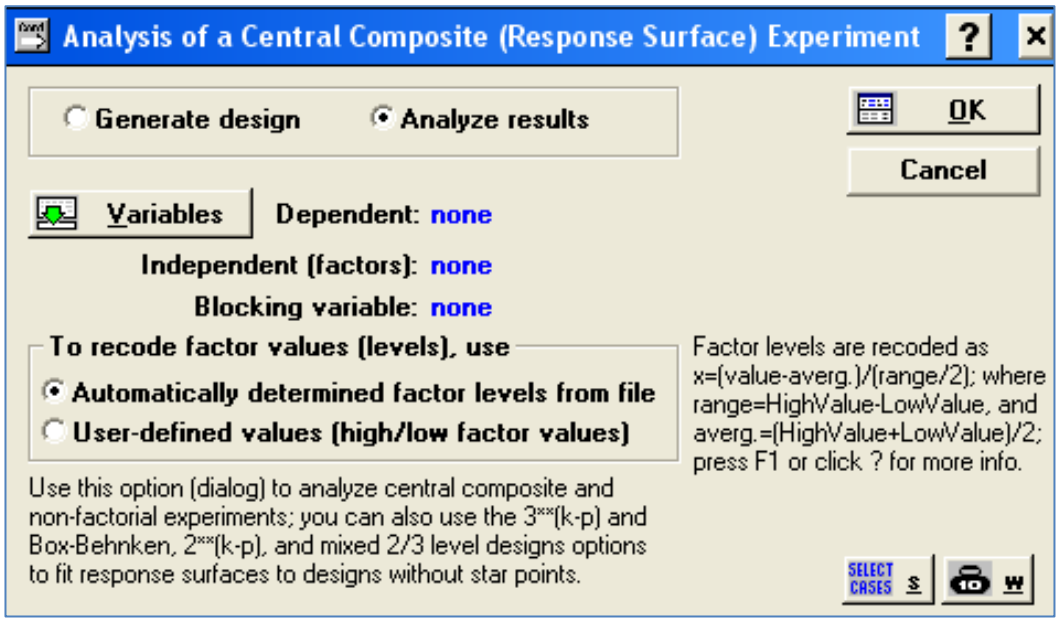


Рисунок 8.6 – Окно для анализа экспериментальных данных

2.3. Для выбора зависимых и независимых переменных щелкните по кнопке Variables (рисунок 8.6). В появившемся окне Select dependent var. and (optional) indep. and blocking var. выберите зависимую и независимые переменные, как показано на рисунке 8.7. Нажмите кнопку OK.

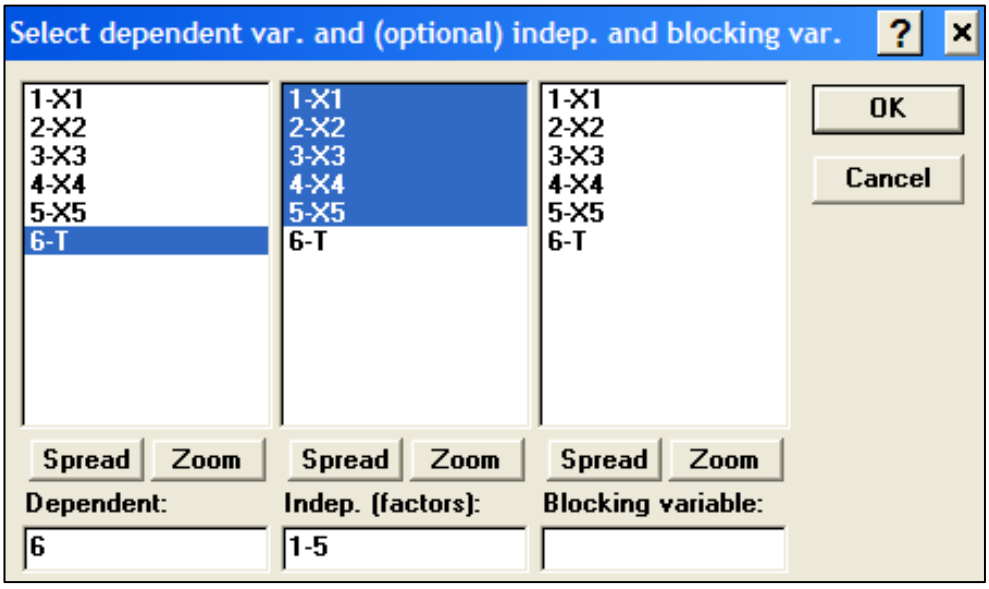


Рисунок 8.7 – Окно для выбора зависимой и независимых переменных

2.4. В окне Analysis of a Central Composite (Response Surface) Experiment нажмите кнопку OK. Окно Analysis of a Central Composite (Response Surface) Experiment примет вид, как показано на рисунке 8.8.

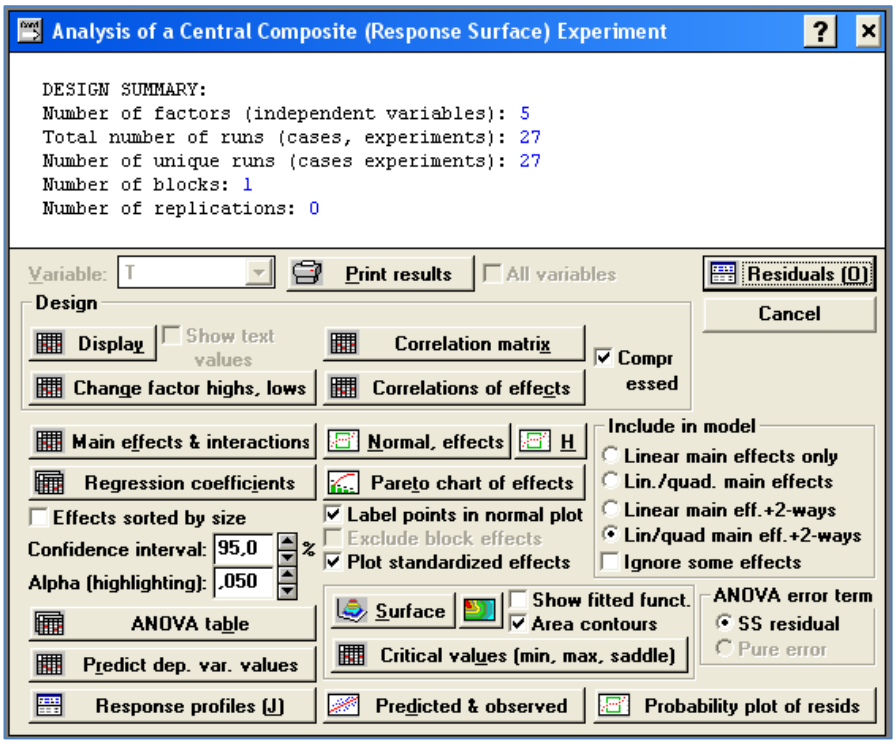


Рисунок 8.8 – Окно Analysis of a Central Composite (Response Surface) Experiment

Опции окна Analysis of a Central Composite (Response Surface) Experiment перечислены в таблице 8.3.

Таблица 8.3 – Назначение опций окна Analysis of a Central Composite (Response Surface) Experiment

Элемент окна	Назначение элемента
1	2
Группа кнопок Design (Планирование)	Позволяет просмотреть план и результаты опыта (Display), изменить уровни факторов (Change factor…), оценить корреляцию каждого коэффициента с откликом (Correlation matrix) и степень смешанности коэффициентов регрессии (Correlations of effects)
Regression coefficients	Открывается окно с результатами расчета коэффициентов регрессии и их статистик (стандартная ошибка, *t*-критерий Стьюдента (причем, значимые коэффициенты выделены красным цветом), вероятность ошибки *p*, верхняя и нижняя границы доверительного интервала, а также коэффициент детерминации)

Окончание таблицы 18.3

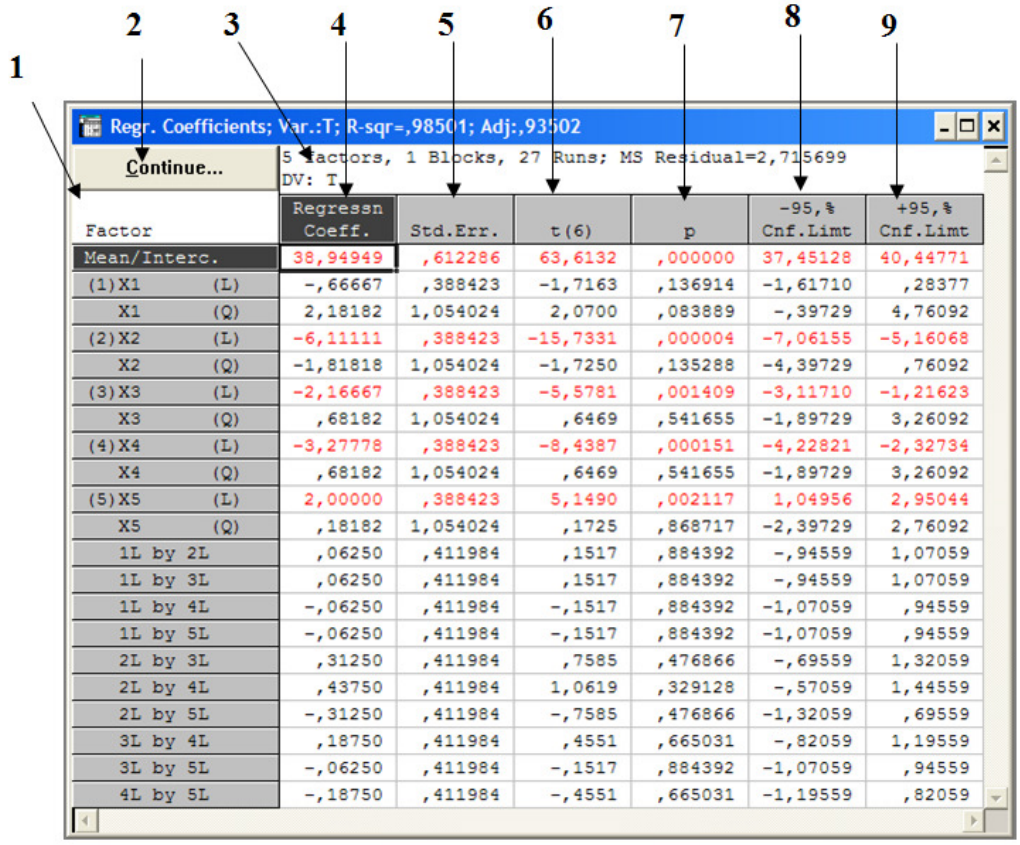
1	2
Main effects & interactions	Показаны эффекты, вносимые каждым коэффициентом, с аналогичными статистиками
Confidence interval	Задает доверительный интервал (по умолчанию 95%)
ANOVA table	Основные статистики (SS – сумма квадратов, df – степени свободы, MS – средний квадрат, F – статистика Фишера, p – вероятность ошибки)
Normal, effects	График эффектов и взаимодействий на нормализованной плоскости
Pareto chart of effects	Диаграмма, отражающая эффект, вносимый в отклик каждым из коэффициентов
Группа Include in model	Позволяет включить в модель не все рассчитанные коэффициенты. После исключения незначимых коэффициентов из модели значимые пересчитываются автоматически
Для анализа полученного уравнения регрессии и визуализации зависимостей служат следующие кнопки:	
Predict dep. var. values	Позволяет вычислить отклик, задаваясь значениями факторов
Response profiles	Позволяет получить графики зависимости отклика от одного фактора при фиксированных в центре остальных факторах, а также двумерные сечения поверхности отклика при тех же условиях
Surface	Поверхности отклика и сечения при выбранных уровнях факторов
Critical values	Значения факторов и отклика в критических точках
Predicted & observed	График наблюдаемых (опытных) и вычисленных значений
Probability plot of resids.	Вероятностный график остатков
Residuals	Для анализа остатков

При активации кнопки Residuals открывается окно Residual statistics, содержащее кнопки, назначение которых описано в таблице 8.4.

Таблица 8.4 – Назначение опций окна Residual statistics

Элемент окна	Назначение элемента
Display predicted and residual values	Показывает наблюдаемые и вычисленные значения отклика, остатки и их расширенные статистики
Группа Plots of predicted and residual values	Позволяет построить гистограммы остатков и вычисленных значений; графики различных комбинаций остатков, вычисленных и наблюдаемых значений
Группа Box-Cox transformation	Использует преобразование Бокса-Кокса для зависимой переменной

2.5. Для расчета коэффициентов уравнения регрессии нажмите кнопку Regression coefficients (см. рисунок 8.8). На экране появится таблица результатов регрессионного анализа (рисунок 8.9).



Factor	Regressn Coeff.	Std.Err.	t(6)	p	-95,% Cnf.Limt	+95,% Cnf.Limt
Mean/Interc.	38,94949	,612286	63,6132	,000000	37,45128	40,44771
(1)X1 (L)	-,66667	,388423	-1,7163	,136914	-1,61710	,28377
X1 (Q)	2,18182	1,054024	2,0700	,083889	-,39729	4,76092
(2)X2 (L)	-6,11111	,388423	-15,7331	,000004	-7,06155	-5,16068
X2 (Q)	-1,81818	1,054024	-1,7250	,135288	-4,39729	,76092
(3)X3 (L)	-2,16667	,388423	-5,5781	,001409	-3,11710	-1,21623
X3 (Q)	,68182	1,054024	,6469	,541655	-1,89729	3,26092
(4)X4 (L)	-3,27778	,388423	-8,4387	,000151	-4,22821	-2,32734
X4 (Q)	,68182	1,054024	,6469	,541655	-1,89729	3,26092
(5)X5 (L)	2,00000	,388423	5,1490	,002117	1,04956	2,95044
X5 (Q)	,18182	1,054024	,1725	,868717	-2,39729	2,76092
1L by 2L	,06250	,411984	,1517	,884392	-,94559	1,07059
1L by 3L	,06250	,411984	,1517	,884392	-,94559	1,07059
1L by 4L	-,06250	,411984	-,1517	,884392	-1,07059	,94559
1L by 5L	-,06250	,411984	-,1517	,884392	-1,07059	,94559
2L by 3L	,31250	,411984	,7585	,476866	-,69559	1,32059
2L by 4L	,43750	,411984	1,0619	,329128	-,57059	1,44559
2L by 5L	-,31250	,411984	-,7585	,476866	-1,32059	,69559
3L by 4L	,18750	,411984	,4551	,665031	-,82059	1,19559
3L by 5L	-,06250	,411984	-,1517	,884392	-1,07059	,94559
4L by 5L	-,18750	,411984	-,4551	,665031	-1,19559	,82059

1 – факторы и их сочетания; 2 – кнопка для продолжения работы; 3 – параметры плана; 4 – коэффициенты регрессионного уравнения; 5 – средние ошибки оценок коэффициентов регрессии; 6 – критерий Стьюдента для каждого коэффициента регрессионного уравнения; 7 – вероятность *p* для квантили критерия Стьюдента *t*; 8, 9 – границы доверительного интервала

Рисунок 8.9 – Таблица результатов регрессионного анализа

2.6. Рассмотрите таблицу результатов регрессионного анализа. Обратите внимание, что в заголовке окна с таблицей результатов регрессионного анализа указано значение коэффициента детерминации R^2: R-sqr = 0,98501 (см. рисунок 8.9).

По таблице результатов регрессионного анализа можно выписать уравнение регрессии для нашей задачи. Но все коэффициенты регрессионного уравнения должны быть значимы на принятом уровне значимости $\alpha = 0{,}05$.

Сделать вывод о значимости коэффициентов уравнения регрессии можно и без определения t_T. В этом случае используют рассчитанную вероятность p для квантили критерия Стьюдента t. Эту вероятность сравнивают с принятым уровнем значимости $\alpha = 0{,}05$. Если $p < \alpha$, то коэффициент уравнения регрессии является значимым на выбранном уровне значимости α. При этом строка с вероятностью $p < \alpha$ выделяется красным цветом (см. рисунок 8.9). Поэтому необходимо исключить незначимые коэффициенты уравнения регрессии. Значимые коэффициенты будут пересчитаны автоматически.

Изменения основных статистик коэффициентов уравнения регрессии показаны в таблице ANOVA table (рисунок 8.10).

Замечание. *При исключении коэффициентов уравнения регрессии необходимо контролировать изменения основных статистик коэффициентов уравнения регрессии.*

ANOVA; Var.:T; R-sqr=,98501; Adj:,93502 (plan_kod.sta)

Continue... | 5 factors, 1 Blocks, 27 Runs; MS Residual=2,715699 DV: T

Factor	SS	df	MS	F	p
(1)X1 (L)	8,000	1	8,0000	2,9458	,136914
X1 (Q)	11,636	1	11,6364	4,2849	,083889
(2)X2 (L)	672,222	1	672,2222	247,5320	,000004
X2 (Q)	8,081	1	8,0808	2,9756	,135288
(3)X3 (L)	84,500	1	84,5000	31,1154	,001409
X3 (Q)	1,136	1	1,1364	,4184	,541655
(4)X4 (L)	193,389	1	193,3889	71,2115	,000151
X4 (Q)	1,136	1	1,1364	,4184	,541655
(5)X5 (L)	72,000	1	72,0000	26,5125	,002117
X5 (Q)	,081	1	,0808	,0298	,868717
1L by 2L	,062	1	,0625	,0230	,884392
1L by 3L	,062	1	,0625	,0230	,884392
1L by 4L	,063	1	,0625	,0230	,884392
1L by 5L	,062	1	,0625	,0230	,884392
2L by 3L	1,563	1	1,5625	,5754	,476866
2L by 4L	3,063	1	3,0625	1,1277	,329128
2L by 5L	1,562	1	1,5625	,5754	,476866
3L by 4L	,562	1	,5625	,2071	,665031
3L by 5L	,063	1	,0625	,0230	,884392
4L by 5L	,563	1	,5625	,2071	,665031
Error	16,294	6	2,7157		
Total SS	1086,667	26			

Рисунок 8.10 – Основные статистики

Для исключения незначимых коэффициентов регрессионного уравнения нужно воспользоваться группой опций Include in model окна Analysis of a Central Composite (Response Surface) Experiment (см. рисунок 8.8).

Замечание. *При исключении незначимых коэффициентов уравнения регрессии необходимо следить за изменениями коэффициента детерминации. Если исключение коэффициентов происходит правильно, то значение коэффициента детерминации изменяется незначительно (изменения во втором-четвертом знаке числа после запятой).*

2.7. В группе опций Include in model окна Analysis of a Central Composite (Response Surface) Experiment выберите опцию Ignore some effects. На экране появится предупреждение (рисунок 8.11). Нажмите кнопку OK. На экране появится окно для выбора факторов и их сочетаний, коэффициенты при которых будут исключены из регрессионной модели (рисунок 8.12).

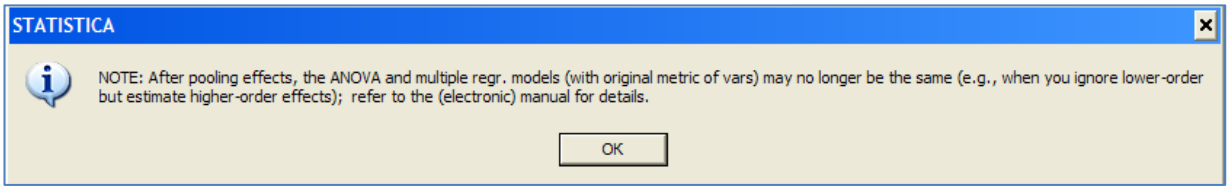


Рисунок 8.11 – Окно с предупреждением

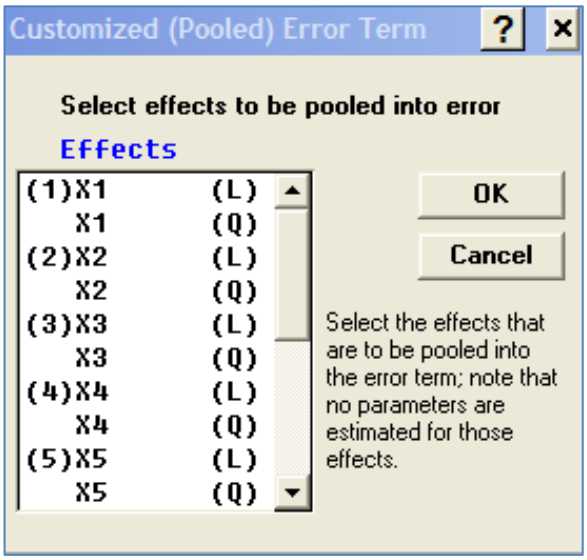


Рисунок 8.12 – Окно для исключения коэффициентов регрессионного уравнения

Факторы и их сочетания, коэффициенты при которых необходимо исключить, выбираются по значениям вероятности p таблицы результатов регрессионного анализа: первым исключается коэффициент, у которого вероятность p самая большая (см. рисунок 8.9). В нашем случае это коэффициенты при сочетаниях факторов: 1 и 2, 1 и 3, 1 и 4, 1 и 5, 3 и 5. Коэффициенты можно исключать по одному или группой. Но при этом обязательно нужно

контролировать изменения коэффициента детерминации и основных статистик (см. рисунки 8.9, 8.10).

2.8. Укажите в окне (рисунок 8.12) те факторы и их сочетания, коэффициенты при которых необходимо исключить. Для этого выделите факторы и их сочетания при нажатой клавише **Ctrl** (рисунок 8.13). Нажмите кнопку OK.

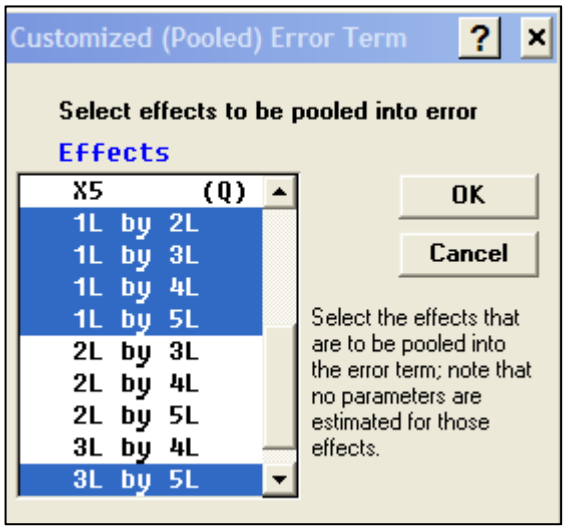


Рисунок 8.13 – Выбор коэффициентов для исключения

2.9. В окне Analysis of a Central Composite (Response Surface) Experiment нажмите кнопку Regression coefficients (см. рисунок 8.8). На экране появится таблица результатов регрессионного анализа, в которой будут отсутствовать исключенные коэффициенты уравнения регрессии и пересчитаны оставшиеся (рисунок 8.14).

Regr. Coefficients; Var.:T; R-sqr=,98472; Adj:,96388

Continue...

5 factors, 1 Blocks, 27 Runs; MS Residual=1,509699
DV: T

Factor	Regressn Coeff.	Std.Err.	t(11)	p	-95,% Cnf.Limt	+95,% Cnf.Limt
Mean/Interc.	38,94949	,456519	85,3184	,000000	37,94470	39,95429
(1)X1 (L)	-,66667	,289607	-2,3020	,041886	-1,30409	-,02925
X1 (Q)	2,18182	,785878	2,7763	,018023	,45211	3,91152
(2)X2 (L)	-6,11111	,289607	-21,1014	,000000	-6,74853	-5,47369
X2 (Q)	-1,81818	,785878	-2,3136	,041040	-3,54789	-,08848
(3)X3 (L)	-2,16667	,289607	-7,4814	,000012	-2,80409	-1,52925
X3 (Q)	,68182	,785878	,8676	,404150	-1,04789	2,41152
(4)X4 (L)	-3,27778	,289607	-11,3180	,000000	-3,91520	-2,64036
X4 (Q)	,68182	,785878	,8676	,404150	-1,04789	2,41152
(5)X5 (L)	2,00000	,289607	6,9059	,000026	1,36258	2,63742
X5 (Q)	,18182	,785878	,2314	,821285	-1,54789	1,91152
2L by 3L	,31250	,307175	1,0173	,330847	-,36359	,98859
2L by 4L	,43750	,307175	1,4243	,182113	-,23859	1,11359
2L by 5L	-,31250	,307175	-1,0173	,330847	-,98859	,36359
3L by 4L	,18750	,307175	,6104	,553997	-,48859	,86359
4L by 5L	-,18750	,307175	-,6104	,553997	-,86359	,48859

Рисунок 8.14 – Таблица результатов регрессионного анализа с пересчитанными коэффициентами

2.10. Обратите внимание на изменения коэффициента детерминации. Выведите на экран таблицу основных статистик (ANOVA table) и обратите внимание на изменения в ней.

2.11. Исключите незначимые коэффициенты уравнения регрессии. Результат сравните с рисунком 8.15.

Regr. Coefficients; Var.:T; R-sqr=,97004; Adj:,96105

EXPERIM. DESIGN

5 factors, 1 Blocks, 27 Runs; MS Residual=1,627778
DV: T

Factor	Regressn Coeff.	Std.Err.	t(20)	p	-95,% Cnf.Limt	+95,% Cnf.Limt
Mean/Interc.	38,88889	,425281	91,4427	,000000	38,00177	39,77601
(1)X1 (L)	-,66667	,300719	-2,2169	,038381	-1,29396	-,03938
X1 (Q)	2,00000	,520861	3,8398	,001023	,91350	3,08650
(2)X2 (L)	-6,11111	,300719	-20,3216	,000000	-6,73840	-5,48382
(3)X3 (L)	-2,16667	,300719	-7,2049	,000001	-2,79396	-1,53938
(4)X4 (L)	-3,27778	,300719	-10,8998	,000000	-3,90507	-2,65049
(5)X5 (L)	2,00000	,300719	6,6507	,000002	1,37271	2,62729

Рисунок 8.15 – Таблица результатов регрессионного анализа

2.12. Выведите на экран таблицу основных статистик. Результат сравните с рисунком 8.16.

ANOVA; Var.:T; R-sqr=,97004; Adj:,96105 (plan_kod.sta)

Continue...

5 factors, 1 Blocks, 27 Runs; MS Residual=1,627778
DV: T

Factor	SS	df	MS	F	p
(1)X1 (L)	8,000	1	8,0000	4,9147	,038381
X1 (Q)	24,000	1	24,0000	14,7440	,001023
(2)X2 (L)	672,222	1	672,2222	412,9693	,000000
(3)X3 (L)	84,500	1	84,5000	51,9113	,000001
(4)X4 (L)	193,389	1	193,3889	118,8055	,000000
(5)X5 (L)	72,000	1	72,0000	44,2321	,000002
Error	32,556	20	1,6278		
Total SS	1086,667	26			

Рисунок 8.16 – Основные статистики

Обратите внимание, что в таблицах на рисунках 8.15,8.16 коэффициенты детерминации равны, факторы и их сочетания совпадают и все коэффициенты значимы, поэтому строки выделены красным цветом.

2.13. Выпишите уравнение регрессии в кодированных факторах (8.1):

$$t = 38{,}88889 - 0{,}66667 \cdot X_1 + 2 \cdot X_1^2 - 6{,}11111 \cdot X_2 - 2{,}16667 \cdot X_3 - 3{,}27778 \cdot X_4 + 2 \cdot X_5.$$

2.14. Сделайте вывод о степени влияния факторов на функцию отклика. Согласно уравнению на увеличение времени созревания растений t влияют факторы X_1, X_5, а на уменьшение – X_2, X_3, X_4.

2.15. Закройте окна с результатами.

3. Построение уравнения регрессии зависимости функции отклика от факторов в именованном виде.

3.1. В стартовой панели Design & Analysis of Experiments модуля Experimental Design (см. рисунок 8.1) в списке планов эксперимента выберите план Central composite, non-factorial, surface design и нажмите кнопку OK. На экране появится диалоговое окно Central Composite (Response Surface) Designs (см. рисунок 8.2).

3.2. В окне Central Composite (Response Surface) Designs в списке **Standard Design** выберите **5/1/27** – 5 факторов, 1 блок, 27 опытов. Нажмите кнопку OK. На экране появится окно Designs of a Response Surface Experiment (Central Composite Designs) для настройки параметров плана (см. рисунок 8.3).

3.3. Для получения матрицы планирования эксперимента с именами и уровнями варьирования факторов в окне Designs of a Response Surface Experiment (Central Composite Designs) нажмите кнопку Factor names, values, ets. На экране появится окно с таблицей для ввода имен факторов и их значений. Отредактируйте таблицу в окне согласно данным задачи. Сравните полученную таблицу с рисунком 8.17.

Summary for Variables (Factors) (plan_kod.sta)

Continue... To change labels, values, etc., type in the desired changes, then click Continue.

Factor	Factor Name	Low Value	Low Label	Center Value	Center Label	High Value	High Label	Star Low Label	Star Hi. Label
A (1)	U	50,00000	Low	125,0000	CenterPt	200,0000	High	StarLow	StarHigh
B (2)	Tp	10,00000	Low	17,5000	CenterPt	25,0000	High	StarLow	StarHigh
C (3)	W	,20000	Low	,5000	CenterPt	,8000	High	StarLow	StarHigh
D (4)	L	100,0000	Low	300,0000	CenterPt	500,0000	High	StarLow	StarHigh
E (5)	G	100,0000	Low	200,0000	CenterPt	300,0000	High	StarLow	StarHigh

Рисунок 8.17 – Окно с отредактированной таблицей имен факторов и их уровней варьирования

3.4. Для продолжения работы нажмите кнопку Continue.

3.5. В окне Designs of a Response Surface Experiment (Central Composite Designs) в разделе Show установите переключатель в положение Mini/maxima, в разделе Denote factors установите переключатель в положение by Names, в разделе Design characteristic установите переключатель в позицию Face centered star points (e.g., ±1). Нажмите кнопку Display/print/save design. На экране появится матрица планирования эксперимента в именованном виде (рисунок 8.18).

3.6. Сохраните матрицу планирования эксперимента в именованном виде в файл с именем **Plan_f.sta**.

3.7. Закройте все открытые окна.

3.8. Выполните команду File – Open Data – Plan_f.sta. На экране появится таблица плана эксперимента для факторов в именованном виде. Отформатируйте данные. Добавьте столбец для зависимой переменной *t* и заполните его значениями (рисунок 8.19). Сохраните файл.

2**(5-1) central composite, nc=16 ns=10 n0=1 Runs=27

Continue...	U	Tp	W	L	G
1	50,0000	10,00000	,200000	100,0000	300,0000
2	50,0000	10,00000	,200000	500,0000	100,0000
3	50,0000	10,00000	,800000	100,0000	100,0000
4	50,0000	10,00000	,800000	500,0000	300,0000
5	50,0000	25,00000	,200000	100,0000	100,0000
6	50,0000	25,00000	,200000	500,0000	300,0000
7	50,0000	25,00000	,800000	100,0000	300,0000
8	50,0000	25,00000	,800000	500,0000	100,0000
9	200,0000	10,00000	,200000	100,0000	100,0000
10	200,0000	10,00000	,200000	500,0000	300,0000
11	200,0000	10,00000	,800000	100,0000	300,0000
12	200,0000	10,00000	,800000	500,0000	100,0000
13	200,0000	25,00000	,200000	100,0000	300,0000
14	200,0000	25,00000	,200000	500,0000	100,0000
15	200,0000	25,00000	,800000	100,0000	100,0000
16	200,0000	25,00000	,800000	500,0000	300,0000
17	50,0000	17,50000	,500000	300,0000	200,0000
18	200,0000	17,50000	,500000	300,0000	200,0000
19	125,0000	10,00000	,500000	300,0000	200,0000
20	125,0000	25,00000	,500000	300,0000	200,0000
21	125,0000	17,50000	,200000	300,0000	200,0000
22	125,0000	17,50000	,800000	300,0000	200,0000
23	125,0000	17,50000	,500000	100,0000	200,0000
24	125,0000	17,50000	,500000	500,0000	200,0000
25	125,0000	17,50000	,500000	300,0000	100,0000
26	125,0000	17,50000	,500000	300,0000	300,0000
27 (C)	125,0000	17,50000	,500000	300,0000	200,0000

Рисунок 8.18 – Окно с матрицей планирования эксперимента для факторов в именованном виде

Data: Plan_f.STA 6v * 27c

NUMERIC VALUES	2**(5-1) central composite, nc=16 ns=10 n0=1					
	1 U	2 TP	3 W	4 L	5 G	6 T
1	50	10,0	,2	100	300	57
2	50	10,0	,2	500	100	44
3	50	10,0	,8	100	100	46
4	50	10,0	,8	500	300	44
5	50	25,0	,2	100	100	38
6	50	25,0	,2	500	300	36
7	50	25,0	,8	100	300	38
8	50	25,0	,8	500	100	29
9	200	10,0	,2	100	100	50
10	200	10,0	,2	500	300	47
11	200	10,0	,8	100	300	50
12	200	10,0	,8	500	100	38
13	200	25,0	,2	100	300	41
14	200	25,0	,2	500	100	31
15	200	25,0	,8	100	100	33
16	200	25,0	,8	500	300	31
17	50	17,5	,5	300	200	42
18	200	17,5	,5	300	200	41
19	125	10,0	,5	300	200	43
20	125	25,0	,5	300	200	32
21	125	17,5	,2	300	200	42
22	125	17,5	,8	300	200	38
23	125	17,5	,5	100	200	43
24	125	17,5	,5	500	200	37
25	125	17,5	,5	300	100	39
26	125	17,5	,5	300	300	40
27 (C)	125	17,5	,5	300	200	36

Рисунок 8.19 – Матрица плана типа Bn для пяти факторов в именованном виде с зависимой переменной

3.9. Выполните команду Analysis – Resume Analysis. В появившемся окне Design & Analysis of Experiments (см. рисунок 8.1) в списке планов эксперимента выберите план Central composite, non-factorial, surface design и нажмите кнопку OK.

3.10. На экране появится диалоговое окно Central Composite (Response Surface) Designs (см. рисунок 8.2). Установите переключатель в позицию Analyze results (см. рисунок 8.6).

3.11. Для выбора зависимой и независимых переменных щелкните по кнопке Variables (см. рисунок 8.6). В появившемся окне Select dependent var. and (optional) indep. and blocking var. выберите зависимую и независимые переменные, как показано на рисунке 8.20. Нажмите кнопку OK.

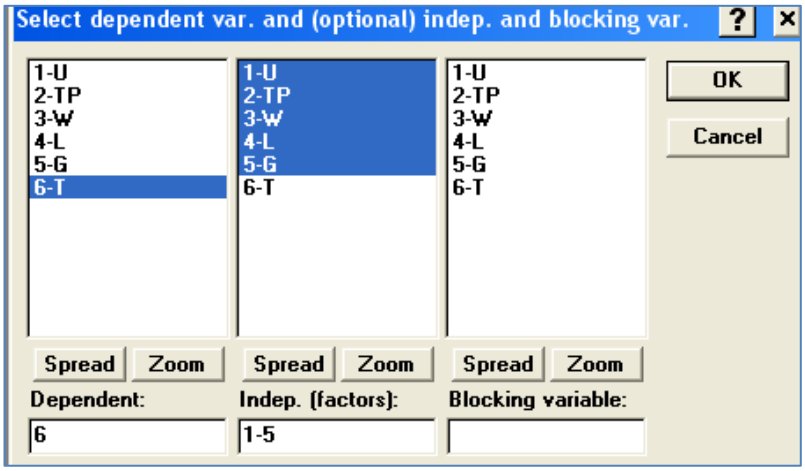


Рисунок 8.20 – Окно для выбора зависимой и независимых переменных

3.12. В окне Analysis of a Central Composite (Response Surface) Experiment нажмите кнопку OK.

3.13. Для расчета коэффициентов уравнения регрессии нажмите кнопку Regression coefficients (см. рисунок 8.8). На экране появится таблица результатов регрессионного анализа (рисунок 8.21).

Regr. Coefficients; Var.:T; R-sqr=,98501; Adj:,93502

Continue...

5 factors, 1 Blocks, 27 Runs; MS Residual=2,715699
DV: T

Factor	Regressn Coeff.	Std.Err.	t(6)	p	-95,% Cnf.Limt	+95,% Cnf.Limt
Mean/Interc.	60,2028	6,00992	10,01723	,000057	45,4970	74,90851
(1)U (L)	-,1063	,05156	-2,06136	,084897	-,2324	,01988
U (Q)	,0004	,00019	2,06999	,083889	-,0001	,00085
(2)TP (L)	,2290	,68442	,33459	,749322	-1,4457	1,90371
TP (Q)	-,0323	,01874	-1,72499	,135288	-,0782	,01353
(3)W (L)	-18,0966	12,88899	-1,40403	,209896	-49,6348	13,44163
W (Q)	7,5758	11,71138	,64687	,541655	-21,0810	36,23248
(4)L (L)	-,0309	,01782	-1,73372	,133665	-,0745	,01271
L (Q)	,0000	,00003	,64687	,541655	-,0000	,00008
(5)G (L)	,0249	,04492	,55469	,599161	-,0850	,13482
G (Q)	,0000	,00011	,17250	,868717	-,0002	,00028
1L by 2L	,0001	,00073	,15170	,884392	-,0017	,00190
1L by 3L	,0028	,01831	,15170	,884392	-,0420	,04758
1L by 4L	-,0000	,00003	-,15170	,884392	-,0001	,00006
1L by 5L	-,0000	,00005	-,15170	,884392	-,0001	,00013
2L by 3L	,1389	,18310	,75852	,476866	-,3092	,58693
2L by 4L	,0003	,00027	1,06193	,329128	-,0004	,00096
2L by 5L	-,0004	,00055	-,75852	,476866	-,0018	,00093
3L by 4L	,0031	,00687	,45511	,665031	-,0137	,01993
3L by 5L	-,0021	,01373	-,15170	,884392	-,0357	,03152
4L by 5L	-,0000	,00002	-,45511	,665031	-,0001	,00004

Рисунок 8.21 – Таблица результатов регрессионного анализа

3.14. Исключите незначимые коэффициенты уравнения регрессии. Вид регрессионного уравнения в именованных факторах будет такой же, как и в кодированных факторах. В результате получите таблицу результатов регрессионного анализа, как показано на рисунке 8.22.

Regr. Coefficients; Var.:T; R-sqr=,97004; Adj:,96105

Continue...

5 factors, 1 Blocks, 27 Runs; MS Residual=1,627778
DV: T

Factor	Regressn Coeff.	Std.Err.	t(20)	p	-95,% Cnf.Limt	+95,% Cnf.Limt
Mean/Interc.	64,34259	1,682068	38,2521	,000000	60,83386	67,85132
(1)U (L)	-,09778	,023494	-4,1618	,000482	-,14679	-,04877
U (Q)	,00036	,000093	3,8398	,001023	,00016	,00055
(2)TP (L)	-,81481	,040096	-20,3216	,000000	-,89845	-,73118
(3)W (L)	-7,22222	1,002398	-7,2049	,000001	-9,31319	-5,13126
(4)L (L)	-,01639	,001504	-10,8998	,000000	-,01953	-,01325
(5)G (L)	,02000	,003007	6,6507	,000002	,01373	,02627

Рисунок 8.22 – Таблица результатов регрессионного анализа

3.15. Выведите на экран таблицу основных статистик. Результат сравните с рисунком 8.23.

ANOVA; Var.:T; R-sqr=,97004; Adj:,96105 (plan_f.sta)

Continue...

5 factors, 1 Blocks, 27 Runs; MS Residual=1,627778
DV: T

Factor	SS	df	MS	F	p
(1)U (L)	8,000	1	8,0000	4,9147	,038381
U (Q)	24,000	1	24,0000	14,7440	,001023
(2)TP (L)	672,222	1	672,2222	412,9693	,000000
(3)W (L)	84,500	1	84,5000	51,9113	,000001
(4)L (L)	193,389	1	193,3889	118,8055	,000000
(5)G (L)	72,000	1	72,0000	44,2321	,000002
Error	32,556	20	1,6278		
Total SS	1086,667	26			

Рисунок 8.23 – Основные статистики

Обратите внимание, что в таблицах на рисунках 8.22, 8.23 коэффициенты детерминации равны, факторы и их сочетания совпадают и все коэффициенты значимы, поэтому строки выделены красным цветом.

3.16. Выпишите уравнение регрессии в именованных факторах (8.2):

$$t = 64{,}34259 - 0{,}09778 \cdot U + 0{,}00036 \cdot U^2 - 0{,}81481 \cdot T_p - 7{,}22222 \cdot W - 0{,}01639 \cdot L + 0{,}02 \cdot G.$$

3.17. Сравните виды уравнений в кодированных и именованных факторах. Виды уравнений в кодированных и именованных факторах совпадают.

4. Оценка значимости коэффициентов уравнения регрессии.

В качестве нулевой гипотезы выдвигается предположение, что коэффициент b_i равен нулю. Для каждого коэффициента уравнения регрессии сравните вычисленный критерий Стьюдента с табличным. Вычисленные значения критерия Стьюдента для коэффициентов уравнения находятся в столбце $t(20)$ (см. рисунок 8.22): для b_0 – $t_{b_0} = 38{,}2521$ и т.д. Теоретическое значение критерия Стьюдента для степени свободы $df_\nu = 20$ и уровня значимости $\alpha = 5\%$ равно $t_{05}(20) = 2{,}09$. Для всех коэффициентов уравнения регрессии выполняется условие $|t_{b_i}| > t_{05}(20)$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: коэффициенты уравнения регрессии отличны от нуля***. Все коэффициенты полученного уравнения значимы.

5. Оценка адекватности и качество регрессионной модели.

5.1. ***В качестве нулевой гипотезы H_0 выдвигается предположение о том, что связь между откликом и факторами отсутствует.*** Вычислите степени свободы для выбора табличного значения функции Фишера.

Число степеней свободы $\mathrm{df_r}$ определяется как разность числа опытов $\mathrm{n} = 27$ и количества значимых коэффициентов уравнения регрессии $\mathrm{d} = 7$, следовательно, $\mathrm{df_r} = 20$.

Число степеней свободы остатков df_o на единицу меньше числа значимых коэффициентов уравнения регрессии $d = 7$, следовательно, $df_o = 6$.

Теоретическая функция $F_{05}(6, 20) = 4{,}35$. Вычисленное значение критерия Фишера $F(6, 20) = 107{,}93$ (значение критерия Фишера вычисляется из таблицы основных статистик – это среднее значений критерия Фишера для каждого коэффициента (см. рисунок 8.23)). Сравниваем критерии Фишера: т.к. опыты проведены без повторений и $\mathrm{F} > \mathrm{F_{05}}$, следовательно, ***нулевая гипотеза отвергается и принимается альтернативная: между откликом и факторами существует зависимость.***

5.2. Оценим качество полученной регрессионной модели. Коэффициент детерминации $R^2 = 0{,}97$, следовательно, 97% значений отклика объяснимы значениями факторов. Вычисленный $R^2 > 0{,}75$, значит, полученная регрессионная модель может использоваться для прогноза.

6. Анализ остатков.

6.1. Выполните анализ остатков (постройте график зависимости остатков регрессионной модели от значений зависимой переменной, вычисленных по регрессионному уравнению; постройте график остатков на нормальной вероятностной бумаге). Для этого в окне Analysis of a Central Composite (Response Surface) Experiment нажмите кнопку Residuals. На экране появится диалоговое окно Residual Statistics (рисунок 8.24).

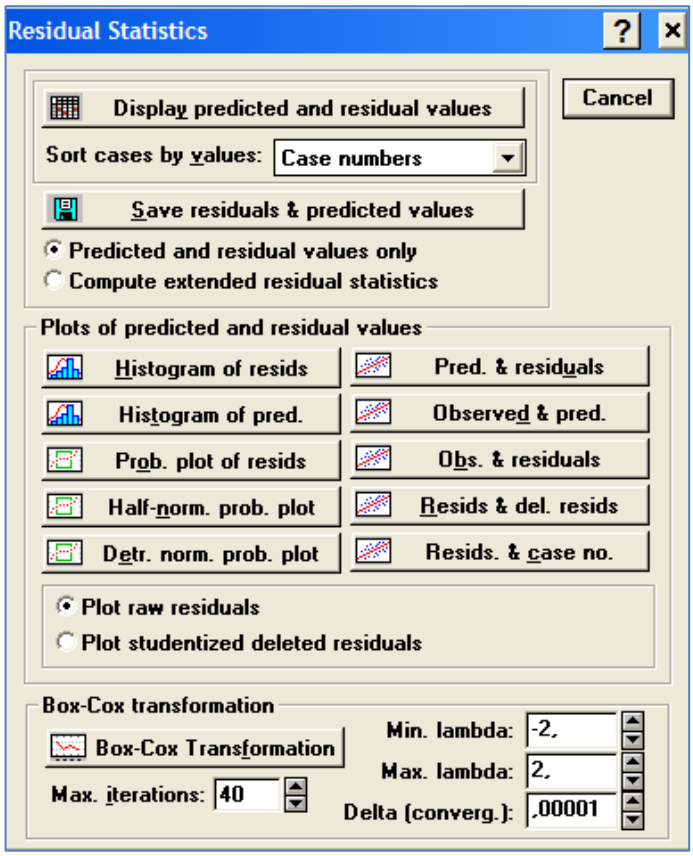


Рисунок 8.24 – Окно для анализа остатков регрессионного уравнения

6.2. В окне Residual Statistics нажмите кнопку Pred. & residuals. В результате на экране появится график зависимости остатков регрессионной модели от значений зависимой переменной (функции отклика), вычисленных по регрессионному уравнению (рисунок 8.25).

6.3. По графику зависимости остатков регрессионной модели от значений зависимой переменной сделайте **вывод**: на рисунке 8.25 точки облака рассеяния разбросаны на координатной плоскости в хаотичном порядке (т.е. не выстраиваются в виде какой-либо линейной или нелинейной зависимости), поэтому связь между остатками регрессионного уравнения и значениями зависимой переменной, вычисленных по регрессионному уравнению, отсутствует. Это говорит о том, что полученная регрессионная модель пригодна для использования.

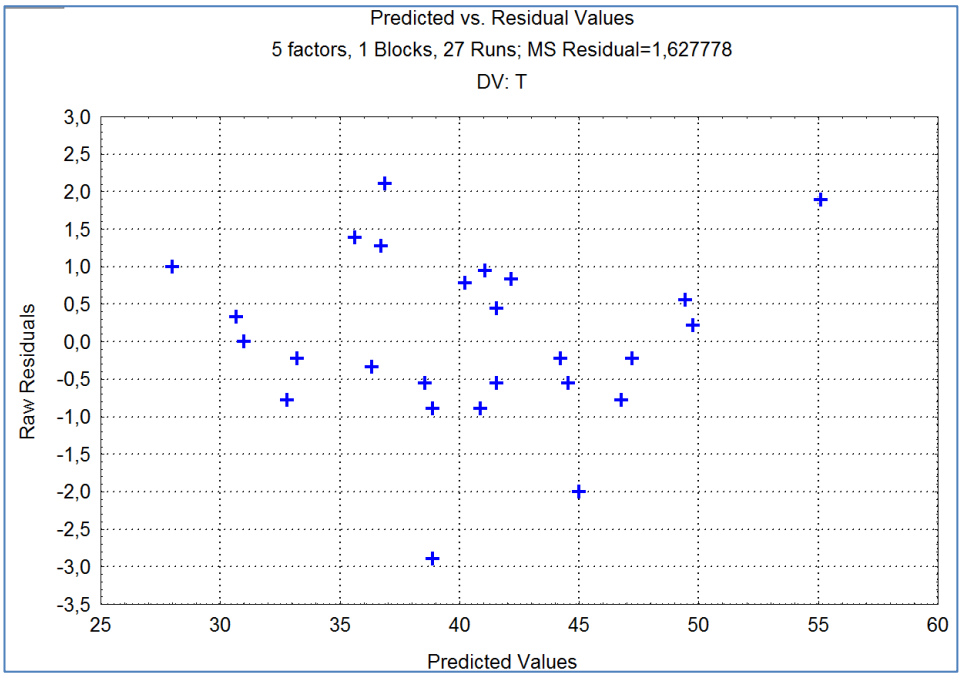


Рисунок 8.25 – График зависимости остатков регрессионной модели от значений зависимой переменной, вычисленных по регрессионному уравнению

О нормальности остатков можно судить по графику остатков на нормальной вероятностной бумаге. Чем ближе распределение к нормальному виду, тем лучше значения остатков укладываются на прямую линию.

6.4. Для построения графика остатков на нормальной вероятностной бумаге нажмите кнопку Prob. plot of resids. На экране появится график остатков (рисунок 8.26).

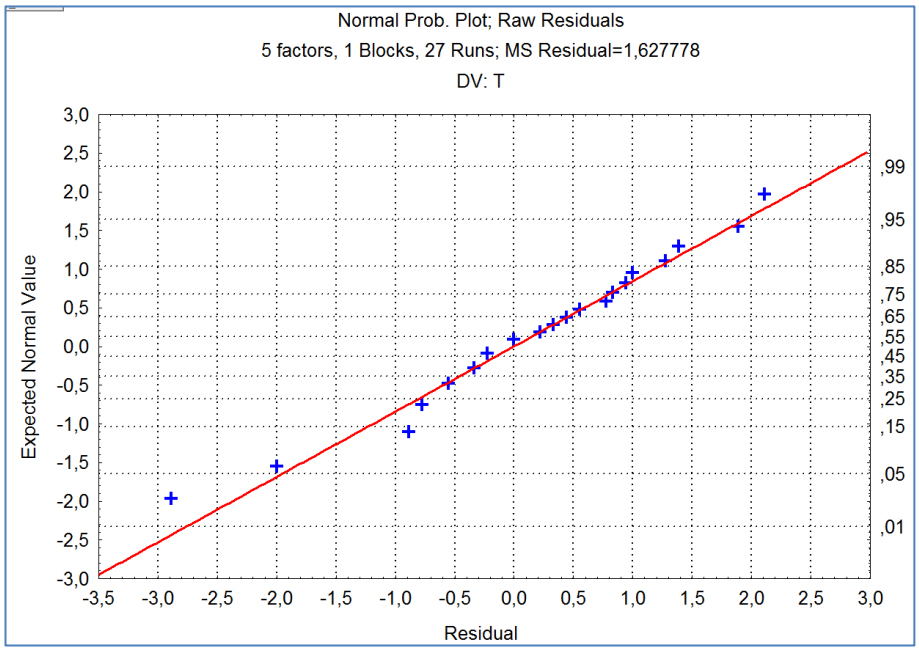


Рисунок 8.26 – График остатков построенный на нормальной вероятностной бумаге

По графику остатков на нормальной вероятностной бумаге (рисунок 8.26) сделаем **вывод**: чем ближе остатки уложены к прямой линии, тем качественнее регрессионная модель отражает исходные данные.

7. Построение поверхности отклика.

7.1. Постройте поверхность отклика. Для этого в окне Analysis of a Central Composite (Response Surface) Experiment (см. рисунок 8.8) нажмите кнопку Surface. На экране появится окно выбора факторов для построения поверхности отклика. Выберите факторы, как показано на рисунке 8.27. Нажмите кнопку OK.

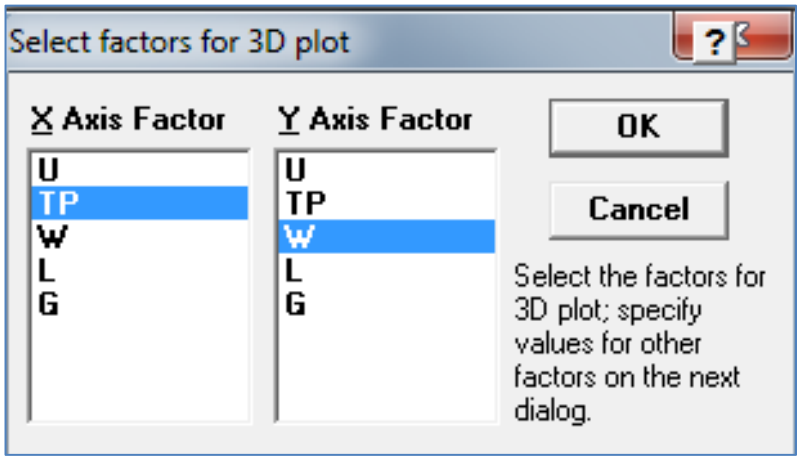


Рисунок 8.27 – Окно выбора факторов для построения поверхности отклика

7.2. На экране появится окно для ввода значений фиксируемых факторов. В регрессионной модели пять факторов, поэтому для построения поверхности, три из них необходимо фиксировать. Зафиксируйте факторы U, L, G сначала на нижнем уровне, затем на среднем и на верхнем уровнях. В появившемся окне сначала укажите нижние уровни варьирования факторов (рисунок 8.28). Нажмите кнопку OK. На экране появится поверхность отклика (рисунок 8.29).

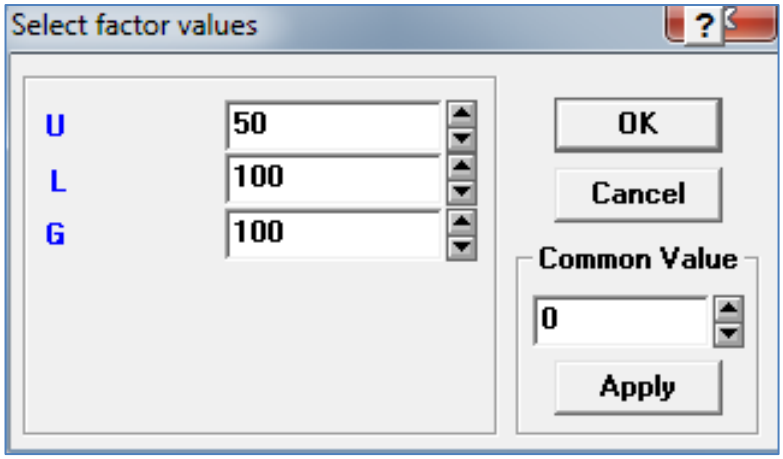


Рисунок 8.28 – Окно для ввода значений фиксируемых факторов

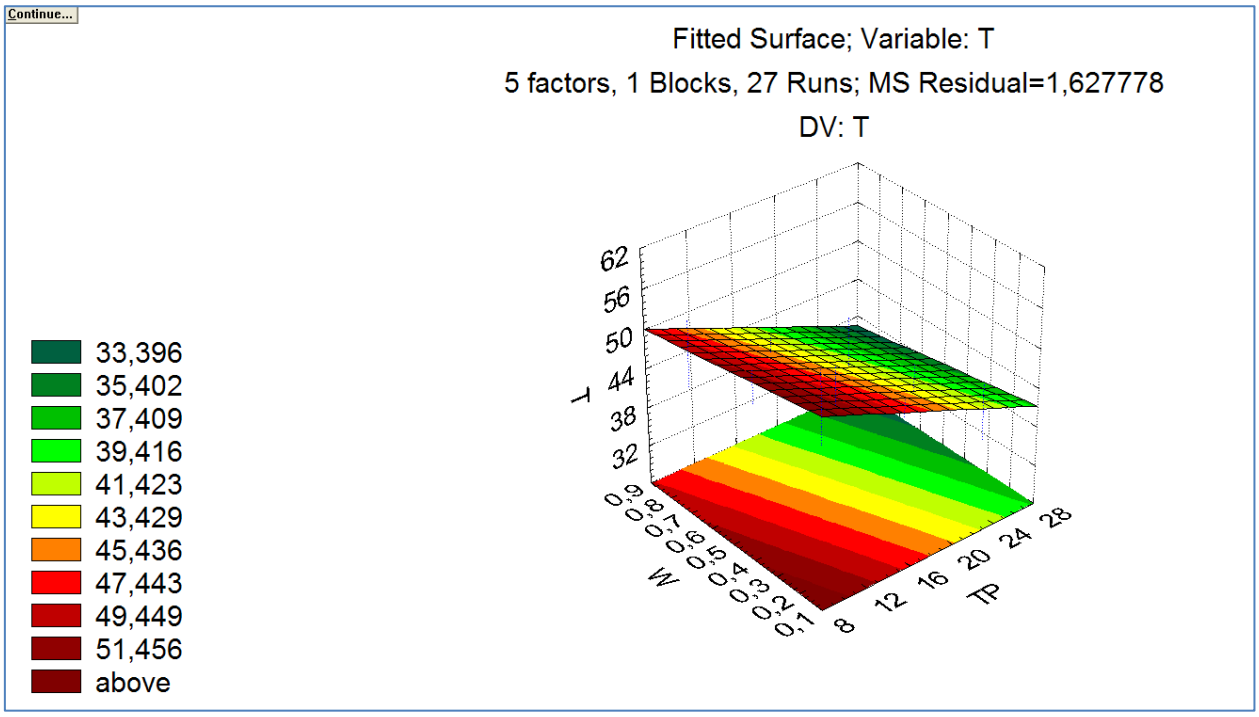


Рисунок 8.29 – Поверхность отклика $\mathbf{t = t(W,\ T_p)}$

7.3. Постройте поверхности отклика, фиксируя переменные на разных уровнях варьирования.

7.4. Постройте поверхности отклика $t = t(W,\ G)$, $t = t(W,\ L)$ и т.д., сделайте выводы.

Замечание. *В процессе анализа поверхностей отклика (или графиков) целевой функции (зависимой переменной, функции отклика), построенных по регрессионным уравнениям, необходимо определить при каких значениях факторов исследуемая целевая функция достигает наилучшего для исследуемого процесса или явления значения.*

7.5. Сохраните данные и закройте программу.

Практическое занятие № 9

КОНТРОЛЬНАЯ РАБОТА № 2

«Прогнозирование изменения состояний объекта исследования. Получение научного результата, имеющего практическое значение»

1. Сгенерируйте ОЦКП для полного трехфакторного эксперимента Факторы именуйте в соответствии с таблицей.

Имя фактора	Нижний уро-	Верхний уровень
X_1	30	60
X_2	15	25
X_3	1,8	2,8

2. Добавьте и заполните столбец зависимой переменной Y.

№ опыта	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Y	112	360	45	300	90	20	180	187	195	105	202	97	45	315	142	150

3. Рассчитайте коэффициенты уравнения регрессии (оставьте таблицу для просмотра преподавателем). Укажите значение коэффициента детерминации модели.

4. Исключите незначимые коэффициенты. Запишите уравнение регрессии. Укажите значение коэффициента детерминации скорректированной модели.

5. Проведите анализ остатков и сравните вычисленные значения функции отклика с наблюдаемыми.

6. Постройте поверхности функции отклика при последовательном фиксировании каждого из факторов в центральной точке.

САМОСТОЯТЕЛЬНАЯ РАБОТА

Практическое занятие № 10

Тема: **НАСТРОЙКА ИНТЕРФЕЙСА ПРОГРАММЫ STATISTICA**

ЦЕЛЬ РАБОТЫ

Изучить:

- элементы интерфейса программы Statistica и их назначение;
- методы настройки элементов интерфейса и восстановления параметров, принятых по умолчанию.

Научиться:

- настраивать интерфейс программы Statistica.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

Для настройки интерфейса программы используют команды Главного меню View (Вид) и Options (Сервис). Меню View (Вид) используется для настройки в текущем сеансе работы или для изменения параметров текущего документа. Меню Options (Сервис) имеет глобальные настройки, которые применяются ко всем последующим сеансам работы и всем открываемым и создаваемым документам.

Все произведенные настройки хранятся в файле **Statist.ini**. Каждое изменение стандартных настроек программы фиксируется в этом файле в текстовом виде. При наличии некоторых навыков можно настраивать программу командами, записываемыми в файле **Statist.ini**. Удаление данного файла ведет к установке настроек по умолчанию. То есть, если удалить файл настройки, то программа будет иметь вид, как при первом запуске после инсталляции (установки). Если Вас не удовлетворяют настройки, произведенные в предыдущем сеансе работы, удалите или переместите в другую папку файл настройки. Можно также сохранять его на дискете для использования личных настроек в последующих сеансах работы.

Рассмотрим более подробно назначение команд меню View (Вид) и Options (Сервис).

2.1 Меню View (Вид)

Назначение команд меню View (Вид) приведено в таблице 10.1.

Шрифты, корректно отображающие текст на кириллице указаны в таблице 10.2.

Таблица 10.1 – Назначение команд меню View (Вид)

Опция меню	Параметр опции	Назначение опции или параметра
1	2	3
Text Values (Текстовые значения)	Задание параметров отображения текстовой переменной	
	Numeric Values (Числовые значения)	Отображение числовых значений текстовых переменных
	Value Labels (Значения меток)	Отображение метки переменной, т.е. комментарий, который может превышать восемь символов (до 128 символов)
Case Names (Имена случаев)	Отображение имен/номеров случаев (строк)	
Full Var Names (Полные имена переменных)	Включение/выключение режима отображения полных имен переменных (столбцов)	
Font (Шрифт)	Открытие диалогового окна для установки параметров шрифта	
	Font (Шрифт)	Определение типа шрифта (Time New Roman, Courier и т.д.). В Statistica не все шрифты правильно отображают начертание русских символов. Рекомендуется использовать шрифты, поддерживающие кириллицу (таблица 3.2)
	Style (Стиль)	Задание начертания шрифта (Обычный, Курсив, Полужирный и Полужирный курсив)
	Size (Размер)	Установка размера (кегля) шрифта
	Save as Default (Сохранить как используемый по умолчанию)	Задание параметров шрифта, используемых по умолчанию при создании новых файлов
Colors (Цвета)	Настройка цветового оформления элементов электронной таблицы. Изменения цветовых настроек отображаются в нижней части окна	
	Text (текст)	Настройка цвета текста
	Background (фон)	Настройка цвета фона

Окончание таблицы 10.1

1	2	3
Colors (Цвета)	Info Box (Информационный блок)	Настройка цвета фона и текста информационного блока
	Titles (Заголовки таблицы)	Настройка цвета фона и текста заголовков таблицы
	Tool Bar (Панель инструментов)	Настройка цвета фона и текста панели инструментов
	Column Names (Названия колонок)	Настройка цвета фона и текста названия колонок
	Row Names (Названия строк)	Настройка цвета фона и текста названий строк
	Data Area (Область данных)	Настройка цвета фона и текста области данных
	Highlight (Выделенный блок)	Настройка цвета фона и текста выделенного блока
	Main (Заголовки)	Выбор цветов линий названий строк и столбцов
	Grid Vert (Сетка вертикальная)	Выбор цветов вертикальных линий сетки
	Grid Horiz (Сетка горизонтальная)	Выбор цветов горизонтальных линий сетки

Таблица 10.2 – Шрифты, правильно отображающие символы кириллицы

Таблицы SpreadSheet и ScrollSheet	Окно отчета	Графики
Courier	Arial Cyr	Fixedsys
Fixedsys	Courier	MS Sans Serif
MS Sans Serif	Courier New Cyr	MS Serif
MS Serif	Fixedsys	System
System	MS Sans Serif	
	MS Serif	
	System	
	Times New Roman Cyr	

2.2 Меню Options (Сервис)

Общее назначение опций меню Options (Сервис) приведено в таблице 1.2 практического занятия №1. Рассмотрим подробнее лишь опции для настройки интерфейса General (Общие) и Display (Отображение). Настройки вида графиков Stat Graphs (Статистические графики) и User Graphs (Пользовательские графики) будут рассмотрены позже.

2.2.1 Команда General (Общие)

Команда General открывает диалоговое окно General (Общие), вид которого представлен на рисунке 10.1. Назначение опций этого окна описано в таблице 10.3.

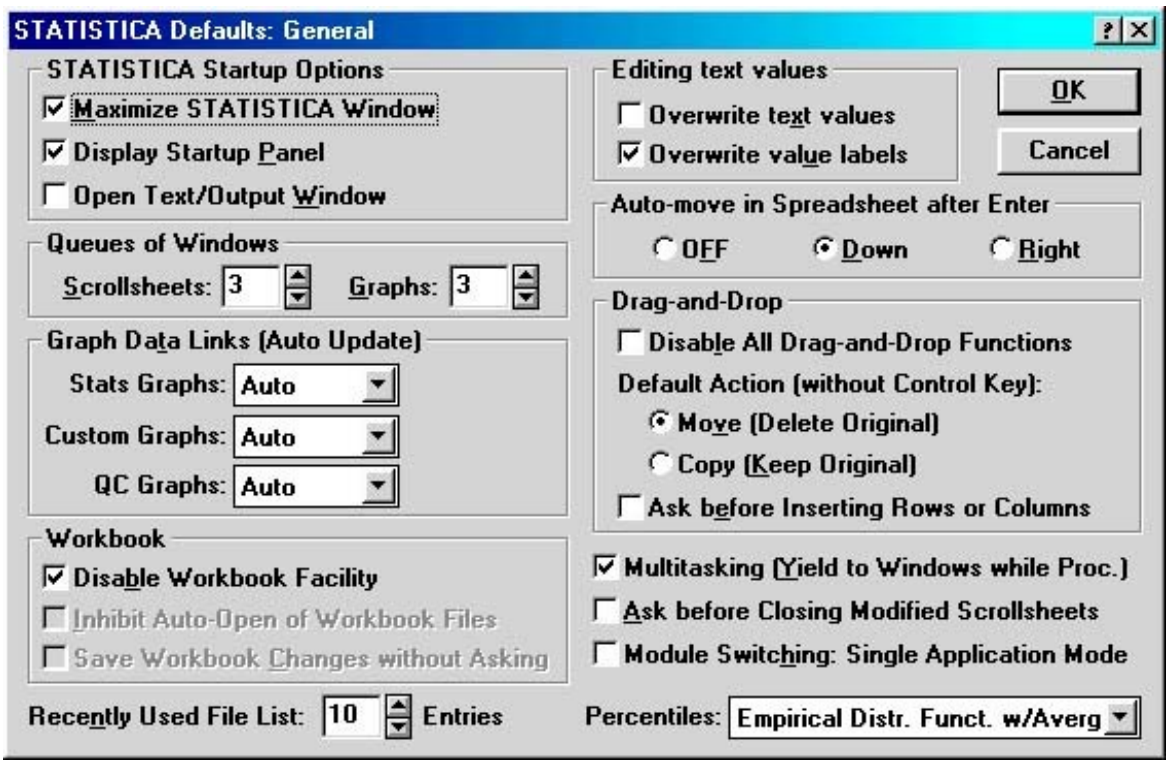


Рисунок 10.1 – Диалоговое окно General (Общие)

Таблица 10.3 – Назначение опций диалогового окна General (Общие)

Группа опций или опция меню	Назначение группы опций или опции
1	2
Statistica Startup Options (Установки запуска)	Группа опций содержит команды, определяющие внешний вид программы при запуске
Maximize Statistica Window (Максимизировать окно Statistica)	Включение/выключение полноэкранного режима отображения окна Statistica

Продолжение таблицы 10.3

1	2
Display Startup Panel (Отображать стартовую панель)	Включение/выключение автоматического отображения стартовой панели при запуске модуля
Open Text/Output Window (Открыть окно отчета)	Управление автоматическим открытием окна отчета при запуске Statistica
Queues of Window (Очередь окон)	Задание количества одновременно открытых окон результатов (ScrollSheets) и графиков (Graphs). Если число окон, открываемых пользователем, превышает указанное число, то окно, открытое первым, автоматически закрывается
Graph Data Links (Связь графика с данными)	Группа опций устанавливает механизм связи графиков с данными рабочей таблицы*
Auto (Автоматическая связь)	Автоматическое изменение графика при редактировании данных исходной таблицы
Manual (Ручное обновление)	Ручное обновление графика при редактировании данных исходной таблицы. Для ручного обновления необходимо выполнить команду Update Graph (Обновить график) из меню File (Файл)
Locked (Блокировать)	Временная отмена механизма связи между графиком и данными
No Link (Нет связи)	Отмена механизма связи между графиком и данными
Workbook (Рабочая книга)	Группа опций управляет способами взаимодействия с рабочими книгами
Disable Workbook Facility (Отменить использование рабочих книг)	Исключение возможности работы с рабочей книгой
Inhibit Auto Open of Workbook Files (Не открывать автоматически файлы рабочей книги)	Выключение/включение функции автоматического открытия файлов рабочей книги (имеется возможность автоматически открывать до 9 файлов книги)
Save Workbook Changes Without Asking (Сохранять изменения в рабочих книгах без уведомления)	Автоматическое сохранение всех файлов рабочей книги при сохранении файла данных
Editing Text Values (Правка текстовых значений)	Группа опций предназначена для управления механизмом исправления текстовых значений

Окончание таблицы 8.3

1	2
Overwrite Text Values (Переписывать текстовые значения)	Изменение всех текстовых значений с соответствующим числовым значением при редактировании одного из значений
Overwrite Values Labels (Переписывать значения меток)	Опция аналогична предыдущей, применительно к меткам
Auto-move in SpreadSheet after Enter (Перемещение в таблице данных после нажатия Enter)	Контроль направления перемещения курсора после нажатия клавиши Enter при редактировании данных. Может иметь три значения: Down (Вниз), Right (Вправо), Off (Выключено)
Drag and Drop (Перетащить и отпустить)	Группа опций предназначена для настройки методов Drag and Drop
Disable All Drag and Drop Functions (Запрет всех функций Drag and Drop)	Запрет всех функций перетаскивания
Default Actions (without Control Key) (Действия по умолчанию (без клавиши Ctrl))	Устанавливает функцию перемещения (Move) / копирования (Copy) при перетаскивании без нажатия на клавишу Ctrl
Ask before Inserting Rows or Columns (Запрашивать перед вставкой строк или столбцов)	Включение запроса на подтверждение при вставке блоков
Recently Used File List (Список открывавшихся файлов)	Задание количества ранее открывавшихся файлов в списке, помещаемом в меню File (Файл)
Multitasking (Многозадачность)	Задание возможности переключения между окнами Windows и выполнения других задач во время анализа данных в Statistica
Ask before Closing Modified ScrollSheets (Запрос перед закрытием измененной таблицы результатов)	Включение запроса на подтверждение перед закрытием измененной таблицы результатов
Module Switching: Single Application Mode (Переключение модулей: режим одного приложения)	Включение режима одного приложения. При этом вызов нового модуля для анализа данных ведет к закрытию текущего модуля

Примечание. Тип связи с данными для конкретного графика можно изменить с помощью команды Graph Data Links (Связь графиков с данными) главного меню Graphs (Графики).

2.2.2 Командаю Display (Экран)

При выборе опции меню Display (Экран) из главного меню Options (Сервис) появляется диалоговое окно Statistica Defaults: Display (Отображение по умолчанию), которое изображено на рисунке 10.2.

Группа опций Display (Отображение на экране) содержит команды, включающие/выключающие отображение следующих элементов: Status Bar (Строка состояния); Full Menus (Полные меню) – сочетания клавиш в меню; Icons in Menus (Значки в меню); Icons in Dialog Boxes (Значки в диалоговых окнах); Tool Bar Button Tool Tips (Назначение кнопок на панели инструментов) – всплывающие подсказки назначения кнопок; Auto Task Buttons (Кнопка автозадач); 3D Dialog Boxes (Объемный вид диалоговых окон).

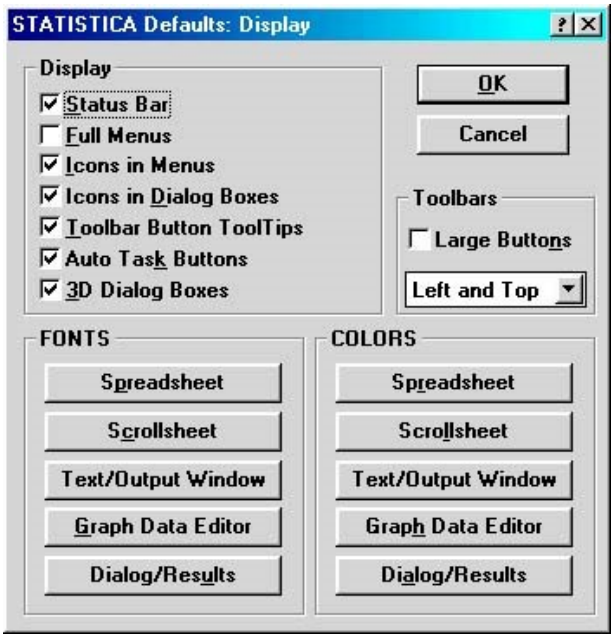


Рисунок 10.2 – Диалоговое окно Statistica Defaults: Display (Отображение по умолчанию)

Группа опций Fonts (Шрифты) и Colors (Цвета) определяет вид шрифта, а также цвет шрифта и фона для таблиц данных (SpreadSheet), таблиц результатов (ScrollSheet), отчетов (Text/Output Window), графиков (Graph Data Editor) и диалогов (Dialog/Results).

Группа опций Tool Bars (Панели инструментов) определяет местоположение и вид значков панелей. В выпадающем меню доступны следующие команды, управляющие положением панелей:

– Left and Top – слева и вверху;

– Right and Top – справа и вверху;
– 1 Lines – в одну строку вверху;
– Lines – в две строки вверху.

Опция Large Buttons (Большие кнопки) служит для увеличения размера кнопок панели.

2.3 Настройка Переключателя модулей (Module Switcher)

Для настройки списка модулей, отображаемого в левой части окна Переключателя модулей (Module Switcher), служит кнопка Customize List... (настройка списка) (рисунок 10.3), которая вызывает окно настройки списка (рисунок 10.4).

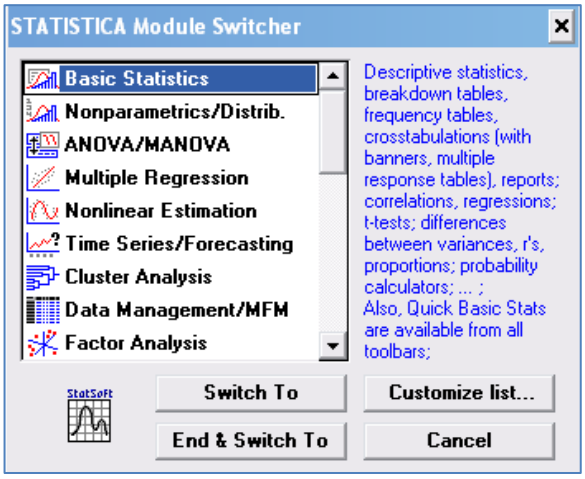


Рисунок 10.3 – Окно переключателей модуля (Module Switcher)

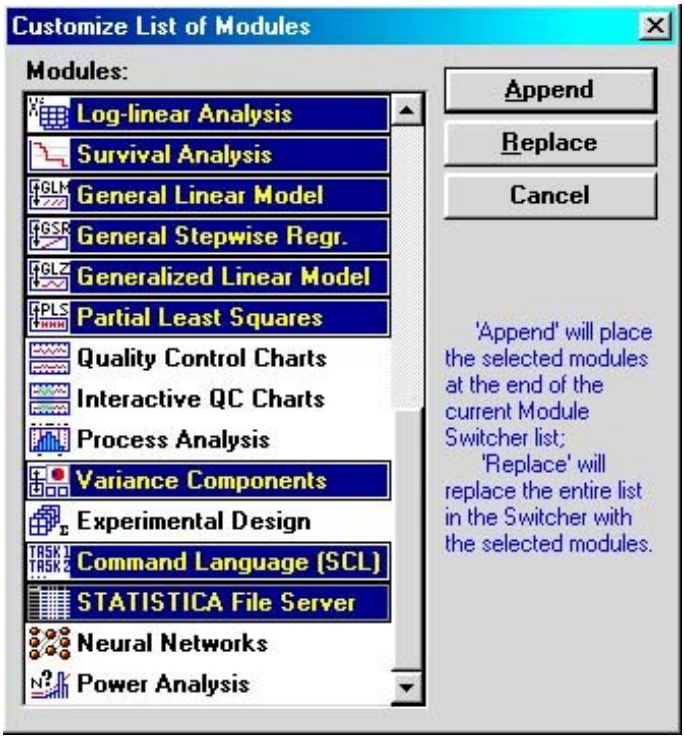


Рисунок 10.4 – Окно настройки списка модулей (Customize List of Modules)

Для настройки списка модулей предназначены следующие кнопки:

– Append (Добавить) – выделенные модули добавляются в конец списка переключателя модулей. С помощью этой кнопки удобно менять порядок следования модулей в списке. Если выделить уже помещенный в список модуль и нажать кнопку Append, то он переместится в нижнюю часть списка.

– Replace (Заменить) – происходит замена списка в переключателе модулей на список модулей, выделенных в диалоговом окне.

Для выделения можно использовать левую кнопку мыши совместно с клавишами **Shift** или **Ctrl**.

Примечание: При первом запуске Переключателя модулей после удаления файла настройки в его окне появляется список всех модулей, находящихся в папке **Stat**.

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

1. Откройте файл **MyAdstudy.sta**.
2. Отключите название случаев командой Case Names (Имена случаев) из меню View (Вид) с помощью команды View – Case Names.
3. Включите режим отображения полных имен случаев (Full Var Names) в меню View (Вид) с помощью команды View – Full Var Names. Обратите внимание на изменения в отображении переменных **Measur**.
4. При помощи меню View (Вид) установите шрифт **MS Sans Serif** размером **9** пунктов. Выберите его в качестве используемого по умолчанию. Для этого выполните команду View – Font... В появившемся диалоговом окне выполните настройки как показано на рисунке 10.5.

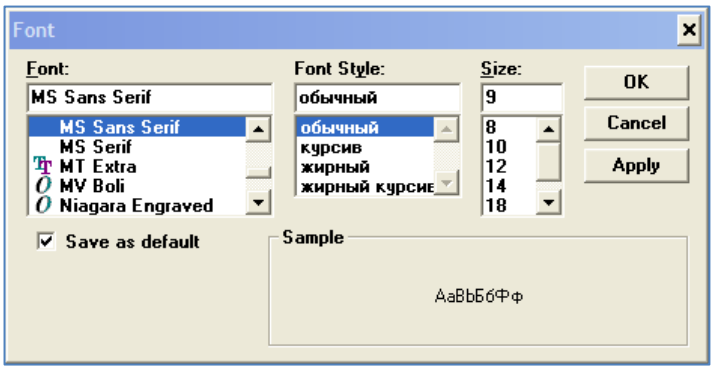


Рисунок 10.5 – Диалоговое окно настройки шрифта

5. Закройте файл **MyAdstudy.sta**, сохранив изменения.
6. Создайте новый файл. Обратите внимание на параметры шрифта и состояние опции Full Var Names.
7. Откройте файл **MyAdstudy.sta**.
8. При помощи команды General (Общие) меню Options (Сервис):

– задайте автоматическое открытие окна отчета при запуске программы;

– установите возможность держать открытыми одновременно 5 таблиц с результатами и 5 графиков;

– запретите использование Рабочих книг;

– задайте направление перемещения курсора вниз при нажатии клавиши **Enter**;

– ограничьте список ранее открывавшихся файлов пятью позициями.

Результат выполнения сравните с рисунком 10.6.

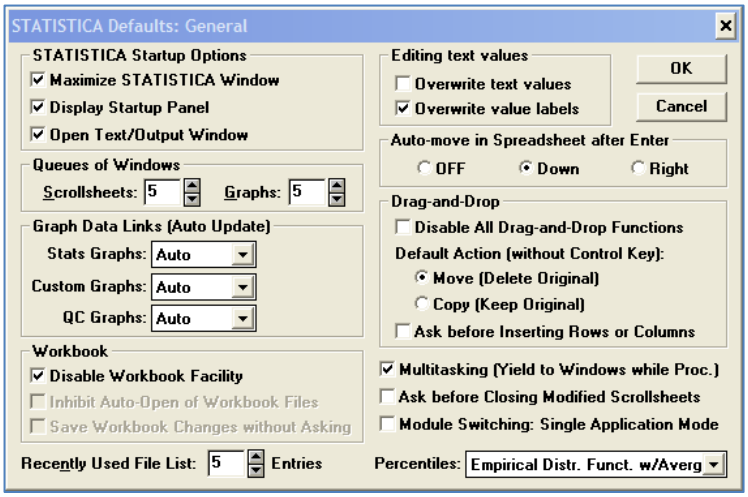


Рисунок 10.6 – Диалоговое окно настройки General

9. При помощи команды Display (Отображение) меню Options (Сервис):

– включите подсказки сочетаний клавиш в **Главном меню**;

– выключите отображение кнопки автозадач;

– разместите панели инструментов в две строки.

Результат выполнения сравните с рисунком 10.7.

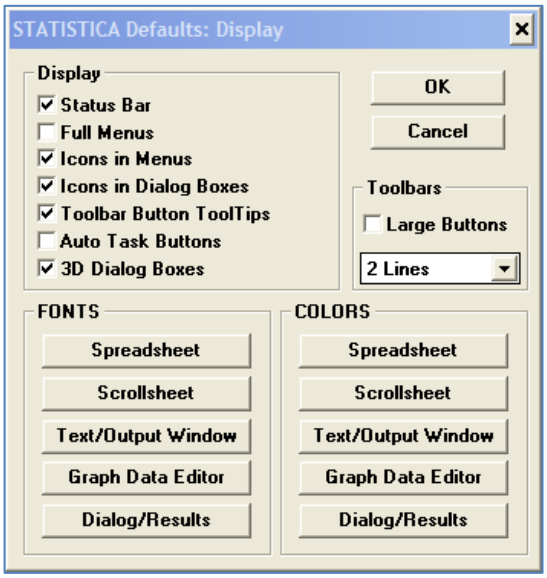


Рисунок 10.7 – Диалоговое окно настройки Display

10. В окне Display (Отображение) просмотрите параметры шрифта (Fonts) таблиц данных (SpreadSheet). Они должны совпадать с параметрами, установленными в пункте 4 задания.

11. Проверьте работоспособность выбранных настроек.

12. Сохраните файл **MyAdstudy.sta**. Завершите работу с программой Statistica.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. Откройте файл **MyAdstudy.sta**.
2. Выберите следующее цветовое оформление элементов таблицы.

Элемент	Цвет	
	текста	фона
Информационный блок	фиолетовый	зеленый
Заголовки	фиолетовый	белый
Названия колонок	белый	синий
Область данных	черный	светло-серый
Вертикальные линии	синий	
Горизонтальные линии	синий	

3. Настройте список переключателя модулей следующим образом:
   - Basic Statistics;
   - Nonparametrics/Distributions;
   - ANOVA/MANOVA;
   - Multiple Regression;
   - Experimental Design.
4. Скопируйте файл **Statist.ini** в личную папку. Используйте этот файл для настройки Statistica на последующих занятиях.
5. Сохраните файл **MyAdstudy.sta**. Завершите работу с программой Statistica.

5 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Назначение меню View и Options. В чем их отличие?
2. Какие параметры можно настроить при помощи главного меню View?
3. Какие параметры можно настроить при помощи главного меню Options?
4. Какие шрифты правильно отображают символы кириллицы во всех типах документов?
5. Назначение файла Statist.ini.
6. Как включить режим отображения полных имен случаев?
7. Как установить параметры шрифта в текущем документе? Как установить параметры шрифта для создаваемых в последующем документов?
8. Как задать размер списка ранее открывавшихся файлов?
9. Как задать количество открытых файлов результатов и графиков?

10. Как настроить перемещение курсора при нажатии клавиши Enter?
11. Как открыть окно отчета? Как установить автоматическое открытие окна отчета при запуске Statistica?
12. Как отключить отображение кнопки автозадач?
13. Как разместить панели инструментов в две строки?
14. Как настроить список Переключателя Модулей?
15. Как удалить все настройки пользователя и восстановить параметры, используемые по умолчанию?

Практическое занятие №11

Тема: **ПЕРВИЧНАЯ ОБРАБОТКА МАССИВА ДАННЫХ**

ЦЕЛЬ РАБОТЫ

Изучить:

– изучить основные статистические показатели, характеризующие массив данных.

Научиться:

– оценивать характеристики массивов опытных данных в программе Statistica.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Основные понятия теории вероятностей

Всякое исследуемое в практике ученого явление связано с анализом множества исходных объектов, которые образуют некоторую совокупность, называемую *генеральной совокупностью*. Например, в полевых опытах общие свойства некоторого варианта обусловлены качествами каждого отдельного растения. Причем, значения параметров отдельных объектов отличаются. Это связано с различием исходных биологических качеств, относительной неравномерностью питательных веществ в почве и многими другими факторами, носящими случайный характер. Поэтому можно сказать, что исследуемые параметры носят так же случайный характер, то есть являются случайными величинами.

Случайной величиной X называется величина, которая в результате опыта может принять то или иное возможное значение, неизвестное заранее, но обязательно одно.

Различают *дискретные* и *непрерывные случайные величины*.

Дискретная случайная величина может принимать отдельные значения, которые можно заранее перечислить. Например, количество стеблей, количество зерен в колосе.

Непрерывная случайная величина может принимать любое значение из некоторого интервала. Например, урожайность, масса 1000 зерен, длина растения.

Появление тех или иных значений дискретной случайной величины можно рассматривать как события, наступлению которых соответствует определенная вероятность.

Получить полное представление о случайной величине можно, установив соответствие между возможными значениями случайной величины и вероятностями их появления.

Всякое соотношение, устанавливающее связь между возможными значениями случайной величины и соответствующими вероятностями, называется *законом распределения дискретной случайной величины*.

Закон распределения случайной величины можно задать при помощи таблицы (таблица 11.1), графически (рисунок 11.1) и аналитически.

Таблица 11.1 – Табличный способ задания закона распределения случайной величины

X	x_1	x_2	x_3	…	x_i	…	x_n
P	p_1	p_2	p_3	…	p_i	…	p_n

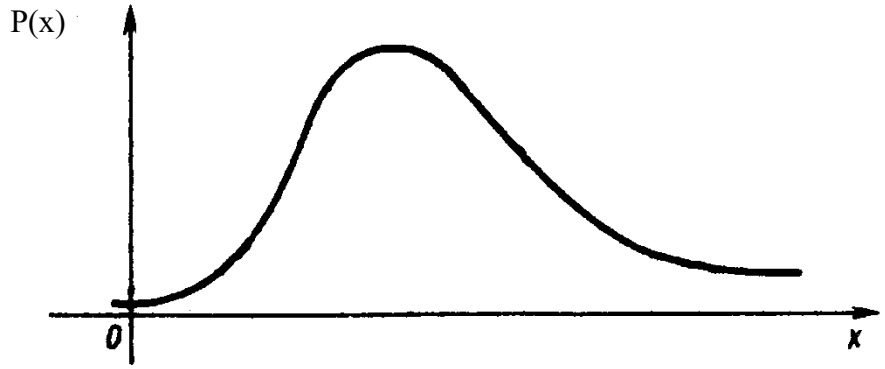


Рисунок 11.1 – Графическое задание закона распределения случайной величины

При задании закона распределения непрерывной случайной величины графический и табличный способы неприемлемы из-за множества значений x_i, заполняющих некоторый промежуток.

Для характеристики поведения непрерывной случайной величины целесообразно использовать не вероятность события $X = x$, а вероятность события $X < x$, где x – некоторое действительное число. В этом случае вероятность $P(X < x)$ является функцией аргумента x. Эту функцию обозначают $F(x)$.

Функцией распределения случайной величины X называется функция $F(x)$, задающая вероятность того, что случайная величина X принимает значение, меньше x, т.е.

$$F(x) = P(X < x). \qquad (11.1)$$

Функцию $F(x)$ еще называют интегральной функцией распределения, ее график приведен на рисунке 11.2.

Функция распределения случайной величины F(x) обладает следующими свойствами:

1. Функция распределения $F(x)$ является неубывающей функцией.

2. Функция распределения $F(x)$ имеет значения на интервале $0 \leq F(x) \leq 1$.

3. Функция распределения $F(x)$ имеет пределы $\lim_{x\to-\infty} F(x) = 0$ и $\lim_{x\to\infty} F(x) = 1$.

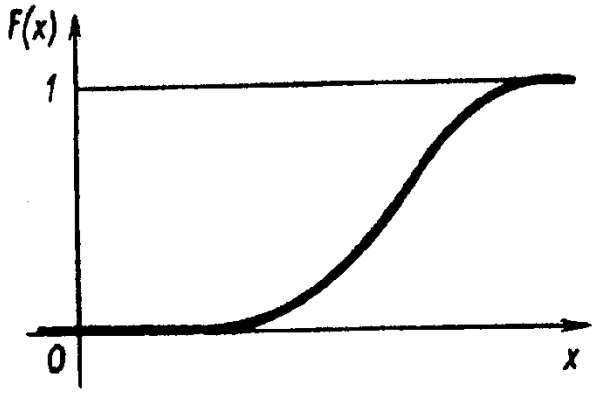


Рисунок 11.2 – График функции распределения
непрерывной случайной величины F(x)

Используя функцию $F(x)$, трудно судить о характере распределения случайной величины в небольшой окрестности той или иной точки числовой оси. Поэтому, часто используют другую функцию, которая называется *дифференциальной функцией распределения* или *функцией плотности распределения вероятности*.

Дифференциальной функцией распределения или *функцией плотности распределения вероятности* $f(x)$ называется первая производная интегральной функции распределения $F(x)$.

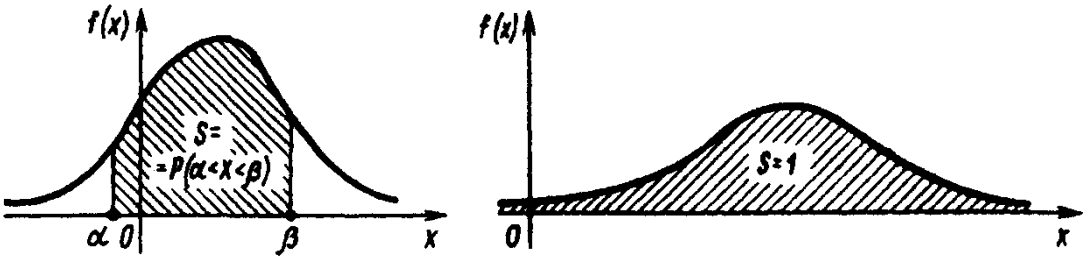
$$f(x) = F'(x). \tag{11.2}$$

Функции плотности распределения вероятности $f(x)$ присущи следующие свойства:

1. Для любых x функция $f(x)$ неотрицательна, т.е. $f(x) \geq 0$.

2. Для $f(x)$ имеет место равенство $P(\alpha < X < \beta) = \int_{\alpha}^{\beta} f(x)dx$. Таким образом, вероятность $P(\alpha < X < \beta)$ численно равна площади криволинейной трапеции, ограниченной графиком функции $f(x)$, осью Ox и прямыми $x = \alpha$ и $x = \beta$ (рисунок 11.3 а).

3. Для функции $f(x)$ имеет место равенство $\int_{-\infty}^{+\infty} f(x)dx = 1$. Т.е. площадь, ограниченная кривой распределения $f(x)$ и осью Ox, равна единице (рисунок 11.3 б).



а) б)

Рисунок 11.3 – Свойства функции плотности распределения вероятности f(x)

Наиболее важными числовыми характеристиками случайной величины являются *математическое ожидание* и *дисперсия*.

Математическим ожиданием MX дискретной случайной величины X называется сумма произведений всех возможных ее значений на соответствующие вероятности появления этих значений, т.е.

$$MX = \sum_{i=1}^{n} x_i p_i \tag{11.3}$$

Математическим ожиданием MX непрерывной случайной величины X, возможные значения которой принадлежат отрезку $[a, b]$, называется определенный интеграл $\int_a^b f(x)dx$, т.е.

$$MX = \int_a^b f(x)dx \tag{11.4}$$

Для оценки рассеивания значений случайной величины вокруг ее математического ожидания используется такая числовая характеристика как *дисперсия*.

Дисперсией DX случайной величины X называют математическое ожидание квадрата отклонения случайной величины X от ее математического ожидания MX, т.е.

$$DX = M(X - MX)^2. \tag{11.5}$$

2.2 Группировка выборочных данных

При проведении опытов исследователь чаще всего не имеет возможности анализировать свойства всех объектов, составляющих некоторую совокупность. Совокупность всех анализируемых однородных объектов называют *генеральной совокупностью*. Поэтому, приходится из генеральной совокупности объектов для обследования отбирать только часть, т.е. проводить

выборочное обследование. При этом совокупность отобранных объектов называют *выборочной совокупностью* или *выборкой*. Причем, для того, чтобы судить о генеральной совокупности выборка должна быть случайной, т.е. все образцы должны иметь одинаковую вероятность попасть в выборку. Если последнее условие выполняется, то выборку называют *репрезентативной*.

Выборочной совокупностью или *выборкой объема n* из совокупности, распределенной по закону $F(x)$, называется n независимых наблюдений с функцией распределения $F(x)$.

Анализ выборочных данных начинают с упорядочения (ранжирования) и группировки. Диапазон значений выборки разбивают на равные интервалы.

Количество интервалов k определяют из выражения

$$k \approx \sqrt{n}. \tag{11.6}$$

и принимают целочисленным.

Длину каждого интервала Δ получают, округляя в сторону увеличения число Δ'.

$$\Delta' = \frac{x_{max} - x_{min}}{k}. \tag{11.7}$$

Для каждого интервала подсчитывают *частоту* m_i – количество наблюдений, попавших в этот интервал.

Частоту, отнесенную к общему числу наблюдений, называют *относительной частотой* или *частостью*

$$p_i = \frac{m_i}{n}. \tag{11.8}$$

Ряд ранжированных значений признака, в котором указана частота или частость их встречаемости в данной совокупности называется *вариационным рядом*.

По сути, вариационный ряд является табличным способом задания закона распределения выборочных данных. Закон распределения, полученный по выборочным данным, называется *эмпирическим законом распределения*.

Выборочной функцией распределения (или функцией распределения выборки) называется функция $F(x)$, задающая для каждого значения x относительную частоту события $X < x$. То есть

$$F(x) = \frac{m_x}{n}. \tag{11.9}$$

где n – объем выборки;

m_x – число выборочных значений $X < x$.

В отличие от выборочной функции распределения интегральную функцию $F(x)$ генеральной совокупности называют *теоретической функцией распределения*.

Графическое представление вариационного ряда в виде столбчатой диаграммы, в которой высоты прямоугольников соответствуют частотам интервалов, называют *гистограммой* (рисунок 11.4).

Если середины соседних столбцов гистограммы соединить отрезками, то получим ломаную линию, которая называется *полигоном* (рисунок 11.4).

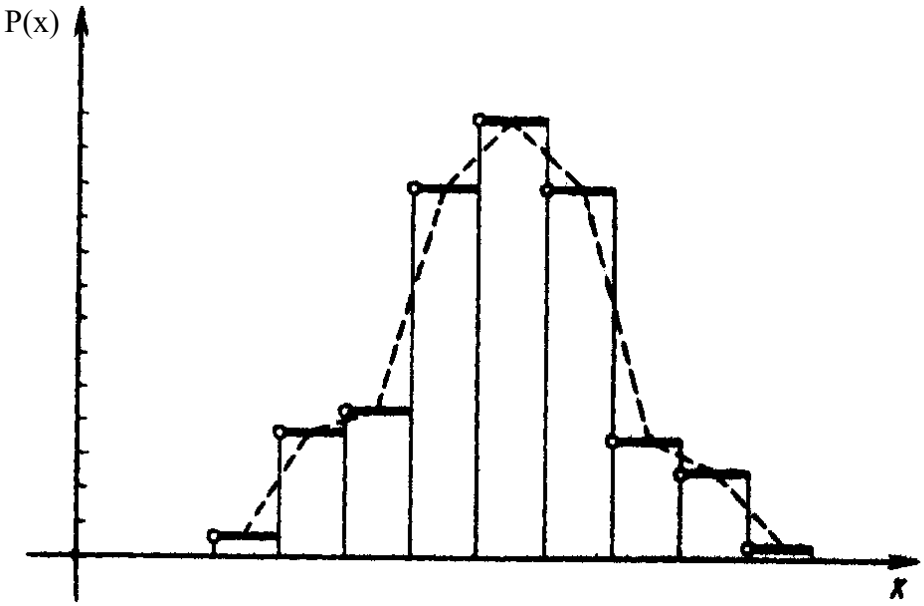


Рисунок 11.4 – Графическое представление вариационного ряда в виде гистограммы и полигона

Сумму частоты (или *частости*) данного интервала с частотами (или *частостями*) всех интервалов, расположенных левее называют *накопленной частотой* (*частостью*).

График накопленных частот приведен на рисунке 11.5. Графическое изображение накопленных частот в виде ломанной или плавной линии называют *кумулятивной кривой* (или *кумулятой*) (рисунок 11.6).

Кумулятивная кривая, построенная с использованием частостей, в сущности, является графическим представлением эмпирической функции распределения вероятности $F(x)$.

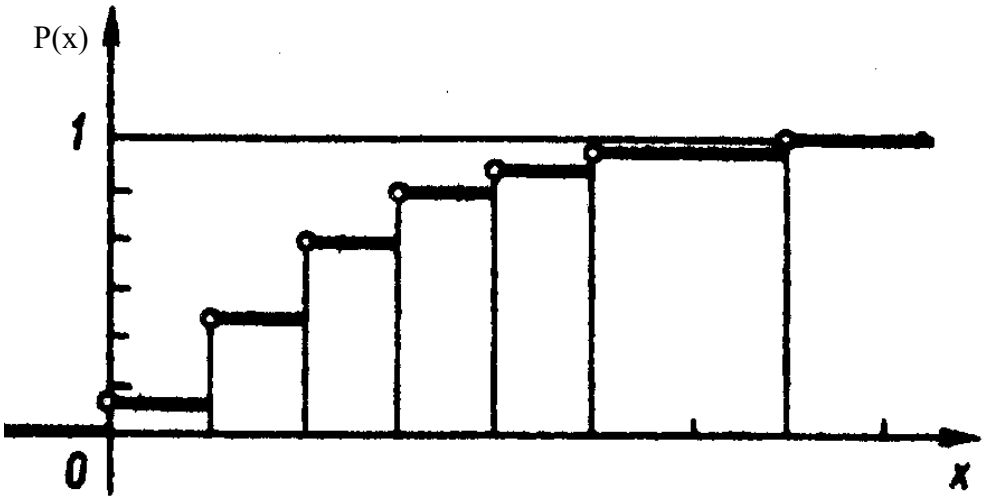


Рисунок 11.5 – График накопленных частот

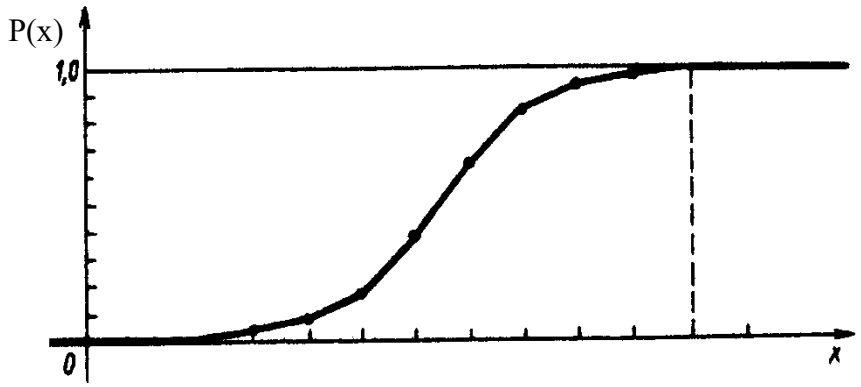


Рисунок 11.6 – Графическое представление группированных данных в виде кумулятивной кривой (кумуляты)

2.3 Числовые характеристики выборочных данных

Выборку описывают следующими числовыми характеристиками:

1. *Объем выборки* n – количество наблюдений представленных в выборке.

2. *Среднее значение* $\bar{X}$ для выборочной совокупности X вычисляется по формуле

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i \tag{11.10}$$

3. *Выборочная дисперсия* S^2 – это средний квадрат отклонений вариантов от их средней величины в данной совокупности.

$$S^2 = \frac{1}{n}\sum_{i=1}^{n} (X_i - \bar{X})^2 \tag{11.11}$$

Выборочная дисперсия S^2 отличается от дисперсии генеральной совокупности D тем, что вместо математических ожиданий M и MX (11.5) используются средние значения величин $\frac{1}{n}\sum_{i=1}^{n} X_i$ и $\bar{X}$ (11.11). Однако, как отмечалось ранее, при увеличении объема выборки среднее значение случайной величины стремиться к ее математическому ожиданию. Поэтому, значение выборочной дисперсии S^2 является приближенным значением дисперсии генеральной совокупности D. И чем больше объем выборки, тем ближе значение выборочной дисперсии S^2 значению дисперсии генеральной совокупности D.

4. *Дисперсия* является мерой изменчивости, вариации признака. Это один из наиболее важных показателей, характеризующих явление.

Для определения выборочной дисперсии чаще используется формула

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad (11.12)$$

В этом случае оценка дисперсии оказывается несмещенной.

5. *Стандартное отклонение* (или *среднее квадратическое отклонение*) S – это квадратный корень из дисперсии.

$$S = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \qquad (11.13)$$

6. *Лимиты* – значения максимальной X_{max} и минимальной X_{min} вариант.

7. *Размах вариации* R – разность максимальной и минимальной вариант данной совокупности

$$R = X_{max} - X_{min}. \qquad (11.14)$$

8. *Выборочной квантилью* называется решение уравнения $F(x) = p$, где p – заданная вероятность. *Квантиль* – это значение переменной, отделяющей заданную часть ранжированного ряда данных. Чаще всего ранжированный ряд делят на две, четыре, десять и сто частей. При этом квантили называют соответственно *медианой*, *квартилями*, *децилями* и *процентилями*.

9. *Медиана* Me – средняя, относительно которой ряд делится на две половины: в обе стороны от медианы располагается одинаковое число членов ряда. Медиана является решением уравнения $F(x) = p$ при $p = 0{,}5$.

10. *Квартиль* – величина, отсекающая $^1/_4$ (25%) членов ряда. Различают *нижнюю* и *верхнюю квартили*. *Нижняя квартиль* – это такое значение, ниже

которого находится 25% случаев выборки ($p = 0{,}25$). *Верхняя квартиль* – это значение выше которого находятся 25% случаев ($p = 0{,}75$).

11. *Интерквартильная широта* – расстояние между верхним и нижним квартилями.

12. *Выборочный коэффициент асимметрии* служит для характеристики смещения вариационного ряда относительно среднего значения по величине и направлению. Если полигон асимметричен, то одна из его ветвей, начиная с вершины, имеет более пологий «спуск», чем другая. В случае отрицательного коэффициента асимметрии более пологий «спуск» полигон наблюдается слева. В противном случае – справа. В первом случае асимметрию называют левосторонней, а во втором – правосторонней. Асимметрия менее 0,5 считается малой.

13. *Выборочный эксцесс* служит для сравнения на «крутость» выборочного распределения с нормальным распределением. Эксцесс нормального распределения равен нулю. Если выборочному распределению соответствует отрицательный эксцесс, то полигон имеет более пологую вершину по сравнению с нормальной кривой. В случае положительного эксцесса – более крутой. Отрицательным пределом величины эксцесса является число –2, положительный предел – не существует.

Часто используются такие относительные показатели как коэффициент вариации и относительная ошибка среднего значения. Программа Statistica не рассчитывает данные показатели, однако они легко могут быть получены из абсолютных показателей.

14. *Коэффициент вариации* V – это отношение стандартного отклонения S к среднему значению $\bar{X}$, выраженное в процентах

$$V = \frac{S}{\bar{X}} \cdot 100\%. \tag{11.15}$$

15. *Стандартная ошибка среднего* $S_{\bar{X}}$ – величина, характеризующая стандартное отклонение выборочного среднего, рассчитанная по выборке n из генеральной совокупности

$$S_{\bar{X}} = \sqrt{\frac{S^2}{n}} = \frac{S}{\sqrt{n}} \tag{11.16}$$

16. *Относительная ошибка среднего* $S_{\bar{X}}\%$ вычисляется как отношение стандартной ошибки среднего к среднему значению, выраженное в процентах

$$S_{\bar{X}}\% = \frac{S_{\bar{X}}}{\bar{X}}. \tag{11.17}$$

3 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

3.1 Вычисление выборочных статистик в модуле Basic Statistics and Tables

Вычисление выборочных статистик производится в модуле Basic Statistics and Tables с помощью команды Descriptive Statistics (Описательные статистики). Эту команду можно выбрать в меню Analysis (Анализ) или на стартовой панели. При выборе команды Descriptive Statistics (Описательные статистики) появляется одноименное окно (рисунок 11.7).

В верхней части окна расположены опции для расчета основных статистик. Нижняя часть занята кнопками для построения графиков, отображающих рассчитанные статистики.

Назначение опций окна Descriptive Statistics (Описательные статистики) представлено в таблице 11.2. Назначение графиков подробно рассмотрено в следующей работе.

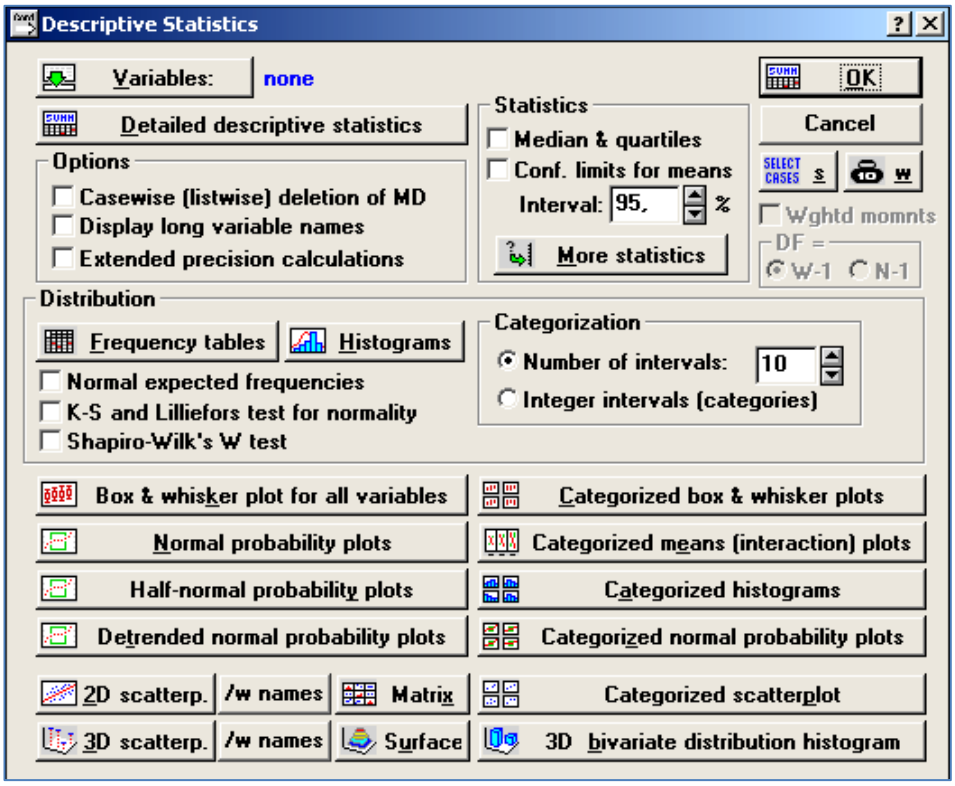


Рисунок 11.7 – Окно для вычисления выборочных статистик Descriptive Statistics (Описательные статистики)

Таблица 11.2 – Назначение опций диалогового окна General (Общие)

Элемент окна	Назначение элемента
1	2
Variables (Переменные)	Задание переменных, статистики которых вычисляются
Detailed Descriptive Statistics (Детальные описательные статистики)	Отображение всех рассчитанных описательных статистик
Options (Параметры)	Группа опций для задания параметров расчетов
Casewise (listwise) Deletion of MD (Удаление случаев с ПД)	Исключение из расчетов случаев с пропущенными данными (ПД)
Display long variable names (Отображать длинные имена переменных)	Отображение длинных имен переменных в таблицах результатов ScrollSheet
Extended precision calculations (Повышенная точность вычислений)	Расчет описательных статистик с повышенной точностью
Distribution (Распределение)	Группа опций для задания параметров подгонки распределений*
Frequency tables (Таблицы частот)	Группировка данных в виде таблицы абсолютных и относительных вариационных рядов и рядов накопленных частот
Histograms (Гистограммы)	Отображение вариационных рядов в виде гистограмм
Categorization (Категоризация)	Группа опций для задания параметров разбиения выборки на интервалы
Number of intervals (Число интервалов)	Задание числа интервалов
Integer intervals (categories) (Целые интервалы (категории))	Задание целочисленных границ интервалов
Normal expected frequencies (Проверка нормальности частот)	Проверка нормальности распределения частот путем наложения на график кривой теоретического нормального распределения с вычисленными по выборке параметрами
K-S and Liliefors test for normality (Тест на нормальность Колмогорова-Смирнова и Лилиефорса)	Отображение на графике результатов теста на нормальность Колмогорова-Смирнова и Лилиефорса*
Shapiro-Wilk's W test (Тест Шаптро-Уилка)	Отображение на графике гистограммы результатов теста на нормальность Шаптро-Уилка

Окончание таблицы 11.2

1	2
Statistics (Статистики)	Группа опций для выбора рассчитываемых статистик
Median & quartiles (Медиана и квартили)	Добавление в рассчитываемые статистики медианы и квартилей
Conf. limits for means Interval (Доверительный интервал для среднего)	Задание границ доверительного интервала для среднего значения
More Statistics (Больше статистик)	Выбор рассчитываемых статистик
Default (По умолчанию)	Выбор статистик, рассчитываемых по умолчанию
All (Все)	Выбор всех статистик для расчета
Valid N (Количество N)	Расчет объема выборки
Mean (Среднее)	Расчет среднего значения
Sum (Сумма)	Расчет суммы
Median (Медиана)	Расчет медианы
Standard Deviation (Стандартное отклонение)	Расчет стандартного отклонения
Variance (Дисперсия)	Расчет дисперсии
Standard error of mean (Стандартная ошибка среднего)	Расчет стандартной ошибки среднего
95% confidence limits of mean (95% доверительный интервал для среднего)	Расчет 95% пределов доверительного интервала для среднего
Minimum & maximum (Минимум и максимум)	Расчет минимального и максимального значений
Lower & upper quartiles (Нижний и верхний квартили)	Расчет нижнего и верхнего квартилей
Range (Размах)	Расчет размаха
Quartile range (Интерквантильная широта)	Расчет интерквантильной широты
Skewness (Асимметрия)	Расчет асимметрии
Kurtosis (Эксцесс)	Расчет эксцесса
Standard error of Kurtosis (Стандартная ошибка эксцесса)	Расчет стандартной ошибки эксцесса

Упражнение 1

В этом упражнении вы научитесь рассчитывать описательные статистики.

1. Откройте файл данных **Visota.sta**.
2. Откройте окно для вычисления выборочных статистик

Descriptive Statistics (см. рисунок 11.7).

3. Выберите переменную для анализа, например **Опыт3**. Для этого щелкните по кнопке Variables: и выделите переменную **Опыт3**.

4. Просмотрите вариационный ряд. Для этого в разделе Distribution щелкните по кнопке Frequency tables. Сохраните таблицу частот в свою папку с именем **Frequency.scr**.

5. Постройте гистограмму для выбранной переменной. Для этого в разделе Distribution щелкните по кнопке Histograms. Сохраните гистограмму в файле **Hist.stg**.

6. Получите все описательные статистики выбранной переменной. Для этого в разделе Statistics щелкните по кнопке More statistics. В появившемся диалоговом окне щелкните по кнопке All, затем по кнопке Ok. Сохраните результаты в файле **Descript.scr**.

7. Просмотрите и проанализируйте значения полученных статистик. Сделайте выводы.

8. Вычислите коэффициент вариации и относительную ошибку среднего.

9. Создайте файл отчета. Поместите в него все результаты анализа и выводы. Сохраните файл с именем **Descript.rtf** и закройте программу.

4 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Проведите сравнительный анализ описательных статистик для переменной заданного варианта и совокупности данных, полученных во всех параллельных опытах, сделайте выводы. Для этого:

1. Откройте файл данных **Visota.sta**. Сохраните его с именем **VisotaAll.sta**.
2. В файле **VisotaAll.sta** подготовьте данные таким образом, чтобы осталось две переменные: первая – заданная по варианту и вторая – объединяющая данные всех опытов.
3. В окне Descriptive Statistics (Описательные статистики) в качестве переменных для анализа укажите обе оставшиеся переменные.
4. Получите необходимые характеристики выборок одновременно для двух переменных, сравните их и сделайте выводы.

5 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. В чем отличие дискретных случайных величин от непрерывных?
2. Что такое закон распределения случайной величины?
3. Какие способы задания закона распределения случайной величины Вам известны?
4. Что такое функция распределения случайной величины (интегральная функция)? Какими она обладает свойствами?
5. Что такое функция плотности распределения вероятности (дифференциальная функция)? Какие свойства ей присущи?

6. Что такое математическое ожидание непрерывной, дискретной случайной величины?
7. Что такое дисперсия случайной величины?
8. Что такое генеральная совокупность, выборка?
9. Что такое вариационный ряд, гистограмма, полигон, кумулятивная кривая?
10. Понятие частоты, относительной частоты, накопленной частоты?
11. Что такое дисперсия, стандартное отклонение?
12. Что называют лимитами, размахом вариации?
13. Понятие медианы, квартилей, интерквартильной широты?
14. Для чего необходимы коэффициент асимметрии и эксцесс?
15. В чем отличие эмпирической функции распределения от теоретической?
16. В чем отличие выборочной дисперсии S^2 от дисперсии генеральной совокупности D?
17. Как рассчитать коэффициент вариации и относительную ошибку среднего?
18. Как задать список переменных для вычисления описательных статистик?
19. Как просмотреть таблицу результатов расчета заданных статистик?
20. Как отобразить на экране вариационный ряд?
21. Как построить гистограмму?
22. Как изменить способ группировки (разбиения на интервалы) данных? Какие способы группировки Вы знаете?
23. Как выбрать расчет всех статистических показателей выборки?

Практическое занятие № 12

Тема: **КОРРЕЛЯЦИЯ**

1 ЦЕЛЬ РАБОТЫ

Получить представление:
- о математическом моделировании;
- о корреляционном анализе;
- о видах корреляционных связей.

Научиться:
- формировать таблицу данных для корреляционного анализа;
- проводить корреляционный анализ в модуле Basic Statistics and Tables.

2 ТЕОРЕТИЧЕСКАЯ ЧАСТЬ

2.1 Математические модели

При исследовании объектов и явлений важным является выяснение их свойств и связей параметров, определяющих свойства исследуемых объектов и явлений, а также прогнозирование изменений их свойств и состояний.

Одним из методов исследования объектов и явлений является *моделирование*, то есть создание и исследование моделей.

Модель – это упрощенное подобие реального объекта, явления или процесса, это такой материальный или мысленно представляемый объект, который замещает объект-оригинал с целью его исследования, сохраняя важные для данного исследования типичные черты и свойства оригинала.

Целью моделирования является:

– постановка экспериментов над моделью с последующей интерпретацией их результатов применительно к объекту исследования;

– прогнозирование будущего состояния или поведения объекта.

Математическая модель дает описание основных свойств объекта исследования в виде математических зависимостей.

Различают *детерминированные* и *стохастические* (*вероятностные*) *модели*.

Если исходные данные известны и достоверны, а случайные воздействия пренебрежимо малы, то есть изменению одного из параметров соответствует точное изменение исследуемого свойства, *модель* называют *детерминированной*.

Если объект исследования подвергается воздействию неучтенных величин, и изменение какого-либо параметра не дает строго определенное изменение значения исследуемого свойства, а имеется некоторый разброс значений, то такую *модель* называют *стохастической* (*вероятностной*). На прак-

тике исследователь чаще всего имеет дело со стохастическими (вероятностными) моделями.

Модели с учетом временного фактора разделяют на *динамические* и *статические*.

Динамическая модель описывает процессы изменения и развития систем. Такие модели часто имеют вид дифференциальных, интегральных или интегро-дифференциальных уравнений.

Статическая модель описывает состояние системы в определенный момент времени. Данные модели выражаются при помощи алгебраических уравнений.

Для получения стохастических математических моделей используют метод «черного ящика».

Черный ящик – система, внутреннее устройство и механизм работы которой очень сложны, неизвестны или неважны в рамках решаемой задачи.

«*Метод черного ящика*» – метод исследования таких систем, когда вместо свойств и взаимосвязей составных частей системы, изучается реакция системы, как целого, на изменяющиеся внешние воздействия.

В рассматриваемой модели объект исследования рассматривается как «черный ящик» с набором входных (X) и выходных (Y) параметров (рисунок 12.1). Реальные связи в таких моделях заменяются математическими зависимостями вида

$$Y = f(X_1, X_2, \dots X_q). \qquad (12.1)$$

Уравнение 12.1 называют *уравнением регрессии*, а модели такого вида – *регрессионными*.

Входные параметры X называют *факторами* или *независимыми переменными*. Выходные параметры Y называют *откликами* или *зависимыми переменными*.



Рисунок 12.1 – Структурная схема объекта исследования

Для получения регрессионных моделей проводят ряд измерений, в которых некоторому набору факторов $X_i = (X_1, X_2, \dots X_q)$ соответствует значение отклика Y_j. Измерения отклика Y_j производят при различных уровнях факторов $X_1, X_2, \dots X_q$. Обрабатывая полученный массив опытных (эмпирических) данных, вычисляют коэффициенты (параметры) регрессионного уравнения 12.1. Процесс вычисления параметров функции $Y = f(X_1, X_2, \dots X_q)$ по опытным данным называют *аппроксимацией*, а саму функцию – *аппроксимирующей*.

Систему операций, воздействий и наблюдений, направленных на получение информации об объекте при исследовательских испытаниях называют *экспериментом*. Отдельную элементарную часть эксперимента называют *опытом*.

На каждый объект воздействует множество факторов. Чаще всего при построении модели невозможно учитывать все факторы, которые могут повлиять на значения величин Y_j. В связи с этим необходимо разделить входные параметры по степени важности влияния их изменений на выходные. От того, насколько правильно выделены важнейшие факторы, зависит успех моделирования. Отбрасывание менее значимых факторов огрубляет объект моделирования, но в то же время способствует пониманию его главных свойств и закономерностей. *Важность фактора определяется на основании априорной (предварительной) информации (литературные источники, предварительные эксперименты).*

Входные параметры можно разделить на следующие категории:

– *контролируемые и управляемые* (исследователь может их изменять);

– *контролируемые, но неуправляемые* (исследователь может только контролировать параметры);

– *неконтролируемые и неуправляемые*.

Когда исследователь может контролировать и изменять входные воздействия *эксперимент* называют *активным*. Если исследователь может только контролировать, но не изменять входные воздействия, то *эксперимент* является *пассивным*.

Как в активном, так и в пассивном экспериментах присутствуют неконтролируемые и неуправляемые факторы ε (рисунок 12.2), которые также изменяют значение выходной величины. Изменения отклика, обусловленные неконтролируемыми факторами, а также погрешностями измерений, называют *ошибкой опыта*.

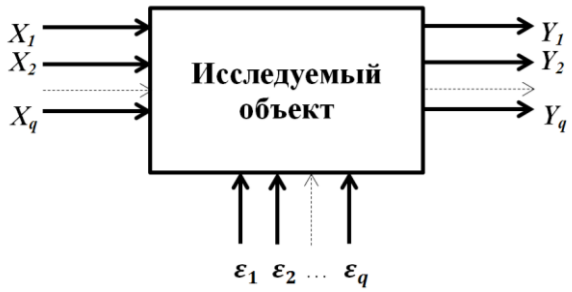


Рисунок 12.2 – Структурная схема объекта исследования с учетом неконтролируемых факторов

С учетом ошибки опыта регрессионное уравнение (12.1) примет вид:

$$Y = f\left(X_1, X_2, \dots X_q\right) + \varepsilon. \qquad (12.2)$$

При исследовании зависимостей выходной величины Y от факторов $X_1, X_2, \dots X_q$ применяют следующие статистические процедуры:

– корреляционный анализ;

– дисперсионный анализ;
– регрессионный анализ.

Корреляционный анализ – метод, позволяющий обнаружить зависимость между несколькими случайными величинами.

Целью корреляционного анализа в этом случае является установление степени влияния каждого из факторов $X_1, X_2, \dots X_q$ на отклик Y, а также выяснение степени взаимосвязи факторов $X_1, X_2, \dots X_q$ между собой.

Дисперсионный анализ предназначен для оценки влияния на отклик Y количественных и неколичественных факторов $(X_1, X_2, \dots X_q)$.

Регрессионный анализ используется для определения аналитической зависимости отклика Y от количественных факторов $(X_1, X_2, \dots X_q)$.

3.2 Корреляционный анализ

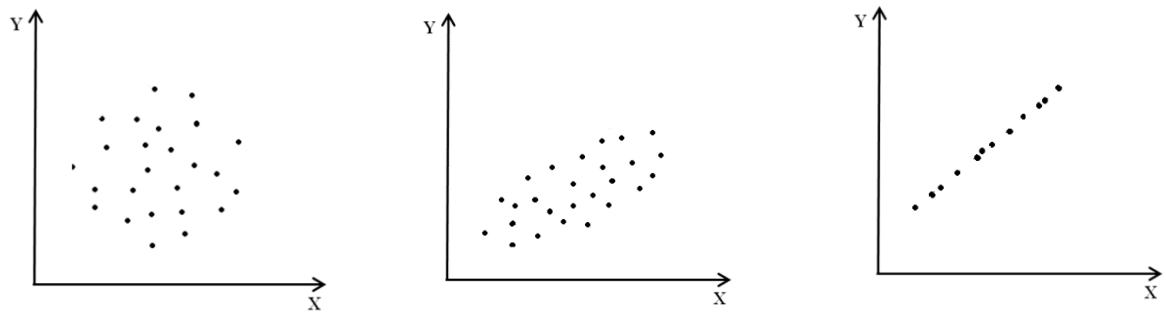
Корреляционный анализ в основном предваряет регрессионный анализ. Целью корреляционного анализа в этом случае является установление степени влияния каждого из факторов $X_1, X_2, \dots X_q$ на отклик Y, а также выяснение степени взаимосвязи факторов $X_1, X_2, \dots X_q$ между собой. При этом интерес вызывает не только степень связи, но и ее форма.

Как и модели, *связи* могут быть:

– *детерминированными* (*функциональными*), когда значению одной величины соответствует строго определенное значение другой величины;

– *вероятностными* (или *стохастическими*), когда значению одной величины может соответствовать множество возможных значений другой величины в некотором диапазоне.

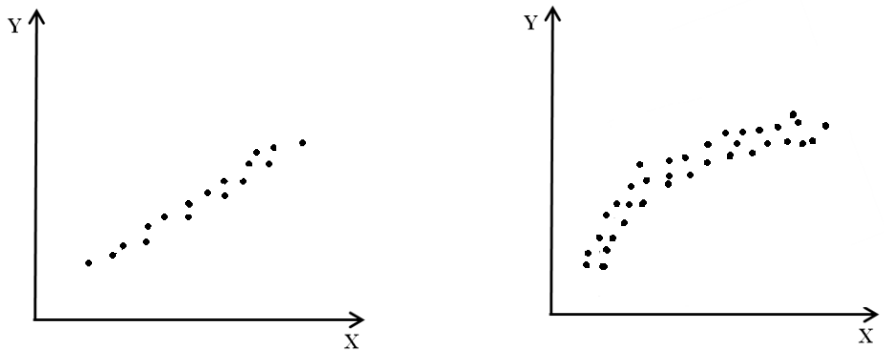
Степень тесноты связи можно оценить визуально, построив диаграмму рассеяния исследуемых переменных (рисунок 12.3). Чем уже облако рассеяния, тем теснее связь. Если облако рассеяния имеет форму круга (рисунок 12.3 *а*), то связь отсутствует. Если облако рассеяния имеет вытянутую форму (рисунок 12.3 *б*), то имеется взаимосвязь величин. Если все точки графика лежат на одной линии (рисунок 12.3 *в*), значит, следует говорить о наличии функциональной связи.



а – связь отсутствует *б* – наличие связи *в* – сильная связь

Рисунок 12.3 – Визуальная оценка тесноты связи

По форме корреляция может быть линейной и нелинейной (криволинейной) (рисунок 12.4).



а – линейная корреляция *б* – нелинейная корреляция

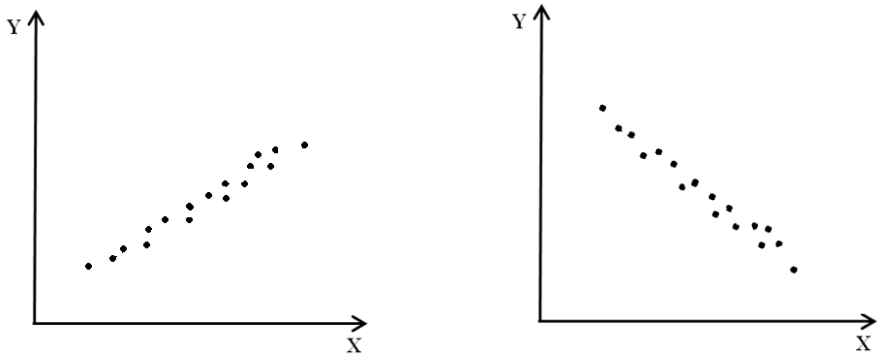
Рисунок 12.4 – Формы корреляционных зависимостей

Под *линейной корреляционной зависимостью* между двумя признаками x и y понимают такую зависимость, которая носит линейный характер и выражается уравнением $y = b_0 + b_1 x$.

Обратите внимание! В дальнейшем будем рассматривать только линейную корреляцию.

По направлению связь может быть прямой и обратной (рисунок 12.5).

Если при увеличении x значение y возрастает, то *корреляция* называется *положительной* или *прямой* (рисунок 12.5 *а*), а если уменьшается – *отрицательной* или *обратной* (рисунок 12.5 *б*).



а – прямая корреляция *б* – обратная корреляция

Рисунок 12.5 – Направления корреляционных связей

Если изучается зависимость между двумя признаками, то такая корреляция называется *простой*. При изучении зависимости между двумя и более признаками – *множественной корреляцией*.

Для получения корреляционной зависимости случайной величины Y от случайной величины X анализируют n парных наблюдений (x_i, y_i).

Оценивают тесноту и направление связи с помощью *линейного коэффициента корреляции Пирсона* r. Коэффициент корреляции Пирсона является безразмерной величиной, изменяющейся в диапазоне $-1 < r < 1$.

Коэффициент корреляции Пирсона рассчитывается по формуле

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 (y_i - \bar{y})^2}}. \quad (12.3)$$

Качественная оценка тесноты связи может быть выявлена на основе шкалы Чеддока (таблица 12.1). Положительный коэффициент корреляции соответствует прямой корреляции, а отрицательный – обратной.

Таблица 12.1 – Оценка тесноты корреляционной связи

Значение $\|r\|$	Теснота связи
$r = 0$	отсутствует
$0{,}1 - 0{,}3$	слабая
$0{,}3 - 0{,}5$	умеренная
$0{,}5 - 0{,}7$	заметная
$0{,}7 - 0{,}9$	высокая
$0{,}9 - 0{,}99$	весьма высокая
$1{,}0$	функциональная

Следует отметить, что коэффициент корреляции близкий к нулю свидетельствует об отсутствии линейной связи между случайными величинами X и Y, хотя нелинейная зависимость может иметь место.

Квадрат коэффициента корреляции r^2 называется *коэффициентом детерминации*. Он показывает долю изменений изучаемого признака, обусловленных влиянием фактора.

Для оценки надежности выборочного коэффициента корреляции вычисляют его *стандартную ошибку* и *критерий значимости*.

Стандартная ошибка коэффициента корреляции определяется по формуле

$$S_r = \sqrt{\frac{1 - r^2}{n - 2}}, \quad (12.4)$$

где n – число парных наблюдений выборки;

r – коэффициент корреляции.

Критерий значимости коэффициента корреляции t_r служит для проверки гипотезы о значимости коэффициента корреляции. В качестве *нулевой гипотезы* H_0 выдвигается предположение о том, что коэффициент корреляции незначим, то есть линейная связь между случайными величинами X и Y отсутствует (H_0: $r = 0$).

Критерий значимости коэффициента корреляции t_r рассчитывается по формуле

$$t_r = \frac{r}{S_r}. \quad (12.5)$$

Рассчитанный критерий значимости t_r сравнивают с табличным критерием Стьюдента $t_{05}(t_{01})$ (приложение Б) при 5% или 1% уровне значимости и $(n-2)$ числе степеней свободы. Если $t_r > t_{05}(t_{01})$, то нулевая гипотеза отвергается, то есть корреляционная связь существенна. Можно также вероятность p, соответствующую квантили распределения Стьюдента (коэффициенты Стьюдента) t_r, сравнить с принятым уровнем значимости. Если $p \leq \alpha$, то коэффициент корреляции значим.

4 ТЕХНОЛОГИЯ ВЫПОЛНЕНИЯ РАБОТЫ

4.1 Корреляционный анализ в модуле Basic Statistics and Tables

Для проведения корреляционного анализа в модуле Basic Statistics and Tables необходимо выполнить команду Analysis – Startup Panel –Correlation matrices. В результате на экране появится диалоговое окно корреляции Пирсона Pearson Product-Moment Correlation (рисунок 12.6).

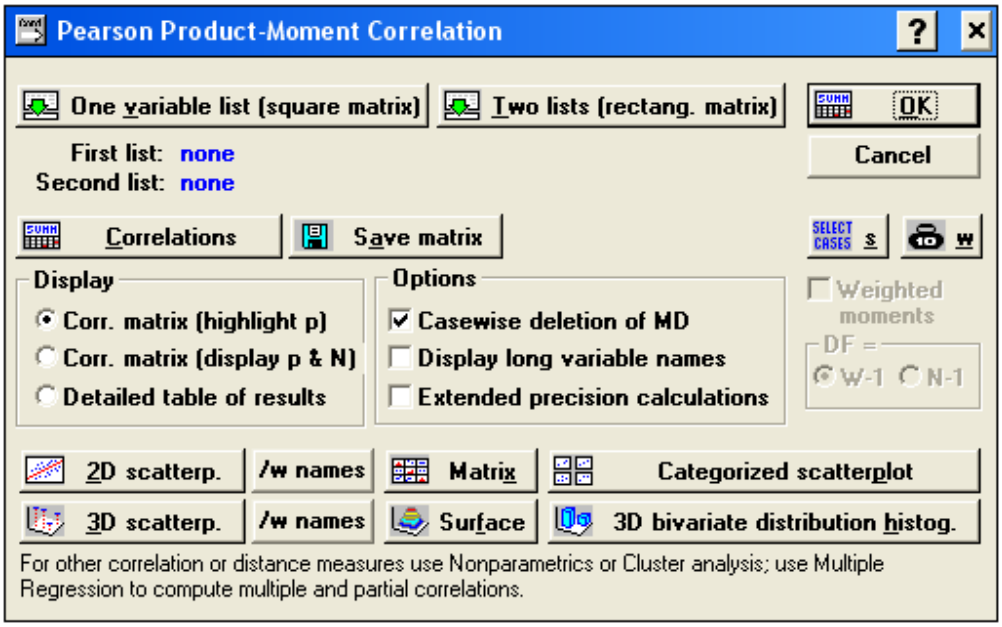


Рисунок 12.6 – Диалоговое окно Pearson Product-Moment Correlation

Назначение кнопок и опций верхней части этого окна представлено в таблице 12.2. Кнопки в нижней части окна предназначены для построения графиков различных типов.

Таблица 12.2 – Назначение опций окна Pearson Product-Moment Correlation

Элемент окна	Назначение элемента
One variable list (square matrix) (Один список переменных (квадратная матрица))	Выбор переменных для анализа. Рассчитываются корреляции между всеми выбранными переменными
Two lists (rectang. Matrix) (Два списка переменных (прямоугольная матрица))	Выбор переменных для анализа. Рассчитываются корреляции между переменными, выбранными в левом листе и в правом листе
Correlations (Корреляции)	Расчет и отображение корреляционной матрицы
Display (Отображение)	Группа опций для выбора вида информации, представленной в корреляционной матрице
Corr. Matrix (highlight p) корреляционная матрица (выделение p)	Отображение корреляционной матрицы с выделенными значимыми корреляциями
Corr. Matrix (display p&N) корреляционная матрица (отображать p и N)	Отображение матрицы значений коэффициентов корреляции и соответствующих им уровней значимости с выделением значимых коэффициентов и выводом числа анализируемых случаев
Detailed table of results (Подробная таблица результатов)	Попарное отображение переменных с подробным представлением полученных результатов корреляционного анализа
Options (Опции)	Группа опций для задания параметров расчета
Casewise deletion of MD (Исключение случаев с ПД)	Исключение из анализа случаев с пропущенными данными (ПД)
Display long variable names (Отображать длинные имена переменных)	Отображение длинных имен переменных
Extended precision calculations (Повышенная точность вычислений)	Расчет значений с повышенной точностью

Для подробного представления результатов корреляционного анализа используются статистические характеристики, приведенные в таблице 12.3.

Таблица 12.3 – Таблица результатов корреляционного анализа

Обозначение	Пояснение
Mean	Среднее значение переменной
Std.Dv. (Standard Deviations)	Стандартное отклонение переменной
r(X,Y)	Коэффициент корреляции пары переменных
rI	Коэффициент детерминации
t	Критерий существенности коэффициента корреляции (Стьюдента)
p	Уровень значимости коэффициента корреляции
N	Число случаев для переменной
Constant dep: Y	Точка пересечения с осью Y
Slope dep: Y	Коэффициент, определяющий угол наклона к оси Y
Constant dep: X	Точка пересечения с осью X
Slope dep: X	Коэффициент, определяющий угол наклона к оси X

Упражнение 1

В этом упражнении вы научитесь выполнять корреляционный анализ данных, представленных в таблице 11.4.

Выполните корреляционный анализ данных по определению относительной влажности *X* и липкости *Y* чернозема.

Таблица 12.4 – Зависимость липкости чернозема ***Y*** от относительной влажности *X*

X, %	19,9	20,9	26,1	29,4	30,5	40,3	44,8	47,8	55,6	58,3	64,5	76,6
Y, г/см^2	0,0	0,6	1,1	1,2	1,7	1,7	2,6	3,4	4,2	5,8	6,3	7,3

1. Изучите назначение опций окна Pearson Product-Moment Correlation (таблицы 12.2, 12.3).
2. Запустите модуль Basic Statistics and Tables.
3. Создайте новую таблицу, содержащую две переменные *X* и *Y* по **12** случаев.
4. Заполните новую таблицу в соответствии с таблицей 12.4. Полученную таблицу (рисунок 12.7) сохраните в файл с именем **Chernozem.sta**.
5. Выполните команду Analysis – Startup Panel – Correlation matrices.
6. В диалоговом окне Pearson Product-Moment Correlation нажмите кнопку Two lists (rectang. matrix) и укажите в качестве первой переменной (First Variable) переменную *X*, второй переменной (Second Variable) – *Y* (рисунок 12.8). Нажмите кнопку OK.

Data: Chern...

	1 X	2 Y
1	19,9	0,0
2	20,9	,6
3	26,1	1,1
4	29,4	1,2
5	30,5	1,7
6	40,3	1,7
7	44,8	2,6
8	47,8	3,4
9	55,6	4,2
10	58,3	5,8
11	64,5	6,3
12	76,6	7,3

Рисунок 12.7 – Таблица данных для определения корреляционной связи между относительной влажностью ***X*** и липкостью чернозема ***Y***

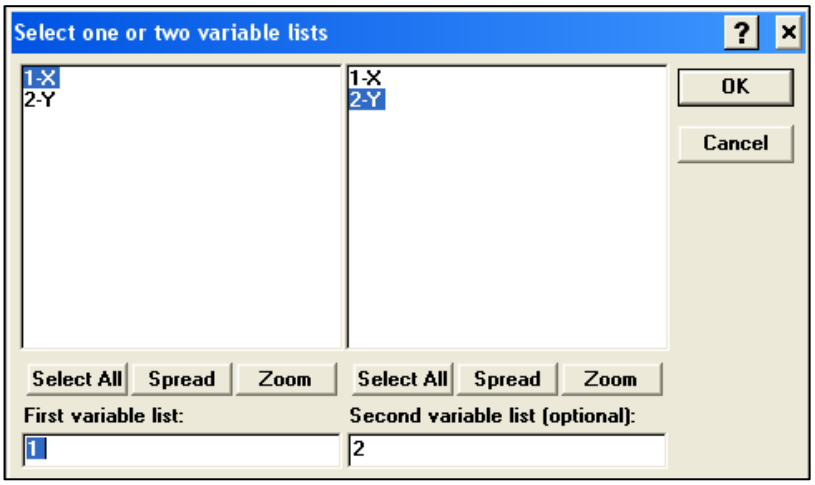


Рисунок 12.8 – Окно для выбора переменных

7. В диалоговом окне Pearson Product-Moment Correlation в группе опций Display установите переключатель Detailed table of results.

8. Нажмите кнопку OK, просмотрите результаты корреляционного анализа (рисунок 12.9).

Correlations (chernozem.sta)

Continue...

Marked correlations are significant at p < ,05000
(Casewise deletion of missing data)

Var. X & Var. Y	Mean	Std.Dv.	r(X,Y)	rI	t	p	N	Constant dep: Y	Slope dep: Y	Constant dep: X	Slope dep: X
X	42,89167	18,25658									
Y	2,99167	2,41151	,977253	,955024	14,57191	,000000	12	-2,54502	,129085	20,75813	7,398395

Рисунок 12.9 – Результаты анализа корреляционной связи между относительной влажностью ***X*** и липкостью чернозема ***Y***

По результатам корреляционного анализа, приведенного на рисунке 12.9, **сделаем следующие выводы:**

1) среднее значение (Mean) переменной X составляет 42,89%, переменной Y – 2,99 г/см^2 при стандартных отклонениях (Std. Dv.) 18,26 и 2,41 соответственно;

2) коэффициент корреляции r(X, Y) равен 0,977, то есть корреляция прямая, корреляционная связь сильная (весьма высокая) (см. таблицу 12.1);

3) значение коэффициента детерминации r^2 (rI) составляет 0,955. Это свидетельствует о том, что 95,5% изменений липкости чернозема вызвано изменением его влажности;

4) критерий значимости коэффициента корреляции t_r (t) равен 14,57. Табличный критерий Стьюдента $t_{0,5}$ на 5%-м уровне значимости при n–2=12–2=10 степенях свободы составляет 2,23 (приложение Б). Так как $t_r = 14{,}57 > t_т = 2{,}23$, считается, что корреляционная связь существенна на 5% уровне значимости. Об этом свидетельствует и практически нулевое значение уровня вероятности p для квантили t. Если рассчитанное значение $p < 0{,}05$, то корреляционная связь значима. Этот факт отражен в заголовке таблицы (см. рисунок 12.9): Marked correlations are significant at p<0,05 – выделенные (красным цветом) корреляции являются значимыми на уровне p<0,05.

9. Сохраните результаты корреляционного анализа в файл с именем **Correlations.scr**.

10. Для перехода в диалоговое окно Pearson Product-Moment Correlation нажмите кнопку Continue.

11. В диалоговом окне Pearson Product-Moment Correlation нажмите кнопку 2D Scatterplot. На экране появится окно с графиком сти $Y(X)$.

12. В параметрах осей Y и X задайте минимальные значения равные **-3** и **0** соответственно.

Для этого выполните команду Layouts – General Layout… В появившемся диалоговом окне General Layout: 2D Graphs выполните настройки, как показано на рисунках 12.10, 12.11.

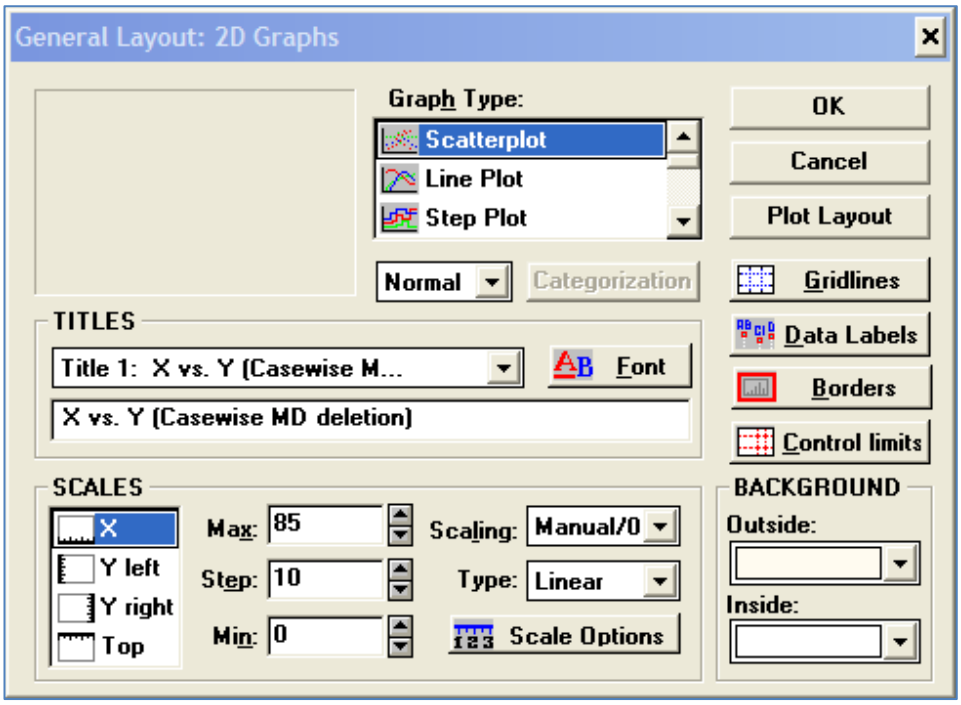


Рисунок 12.10 – Настройка параметров оси X

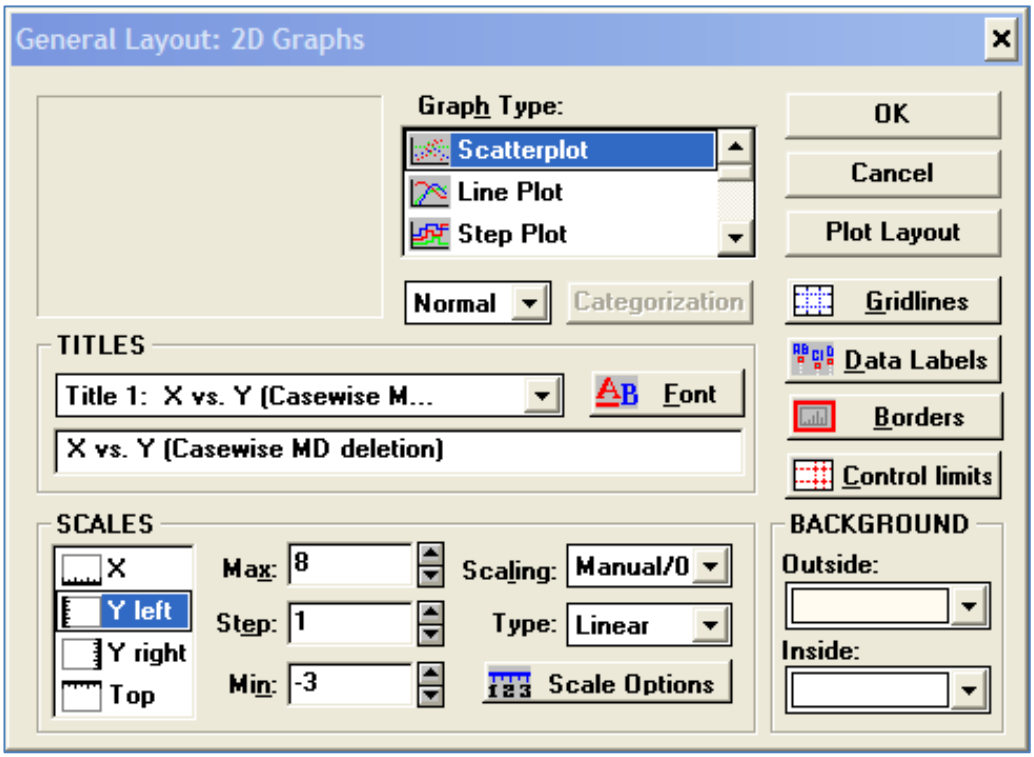


Рисунок 12.11 – Настройка параметров оси Y

Полученный график пересекает ось *Y* в точке примерно **-2,5**, а ось *X* – в точке около **20** (рисунок 12.12), что подтверждают значения, приведенные на рисунке 12.9.

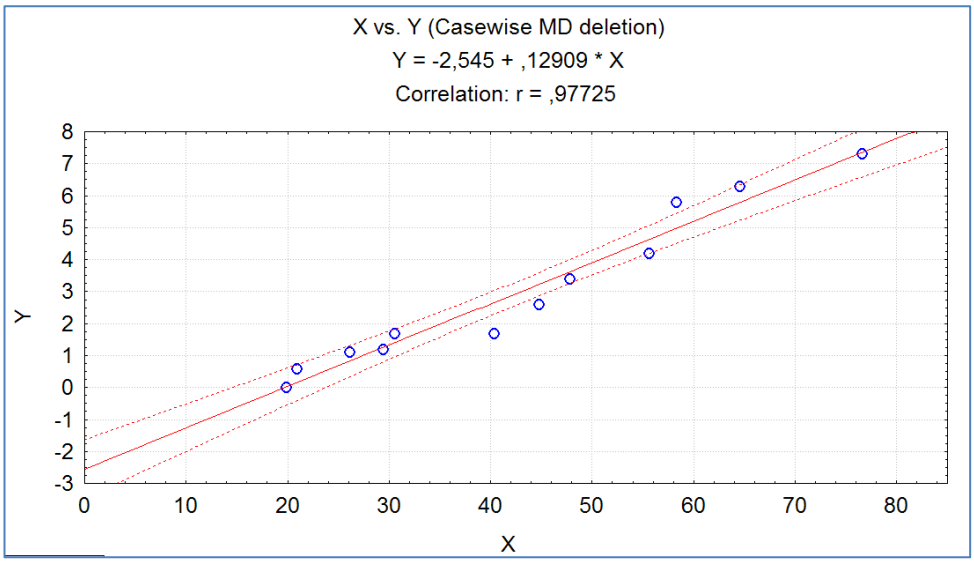


Рисунок 12.12 – Двумерная зависимость (2D Scatterplot) липкости чернозема ***Y*** от его относительной влажности ***X***

По графику двумерной зависимости (рисунок 12.12) можно сделать следующий **вывод:** корреляционная связь между липкостью чернозема *Y* и относительной влажностью чернозема *X* сильная (весьма высокая) (см. рисунок 12.3), прямая (угол наклона прямой к положительному направлению оси X меньше 90°), линейная (точки практически лежат на прямой линии (см. рисунок 12.4).

Выводы, выполненные по графику (рисунок 12.12) полностью подтверждают выводы выполненные по таблице корреляционного анализа (см. рисунок 12.9).

5 УПРАЖНЕНИЯ ДЛЯ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

1. В модуле Basic Statistics and Tables создайте новую таблицу, содержащую две переменные *F* и *M* по **20** случаев.

2. Заполните новую таблицу данными в соответствии с таблицей 12.5. Сохраните файл с именем **Massa.sta**.

3. Проведите детальный анализ корреляционной связи между массой зерна и влажностью зерна.

4. Выводы о результатах анализа запишите в отчете **CorrMassa.docx**.

5. Откройте файл **Visota.sta**. Проведите детальный корреляционный анализ. Результаты сохраните в отчете с именем **CorrVisota.docx**.

Таблица 12.5 – Результаты измерений массы 1000 семян при различной влажности зерна

φ, %	13	17	23	19	20	15	20	21	23	15
m, г	41,1	43,6	46,5	44,2	44,4	42,4	45,4	44,6	45,2	41,7
φ, %	16	24	14	13	22	18	18	14	20	16
m, г	43,5	46,8	42,5	41,9	46,0	44,5	43,0	41,9	43,5	42

6 КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что представляет собой математическая модель? Какие виды моделей вы знаете?
2. Что называют факторами, откликом?
3. Чем активный эксперимент отличается от пассивного?
4. Что такое ошибка опыта? Чем она вызвана?
5. Какие статистические процедуры используют при анализе связей? Назначение каждой из них.
6. Что такое детерминированная (функциональная) и вероятностная (стохастическая) связи?
7. Какие виды корреляционных связей вы знаете?
8. Какие различают виды корреляционных связей по направлению? Как определить направление связи по значению коэффициента корреляции?
9. Какие различают виды корреляционных связей по тесноте? Укажите диапазоны значений коэффициентов корреляции для каждого из видов.
10. Что такое коэффициент детерминации?
11. Что такое критерий значимости коэффициента корреляции? Как он рассчитывается?
12. Сформулируйте нулевую гипотезу при проверке значимости (существенности) коэффициента корреляции.
13. Как оценить значимость коэффициента корреляции?
14. Как вызвать окно для проведения корреляционного анализа?
15. Какие статистические характеристики получают в результате детального корреляционного анализа? Для чего нужны?
16. Как оценить значимость коэффициента корреляции по значению вероятности p квантили распределения t?

ЛИТЕРАТУРА

1. Experimental studies to identify ultraviolet radiation impact on tomato seeds «Rozoviy Novichok» seeding quality / N. E. Ponomareva, G. V. Stepanchuk, N. N. Gracheva [et al.] // IOP Conference Series: Earth and Environmental Science, Zernograd, Rostov Region, 27–28 августа 2020 года. – Zernograd, Rostov Region, 2021. – P. 012035. – DOI 10.1088/1755-1315/659/1/012035.

2. Study of the effect of ultraviolet radiation of the UV-B and UV-C range on the sowing qualities of wheat seeds «Zernogradka-9» / V. N. Belenov, N. N. Gracheva, N. Y. Ponomareva [et al.] // Journal of Agriculture and Environment. – 2024. – No. 11(51). – DOI 10.60797/JAE.2024.51.20.

3. Vasilyev, A.N. Elektrotechnologies as controlling influence on biological objects in agricultural production / A.N. Vasilyev // Research in Agricultural Electric Engineering. – 2013. – Vol. 3. – № 3. – P. 96-102.

4. Боровиков, В.П. Statistica. Искусство анализа данных на компьютере: для профессионалов (+ CD). / В.П. Боровиков. – 2-е изд. – Санкт-Петербург: Питер, 2003. – 688 с.

5. Будников, Д.А. Регрессионные модели ЭАВ сушки зерна // Д.А. Будников, Н.Н. Грачева // Новейшие направления развития аграрной науки в работах молодых учёных: сб. тр. V Междунар. науч.-практ. конф. мол. учёных, посвящ. 10-летию её проведения (20 апреля 2012 г., пос. Краснообск) / СО Россельхозакадемии. – Новосибирск, 2012. – С. 14–16.

6. Буханцов, К.Н. Исследование процесса обеззараживания семян подсолнечника сорта Флагман от плесневых грибов Phomopsis helianthi Munt., Cvet. et al. в электромагнитном поле сверхвысокой частоты / К.Н. Буханцов, А.А. Парапонов // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – Краснодар: КубГАУ, 2013. – № 91(07). – С. 812–840. – Режим доступа: ej.kubagro.ru /2013/07/pdf/22.pdf, 1,813 у.п.л.

7. Буханцов, К.Н. Методика проведения экспериментальных исследований конвективно-озоновоздушной сушки зерновых материалов / К.Н. Буханцов // Механизация технологических процессов в животноводстве: технологии, машины, оборудование: сб. науч. тр. 4-й Междунар. науч.-техн. конференции: «Ресурсосберегающие технологии и инновационные проекты в АПК» (г. Зерноград Ростовской обл., ВНИПТИМЭСХ, 14–15 апреля 2009 г.). – Зерноград, 2009. – С. 265–276.

8. Вакуумные управляемые пульсаторы, доильные автоматы и системы: монография / И.К. Винников, О.Б. Забродина, О.Н. Бахчевников, В.Н. Литвинов. – Зерноград: Азово-Черноморский инженерный институт ФГБОУ ВПО ДГАУ, 2015. – 188 с.

9. Васильев, А.Н. Планирование эксперимента по определению сопротивления плотного зернового слоя воздушному потоку / А.Н. Васильев, Н.Б. Руденко // Электротехнологии и электрооборудование в сельскохозяйственном производстве: сборник научных трудов. – Зерноград: ФГОУ ВПО АЧГАА, 2007. – Т. 1. – Вып. 7. – С. 84–87.

10. Васильев, А.Н. Регрессионные модели нагрева в поле действия СВЧ нагрева зерна при рециркуляционной сушке активным вентилированием / А.Н. Васильев, Д.А. Будников, Н.Б. Руденко // Инновационные технологии и техника нового поколения – основа модернизации сельского хозяйства: сб. науч. трудов / ВИМ. – Москва: ВИМ, 2011. – Ч. 1. – С. 407–411.

11. Васильев, А.Н. Экспериментальное исследование сушки зерна электроактивированным воздухом / А.Н. Васильев, Д.А. Будников, Н.Н. Грачева // Вестник Орловского государственного аграрного университета. – 2012. – № 5 (38) – С. 160–162.

12. Гмурман, В.Е. Теория вероятностей и математическая статистика: учебное пособие для вузов / В.Е. Гмурман. – 9-е изд., стер. – Москва: Высшая школа, 2003. – 479 с.

13. Грачева, Н. Н. Использование прикладного программного обеспечения для моделирования процессов в теплоэнергетических системах сельского хозяйства / Н. Н. Грачева, А. Н. Токарева, М. С. Демченко // Цифровые технологии и математическое моделирование в науке, образовании и производстве: материалы Всероссийской научно-практической конференции для преподавателей и научных сотрудников, Иркутск, 26–27 сентября 2024 года. – Иркутск: Иркутский государственный аграрный университет им. А.А. Ежевского, 2024. – С. 25-32.

14. Доспехов, Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований) / Б.А. Доспехов. – Москва: Колос, 1973. – 336 с.

15. Исупова, А. М. Использование методов оптимизации для формирования резервного фонда электросетевых предприятий / А. М. Исупова, В. Я. Хорольский, Н. Б. Руденко // Вестник аграрной науки Дона. – 2022. – Т. 15, № 4(60). – С. 105-115. – DOI 10.55618/20756704_2022_15_4_105-115.

16. Красовский, Г.И. Планирование эксперимента / Г.И. Красовский, Г.Ф. Филаретов. – Минск: Изд-во БГУ, 1982. – 302 с.

17. Мельников, С.В. Планирование эксперимента в исследованиях сельскохозяйственных процессов / С.В. Мельников, В.Р. Алешкин, П.М. Рощин. – 2-е изд., перераб. и доп. – Ленинград: Колос, Ленингр. отделение, 1980. – 168 с., ил.

18. Методология выбора технических мероприятий по снижению потерь электрической энергии в практике эксплуатации сельских электрических сетей / К. М. Юндин, В. Я. Хорольский, А. М. Исупова, Н. Б. Руденко // Энергобезопасность и энергосбережение. – 2023. – № 4. – С. 10-17.

19. Моделирование организации аварийно-восстановительных работ в распределительных электрических сетях в результате воздействия стихийных явлений / К. М. Юндин, А. М. Исупова, В. Я. Хорольский [и др.] // Электрические станции. – 2025. – № 3(1124). – С. 44-48. – DOI 10.71841/EP.elst.2025.1124.3.07.

20. Обоснование параметров водоподготовительной установки в котельной / И. В. Юдаев, И. Э. Липкович, А. Н. Токарева [и др.] // Сельский механизатор. – 2022. – № 5. – С. 22-23.

21. Обоснование схемы расположения инфракрасных излучателей в помещении ремонтного цеха с использованием программы STATISTICA / А. А. Серегин, А. Н. Токарева, Н. Н. Грачева, А. А. Недопекина // Энергосбережение и водоподготовка. – 2021. – № 3(131). – С. 57-63.

22. Определение рациональных параметров цикла теплонасосной установки для теплоснабжения животноводческого комплекса / С. В. Брагинец, А. Н. Токарева, Н. Н. Грачева [и др.] // Известия Горского государственного аграрного университета. – 2018. – Т. 55, № 2. – С. 132-137.

23. Планирование эксперимента по предпосевной обработке семян переменным электромагнитным полем промышленной частоты / М.В. Жолобова, М.Г. Федорищенко, А.С. Казакова, Н.Н. Грачева // Политематический сетевой электронный научный журнал Кубанского государственного аграрного университета. – Краснодар: КубГАУ, 2013. – № 91(07). – С. 929–938. – Режим доступа: http://ej.kubagro.ru/2013/07/pdf/40.pdf.

24. Пономарева, Н. Е. Оценка влияния освещённости на показатели характеризующие рассаду огурца / Н. Е. Пономарева, Н. Н. Грачева, А. А. Герасименко // Агротехника и энергообеспечение. – 2020. – № 1(26). – С. 60-66.

25. Попов, М. Ю. Результаты экспериментов по выработке электрической энергии фотоэлектрическими модулями для автономного энергоснабжения био-вегетационного комплекса / М. Ю. Попов, Н. Н. Грачева // Вестник Курганской ГСХА. – 2024. – № 3(51). – С. 78-85.

26. Спирин, Н.А. Методы планирования и обработки результатов инженерного эксперимента: конспект лекций Н.А. Спирина / Н.А. Спирин, В.В. Лавров. – Екатеринбург: ГОУ ВПО УГТУ-УПИ, 2004. – 257 с.

27. Черноволов, В.А. Методика моделирования процесса дождевания дальнеструйными аппаратами при работе по кругу / В.А. Черноволов, Л.В. Кравченко // Вестник АПК Ставрополья. – 2014. – № 3 (15). – С. 68–72.

28. Черноволов, В.А. Основы научных исследований: практикум / В.А. Черноволов. – Зерноград, 2014. – 113 с.

29. Юдаев, И. В. Автоматизация процесса электроимпульсной обработки растительного сырья / И. В. Юдаев, Р. Г. Кокурин, Н. Н. Грачева // Вестник аграрной науки Дона. – 2022. – № 1(57). – С. 14-23. – DOI 10.55618/20756704_2022_15_1_14-23.

ПРИЛОЖЕНИЯ

ПРИЛОЖЕНИЕ А

Двумерные статистические графики

1 Гистограммы

Гистограммы (Histogramms) предназначены для отображения распределения значений переменных по интервалам группировки. На гистограмме хорошо видно, какие значения или диапазоны значений исследуемого признака являются наиболее частыми, насколько они различаются между собой, являются распределения симметричными или нет. Поэтому, часто визуальный анализ данных начинают с построения гистограмм для всех переменных.

В графической галерее имеются следующие типы гистограмм:

– Простые (Regular) – наиболее распространенный тип гистограмм;

– Составные (Multiple) – гистограммы нескольких переменных;

– С двойной осью Y (Double-Y) – гистограммы нескольких переменных с двумя осями Y;

– Висячие столбцы (Hanging Bars) – гистограммы, верхняя часть столбцов которых связана с кривой нормального распределения. Отклонение нижней части столбцов от нулевой отметки оси Y говорит об отличии от нормального распределения.

Выбрать тип гистограммы можно в списке Graph Type (Тип графика) окна 2D Histogramms (рисунок А.1), появляющемся при выборе группы Histogramms в графической галерее.

В данном окне выбирается переменная (или несколько переменных) для построения гистограммы с помощью кнопки Variables (Переменные)

Гистограмму можно представить в виде графика накопленных частот. Для этого необходимо выбрать опцию Cumulative Counts (Накопленные частоты).

Столбцы гистограммы могут располагаться вплотную друг к другу или иметь незаполненные промежутки. Это зависит от того, выбрана ли опция Breaks between Columns (Разрывы между столбцами).

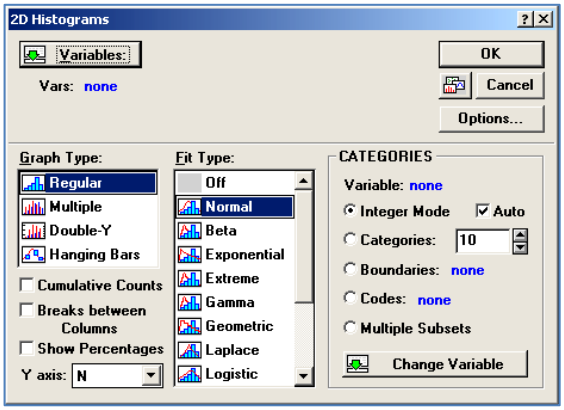


Рисунок А.1 – Окно настройки параметров двумерных гистограмм

С помощью опции Show Percentages (Отображать проценты) над каждым столбцом отображается процент случаев, попавших в заданный интервал.

По оси Y откладывается либо количество случаев (N), попавших в интервал, либо их процентное отношение к общему числу испытаний (%). Возможна так же комбинация этих показателей.

На гистограмму может быть наложена кривая теоретического распределения вероятностей. Вид теоретического распределения выбирается в прокручиваемом списке Fit Type (Тип подгонки).

Группа опций Categories (Категории) предназначена для выбора способа разбиения массива данных на интервалы (категории). Назначение кнопки внутри данной группы опций изменяется в зависимости от положения переключателя способа разбиения.

Могут быть заданы следующие способы разбиения на категории:

– Integer Mode (Целочисленный режим) – серединами интервалов гистограммы являются целые числа;

– Categories (Категории) – в поле ввода задается число интервалов разбиения;

– Boundaries (Границы) – задаются границы интервалов при помощи кнопки Specify Boundaries (Определение границ). При нажатии на эту кнопку появляется одноименное окно в котором указывают границы интервалов через пробел (1 3 5 8) или с помощью логических выражений (например, (1) x<=1, (2) 1<x<=3, (3) 3<x<=5, (4) 5<x<=8, (5) 8<x);

– Codes (Коды) – задаются список значений переменной, для которых будут подсчитаны и отображены графически в виде гистограммы частоты встречаемости. Список значений можно задать в диалоговом окне Category Codes (Категории кодов), вызываемом кнопкой Specify Codes (Определение кодов). Список значений задается в окне ввода с клавиатуры либо выбирается из набора данных в окне, вызываемом кнопкой View (Просмотр);

– Multiple Subsets (Множество подгрупп) – категории задаются отбором данных из совокупности при помощи логических операций. При нажатии на кнопку Specify Subsets (Определение подгрупп) появляется окно Specify Multiple Subsets, в котором для каждой категории задается логическое выражение – условие включения (Include) или невключения (Exclude).

В окне 2D Histogramms имеется так же кнопка Графической галереи  для быстрого перехода к другим типам графиков и кнопка Options (Опции) назначение которой будет описано ниже.

2 Диаграммы рассеяния

Диаграммы рассеяния (Scatterplots) применяют для визуального анализа зависимости между двумя переменными, значения которых откладываются в виде точек с координатами, соответствующими значениям двух пере-

менных. Когда переменные связаны, то множество точек принимает определенную форму, в противном случае точки образуют облако рассеяния.

Если на диаграмме рассеяния через полученные точки провести *сглаживающую* (аппроксимирующую) линию (Fit), то можно приблизительно оценить зависимость одной отображаемой переменной от другой. Процесс поиска зависимости (уравнения), связывающей несколько величин называют *подгонкой* (Fitting).

Диаграммы рассеяния позволяют так же увидеть так называемые «выбросы» – точки, имеющие значительное отклонение от других значений. Часто такие точки являются результатом ошибки в наблюдении. «Выбросы» могут значительно искажать зависимости между переменными.

Различают следующие Типы диаграммы рассеяния (Graph Type):

– Простые (Regular) – отображение зависимости переменной Y от переменой X;

– Составные (Multiple) – отображение зависимости нескольких переменных Y от переменой X;

– С двойной осью Y (Double-Y) – отображение зависимости двух групп переменных Y от переменой X, причем, для каждой группы переменных Y отображается своя ось;

– Частоты (Frequency) – если для одного значения X имеется несколько точек Y, то их частота отображается на данной диаграмме;

– Квантили (Quantile) – на данной диаграмме по обеим осям отображаются квантили двух переменных, в случае сходства распределений частот переменных точки графика ложатся на линию;

– Диаграммы Вороного (Voronoy) – каждая точка графика расположена внутри зоны, размер которой зависит от значения переменной.

Выбрать тип диаграммы рассеяния и настроить некоторые параметры отображения графика можно в окне 2D Scatterplots (Двумерные диаграммы рассеяния), изображенном на рисунке А.2.

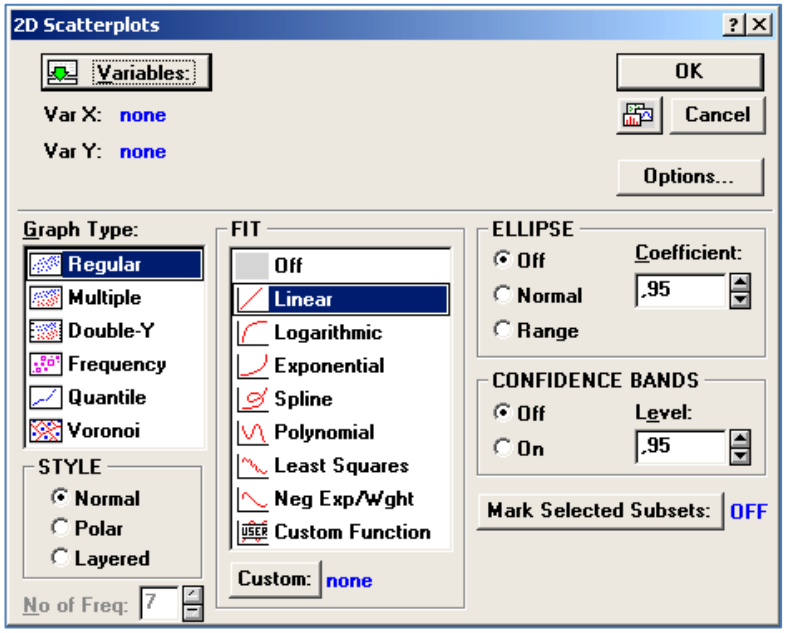


Рисунок А.2 – Окно настройки параметров двумерных диаграмм рассеяния

В данном окне выбираются отображаемые на диаграмме Переменные (Variables); тип Сглаживающей линии (Fit): Линейная (Linear), Логарифмическая (Logarithmic), Экспоненциальная (Exponential), Сглаживание сплайнами (Spline), Полиномиальная (Polynomial), Сглаживание методом наименьших квадратов (Least Squares), Функция пользователя (Custom Function); Стиль диаграммы (Style): Нормальный (Normal), С полярными координатами (Polar), Со сжатием (Layered). Имеется так же возможность отобразить Доверительный интервал (Confidence Bands), Эллипс (Ellipse) Нормально распределенных значений (Normal) и Диапазонов (Range), Пометить выбранные подгруппы (Mark Selected Subsets).

3 Нормальные вероятностные графики

Нормальные вероятностные графики (Normal Probability Plots) позволяют визуально оценить, насколько распределение исследуемого признака близко к нормальному.

Различают:

– Стандартные нормальные вероятностные графики (Normal Probability Plots);

– Полунормальные вероятностные графики (Half-Normal Probability Plots);

– Нормальные вероятностные графики с исключенным трендом (Detrended Probability Plots).

В Стандартном нормальном вероятностном графике (Normal Probability Plots) все значения переменной ранжируются. По рангам рассчитываются значения нормального распределения, которые откладываются по оси Y. По оси X откладываются наблюдения. Если наблюдаемый признак распределен по нормальному закону, то точки графика ложатся на прямую линию. В противном случае наблюдаются отклонения от прямой.

Полунормальный вероятностный график (Half-Normal Probability Plots) строится как и стандартный, только в нем рассматривается лишь положительная часть нормальной кривой.

Нормальный вероятностный график с исключенным трендом (Detrended Probability Plots) отличается от нормального тем, что в нем удален линейный тренд.

Нормальные вероятностные графики широко используются как для проверки нормальности, так и для анализа остатков в регрессионном анализе.

4 Графики квантиль-квантиль и вероятность-вероятность

Графики квантиль-квантиль (Quantile-Quantile Plots) используются при подгонке вероятностных распределений. По оси X откладываются теоретические квантили выбранного вероятностного распределения, а по оси Y –

наблюдаемые значения. Если точки графика ложатся на прямую линию, значит вероятностное распределение наблюдаемого признака (эмпирическое распределение) совпадает с выбранным теоретическим распределением.

Графики вероятность-вероятность имеют то же назначение. Отличие состоит лишь в том, что по осям откладываются не квантили, а вероятности эмпирического и теоретического распределений.

5 Диаграммы диапазона значений

На диаграммах диапазонов (Range Plots) отображают диапазоны изменения наблюдаемой величины, относящиеся к определенной точке. Чаще всего по оси X откладываются некоторые промежутки времени. Например, с помощью такой диаграммы могут быть представлены интервалы колебаний температуры для каждого дня в течение недели или месяца. Такие диаграммы могут иметь вид линейных отрезков (Whiskers), прямоугольников (Boxes), столбцов (Columns) или прямоугольных скобок (High-Low Close).

Конечный вид диаграммы диапазонов можно выбрать в окне 2D Range Plots (Двумерные диаграммы размаха) рисунок А.3.

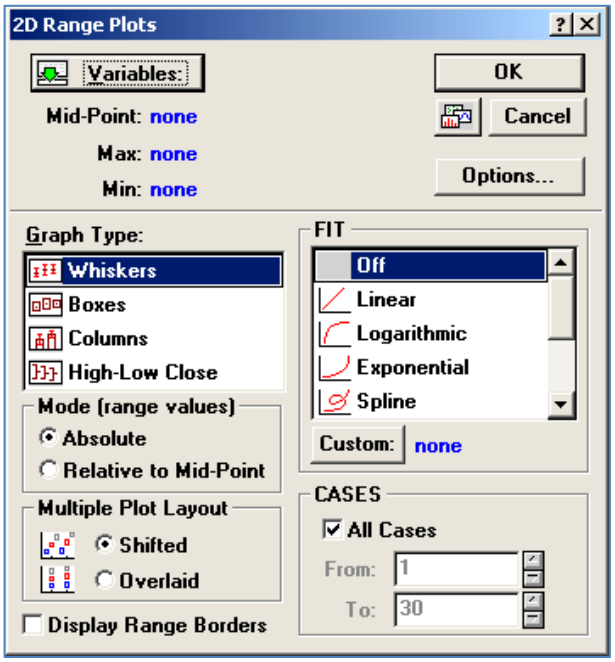


Рисунок А.3 – Окно настройки параметров двумерных диаграмм диапазонов

В окне имеется возможность задать вид отображения диапазонов – абсолютные диапазоны (Absolute) или диапазоны относительно средней точки (Relative to Mid-Point); вид множественных графиков (Multiple Plot Layout) – со сдвигом (Shifted) или с перекрытием (Overlaid); тип сглаживающей линии (Fit); отображаемые случаи (Cases).

6 Диаграммы размаха

Диаграммы размаха (Box Plots) называют так же диаграммами типа прямоугольники-отрезки или типа ящики-усы. На таких диаграммах отображают диапазоны значений выбранной переменной (или нескольких переменных). Отличие от предыдущего типа диаграмм заключается в том, что часть данных для построения диаграмм размаха (медиана, среднее значение, квартили и т.д.) вычисляется по значениям переменной, в то время как диаграммы диапазонов значений строятся непосредственно по наблюдаемым точкам.

Различают:

– Простые (Regular) диаграммы размаха;
– Составные (Multiple) диаграммы размаха.

Имеется возможность построить следующие типы диаграмм размаха (Graph Type):

– Прямоугольники-отрезки (Box-Whiskers);
– Прямоугольники (Boxes);
– Отрезки (Whiskers);
– Столбцы (Columns).

Внешний вид диаграммы в соответствии с выбранными типом и параметрами отображается в центре окна Двумерные диаграммы размаха (2D Box Plots) (рисунок А.4).

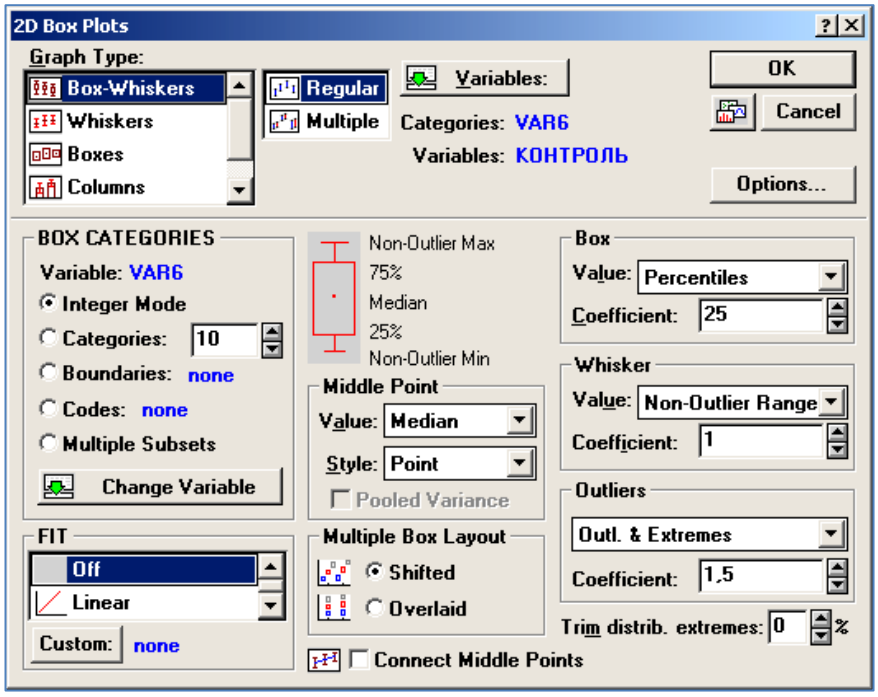


Рисунок А.4 – Окно настройки параметров двумерных диаграмм размаха

В данном окне выбирается Тип группировки (Box Categories) (по аналогии с группировкой для гистограмм); тип подгонки (Fit); расположение множественных графиков (Multuple Box Layout) – со сдвигом (Shifted) или с перекрытием (Overlaid). Имеется возможность настраивать параметры Средних

точек (Middle Point), Прямоугольников (Box), Отрезков (Whisker) и Выбросов (Outliers).

В качестве Средней точки (Middle Point) может использоваться Среднее значение (Mean) или Медиана (Median) в виде Точки (Point) или Линии (Line).

Прямоугольник (Box) может отображать значения (Value) Процентилей (Percentiles), Размах вариации (Min-Max) и Постоянные (Constant) с учетом коэффициента (Coefficient).

Отрезки (Whisker) могут отображать значения Процентилей (Percentiles), Диапазон без выбросов (Non-Outlier Range), Размах вариации (Min-Max) и Постоянные (Constant) с учетом коэффициета (Coefficient).

Группа опций Выбросы (Outliers) служит для отображения Выбросов (Outliers), Критических точек (Extremes), Выбросов и критических точек (Outl. & Extremes). Выбросы и критические точки могут не отображаться на диаграммах размаха (значение Off).

При выборе различных типов диаграмм размаха часть опций диалогового окна 2D Box Plots становится неактивной.

7 Столбчатые диаграммы

Столбчатые диаграммы (Bar/Column Plots) отображают последовательности значений переменной в виде столбцов. Отличие от гистограмм заключается в том, что в виде столбца отображается каждое значение (случай) переменной, в то время как на гистограмме каждый столбец представляет группу значений.

Различают Простые (Regular) и Составные (Multiple) столбчатые диаграммы. В окне настройки можно так же выбрать Ориентацию диаграммы (Orientation) и отображаемые случаи (Cases).

8 Линейные графики (переменных)

Линейные графики (переменных) (Line Plots (Variables)) представляют собой изображения значений переменной для каждого случая в виде точек, соединенных прямыми линиями. По оси X откладываются имена случаев, по оси Y – значения переменной.

Различают Линейные графики (переменных) следующих типов:

- Простые (Regular);
- Составные (Multiple);
- С двойной осью Y (Double-Y);
- Трассировочные XY-графики (XY Trace);
- Агрегированные (Aggregated).

Построение первых трех типов графиков подобно соответствующим типам гистограмм и диаграмм рассеяния.

На Трассировочных XY-графиках отображается диаграмма рассеяния двух переменных X и Y, точки которой соединяются линиями в порядке следования случаев.

Агрегированные графики изображают последовательности средних значений для последовательных подмножеств переменной. При построении данного графика вся совокупность значений переменных разбивается на подгруппы с указанным числом случаев (Aggr. Index), для которых вычисляются средние значения. На графике можно отобразить диапазон значений для каждой подгруппы (Value Range Bars).

В окне выбора типа и настроек Линейных графиков (переменных) можно дополнительно указать Тип координат (Style), Тип сглаживания (Fit), отображаемые Случаи (Cases), вид отображения точек (Display Points).

9 Линейные графики (профили наблюдений)

Линейные графики (профили наблюдений) (Line Plots (Cases Profiles)) аналогичны предыдущим графикам и отличаются от них тем, что по оси X откладываются имена переменных, а по оси Y – значения переменных для каждого из случаев. Одноименные наблюдения (случаи) различных переменных соединяются линиями, образуя профили.

Различают Простые (Regular) и Составные (Multiple) Линейные графики (профили наблюдений). Настройки вида графиков аналогичны предыдущим.

10 Последовательные/наложенные графики

Последовательные/наложенные графики (Sequential/Stacked Plots) используются для визуализации последовательностей значений нескольких переменных. При выборе одной переменной эти графики сходны с Линейными графиками (переменных). Однако, значения каждой последующей выбранной переменной складываются со значениями предыдущей переменной для каждого случая. Таким образом, линии графиков каждой переменной не пересекаются, если значения переменной больше ноля.

Существуют следующие типы Последовательных/ наложенных графиков:

– Линейный (Line) – значения переменных соединяются линиями;

– Зонный (Area) – зоны под линиями заштриховываются;

– Смешанный линейный (Mixed Line) – зоны под линиями переменных из первого списка заштриховываются, а из второго – нет;

– Ступенчатый (Step) – последовательности значений представляются ступенчатыми линиями;

– Ступенчатый зонный (Step Area) – зоны под ступенчатыми линиями заштриховываются;

– Смешанный ступенчатый (Mixed Step) – переменные из первого списка отображаются в виде зон, переменные из второго списка – в виде линий;

– Столбчатая диаграмма (Column) – последовательности значений переменных отображаются в виде сегментов вертикальных столбцов.

11 Круговые диаграммы

Круговые диаграммы (Pie Charts) служат для представления пропорций (долей) от общего числа.

Имеется возможность построить круговые диаграммы следующих типов:

– Круговые диаграммы частот (Pie Charts-Count);
– Круговые диаграммы значений (Pie Charts-Values);
– Многоцветные столбчатые диаграммы.

При настройке параметров диаграммы можно указать способ группировки (Pie Categories) – только для диаграмм частот; задать случаи (Cases); вид легенды (Pie Legend) – проценты (Percent), значения (Values), текстовые метки (Text Labels) и их комбинации; стиль (Style) – двумерный (2D), трехмерный (3D); фигуру (Shape) – круг (Circle) или эллипс (Ellipse).

12 Графики функций, заданных пользователем

График функции, заданной пользователем (Custom Function Plot) позволяет построить зависимость одной переменной от другой. Для этого нет необходимости набирать значения переменной, достаточно ввести функцию (Custom Function) и указать пределы изменения переменной X (X Min и X Max).

Следует отметить, что имеется возможность наложить график функции пользователя на все графики, где имеется опция сглаживания (Fit). Часто используются графики пользовательских функций совместно с диаграммами рассеяния для проверки соответствия теоретических положений или полученных зависимостей эмпирическим (опытным) данным.

ПРИЛОЖЕНИЕ Б

Значения критерия Стьюдента, t на 5, 1 и 0,1%-ном уровне значимости

Число степеней свободы	Уровень значимости		
	0,05	0,01	0,001
1	12,71	63,66	-
2	4,30	9,93	31,60
3	3,18	5,84	12,94
4	2,78	4,60	8,61
5	2,57	4,03	6,86
6	2,45	3,71	5,96
7	2,37	3,50	5,41
8	2,31	3,36	5,04
9	2,26	3,25	4,78
10	2,23	3,17	4,59
11	2,20	3,11	4,44
12	2,18	3,06	4,32
13	2,16	3,01	4,22
14	2,15	2,98	4,14
15	2,13	2,95	4,07
16	2,12	2,92	4,02
17	2,11	2,90	3,97
18	2,10	2,88	3,92
19	2,09	2,86	3,88
20	2,09	2,85	3,85
21	2,08	2,83	3,82
22	2,07	2,82	3,79
23	2,07	2,81	3,77
24	2,06	2,80	3,75
25	2,06	2,79	3,73
26	2,06	2,78	3,71
27	2,05	2,77	3,69
28	2,05	2,76	3,67
29	2,05	2,76	3,66
30	2,04	2,75	3,65
50	2,01	2,68	3,50
100	1,98	2,63	3,39
∞	1,96	2,58	3,29

ПРИЛОЖЕНИЕ В

Значения критерия Фишера, F на 5%-ном уровне значимости (вероятность 95%)

Степень свободы для меньшей дисперсии (знаменателя)	Степени свободы для большей дисперсии (числителя)													
	1	2	3	4	5	6	7	8	9	10	12	24	50	100
1	161	200	216	225	230	234	237	239	241	242	244	249	252	253
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,45	19,47	19,49
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,64	8,58	8,56
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,77	5,70	5,66
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68	4,53	4,44	4,40
6	5,99	5,14	4,76	4,53	4,39	4,27	4,21	4,15	4,10	4,06	4,00	3,84	3,75	3,71
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57	3,41	3,32	3,28
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,12	3,03	2,98
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,90	2,80	2,76
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91	2,74	2,64	2,59
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,61	2,50	2,45
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,69	2,50	2,40	2,35
13	4,64	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,42	2,32	2,26
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,35	2,24	2,19
15	4,54	3,60	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,29	2,18	2,12
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,24	2,13	2,07
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,38	2,19	2,08	2,02
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,15	2,04	1,98
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,11	2,00	1,94
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,08	1,96	1,90
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,05	1,93	1,87
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,23	2,03	1,91	1,84
23	4,28	3,42	3,03	2,80	2,64	2,53	2,45	2,38	2,32	2,28	2,20	2,00	1,88	1,82
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	1,98	1,86	1,80
25	4,24	3,38	2,99	2,76	2,60	2,49	2,41	2,34	2,25	2,24	2,16	1,96	1,84	1,77
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	1,95	1,82	1,76
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,12	1,91	1,78	1,72
30	4,17	3,32	2,92	2,69	2,53	2,42	2,34	2,27	2,21	2,12	2,09	1,89	1,76	1,69
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,79	1,66	1,59
50	4,03	3,18	2,79	2,56	2,40	2,29	2,20	2,13	2,07	2,02	1,95	1,74	1,60	1,52
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,63	1,48	1,39

ПРИЛОЖЕНИЕ Г

Значения критерия χ^2

Число степеней свободы	Уровень значимости							
	0,99	0,95	0,75	0,50	0,25	0,10	0,05	0,01
1	-	-	0,10	0,45	1,32	2,71	3,84	6,63
2	0,02	0,10	0,58	1,39	2,77	4,61	5,99	9,21
3	0,11	0,35	1,21	2,37	4,11	6,25	7,81	11,34
4	0,30	0,71	1,92	3,36	5,39	7,78	9,49	13,28
5	0,55	1,15	2,67	4,35	6,63	9,24	11,07	15,09
6	0,87	1,64	3,45	5,35	7,84	10,64	12,59	16,81
7	1,24	2,17	4,25	6,35	9,04	12,02	14,07	18,48
8	1,65	2,73	5,07	7,34	10,22	13,36	15,51	20,09
9	2,09	3,33	5,90	8,34	11,39	14,68	16,92	21,67
10	2,56	3,94	6,74	9,34	12,55	15,99	18,31	23,21
11	3,05	4,57	7,58	10,34	13,70	17,28	19,68	24,72
12	3,57	5,23	8,44	11,34	14,85	18,55	21,03	26,22
13	4,11	5,89	9,30	12,34	15,98	19,81	22,36	27,69
14	4,66	6,57	10,17	13,34	17,12	21,06	23,68	29,14
15	5,23	7,26	11,04	14,34	18,25	22,31	25,00	30,58
16	5,81	7,96	11,91	15,34	19,37	23,54	26,30	32,00
17	6,41	8,67	12,79	16,34	20,49	24,77	27,59	33,41
18	7,01	9,39	13,68	17,34	21,60	25,99	28,87	34,81
19	7,63	10,12	14,56	18,34	22,72	27,20	30,14	36,19
20	8,26	10,85	15,45	19,34	23,83	28,41	31,41	37,57
21	8,90	11,59	16,34	20,34	24,93	29,62	32,67	38,93
22	9,54	12,34	17,24	21,34	26,04	30,81	33,92	40,29
23	10,20	13,09	18,14	22,34	27,14	32,01	35,17	41,64
24	10,86	13,85	19,04	23,34	28,24	33,20	36,42	42,98
25	11,52	14,61	19,94	24,34	29,34	34,38	37,65	44,31
26	12,20	15,38	20,84	25,34	30,43	35,56	38,89	45,64
27	12,88	16,15	21,75	26,34	31,53	36,74	40,11	46,93
28	13,56	16,93	22,66	27,34	32,62	37,92	41,34	48,28
29	14,26	17,71	23,57	28,34	33,71	39,09	42,56	49,59
30	14,95	18,49	24,48	29,34	34,80	40,26	43,77	50,89
40	22,16	26,51	33,66	39,34	45,62	51,80	55,76	63,69
50	29,71	34,76	42,94	49,33	56,33	63,17	67,50	76,15
60	37,48	43,19	52,29	59,33	66,98	74,40	79,08	88,38
70	45,44	51,74	61,70	69,33	77,58	85,53	90,53	100,42
80	53,54	60,39	71,14	79,33	88,13	96,58	101,88	112,33
90	61,75	69,13	80,62	89,33	98,64	107,56	113,14	124,12
100	70,06	77,93	90,13	99,33	109,14	118,50	124,34	135,81

Учебное издание

Грачева Наталья Николаевна
кандидат технических наук, доцент

Руденко Нелли Борисовна
кандидат технических наук, доцент

Литвинов Владимир Николаевич
кандидат технических наук, доцент

Украинцев Максим Михайлович
кандидат технических наук, доцент

СПЕЦИАЛИЗИРОВАННОЕ ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ НАУЧНЫХ ИССЛЕДОВАНИЙ

Практикум

Издается в авторской редакции

Объем ЭИ: 8,7 Мб
Формат ЭИ: Portable Document Format (PDF)

Вернуться к содержанию