

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Оренбургский государственный университет»

Кафедра вычислительной техники и защиты информации

Ю.И. Сеницын

ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДААННЫХ НА ЭВМ

Методические указания

Рекомендовано к изданию редакционно-издательским советом федерального государственного бюджетного образовательного учреждения высшего образования «Оренбургский государственный университет» для обучающихся по образовательной программе высшего образования по направлению подготовки 10.03.01 Информационная безопасность

Оренбург
2019

УДК 004.42(076.5)
ББК 32.971.3я7
С 38

Рецензент – доцент, кандидат педагогических наук Е. И. Ряполова

Синицын, Ю.И.
С 38 Обработка экспериментальных данных на ЭВМ: методические указания /
Ю. И. Синицын; Оренбургский гос. ун-т. - Оренбург: ОГУ, 2019.

Методические указания предназначены для проведения лабораторных работ студентами, изучающими дисциплину «Обработка экспериментальных данных на ЭВМ». Предложены лабораторные работы по обработке данных с использованием MS Excel и пакета прикладных программ Statistica.

Методические указания предназначены для студентов направления подготовки 10.03.01 Информационная безопасность.

УДК 004.42(076.5)
ББК 32.971.3я7

© Синицын Ю.И., 2019
© ОГУ, 2019

Содержание

1 Общая часть	4
1.1 Общие сведения.....	4
1.2 Использование MS Excel для обработки данных.....	6
2 Лабораторные работы	11
2.1 Лабораторная работа № 1. Использование функций для обработки данных в среде MS Excel.....	11
2.2 Лабораторная работа № 2. Решение систем линейных уравнений в MS Excel	19
2.3 Лабораторная работа № 3. Обработка данных средствами электронных таблиц MS Excel. Применение итоговых функций	23
2.4 Лабораторная работа № 4. Создание сводной таблицы в MS Excel	28
2.5 Лабораторная работа № 5. Расчеты коэффициента корреляции и уравнения регрессии в MS Excel	31
2.6 Лабораторная работа № 6. Аппроксимация экспериментальных данных в MS Excel	34
2.7 Лабораторная работа № 7. Использование инструментов Пакета анализа для статистической обработки данных в MS Excel	41
2.8 Лабораторная работа № 8. Расчет описательных статистик с помощью электронных таблиц MS Excel	46
2.9 Лабораторная работа № 9. Многомерная обработка данных с использованием интегрированной системы Statistica	56
2.10 Лабораторная работа № 10. Компонентный анализ	62
2.11 Лабораторная работа № 11. Работа со списками в MS Excel	66
Список использованных источников	77

1 Общая часть

1.1 Общие сведения

Для сбора и обработки экспериментальных данных в настоящее время широко используется компьютерная техника. Ее использование позволяет увеличить скорость обработки исходной экспериментальной информации, выявить закономерности. В начальный период развития компьютерной техники ее использование требовало обязательного знания основ программирования и самостоятельного составления соответствующих программ, что существенно ограничивало широкое использование вычислительной техники при обработке экспериментальных данных в естественных науках [5].

Широкое применение методов обработки экспериментальной информации связано:

- с математизацией биологических, социологических и других наук;
- с совершенствованием техники экспериментальных исследований;
- с широким применением ЭВМ, в т. ч. встроенных в измерительную аппаратуру и снабженных математическим обеспечением для обработки данного класса экспериментов;
- с повышением требований к качеству данных, диктуемым развитием теории.

Целью обработки экспериментальных данных является выявление закономерностей в характеристиках исследуемых объектов и процессов. Результаты обработки экспериментальных данных позволяют оценить качество объекта, они необходимы для оперативного управления процессами, решения задач адаптации объекта к изменившимся условиям или формирования требований ко вновь создаваемым системам [2].

Объект исследования – это объект любого характера (технического, социального, экономического, астрономического и т.д. и т.п.), который изучается экспериментальным путем.

Эксперимент – это специальным образом спланированная и организованная **процедура изучения** некоторого объекта исследования, при которой на этот объект оказывают запланированные воздействия и регистрируют его реакции на эти воздействия [6].

Факторы – это воздействия на объект. Мы будем обозначать факторы величинами x_1, x_2, \dots, x_k .

Откликами объекта исследования называют его реакции на воздействия; будем обозначать их символом y_g .

Эксперимент состоит из ряда **опытов** (или **наблюдений**), при которых каждый из факторов x_1, x_2, \dots, x_k имеет разное значение.

Экспериментальные данные – все исходные и выходные числовые данные эксперимента, сведенные в **таблицу экспериментальных данных**.

Обработка экспериментальных данных – различные методы построения математической модели объекта по **таблице экспериментальных данных**.

Регрессионный анализ – наиболее распространенный метод обработки данных, который включает в себя метод наименьших квадратов. При регрессионном анализе таблица экспериментальных данных обычно отражается алгебраическими степенными полиномами, которые называют **полиномами или уравнениями регрессии**. Отсюда термины – **задача регрессии, коэффициенты регрессии** и т.п. Сам термин **регрессия** отражает тот факт, что с увеличением степени полинома точность отражения **таблицы экспериментальных данных** обычно возрастает, а ошибка отражения соответственно уменьшается, **регрессирует**.

Управляемые факторы – это такие воздействия на объект исследования, численные значения которых определяются и контролируются самим экспериментатором.

Активный эксперимент – это эксперимент, в котором задействованы только управляемые факторы. Пример – изучение зависимости урожайности какой-либо сельскохозяйственной культуры от объемов орошения. Эти объемы для различных экспериментальных полей посева назначаются самим исследователем.

Контролируемые факторы – это такие воздействия на объект исследования,

численные значения которых экспериментатором не устанавливаются, но значения их исследователь может измерять, контролировать и фиксировать.

Пассивный эксперимент – это эксперимент, в котором задействованы только контролируемые факторы. Пример – изучение зависимости урожайности сельскохозяйственной культуры от объемов атмосферных осадков, которыми экспериментатор управлять не может.

Активно-пассивный (или пассивно-активный) эксперимент – это совмещение обоих видов эксперимента, когда зависимость урожайности изучается от совместного объема и орошения и атмосферных осадков.

Основным «рабочим инструментом» и эксперимента и обработки экспериментальных данных является численное значение факторов воздействия и откликов объекта исследования, т.е. **число**. Какова ни была бы природа факторов и откликов, включая, в том числе, эмоции или впечатления, они должны быть выражены количественно, числом.

Числа при экспериментировании получают тремя способами:

- подсчетом;
- измерением;
- методом экспертных оценок [6].

1.2 Использование MS Excel для обработки данных

Microsoft Excel входит в состав пакета MS Office и является одним из самых популярных сегодня табличных процессоров - программ для работы с электронными таблицами. В отличие от текстового процессора Word, предназначенного для оформления текстовых документов, Excel специализирован для выполнения вычислений с табличными данными. Excel имеет большое количество встроенных функций для математических, статистических, финансовых и других вычислений. С другой стороны, Excel - это среда, ориентированная на непрограммирующего пользователя, что делает его популярным среди экономистов, бухгалтеров и других специалистов, обрабатывающих табличные данные.

Документ Excel называется рабочей книгой, состоящей из набора рабочих листов. Книга хранится в виде файла с расширением *.xls. Одна книга может содержать до 256 рабочих листов.

Excel имеет стандартный интерфейс Windows, показанный на рисунке 1. Поэтому рассмотрим только те элементы рабочего окна, которые отличают Excel от Word [7].

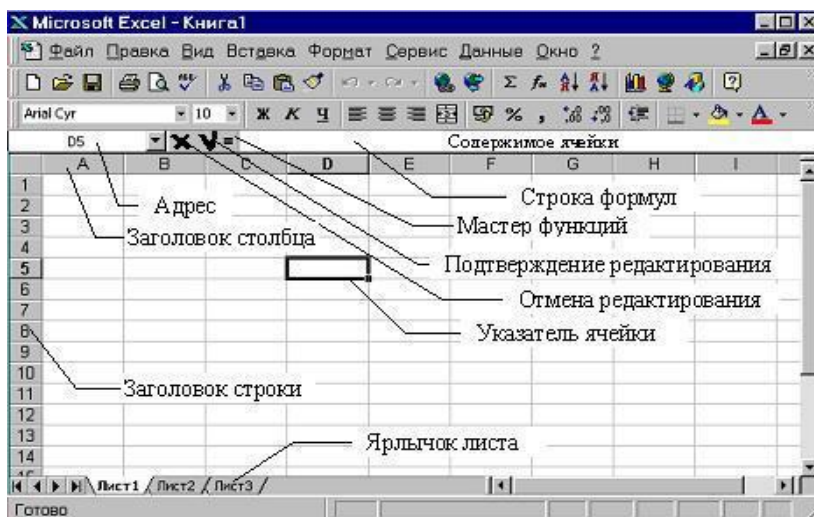


Рисунок 1 - Элементы рабочего окна MS Excel

Строка меню отличается от меню в Word заменой пункта «Таблица» на пункт «Данные», предназначенный для выполнения операций сортировки, фильтрации и некоторых других над табличными данными.

На панели форматирования следует отметить кнопку «Объединить и поместить в центре», а также «Денежный формат» и «Процентный формат».

Строка формул состоит из поля адреса, управляющих кнопок и поля содержимого ячейки. При активизации ячейки таблицы в этих полях появляется соответствующая информация. Управляющие кнопки выполняют: отмену редактирования ячейки (кнопка с крестом); принятие редактирования (кнопка с галочкой - соответствует нажатию Enter); вызов «Мастера функций» (кнопка "="). Редактировать содержимое ячейки можно непосредственно в самой ячейке или в поле содержимого ячейки. В первом случае надо выполнить двойной щелчок по

ячейке или нажать клавишу **F2**, во втором - активизировать ячейку и щелкнуть в поле содержимого строки формул [7].

Столбцы и строки таблицы имеют заголовки. Для столбцов заголовки - буквы латинского алфавита; для строк - целые числа. При большом количестве столбцов их заголовки состоят из двух латинских букв, например: **AD**, **BF**. Всего на рабочем листе можно разместить 256 столбцов и 65536 строк.

Указатель ячейки — рамка, выделяющая ячейку.

Ячейки Excel. Одним из центральных понятий в Excel является ссылка на ячейку. Ссылки используются для записи формул. По форме записи ссылки могут быть двух типов: **A1** и **R1C1**. В типе **A1** первым указывается заголовок столбца, вторым — заголовок строки. Например: **D5**; **G24**; **AF13**. В типе **R1C1** первой указывается строка с префиксом R, а вторым — числовой номер столбца с префиксом C, например, **R12C5**. По своим свойствам ссылки могут быть абсолютными и относительными. Во втором случае говорят просто "ссылка". Относительные ссылки обладают свойством автоматической коррекции координат ячейки при переносе (копировании) формулы в другую ячейку. Абсолютные ссылки таким свойством не обладают. В приведенных выше примерах все ссылки — относительные. Для задания абсолютной ссылки в типе A1 добавляется знак "\$" перед заголовком столбца и (или) перед заголовком строки, например: **\$A\$10**; **D\$25**; **\$AF16**. Для типа **R1C1** номер строки и (или) столбца при задании абсолютной ссылки заключается в квадратные скобки, например, для приведенных выше примеров можно записать: **R[10]C[1]**; **R[25]C4**; **R16C[34]**.

Допускаются ссылки на другой лист, другую книгу (внешние ссылки) и другое приложение (удаленные ссылки).

Ячейка таблицы, на которой стоит указатель ячейки является активной. Для активизации необходимо щелкнуть кнопкой мыши по ячейке [7].

Группа ячеек может быть выделена для выполнения различных операций. Если выделяемая область таблицы имеет прямоугольную форму, то технология ее выделения обычная - либо мышью с удержанием левой кнопки, либо клавиатурой - при нажатой клавише **Shift**. Если выделяются несмежные ячейки таблицы, то

сначала следует выделить первую область, а затем, удерживая **Ctrl**, мышью выделить другие.

При необходимости ссылки на прямоугольную область таблицы говорят о диапазоне ячеек. В общем случае диапазон записывается как две ссылки на верхнюю левую и нижнюю правую ячейки прямоугольной области, разделенные двоеточием, например **D5:F13**. Диапазоном может быть частичный столбец (например, **R4:R25**) или частичная строка (например, **A12:Z12**). Если двоеточием разделены только номера строк или столбцов (например, **A:F** или **5:13**), то в диапазон входят все ячейки таблицы, ограниченные этими строками или столбцами.

Для ввода данных в ячейку, ее надо активизировать, набрать данные с клавиатуры и нажать **Enter** или щелкнуть мышью в другой ячейке.

Данные, вводимые в ячейку, могут быть текстом, формулой или ссылкой. Excel может определять тип данных автоматически. Если введен текст, он обычно выравнивается по левому краю ячейки. Если длина текста превышает ширину ячейки, но ячейка справа пуста, текст на экране будет занимать эту ячейку. Если же ячейка справа занята, то на экране текст ограничивается размером ячейки. Фактически же ячейка содержит полный текст, в чем можно убедиться, просмотрев ее содержимое в строке формул.

Вводимые числа выравниваются по правому краю ячейки. Для разделения целой и дробной частей десятичных чисел используется запятая. Если целая часть числа не умещается по ширине ячейки, на экран выводятся знаки "#####".

Запись формул и ссылок на другие ячейки начинается со знака "=", за которым следует выражение или ссылка. Если выражение синтаксически правильно, то после завершения ввода формулы в ячейке размещается результат вычисления, иначе - сообщение типа "#ИМЯ?".

Достоинство использования ссылок и формул со ссылками в том, что при изменении данных в исходных ячейках, результат в ячейке с формулой будет скорректирован автоматически [7].

Функции в Excel. Для выполнения вычислений в таблицах нужны формулы. Поскольку некоторые формулы и их комбинации встречаются очень часто, то

программа Excel предлагает более 200 заранее запрограммированных формул, которые называются функциями. В программе имеется упорядоченный по алфавиту полный список всех функций, в котором можно легко найти функцию, если известно ее имя; в противном случае следует производить поиск по категориям (все функции разделены по категориям, чтобы в них было проще ориентироваться). Многие функции различаются очень незначительно, поэтому при поиске по категориям полезно воспользоваться краткими описаниями функций, которые предлагает Конструктор функций. Функция оперирует некоторыми данными, которые называются ее аргументами. Аргумент функции может занимать одну ячейку или размещаться в целой группе ячеек. Конструктор функций оказывает помощь в задании любых типов аргументов, он помогает на всех этапах работы правильно применять функции.

Выделите ту ячейку, в которой должен появиться результат вычислений. Затем нажмите мышкой на пиктограмму Конструктора функций со значком f_x и откройте диалоговое окно Конструктора «Insert Function».

В поле «Or select a category» этого окна перечислены категории функций:

Most Recently Used – функции, которые недавно использовались, это функции, которые показываются в нижнем окне «Select a function:» сразу же при открытии диалогового окна Конструктора «Insert Function»:

- All – список всех функций по алфавиту;
- Financial – финансовые функции;
- Date & Time – функции даты и времени;
- Math & Trig – математические и тригонометрические функции;
- Statistical – статистические функции;
- Lookup & Reference – функции ссылок и сносок;
- Database - функции баз данных;
- Text – текстовые функции;
- Logical – логические функции;
- Information – информационные функции;
- User Defined – функции, определенные пользователем.

При выборе категории функции в поле «Or select a category» в нижнем поле «Select a function:» появляются функции, соответствующие выбранной категории.

Построения графиков с помощью MS Excel.

На рисунке 2 показано получение графика координаты от номера ячейки X в столбце (красный график) и времени t от номера ячейки [7].

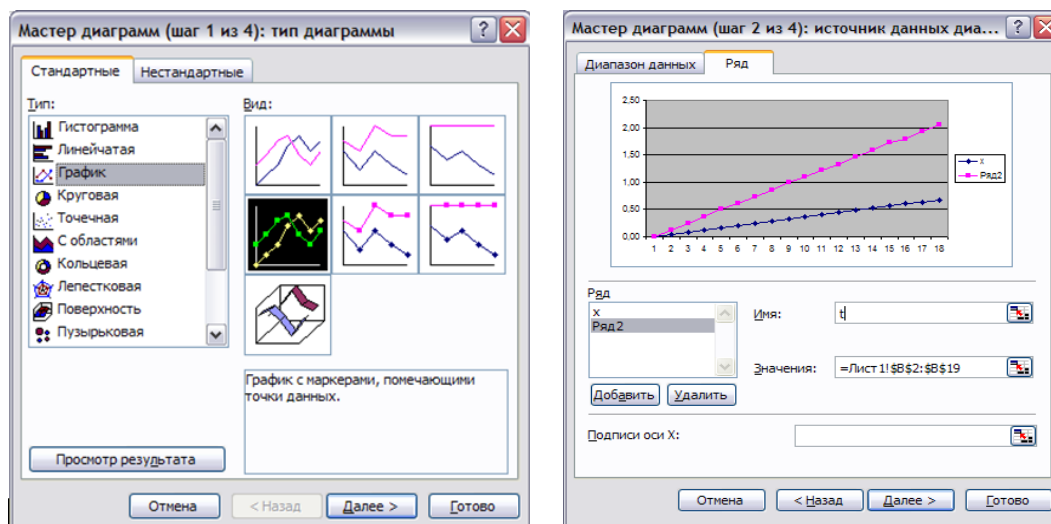


Рисунок 2 – Примеры графиков

Для построения графиков при обработке экспериментальных результатов следует выбирать в Мастере диаграмм опцию **Точечная** и только ее.

2 Лабораторные работы

2.1 Лабораторная работа № 1. Использование функций для обработки данных в среде MS Excel

Цели работы:

- 1 Выполнить сортировку данных таблицы по возрастанию кода предмета, даты проведения занятия, номера группы.
- 2 Ознакомиться со способом фильтрации записей списка, автофильтрации, работой с формой данных.

3 Закрепить навыки использования логических функций в расчетах.

Задание на выполнение лабораторной работы.

Задание №1. Сортировка данных в списке:

1 Выполнить сортировку данных таблицы 1 по возрастанию кода предмета, даты проведения занятия, номера группы.

2 Выполнить сортировку данных таблицы 1 по возрастанию, используя сочетания признаков: код предмета и дату проведения занятия; код предмета и номер группы; номер группы и дату проведения занятия, а также сочетание всех трех признаков [7].

Порядок выполнения работы:

1 Создайте новую рабочую книгу (меню *Файл* команда *Создать*) и сохраните ее под именем Sort.xls в рабочем каталоге (меню *Файл* команда *Сохранить как*).

2 Сформируйте таблицу результатов занятий, как показанную в таблице 1.

Таблица 1 – Результаты занятий

А	Б	С	Д	Е	Ф	Г	Н
1	№ группы	№ зачетной книжки	Код предмета	Табл. Препод.	Вид занятия	Дата	Оценка
2	3-3230	3-3230/03	Ш	А1	Практика	26.05.17	3
3	3-3230	3-3230/12	П2	А2	Лекиия	26.05.17	4
4	3-3230	3-3230/06	Ш	А1	Лекиия	11.06.17	4
5	3-3330	3-3230/0S	Ш	А2	Лекиия	11.06.17	5
6	3-3330	3-3230/1S	П2	А1	Практика	16.05.17	2
7	3-6230	3-6230/03	П2	А3	Лекиия	20.05.17	3
8	3-3230	3-3230/09	Ш	А1	Лекция	16.05.17	3
9	3-3230	3-3230/1S	Ш	А3	Лекция	16.05.17	4
10	3-3330	3-3330/03	Ш	А2	Лекция	26.05.17	4
11	3-3531	3-3531/02	П2	А1	Лекция	11.06.17	2
12	3-3532	3-3532/03	Ш	А2	Практика	20.05.17	5
13	3-3230	3-3230/20	П2	А1	Лекция	26.05.17	5

3 Отформатируйте шапку таблицы следующим образом:

- шрифт Times New Roman;
- размер шрифта 12 пт., курсив;

- выравнивание по горизонтали — *По значению*;
- выравнивание по вертикали — *По верхнему краю*;
- установите ключ *“Переносить по словам”* (выделить соответствующие ячейки и выполнить команду *Формат —> Ячейки*).

4 Выполните сортировку по столбцу “Код предмета”, расположив коды предметов по возрастанию. Для этого нужно:

- выделить таблицу с одной строкой заголовка;
- выполнить команду меню *Данные —> Сортировка*;
- в окне *Сортировка диапазона* в строке *Сортировать по* “коду предмета”.

5 Результат сортировки скопируйте на Лист 2:

- выделите всю таблицу, выполнить команду *Правка —> Копировать*;
- затем на Листе 2 установите курсор в ячейку **A1** и выполните команду *Правка —> Вставить*.

6 Переименуйте Лист 2, дав ему имя – *Сортировка*:

- указатель мыши установите на ярлычке Лист 2;
- правой клавишей мыши вызовите контекстное меню;
- выполните команду *Переименовать*.

7 Выполните сортировку по столбцу “Дата”, расположив данные возрастанию. Для этого следует установить курсор в любую ячейку поля “Дата” и ввести команду *Сортировка* из меню *Данные*, при этом должна выделиться вся область списка, а в окне *Сортировка Диапазона* в строке *Сортировать по* – столбец **G**. Если этого не произошло, то предварительно выделите весь список, а затем выполните указанную команду.

8 Выполните сортировку по сочетанию признаков “Дата”, “№ группы”, “Код предмета”. Для этого следует выделить всю таблицу и в диалоговом окне *Сортировка* установить:

- в строке *Сортировать по* — поле “Дата” по возрастанию;
- в строке *Затем* — поле “№ группы”, по возрастанию;
- в следующей строке *Затем* — поле “Код предмета” по возрастанию;

– установите флажок *Строка меток столбцов*.

Результат сортировки скопировать на Лист 3 и переименовать его в *Сортировка 2*.

Задание № 2. Фильтрация записей.

Провести фильтрацию записей списка, автофильтрацию, работу с формой данных [7].

Порядок выполнения работы:

1 Создайте новую рабочую книгу с названием “Фильтрация”.

2 Скопируйте в новую рабочую книгу таблицу, созданную в задании № 1 и показанной в таблице 2.

Таблица 2 – Рабочая книга

А	Б	С	Д	Е	Ф	Г	Н
1	№ группы	№ зачетной книжки	Код предмета	Табл. Препод.	Вид занятия	Дата	Оценка
2	3-3230	3-3230/03	Ш	А1	Практика	26.05.17	3
3	3-3230	3-3230/12	П2	А2	Лекиия	26.05.17	4
4	3-3230	3-3230/06	Ш	А1	Лекиия	11.06.17	4
5	3-3330	3-3230/0S	Ш	А2	Лекиия	11.06.17	5
6	3-3330	3-3230/1S	П2	А1	Практика	16.05.17	2
7	3-6230	3-6230/03	П2	А3	Лекиия	20.05.17	3
8	3-3230	3-3230/09	Ш	А1	Лекция	16.05.17	3
9	3-3230	3-3230/1S	Ш	А3	Лекция	16.05.17	4
10	3-3330	3-3330/03	Ш	А2	Лекция	26.05.17	4
11	3-3531	3-3531/02	П2	А1	Лекция	11.06.17	2
12	3-3532	3-3532/03	Ш	А2	Практика	20.05.17	5
13	3-3230	3-3230/20	П2	А1	Лекция	26.05.17	5

3 Переименуйте Лист1, присвоив ему имя “Автофильтр №1”.

4 Чтобы применить *Автофильтрацию*, установите курсор в область шапки таблицы и выполните команду *Данные—>Фильтр—>Автофильтр*.

5 Сформируйте условия отбора: для преподавателя **A1** выбрать сведения о сдаче экзамена на положительную оценку, вид занятий – Лекция. Для этого выполните следующие действия:

– в столбце *Таб № препод.* нажмите кнопку *Фильтр*, из списка условий отбора выберите **A1**;

– в столбце *Оценка* нажмите кнопку *Фильтр*, из списка условий отбора выберите *Условие* и в диалоговом окне сформируйте условие отбора *>2*;

– в столбце *Вид занятий* нажмите кнопку *Фильтр*, из списка условий отбора выберите *Лекция*.

6 Результат фильтрации скопируйте на новый лист, присвоив ему имя “*Автофильтр №2*”.

7 На листе “*Автофильтр №1*” результат автофильтрации отмените, установив указатель мыши в область списка и выполнив команду *Данные —> Фильтр —> Автофильтр*.

8 Сформулируйте выборку: для группы 3-3230 получите сведения о сдаче экзамена по предмету П2 на оценки 3 и 4.

9 Результат сохраните на новом листе, присвоив ему имя “*Автофильтр №3*”.

10 Скопируйте исходную таблицу на новый рабочий лист, переименовав его в *Форма данных*.

11 Установите курсор в область шапки таблицы и выполните команду *Данные —>Форма*.

12 В окне *Форма данных* просмотрите записи списка и внесите необходимые изменения по своему усмотрению с помощью кнопок *<Предыдущая>* и *<Следующая>*.

13 С помощью кнопки *<Создать>* добавьте новые записи.

14 В окне *Форма данных* сформируйте условия отбора записей.

Для этого нажмите кнопку *<Критерии>*, название которой поменяется на *<Правка>*. В пустых строках имен полей списка введите критерии:

- в строку *Табл. № препод.* введите **A1**;
- в строку *вид занятия* введите *Лекция*;
- в строку *оценка* введите условие > 2 .

15 Просмотрите отобранные записи нажатием на кнопку *<Предыдущая>* или *<Следующая>*.

16 По аналогии сформулируйте условия отбора записей, указанные в п. 8.

Задание № 3. Использование логических функций:

1 Подсчитайте количество отличных, хороших и т. д. оценок на основании зачетной ведомости, представленной в таблице 1 [7].

2 Произведите расчет, используя операцию “Присвоение имени блоку ячеек”.

Порядок выполнения работы:

1 На новом листе рабочей книги создайте таблицу по образцу таблицы 3.

2 Заполните данными столбцы **A, B, C, D**.

Таблица 3 – Зачетная ведомость

A	B	C	D	E	F	G	H	I
№ п/п	Фам.. имя. отчество	№ зач. книжки	Оценка	Кол- во	Кол- во	Кол- во	Кол- во	Неявка
1	Демидов М.И.	3-3230/04	5					
2	Иванов И. П.	3-3230/05	4					
3	Кукушкин В. Л.	3-3230/07	3					
4	Орлов А. П.	3-3230/11	4					
5	Петров К.Н.	3-3230/13	5					
6	Сидоров В.О.	3-3230/15	2					
7	Фролов В А.	3-3230/18	0					

3 В столбцы **E, F, G, H, I** введите формулы, для этого:

- установите курсор в первую ячейку столбца количества отличных оценок (**E2**) и выполните *Вставка* → *Функция* → категория *Логические* → функция =

ЕСЛИ, убрать появившееся диалоговое окно *Аргументы функции* с таблицы 3.7 (если оно ее закрывает);

– в диалоговом окне *Аргументы функции* установите курсор в поле *Лог.выражение* и щелкните мышью в рабочей области Excel на ячейке **D2**;

– появится адрес ячейки **D2** и с клавиатуры введите ≤ 5 , т.е. сформируется логическое выражение **D2=5**;

– в поле *Значение_если_истина* введите $\langle 1 \rangle$;

– в поле *Значение_если_ложь* введите $\langle 0 \rangle$;

– щелкните на кнопке **ОК**.

Методом протягивания скопируйте формулу по столбцу **E** - “Кол-во 5”.

4 С помощью **Мастера функций** аналогичным способом введите формулы в столбцы “Кол-во 4”, “Кол-во 3” и т. д., изменяя соответственно значение поля *Логическое выражение* на **D2 = 4**, **D2 = 3**, **D2 = 2**. Для подсчета количества неявившихся на экзамен необходимо задавать логическое выражение **D2=0**;

5 Чтобы подсчитать сумму всех пятерок, четверок и т. д. и результаты представить в виде отдельной таблицы, нужно по каждому столбцу “Кол-во оценок” задать имена блокам соответствующих ячеек. Для этого выполните следующие действия:

– выделите блок ячеек **E2:E8** столбца “Кол-во 5”;

– выполните команду меню *Вставка* —> *Имя* —> *Присвоить*;

– в диалоговом окне *Присвоение имени* в строке *Имя* введите слово *Отлично* и щелкните на кнопке *Добавить и затем* $\langle \text{ОК} \rangle$;

– далее выделите ячейки **F2:F8** столбца “Кол-во 4” и выполните команду *Вставка* —> *Имя* —> *Присвоить*;

– в диалоговом окне *Присвоение имени* в строке *Имя* введите слово *Хорошо* и щелкните на кнопке $\langle \text{Добавить} \rangle$ и затем $\langle \text{ОК} \rangle$;

– аналогичные действия выполните с остальными столбцами табл. 3.7, создав имена блоков ячеек *Удовлетворительно*, *Неудовлетворительно*, *Неявка*.

6 Создайте таблицу *Итоги сессии*, бланк которой показан в таблице 4 [7].

Таблица 4 - Итоги сессии

Итоги сессии	
Количество отличных оценок	
Количество хороших оценок	
Количество удовлетворительных оценок	
Количество неудовлетворительных оценок	
Неявки	
ИТОГО	

7 Введите формулу подсчета количества полученных оценок определенного вида:

– установите курсор в ячейку подсчета количества отличных оценок и выполните *Вставка* → *Функция* → категория *Математические* → функция = **СУММ**;

– щелкните на кнопке **ОК**;

– в диалоговом окне *Аргументы функции* установите курсор в строку *Число1* и выполните *Вставка* → *Имя* → *Вставить*;

– в диалоговом окне *Вставка имени* выберите имя блока ячеек *Отлично* и щелкните на кнопке **ОК**;

– повторите аналогичные действия для подсчета количества других оценок.

8 Подсчитайте **ИТОГО** – количество всех полученных оценок, используя кнопку *Автосумма* на стандартной панели инструментов.

Кнопка *Автосумма* – это символ Σ на стандартной панели инструментов.

Если его нет, необходимо выполнить *Вид* → *Панели инструментов* → поставить флажок на кнопку *Стандартная*.

Для подсчета **ИТОГО**:

– щелкнуть на ячейку, где должен быть результат **ИТОГО**;

– щелкнуть на кнопку *Автосумма*;

– выделить группу ячеек с количеством всех оценок;

– **Enter**.

Содержание отчёта:

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.2 Лабораторная работа № 2. Решение систем линейных уравнений в MS Excel

Цель работы. Изучение возможностей пакета MS Excel при решении задач линейной алгебры. Приобретение навыков решения систем линейных алгебраических уравнений и выполнение действий над матрицами.

Задание на выполнение лабораторной работы.

Решить систему методом обратной матрицы [3]:

$$\begin{cases} x_2 - 13x_3 + 4x_4 = -5 \\ x_1 - 2x_3 + 3x_4 = -4 \\ 3x_1 + 21x_2 - 5x_4 = 2 \\ 4x_1 + 3x_2 - 5x_3 = 5 \end{cases} \quad (1)$$

Порядок выполнения работы.

Введём матрицу **A** и вектор **b** в рабочий лист MS Excel, как показано на рисунке 3.

	A	B	C	D	E	F	G	H
1	A=	0	1	-13	4	b=	-5	
2		1	0	-2	3		-4	
3		3	21	0	-5		2	
4		4	3	-5	0		3	
5								

Рисунок 3 - Ввод основной матрицы системы линейных уравнений и столбца свободных членов

В нашем случае матрица **A** находится в ячейках **B1:E4**, а вектор **b** в диапазоне **G1:G4**. Для решения системы методом обратной матрицы необходимо вычислить

матрицу, обратную к **A**. Для этого выделим ячейки для хранения обратной матрицы и пусть в нашем случае это будут ячейки **B6:E9**. Теперь обратимся к мастеру функций, экранная форма на рисунке 4, и в категории **Математические** выберем функцию **МОБР**, предназначенную для вычисления обратной матрицы, щелкнув по кнопке **ОК**, перейдём ко второму шагу мастера функций. В диалоговом окне, появляющемся на втором шаге мастера функций, необходимо заполнить поле ввода Массив, как на рисунке 5. Это поле должно содержать диапазон ячеек, в котором хранится исходная матрица - в нашем случае **B1:E4**. Данные в поле ввода Массив можно ввести, используя клавиатуру или выделив их на рабочем листе, удерживая левую кнопку мыши [3].

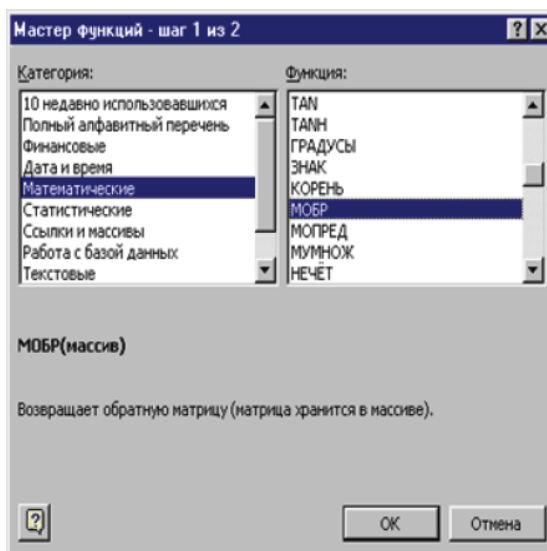


Рисунок 4 - Выбор функции с помощью инструмента «Мастер функций»

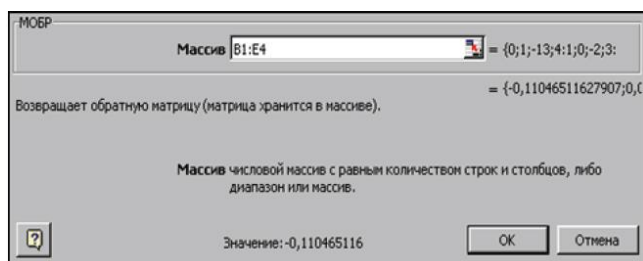


Рисунок 5 - Работа с функцией МОБР

Если поле Массив заполнено, можно нажать кнопку **ОК**. В первой ячейке, выделенного под обратную матрицу диапазона, появится некоторое число. Для того

чтобы получить всю обратную матрицу, необходимо нажать клавишу **F2** для перехода в режим редактирования, а затем одновременно клавиши **Ctrl+Shift+Enter**. В нашем случае рабочая книга MS Excel примет вид, показанный на рисунке 6 [3].

	A	B	C	D	E	F	G
1		0	1	-13	4		-5
2	A=	1	0	-2	3	b=	-4
3		3	21	0	-5		2
4		4	3	-5	0		3
5							
6		-0,11047	0,096899	-0,03023	0,24845		
7		0,011628	0,077519	0,055814	-0,06124		
8		-0,0814	0,124031	0,009302	-0,03798		
9		-0,01744	0,383721	0,016279	-0,10814		

Рисунок 6 - Вычисление обратной матрицы

Теперь необходимо умножить полученную обратную матрицу на вектор **b**. Выделим ячейки для хранения результирующего вектора, например **H6:H9**. Обратимся к мастеру функций, и в категории **Математические** выберем функцию **МУМНОЖ**, которая предназначена для умножения матриц. Напомним, что умножение матриц происходит по правилу строка на столбец и матрицу **A** можно умножить на матрицу **B** только в том случае, если количество столбцов матрицы **A** равно количеству строк матрицы **B**.

Кроме того, при умножении матриц важен порядок сомножителей, т.е. **AB** не равно **BA**.

Перейдём ко второму шагу мастера функций. Появившееся диалоговое окно (рисунок 7) содержит два поля ввода **Массив1** и **Массив2**. В поле **Массив1** необходимо ввести диапазон ячеек, в котором содержится первая из перемножаемых матриц, в нашем случае **B6:E9** (обратная матрица), а в поле **Массив2** ячейки, содержащие вторую матрицу, в нашем случае **G1:G4** (вектор **b**).

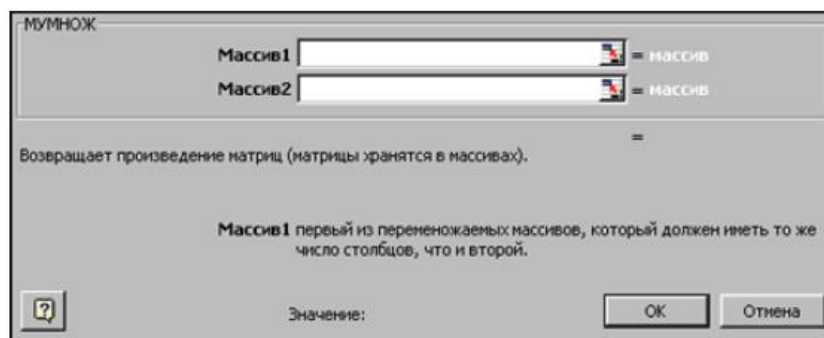


Рисунок 7 - Работа с функцией МУМНОЖ [3]

Если поля ввода заполнены, можно нажать кнопку **ОК**. В первой ячейке выделенного диапазона появится соответствующее число результирующего вектора. Для того чтобы получить весь вектор, необходимо нажать клавишу **F2**, а затем одновременно клавиши **Ctrl+Shift+Enter**. В нашем случае результаты вычислений (вектор x), находится в ячейках **H6:H9**.

Для того чтобы проверить, правильно ли решена система уравнений, необходимо умножить матрицу A на вектор x и получить в результате вектор b . Умножение матрицы A на вектор x осуществляется при помощи функции **МУМНОЖ(B1:E4;H6:H9)**, так как было описанной выше. В результате проведенных вычислений рабочий лист примет вид, изображенный на рисунке 8.

	A	B	C	D	E	F	G	H	I	J
1			0	1	-13	4		-5		-5
2			1	0	-2	3		-4		-4
3	A=		3	21	0	-5	b=	2	Проверка	2
4			4	3	-5	0		3		3
5										
6			-0,11047	0,096899	-0,03023	0,24845		0,849612		
7			0,011628	0,077519	0,055814	-0,06124		-0,44031		
8			-0,0814	0,124031	0,009302	-0,03798		-0,1845		
9			-0,01744	0,383721	0,016279	-0,10814		-1,73953		

Рисунок 8 - Результат решения системы линейных уравнений методом обратной матрицы

Содержание отчёта:

1 Цель работы.

- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.3 Лабораторная работа № 3. Обработка данных средствами электронных таблиц MS Excel. Применение итоговых функций

Цель работы. Ознакомиться с приемами обработки данных с помощью электронных таблиц MS Excel.

Задание на выполнение лабораторной работы.

Задание № 1. Обработка данных.

Порядок выполнения работы:

- 1 Запустите программу *Excel* (*Пуск / Программы / Microsoft Excel*).
- 2 Создайте новую рабочую книгу (кнопка *Создать* на стандартной панели инструментов).
- 3 Дважды щелкните на ярлычке текущего рабочего листа и дайте этому рабочему листу имя *Данные*.
- 4 Дайте команду *Файл / Сохранить как* и сохраните рабочую книгу под именем *book.xls*.
- 5 Сделайте текущей ячейку **A1** и введите в нее заголовок *Результаты измерений*.
- 6 Введите произвольные числа в последовательные ячейки столбца **A**, начиная с ячейки **A2**.
- 7 Введите в ячейку **B1** строку *Удвоенное значение*.
- 8 Введите в ячейку **C1** строку *Квадрат значения*.
- 9 Введите в ячейку **D1** строку *Квадрат следующего числа*.
- 10 Введите в ячейку **B2** формулу $=2*A2$.
- 11 Введите в ячейку **C2** формулу $=A2*A2$.
- 12 Введите в ячейку **D2** формулу $=B2+C2+1$.

13 Выделите протягиванием ячейки **B2**, **C2** и **D2**.

14 Наведите указатель мыши на маркер заполнения в правом нижнем углу рамки, охватывающей выделенный диапазон. Нажмите левую кнопку мыши и перетащите этот маркер, чтобы рамка охватила столько строк в столбцах **B**, **C** и **D**, сколько имеется чисел в столбце **A**.

15 Убедитесь, что формулы автоматически модифицируются так, чтобы работать со значением ячейки в столбце **A** текущей строки.

16 Измените одно из значений в столбце **A** и убедитесь, что соответствующие значения в столбцах **B**, **C** и **D** в этой же строке были автоматически пересчитаны.

17 Введите в ячейку **E1** строку *Масштабный множитель*.

18 Введите в ячейку **E2** число 5.

19 Введите в ячейку **F1** строку *Масштабирование*.

20 Введите в ячейку **F2** формулу $=A2 * E2$.

21 Используйте метод автозаполнения, чтобы скопировать эту формулу в ячейки столбца **F**, соответствующие заполненным ячейкам столбца **A**.

22 Убедитесь, что результат масштабирования оказался неверным так как адрес **E2** в формуле задан относительной ссылкой.

23 Щелкните на ячейке **F2**, затем в строке формул. Установите текстовый курсор на ссылку **E2** и нажмите клавишу **F4**. Убедитесь, что формула теперь выглядит как $=A2 * \$E\2 , и нажмите клавишу **Enter**.

24 Повторите заполнение столбца **F** формулой из ячейки **F2** [3].

25 Убедитесь, что благодаря использованию абсолютной адресации значения ячеек столбца **F** теперь вычисляются правильно. Сохраните рабочую книгу book.xls.

Задание № 2. Применение итоговых функций.

Порядок выполнения работы:

1 Запустите программу *Excel* (*Пуск / Программы / Microsoft Excel*) и откройте рабочую книгу book.xls, созданную ранее [3].

2. Выберите рабочий лист *Данные*.

3 Сделайте текущей первую свободную ячейку в столбце **A**.

4. Щелкните на кнопке **Автосумма** на стандартной панели инструментов.

5 Убедитесь, что программа автоматически подставила в формулу функцию **СУММ** и правильно выбрала диапазон ячеек для суммирования. Нажмите клавишу **Enter**.

6 Сделайте текущей следующую свободную ячейку в столбце **A**.

7 Щелкните на кнопке **Вставка функции** на стандартной панели инструментов.

8 В списке *Категория* выберите пункт *Статистические*.

9 В списке *Функция* выберите функцию **СРЗНАЧ** и щелкните на кнопке **ОК**.

10 Переместите методом перетаскивания палитру формул, если она заслоняет нужные ячейки. Обратите внимание, что автоматически выбранный диапазон включает все ячейки с числовым содержимым, включая и ту, которая содержит сумму. Выделите правильный диапазон методом протягивания и нажмите клавишу **Enter**.

11 Используя порядок действий, описанный в пп. 6-10, вычислите минимальное число в заданном наборе (функция **МИН**), максимальное число (**МАКС**), количество элементов в наборе (**СЧЕТ**).

12 Сохраните рабочую книгу book.xls.

Задание № 3. Подготовка и форматирование документов.

Порядок выполнения работы:

1 Запустите программу *Excel* (*Пуск / Программы / Microsoft Excel*) и откройте рабочую книгу book.xls.

2 Выберите щелчком на ярлычке неиспользуемый рабочий лист или создайте новый (*Вставка / Лист*). Дважды щелкните на ярлычке нового листа и переименуйте его как *Прейскурант*.

3 В ячейку **A1** введите текст *Прейскурант* и нажмите клавишу **Enter**.

4 В ячейку **A2** введите текст *Курс пересчета:* и нажмите клавишу **Enter**. В ячейку **B2** введите текст *1 у.е.=* и нажмите клавишу **Enter**, В ячейку **C2** введите текущий курс пересчета и нажмите клавишу **Enter**.

5 В ячейку **A3** введите текст *Наименование товара* и нажмите клавишу **Enter**. В ячейку **B3** введите текст *Цена (у.е.)* и нажмите клавишу **Enter**. В ячейку **C3** введите текст *Цена (руб.)* и нажмите клавишу **Enter**.

6 В последующие ячейки столбца **A** введите названия товаров, включенных в прейскуронт.

7 В соответствующие ячейки столбца **B** введите цены товаров в условных единицах.

8 В ячейку **C4** введите формулу: **=B4*\$C\$2**, которая используется для пересчета цены из условных единиц в рубли.

9 Методом автозаполнения скопируйте формулы во все ячейки столбца **C**, которым соответствуют заполненные ячейки столбцов **A** и **B** [3].

10 Измените курс пересчета в ячейке **C2**. Обратите внимание, что все цены в рублях при этом обновляются автоматически.

11 Выделите методом протягивания диапазон **A1:C1** и дайте команду *Формат / Ячейки*. На вкладке *Выравнивание* задайте выравнивание по горизонтали *По центру* и установите флажок *Объединение ячеек*.

12 На вкладке *Шрифт* задайте размер шрифта в 14 пунктов и в списке *Начертание* выберите вариант *Полужирный*. Щелкните на **ОК**.

13 Щелкните правой кнопкой мыши на ячейке **B2** и выберите в контекстном меню команду *Формат ячеек*. Задайте выравнивание по горизонтали *По правому краю* и щелкните на кнопке **ОК**.

14 Щелкните правой кнопкой мыши на ячейке **C2** и выберите в контекстном меню команду *Формат ячеек*. Задайте выравнивание по горизонтали *По левому краю* и щелкните на кнопке **ОК**.

15 Выделите методом протягивания диапазон **B2:C2**. Щелкните на раскрывающей кнопке рядом с кнопкой *Границы* на панели инструментов *Форматирование* и задайте для этих ячеек широкую внешнюю рамку (кнопка в правом нижнем углу открывшейся палитры).

16 Дважды щелкните на границе между заголовками столбцов **A** и **B**, **B** и **C**, **C** и **D**. Обратите внимание, как при этом изменяется ширина столбцов **A**, **B** и **C**.

17 Посмотрите, устраивает ли вас полученный формат таблицы. Щелкните на кнопке *Предварительный просмотр* на стандартной панели инструментов, чтобы увидеть, как документ будет выглядеть при печати.

18 Сохраните рабочую книгу *book.xls* [3].

Задание № 4. Построение экспериментального графика.

Порядок выполнения работы:

1 Запустите программу *Excel* (*Пуск / Программы/ Microsoft Excel*) и откройте рабочую книгу *book.xls*, созданную ранее.

2 Выберите щелчком на ярлычке неиспользуемый рабочий лист или создайте новый (*Вставка / Лист*). Дважды щелкните на ярлычке листа и переименуйте его как *Обработка эксперимента*.

3 В столбец **A**, начиная с ячейки **A1**, введите произвольный набор значений независимой переменной.

4 В столбец **B**, начиная с ячейки **B1**, введите произвольный набор значений функции.

5 Методом протягивания выделите все заполненные ячейки столбцов **A** и **B**.

6 Щелкните на значке **Мастер диаграмм** на стандартной панели инструментов.

7 В списке *Тип* выберите пункт **Точечная** (для отображения графика, заданного парами значений). В палитре *Вид* выберите средний пункт в первом столбце (маркеры, соединенные гладкими кривыми). Щелкните на кнопке *Далее*.

8 Так как диапазон ячеек был выделен заранее, мастер диаграмм автоматически определяет расположение рядов данных. Убедитесь, что данные на диаграмме выбраны правильно. На вкладке *Ряд* в поле *Имя* укажите: *Результаты измерений*. Щелкните на кнопке *Далее*.

9 Выберите вкладку *Заголовки*. Убедитесь, что заданное название ряда данных автоматически использовано как заголовок диаграммы. Замените его, введя в поле *Название диаграммы* заголовок *Экспериментальные точки*. Щелкните на кнопке *Далее*.

10 Установите переключатель *Отдельном*. По желанию, задайте произвольное имя добавляемого рабочего листа. Щелкните на кнопке *Готово*.

11 Убедитесь, что диаграмма построена и внедрена в новый рабочий лист. Рассмотрите ее и щелкните на построенной кривой, чтобы выделить ряд данных.

12 Дайте команду *Формат / Выделенный ряд*. Откройте вкладку *Вид*.

13 На панели *Линия* откройте палитру *Цвет* и выберите красный цвет. В списке *Тип линии* выберите пунктир.

14 На панели *Маркер* выберите в списке *Тип маркера* треугольный маркер. В палитрах *Цвет* и *Фон* выберите зеленый цвет.

15 Щелкните на кнопке **ОК**, снимите выделение с ряда данных и посмотрите, как изменился вид графика.

16 Сохраните рабочую книгу [3].

Содержание отчёта:

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.4 Лабораторная работа № 4. Создание сводной таблицы в MS Excel

Цели работы:

- 1 Создание сводной таблицы на основе списка базы данных Excel.
- 2 Получение практических навыков работы со сводными отчетами.

Задание на выполнение лабораторной работы.

На основании «Ведомости начисления заработной платы» создать расчетный лист для каждого работника предприятия с использованием **Мастера сводных таблиц**.

Порядок выполнения работы:

- 1 Открыть рабочую книгу «Зарплата».

2 Активизировать рабочий лист «март».

3 Создать расчетный лист для каждого работника предприятия, как показанный на рисунке 8.

4 Для этого воспользоваться **Мастером сводных таблиц**:

- выделить область списка, например весь список таблицы **ведомость за март**.

- выполнить **Данные** надо выбрать команду **Сводная таблица**, по которой на экран выводится окно **Мастера сводных таблиц**.

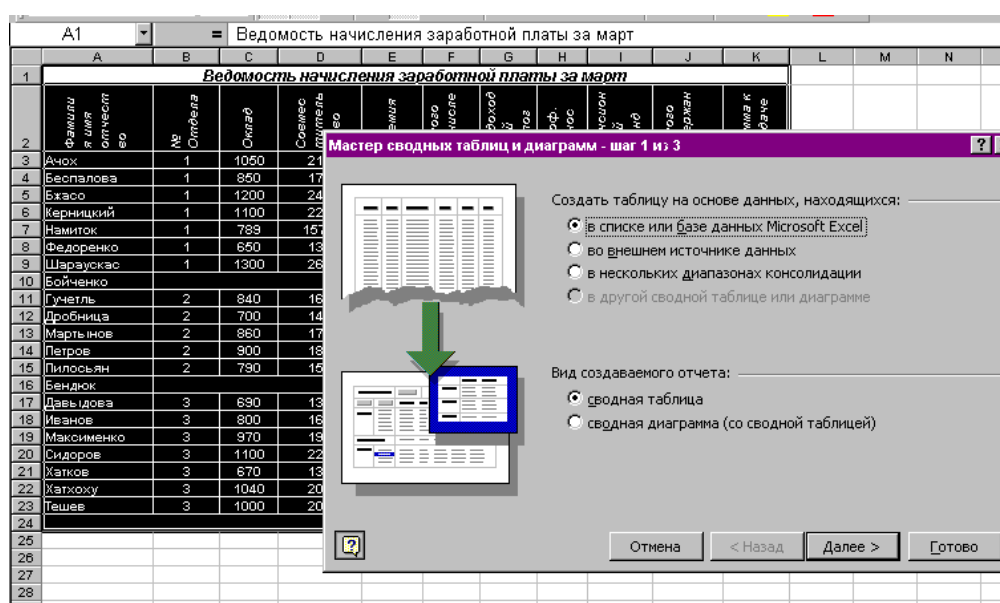


Рисунок 8 - Диалоговое окно Мастера сводных таблиц [3]

Сводная таблица — это еще один инструмент обработки больших списков с данными. Поскольку в этом случае сразу подводятся итоги, выполняется сортировка и фильтрация списков, то сводная таблица обслуживается более мощным инструментом обработки данных, который называется **Мастером сводных таблиц**, показанный на рисунке 9. Он состоит из 4 шагов.

На 1 шаге Выбираем основу создания сводной таблицы. Щелкнув кнопку **Далее**, перейти к следующему шагу.

	Фамилия я для отчет го	№ Отдела	Оклад	Совмес титель ство	Премия	Итого начисле но	Подход ный налог	Проф. Взнос	Пенсион ный фонд	Итого удержан но	Сумма к выдаче
2											
3	Анох	1	1050	210	336	1596	136,5	21	10,5	168	1428
4	Беспалова	1	860	170	272	1292	110,5	17	8,5	136	1156
5	Бжасо	1	1200	240	384	1824	156	24	12	192	1632
6	Керницкий	1	1100	220	352	1672	143	22	11	176	1496
7	Намиток	1	789	157,8	252,48	1199,3	102,57	15,78	7,89	126,24	1073,04
8	Федоренко	1	650	130	208	988	84,5	13	6,5	104	884
9	Шараускас	1	1300	260	416	1976	168	28	14	224	1752
10	Бойченко				2						
11	Гучель	2	840	168	268,8						
12	Дробница	2	700	140	224						
13	Марьинов	2	860	172	275,2						
14	Петров	2	900	180	288						
15	Пилосьян	2	790	158	252,8						
16	Бендюк				3						
17	Давыдова	3	690	138	220,8						
18	Иванов	3	800	160	256						
19	Максименко	3	970	194	310,4						
20	Сидоров	3	1100	220	352	1672	143	22	11	176	1496
21	Хатков	3	670	134	214,4	1018,4	87,1	13,4	6,7	107,2	911,2
22	Хатхоху	3	1040	208	332,8	1580,8	135,2	20,8	10,4	166,4	1414,4
23	Тешев	3	1000	200	320	1520	130	20	10	160	1360
24											

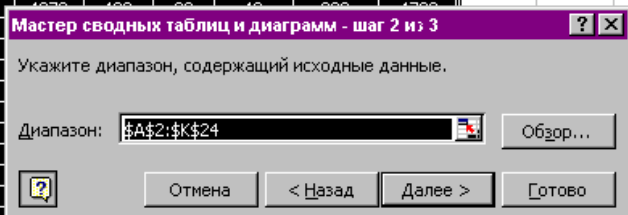


Рисунок 9 - Мастер сводных таблиц

Второй шаг служит для определения диапазона. Кнопка **Обзор** служит для подключения данных из других книг или файлов.

Третий шаг, экранная форма показана на рисунке 10, служит для создания макета сводной таблицы и определяет, где будет размещена сводная таблица. Рекомендуется размещать ее на отдельном рабочем листе.

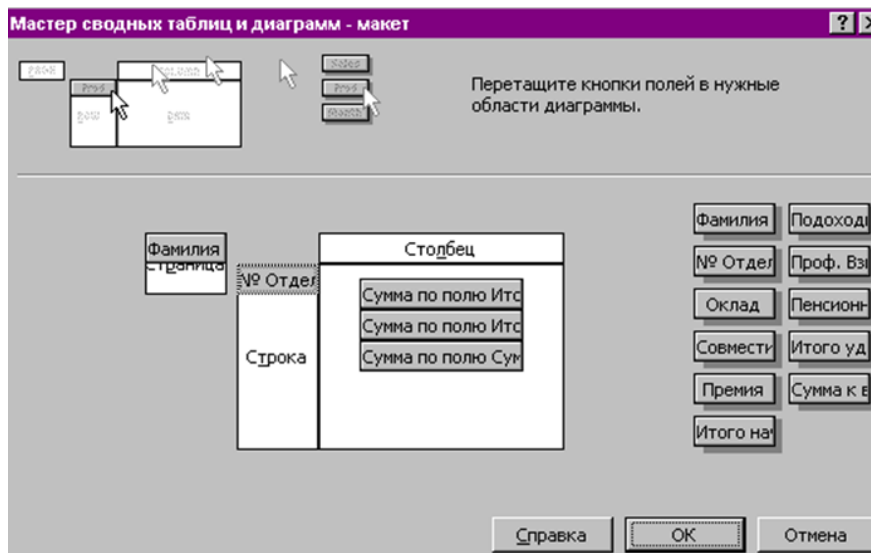


Рисунок 10 – Мастер сводных таблиц и диаграмм [3]

С помощью мыши перетащить поля списка из правой части окна **Мастера**

сводных таблиц в области **Строк, Столбцов и Данных** в том же окне как указано на рисунке 11. Таким образом формируется вид расчетного листа.

	А	В	С	D	E	F	G	H	I
1	Фамилия	Ачох							
2									
3	№ Отде...	Данные	Всего						
4		Сумма по полю Итого начислено	1596						
5		Сумма по полю Итого удержанно	168						
6		Сумма по полю Сумма к выдаче	1428						
7	Итого	Сумма по полю Итого начислено	1596						
8	Итого	Сумма по полю Итого удержанно	168						
9	Итого	Сумма по полю Сумма к выдаче	1428						

Рисунок 11 – Вид расчетного листа

Изучить назначение кнопок панели инструментов сводной таблицы.

Содержание отчёта:

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.5 Лабораторная работа № 5. Расчеты коэффициента корреляции и уравнения регрессии в MS Excel

Цель работы. При помощи возможностей Microsoft Excel научиться обрабатывать статистические данные и находить между ними зависимость.

Краткие теоретические сведения.

Для выполнения заданий необходимо воспользоваться уравнением среднеквадратической регрессии [3]:

$$Y = m_y + r \frac{\sigma_y}{\sigma_x} (X - m_x), \quad (2)$$

где m_x - математическое ожидание X;

m_y - математическое ожидание Y;

r - коэффициент корреляции величин X и Y.

$$r = \frac{\mu_{xy}}{\sigma_x \sigma_y}, \quad (3)$$

где μ_{xy} - корреляционный момент величин X и Y;

σ_x, σ_y - дисперсии X и Y.

Порядок выполнения работы.

Для приведенных в таблице результатов эксперимента определить параметры линейной регрессии, т.е. коэффициенты уравнения $y = ax + b$. Построить на одной координатной плоскости два графика зависимости y от x: экспериментальный и полученный в результате линейного приближения, т.е. прямую $y = ax + b$.

1 Согласно варианту задания построить два столбца экспериментальных данных.

2 При помощи функции **КОРРЕЛ()**, которая находится в разделе *Статистические*, определить коэффициент корреляции между величинами X и Y. Сделать вывод о зависимости случайных величин.

3 При помощи функции **ЛИНЕЙН()** определить параметры линейного приближения по методу наименьших квадратов. Для этого необходимо выделить две ячейки, в которых будут находиться значения параметров a и b, вызвать функцию **ЛИНЕЙН()**, подставить значения X, Y, константа - 1, Стат - 0. Нажать одновременно клавиши **Shift, Ctrl, Enter** для получения коэффициентов уравнения $y = ax + b$.

Таблица 5 – Результаты корреляции

Коэффициент корреляции	Параметр a	Параметр b
-0,7144	-0,576	5,994

4 Построить столбец теоретических значений Y, подставив полученные параметры в уравнение линейной зависимости $y = ax + b$ [3]. Результаты экспериментальных и теоретических значений показаны в таблице 6.

Таблица 6 – Результаты экспериментальных и теоретических значений Y

X	Экспериментальные значения Y	Теоретические значения Y
0,1	5,998	5,9364
0,2	5,82	5,8788
0,3	5,754	5,8212
0,4	5,828	5,7636

5 Построить на одной координатной плоскости два графика зависимости y от x - экспериментальный и полученный в результате линейного приближения, как показано на рисунке 12. Сделать вывод о качестве линейного приближения.

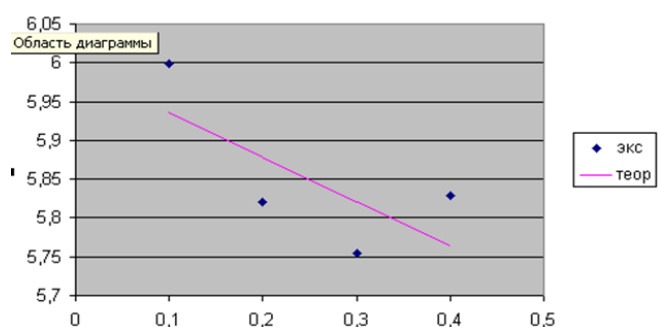


Рисунок 12 - Графики зависимости y от x - экспериментальный и полученный в результате линейного приближения

6 Спрогнозируйте значение функции Y еще на два шага вперед от максимального X .

Варианты заданий показаны в таблице 7.

Таблица 7 – Варианты заданий

i	Значения $y_i = y(x_i)$									
	1	2	3	4	5	6	7	8	9	10
0										
1	5.998	6.030	5.85	6.310	5.650	6.323	3.88	4.08	3.90	4.03
2	5.820	6.072	5.619	6.308	5.431	6.523	3.86	4.18	3.83	4.23
3	5.754	6.297	5.569	6.546	5.250	6.646	3.84	4.38	3.60	4.49
4	5.828	6.428	5.426	6.855	5.000	7.256	3.91	4.46	3.47	4.71
5	5.627	6.425	5.237	7.073	4.790	7.487	3.71	4.44	3.31	5.00
6	5.597	6.473	5.025	7.770	4.569	7.827	3.49	4.55	3.05	5.26

Продолжение таблицы 7

0	1	2	3	4	5	6	7	8	9	10
7	5.693	6.592	4.988	7.225	4.296	8.133	3.51	4.66	3.14	5.36
8	5.469	6.815	5.037	7.739	4.065	8.402	3.68	4.89	2.83	5.87
9	5.413	6.786	4.586	7.995	3.837	8.581	3.74	4.86	2.66	5.67
10	5.526	6.925	4.575	8.063	3.519	9.014	3.47	5.04	2.53	5.89
11	5.344	7.116	4.445	8.247	3.281	9.049	3.60	5.22	2.35	6.16
12	5.304	7.053	4.353	8.472	2.926	9.571	3.51	4.99	2.49	6.65
13	5.352	7.224	3.933	8.627	2.801	9.891	3.48	5.39	2.19	6.39
14	5.301	7.439	3.899	8.936	2.546	10.073	3.30	5.56	1.82	6.81
15	5.424	7.302	3.793	9.082	2.232	10.406	3.23	5.42	1.69	7.08
16	4.996	7.426	3.473	9.076	2.016	10.821	3.26	5.85	1.54	7.24
17	5.080	7.797	3.551	9.363	1.794	11.151	3.14	5.99	1.22	7.61
18	5.256	7.871	3.171	9.679	1.663	11.232	3.17	5.85	1.17	7.64
19	5.090	7.929	3.330	9.846	1.375	11.655	2.96	6.01	1.04	8.03
20	5.053	8.060	3.044	10.013	1.217	11.952	2.81	5.97	1.12	10.013

Содержание отчёта:

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.6 Лабораторная работа № 6. Аппроксимация экспериментальных данных в MS Excel

Цель работы. Исследовать принцип проведения аппроксимации экспериментальных данных.

Краткие теоретические сведения.

Задача аппроксимации распадается на две составляющие. Сначала устанавливают вид зависимости $y = f(x)$ и, соответственно, вид эмпирической формулы, то есть решают, является ли она линейной, квадратичной, логарифмической или какой-либо другой. После этого определяют численные значения неизвестных параметров выбранной эмпирической формулы, для которых приближение к заданной функции оказывается наилучшим. Если нет каких-либо

теоретических соображений для подбора вида формулы, обычно выбирают функциональную зависимость из числа наиболее простых, сравнивая их графики с графиком заданной функции [1].

После выбора вида формулы определяют ее параметры. Для наилучшего выбора параметров задают меру близости аппроксимации экспериментальных данных. Во многих случаях, в особенности если функция $f(x)$ задана графиком или таблицей (на дискретном множестве точек), для оценки степени приближения рассматривают разности $f(x_i) - \varphi(x_i)$ для точек x_0, x_1, \dots, x_n . Существуют различные меры близости и, соответственно, способы решения этой задачи. Некоторые из них очень просты, быстро приводят к результату, но результат этот является сильно приближенным. Другие более точные, но и более сложные. Обычно оценку определения параметров при известном виде зависимости осуществляют по методу наименьших квадратов. При этом функция $\varphi(x)$ считается наилучшим приближением к $f(x)$, если для нее сумма квадратов отклонений «теоретических» значений $\varphi(x_i)$, найденных по эмпирической формуле, от соответствующих опытных значений y_i :

$$Z = \sum_{i=0}^n [f(x_i) - \varphi(x_i)]^2 \rightarrow \min \quad (4)$$

Метод наименьших квадратов формулирует аналитические условия достижения суммой квадратов отклонений (4) своего наименьшего значения.

В простейшем случае задача аппроксимации экспериментальных данных выглядит следующим образом.

Пусть есть какие-то данные, полученные практическим путем (в ходе эксперимента или наблюдения), которые можно представить парами чисел $(x; y)$. Зависимость между ними отражает следующая таблица на рисунке 13.

x	x_1	...	x_n
y	y_1	...	y_n

Рисунок 13 – Зависимость между парами чисел

На основе этих данных требуется подобрать функцию $y = \varphi(x)$, которая наилучшим образом сглаживала бы экспериментальную зависимость между переменными и по возможности точно отражала общую тенденцию зависимости между x и y , исключая погрешности измерений и случайные отклонения. Это значит, что отклонения $y_i - y_i(x_i)$ должны быть наименьшими [1].

Выяснить вид функции можно либо из теоретических соображений, либо анализируя расположение точек $(x_i; y_i)$ на координатной плоскости.

Например, пусть точки расположены так, как показано на рисунке 14.

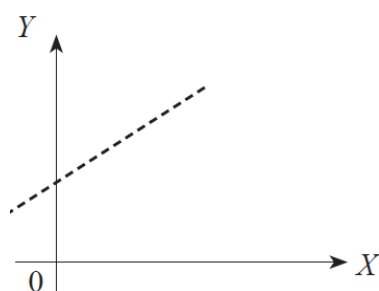


Рисунок 14 - Возможный вариант расположения экспериментальных точек

Учитывая то, что практические данные получены с некоторой погрешностью, обусловленной неточностью измерений, необходимостью округления результатов и т. п., естественно предположить, что здесь имеет место линейная зависимость $y = ax + b$.

Чтобы функция приняла конкретный вид, необходимо вычислить a и b .

Построение эмпирической функции сводится к вычислению входящих в нее параметров так, чтобы из всех функций такого вида выбрать ту, которая лучше других описывает зависимость между изучаемыми величинами. То есть сумма квадратов разности между табличными значениями функции в некоторых точках и значениями, вычисленными по полученной формуле, должна быть минимальна [1].

В MS Excel аппроксимация экспериментальных данных осуществляется путем построения их графика (x – отвлеченные величины) или точечного графика (x – имеет конкретные значения) с последующим подбором подходящей аппроксимирующей функции (линии тренда). Возможны следующие варианты

функций:

1 Линейная: $y = ax + b$. Обычно применяется в простейших случаях, когда экспериментальные данные возрастают или убывают с постоянной скоростью.

2 Полиномиальная: $y = a_0 + a_1x + a_2x_2 + \dots + a_nx_n$ до шестого порядка включительно ($n \leq 6$), a_i – константы. Используется для описания экспериментальных данных, попеременно возрастающих и убывающих. Степень полинома определяется количеством экстремумов (максимумов или минимумов). Полином второй степени может описать только один максимум или минимум, полином третьей степени может иметь один или два экстремума, четвертой степени – не более трех экстремумов и т. д.

3 Логарифмическая: $y = a \ln x + b$, где a и b – константы, \ln – функция натурального логарифма. Функция применяется для описания экспериментальных данных, которые вначале быстро растут или убывают, а затем постепенно стабилизируются.

4 Степенная: $y = bx^a$, где a и b – константы. Аппроксимация степенной функцией используется для экспериментальных данных с постоянно величивающейся (или убывающей) скоростью роста. Данные не должны иметь нулевых или отрицательных значений.

5 Экспоненциальная: $y = be^{ax}$, где a и b – константы, e – основание натурального логарифма. Применяется для описания экспериментальных данных, которые быстро растут или убывают, а затем постепенно стабилизируются. Часто ее использование вытекает из теоретических соображений.

Степень близости аппроксимации экспериментальных данных выбранной функцией оценивается коэффициентом детерминации (R^2). Таким образом, если есть несколько подходящих вариантов типов аппроксимирующих функций, можно выбрать функцию с большим коэффициентом детерминации (стремящимся к 1).

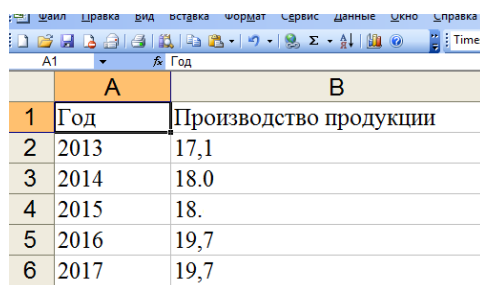
Для осуществления аппроксимации на диаграмме экспериментальных данных в случае использования пакета Microsoft Excel необходимо щелчком правой кнопки мыши вызвать контекстное меню и выбрать пункт **Добавить линию тренда**.

В появившемся диалоговом окне Линия тренда на вкладке Тип выбирается вид

аппроксимирующей функции, а на вкладке Параметры задаются дополнительные параметры, влияющие на отображение аппроксимирующей кривой [1].

Задание на выполнение лабораторной работы.

Исследовать характер изменения с течением времени уровня производства некоторой продукции и подобрать аппроксимирующую функцию, располагая следующими данными, показанными на рисунке 15.



	A	B
1	Год	Производство продукции
2	2013	17,1
3	2014	18,0
4	2015	18,
5	2016	19,7
6	2017	19,7

Рисунок 15 – Данные производства некоторой продукции

Порядок выполнения работы:

1 Для построения диаграммы прежде всего необходимо ввести данные в рабочую таблицу.

2 Далее по введенным в рабочую таблицу данным необходимо построить диаграмму, как показано на рисунке 16. Поскольку здесь необходимо показать динамику изменений производства продукции, не привязываясь к конкретному году, а от отвлеченных переменных, выберем диаграмму График [1].

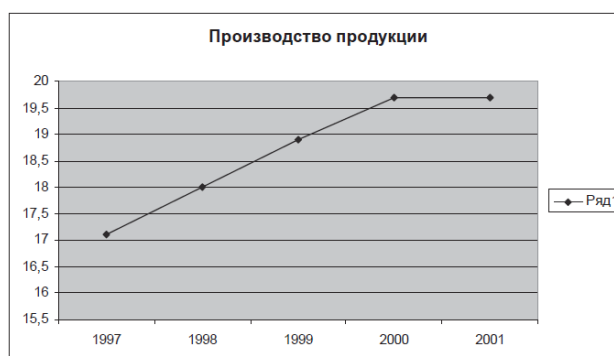


Рисунок 16 - График экспериментальных данных

3 Осуществим аппроксимацию полученной кривой полиномиальной функцией второго порядка, поскольку кривая довольно гладкая и не сильно отличается от прямой линии. Для этого указатель мыши устанавливаем на одну из точек графика и щелкаем правой кнопкой. В появившемся контекстном меню выбираем пункт **Добавить линию тренда**. Появляется диалоговое окно **Линия тренда**, как показано на рисунке 17.

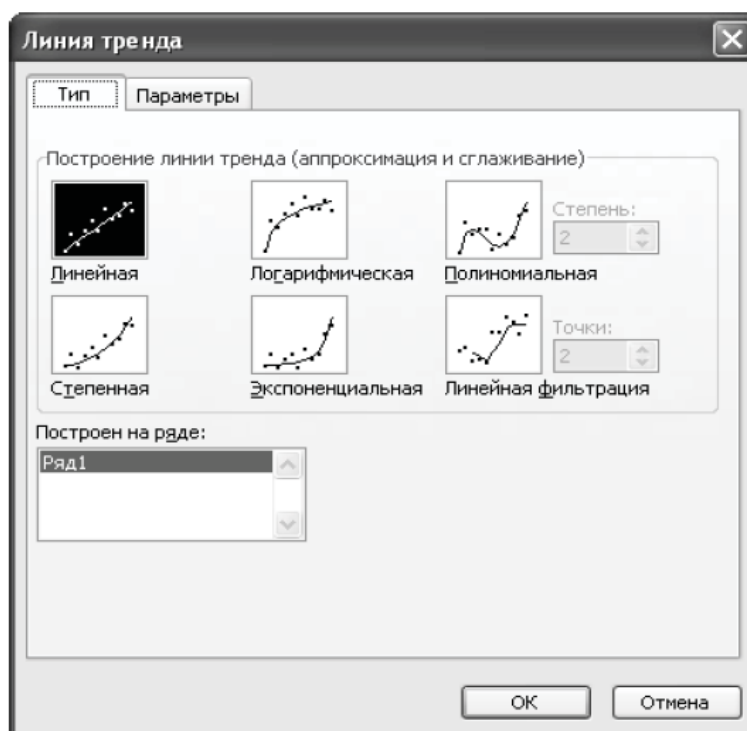


Рисунок 17 - Вкладка Тип диалогового окна **Линия тренда**

В этом окне на вкладке **Тип** выбираем тип линии тренда – **Полиномиальная** – и устанавливаем степень – 2. Затем открываем вкладку **Параметры**, показанный на рисунке 18 и устанавливаем флажки в поля показывать уравнение на диаграмме и поместить на диаграмму величину достоверности аппроксимации (R^2) [1].

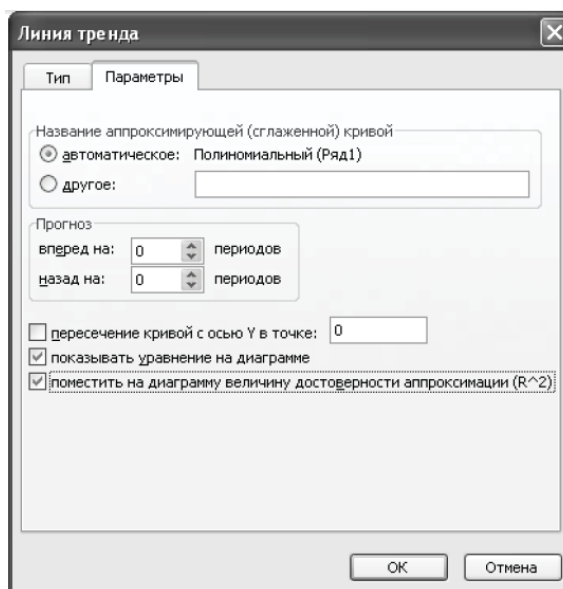


Рисунок 18 - Вкладка Параметры диалогового окна **Линия тренда**

После чего нужно щелкнуть по кнопке **ОК**. В результате получим на диаграмме аппроксимирующую кривую, показанную на рисунке 19 [1].

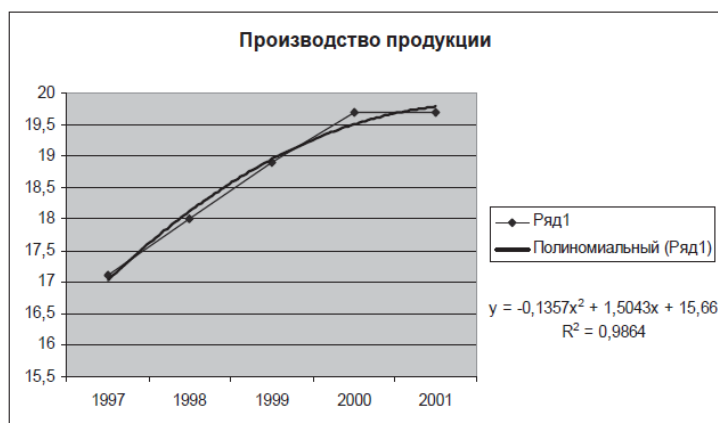


Рисунок 19 - Экспериментальные данные, аппроксимированные полиномиальной кривой

Как видно из рисунка 19, уравнение наилучшей полиномиальной аппроксимирующей функции для некоторых отвлеченных значений x (1, 2, 3, ...) выглядит как $y = -0,14x^2 + 1,5x + 15,66$.

При этом точность аппроксимации достаточно высока – $R^2 = 0,986$.

Содержание отчёта.

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.7 Лабораторная работа № 7. Использование инструментов Пакета анализа для статистической обработки данных в MS Excel

Цель работы. Изучение возможностей пакета MS Excel при использовании инструментов Пакета анализа для статистической обработки данных.

Краткие теоретические сведения.

В пакете Excel помимо **Мастера функций** имеется набор более мощных инструментов для работы с несколькими выборками и углубленного анализа данных, называемый **Пакет анализа**, который может быть использован для решения задач статистической обработки выборочных данных.

Для установки **Пакета анализа** в Excel выполните следующее [1]:

- в меню **Сервис** выберите команду **Надстройки**;
- в появившемся списке установите флажок **Пакет анализа**.

Ввод данных. Исследуемые данные следует представить в виде таблицы, где столбцами являются соответствующие показатели. При создании таблицы Excel информация вводится в отдельные ячейки. Совокупность ячеек, содержащих анализируемые данные, называется входным диапазоном.

Последовательность обработки данных. Для использования статистического пакета анализа данных необходимо:

- выполнить команду **Сервис – Анализ данных**;
- выбрать необходимую строку в появившемся списке **Инструменты анализа**;
- ввести входной и выходной диапазоны и выбрать необходимые параметры.

Нахождение основных выборочных характеристик. Для определения

характеристик выборки используется процедура **Описательная статистика**. Процедура позволяет получить статистический отчет, содержащий информацию о центральной тенденции и изменчивости входных данных. Для выполнения процедуры необходимо:

- выполнить команду **Сервис – Анализ данных**;
- в появившемся списке **Инструменты анализа** выбрать строку **Описательная статистика** и нажать кнопку **ОК**, как на рисунке 20.;
- в появившемся диалоговом окне указать входной диапазон, то есть ввести ссылку на ячейки, содержащие анализируемые данные;
- указать выходной диапазон, то есть ввести ссылку на ячейки, в которые будут выведены результаты анализа;
- в разделе **Группировка** переключатель установить в положение по столбцам;
- установить флажок в поле **Итоговая статистика**;
- нажать кнопку **ОК** [1].

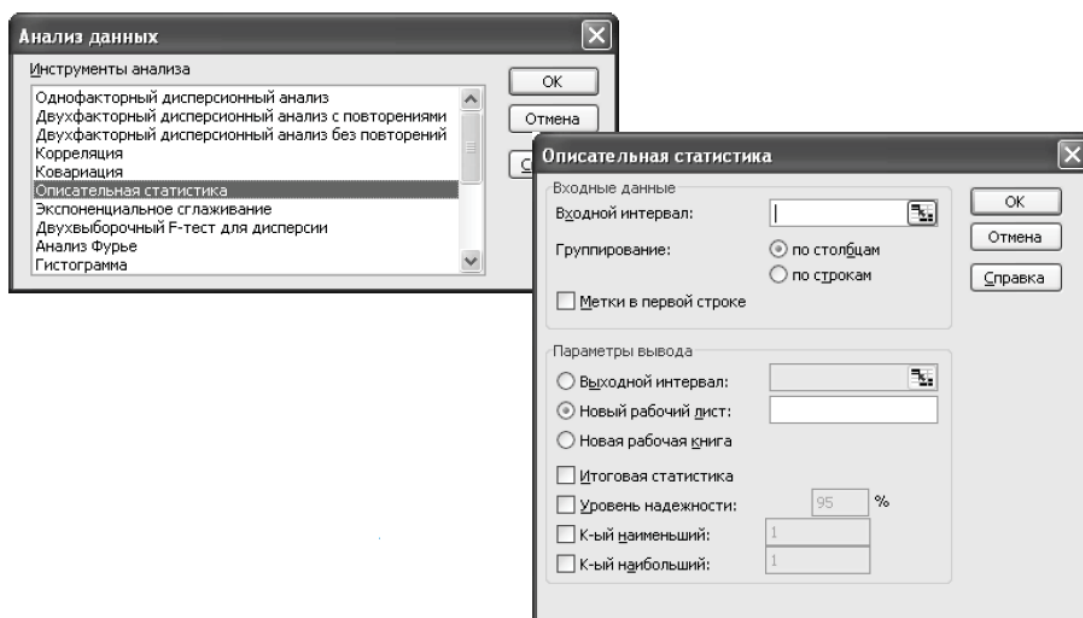


Рисунок 20 - Окно выбора метода обработки данных и
окно **Описательной статистики**

В результате анализа в указанном выходном диапазоне для каждого столбца данных выводятся следующие статистические характеристики: среднее, стандартная ошибка (среднего), медиана, мода, стандартное отклонение, дисперсия выборки, эксцесс, асимметричность, интервал, минимум, максимум, сумма, счет, наибольшее, наименьшее, уровень надежности.

Задание на выполнение лабораторной работы.

Рассматривается зарплата основных групп работников гостиницы: администрации, обслуживающего персонала и работников ресторана. Были получены следующие данные, показанные на рисунке 21.

Администрация	Персонал	Ресторан
4500	2100	3200
4000	2100	3000
3700	2000	2500
3000	2000	2000
2500	2000	1900
	1900	1800
	1800	
	1800	

Рисунок 21 – Задание на выполнение лабораторной работы [1]

Необходимо определить основные статистические характеристики в группах данных.

Порядок выполнения работы:

1 Для использования инструментов анализа исследуемые данные следует представить в виде таблицы, где столбцами являются соответствующие показатели. Значения зарплат сотрудников администрации введите в диапазон **A1:A5**, обслуживающего персонала – в диапазон **B1:B8** и т. д. В результате получится таблица, представленная на рисунке 22.

	A	B	C
1	4500	2100	3200
2	4000	2100	3000
3	3700	2000	2500
4	3000	2000	2000
5	2500	2000	1900
6		1900	1800
7		1800	
8		1800	

Рисунок 22 - Таблица из задания [1]

2 Далее необходимо провести элементарную статистическую обработку. Для этого выполните команду **Сервис – Анализ данных**. Затем в появившемся списке **Инструменты анализа** выберите строку **Описательная статистика**.

3 В появившемся диалоговом окне, как на рисунке 23, в рабочем поле **Входной интервал** укажите входной диапазон – **A1:C8**. Активировав переключателем рабочее поле **Выходной интервал**, укажите выходной диапазон – ячейку **A9**. В разделе **Группировка** переключатель установите в положение по столбцам. Установите флажок в поле **Итоговая статистика** и нажмите кнопку **ОК**.

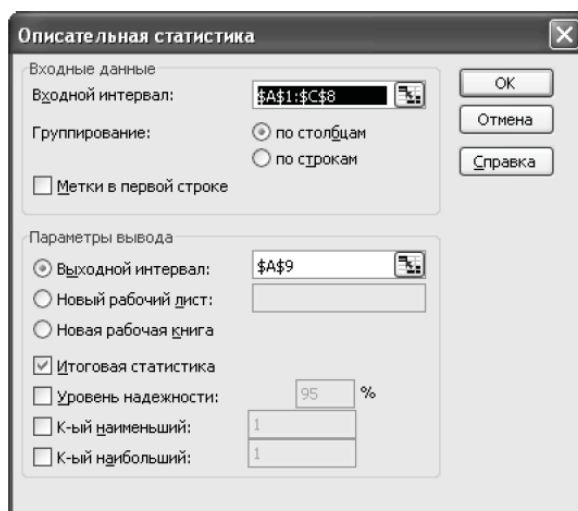


Рисунок 23 - Пример заполнения диалогового окна **Описательная статистика**

В результате анализа в указанном выходном диапазоне для каждого столбца данных получим соответствующие результаты, как показанные на рисунке 24 [1].

9	Столбец1		Столбец2		Столбец3	
10						
11	Среднее	3540	Среднее	1962,5	Среднее	2400
12	Стандартная ошибка	355,8089375	Стандартная ошибка	41,99277	Стандартная ошибка	243,5843
13	Медиана	3700	Медиана	2000	Медиана	2250
14	Мода	#Н/Д	Мода	2000	Мода	#Н/Д
15	Стандартное отклонение	795,6129712	Стандартное отклонение	118,7735	Стандартное отклонение	596,6574
16	Дисперсия выборки	633000	Дисперсия выборки	14107,14	Дисперсия выборки	356000
17	Эксцесс	-1,29384635	Эксцесс	-1,22929	Эксцесс	-2,06887
18	Асимметричность	-0,245024547	Асимметричность	-0,39433	Асимметричность	0,457606
19	Интервал	2000	Интервал	300	Интервал	1400
20	Минимум	2500	Минимум	1800	Минимум	1800
21	Максимум	4500	Максимум	2100	Максимум	3200
22	Сумма	17700	Сумма	15700	Сумма	14400
23	Счет	5	Счет	8	Счет	6

Рисунок 24 - Результаты работы инструмента **Описательная статистика**

Все полученные характеристики были рассмотрены ранее в разделе «Выборочные характеристики», за исключением последних четырех:

- минимум – значение минимального элемента выборки;
- максимум – значение максимального элемента выборки;
- сумма – сумма значений всех элементов выборки;
- счет – количество элементов в выборке.

Среди этих характеристик наиболее важными являются показатели Среднее, Стандартная ошибка (среднего) и Стандартное отклонение.

Содержание отчёта:

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.8 Лабораторная работа № 8. Расчет описательных статистик с помощью электронных таблиц MS Excel

Цель работы. Приобретение практических навыков обработки статических выборок с помощью электронных таблиц Excel.

Краткие теоретические сведения.

Это средство является наиболее часто используемым из всего пакета анализа, поскольку быстро и просто вычисляет основные статистические характеристики одномерных выборок.

В большинстве таких диалоговых окон (на рисунке 25 показано диалоговое окно средства *Описательная статистика*) выделены области *Входные данные* и *Параметры вывода*.

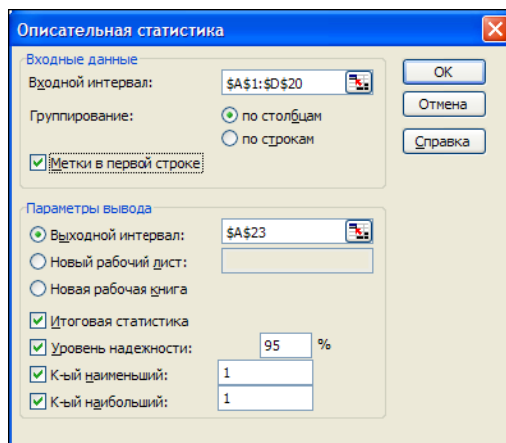


Рисунок 25 – Диалоговое окно средства Описательная статистика

В области *Входные данные* указывается диапазон ячеек, в котором содержатся данные (поле *Входной интервал*), указывается, сгруппированы ли данные, и если сгруппированы, то по столбцам или по строкам (переключатели *по* и *столбцам по строкам*). Если задаётся входной диапазон данных вместе с заголовками, то устанавливается флажок опции *Метки в первой строке (столбце)*. (если заголовки не задаются, то данным автоматически присваиваются заголовки *Столбец1*, *Столбец2* и т.д. или *Строка1*, *Строка2* и т.д. в зависимости от того, расположены

данные в столбцах или в строках). В некоторых диалоговых окнах в области *Входные данные* необходимо указать несколько входных диапазонов (например, в окне *Регрессия*) либо дополнительные параметры для проведения выбранной статистической процедуры, например доверительный уровень для проведения тестов.

В области *Параметры вывода*, как правило, надо указать, куда будут выводиться результаты расчётов. Предусмотрено три возможности: на текущий рабочий лист (переключатель *Выходной интервал*), при этом необходимо указать выходной интервал (достаточно указать адрес одной ячейки, которая определяет верхний левый угол выходного диапазона); на новый рабочий лист текущей рабочей книги начиная с ячейки **A1** (переключатель *Новый рабочий лист*), при этом можно сразу задать имя этому листу; в новую рабочую книгу (переключатель *Новая рабочая книга*), в этом случае автоматически открывается новая рабочая книга. Также в этой области часто имеются опции, которые указывают, что именно необходимо вывести из возможного набора выходных результатов (например, графики, либо дополнительные статистические характеристики).

На рисунке 26 показан рабочий лист, содержащий три ряда данных (три независимые выборки, имеющие разные распределения) и диалоговое окно *Описательная статистика*.

В данном случае имеются выборки разных размеров. Средство *Описательная статистика* правильно определяет размеры выборок, игнорируя пустые ячейки. На рисунке 27 показан рабочий лист с результатами расчётов.

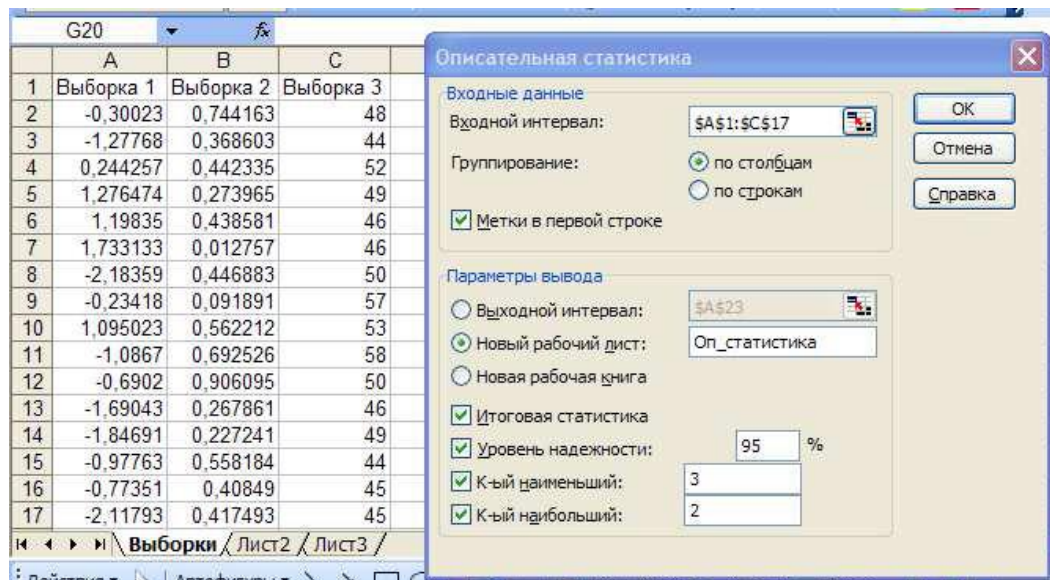


Рисунок 26 – Три выборки и диалоговое окно **Описательная статистика**

	A	B	C	D	E	F	G
1	Выборка 1		Выборка 2		Выборка 3		
2							
3	Среднее	-0,47698	Среднее	0,428705	Среднее	48,875	
4	Стандартная ошибка	0,317535	Стандартная ошибка	0,05821	Стандартная ошибка	1,0831936	
5	Медиана	-0,73186	Медиана	0,428037	Медиана	48,5	
6	Мода	#Н/Д	Мода	#Н/Д	Мода	46	
7	Стандартное отклонение	1,27014	Стандартное отклонение	0,232841	Стандартное отклонение	4,3339743	
8	Дисперсия выборки	1,613255	Дисперсия выборки	0,054215	Дисперсия выборки	18,783333	
9	Эксцесс	-1,04365	Эксцесс	0,121328	Эксцесс	0,0938121	
10	Асимметричность	0,391197	Асимметричность	0,221317	Асимметричность	0,9169154	
11	Интервал	3,916723	Интервал	0,893338	Интервал	14	
12	Минимум	-2,18359	Минимум	0,012757	Минимум	44	
13	Максимум	1,733133	Максимум	0,906095	Максимум	58	
14	Сумма	-7,63175	Сумма	6,85928	Сумма	782	
15	Счет	16	Счет	16	Счет	16	
16	Наибольший(3)	1,19835	Наибольший(3)	0,692526	Наибольший(3)	53	
17	Наименьший(2)	-2,11793	Наименьший(2)	0,091891	Наименьший(2)	44	
18	Уровень надежности(95,0%)	0,67681	Уровень надежности(95,0%)	0,124072	Уровень надежности(95,0%)	2,3094119	
19							

Рисунок 27 – Результаты работы средства **Описательная статистика**

Установка флажка опции *Итоговая статистика* указывает, что в итоговом отчёте этого средства будут вычислены все статистические характеристики выборки, за исключением границы доверительного интервала для среднего и К-х наибольших и наименьших значений, для которых имеются отдельные опции *Уровень надёжности*, *К-ый наименьший* и *К-ый наибольший*. Если флажок опции *Итоговая статистика* не установлен, то выводится только то, что задаётся с

помощью опций *Уровень надёжности*, *K-ый наименьший* и *K-ый наибольший*.

Опция *Уровень надёжности* указывает, надо ли вычислять границу доверительного интервала для среднего. В поле ввода рядом с этой опцией задаётся доверительный уровень в процентах.

В полях ввода рядом с опциями *K-ый наибольший* и *K-ый наименьший* указываются порядки выводимых наибольшего и наименьшего значений. Если эти порядки равны 1, то выводятся соответственно максимальное и минимальное выборочные значения.

Для выполнения лабораторной работы используются статистические функции и функции баз данных.

Статистические функции:

– *СРОТКЛ* - возвращает среднее абсолютных значений отклонений точек данных от среднего;

– *СРЗНАЧ* - возвращает среднее арифметическое аргументов;

– *СРЗНАЧА* - возвращает среднее арифметическое аргументов, включая числа, текст и логические значения;

– *БЕТАРАСП* - возвращает интегральную функцию плотности бета вероятности;

– *БЕТАОБР* - возвращает обратную функцию к интегральной функции плотности бета-вероятности;

– *БИНОМРАСП* - возвращает отдельное значение биномиального распределения;

– *ХИ2РАСП* - возвращает одностороннюю вероятность распределения хи-квадрат;

– *ХИ2ОБР* - возвращает обратное значение односторонней вероятности распределения хи-квадрат;

– *ХИ2ТЕСТ* - возвращает тест на независимость;

– *ДОВЕРТИ* - возвращает доверительный интервал для среднего значения по генеральной совокупности;

– *КОРРЕЛ* - возвращает коэффициент корреляции между двумя множествами

данных;

- *СЧЁТ* - подсчитывает количество чисел в списке аргументов;
- *СЧЁТЗ* - подсчитывает количество значений в списке аргументов;
- *СЧИТАТЬПУСТОТЫ* - подсчитывает количество пустых ячеек в заданном

диапазоне;

- *СЧЁТЕСЛИ* - подсчитывает количество непустых ячеек, удовлетворяющих заданному условию внутри диапазона;

– *КОВАР* - возвращает ковариацию, то есть среднее произведений отклонений для каждой пары точек;

– *КРИТБИНОМ* - возвращает наименьшее значение, для которого биномиальная функция распределения меньше или равна заданному значению;

- *КВАДРОТКЛ* - возвращает сумму квадратов отклонений;

- *ЭКСПРАСП* - возвращает экспоненциальное распределение;

- *ФРАСП* - возвращает F-распределение вероятности;

– *ФРАСПОБР* - возвращает обратное значение для F-распределения вероятности;

- *ФИШЕР* - возвращает преобразование Фишера;

- *ФИШЕРОБР* - возвращает обратное преобразование Фишера;

- *ПРЕДСКАЗ* - возвращает значение линейного тренда;

– *ЧАСТОТА* - возвращает распределение частот в виде вертикального массива;

- *ФТЕСТ* - возвращает результат F-теста;

- *ГАММАРАСП* - возвращает гамма-распределение;

- *ГАММАОБР* - возвращает обратное гамма-распределение;

- *ГАММАНЛОГ* - возвращает натуральный логарифм гамма функции, $\Gamma(x)$;

- *СРГЕОМ* - возвращает среднее геометрическое;

- *РОСТ* - возвращает значения в соответствии с экспоненциальным трендом;

- *СРГАРМ* - возвращает среднее гармоническое;

- *ГИПЕРГЕОМЕТ* - возвращает гипергеометрическое распределение;

- *ОТРЕЗОК* - возвращает отрезок, отсекаемый на оси линией линейной

регрессии;

– *ЭКЦЕСС* - возвращает эксцесс множества данных;

– *НАИБОЛЬШИЙ* - возвращает k-ое наибольшее значение из множества данных;

– *ЛИНЕЙН* - возвращает параметры линейного тренда;

– *ЛГРФПРИБЛ* - возвращает параметры экспоненциального тренда;

– *ЛОГНОРМОБР* - возвращает обратное логарифмическое нормальное распределение;

– *ЛОГНОРМРАСП* - возвращает интегральное логарифмическое нормальное распределение;

– *МАКС* - возвращает максимальное значение из списка аргументов;

– *МАКСА* - возвращает максимальное значение из списка аргументов, включая числа, текст и логические значения;

– *МЕДИАНА* - возвращает медиану заданных чисел;

– *МИН* - возвращает минимальное значение из списка аргументов;

– *МИНА* - возвращает минимальное значение из списка аргументов, включая числа, текст и логические значения;

– *МОДА* - возвращает значение моды множества данных;

– *ОТРБИНОМРАСП* - возвращает отрицательное биномиальное распределение;

– *НОРМРАСП* - возвращает нормальную функцию распределения;

– *НОРМОБР* - возвращает обратное нормальное распределение;

– *НОРМСТРАСП* - возвращает стандартное нормальное интегральное распределение;

– *НОРМСТОБР* - возвращает обратное значение стандартного нормального распределения;

– *ПИРСОН* - возвращает коэффициент корреляции Пирсона;

– *ПЕРСЕНТИЛЬ* - возвращает k-ую перцентиль для значений из интервала;

– *ПРОЦЕНТРАНГ* - возвращает процентную норму значения в множестве данных;

- *ПЕРЕСТ* - возвращает количество перестановок для заданного числа объектов;
- *ПУАССОН* - возвращает распределение Пуассона;
- *ВЕРОЯТНОСТЬ* - возвращает вероятность того, что значение из диапазона находится внутри заданных пределов;
- *КВАРТИЛЬ* - возвращает квартиль множества данных;
- *РАНГ* - возвращает ранг числа в списке чисел;
- *КВПИРСОН* - возвращает квадрат коэффициента корреляции Пирсона;
- *СКОС* - возвращает асимметрию распределения;
- *НАКЛОН* - возвращает наклон линии линейной регрессии;
- *НАИМЕНЬШИЙ* - возвращает k -ое наименьшее значение в множестве данных;
- *НОРМАЛИЗАЦИЯ* - возвращает нормализованное значение;
- *СТАНДОТКЛОН* - оценивает стандартное отклонение по выборке;
- *СТАНДОТКЛОНА* - оценивает стандартное отклонение по выборке, включая числа, текст и логические значения;
- *СТАНДОТКЛОНП* - вычисляет стандартное отклонение по генеральной совокупности;
- *СТАНДОТКЛОНПА* - вычисляет стандартное отклонение по генеральной совокупности, включая числа, текст и логические значения;
- *СТОШУХ* - возвращает стандартную ошибку предсказанных значений y для каждого значения x в регрессии;
- *СТЬЮДРАСП* - возвращает t -распределение Стьюдента;
- *СТЬЮДРАСПОБР* - возвращает обратное t -распределение Стьюдента;
- *ТЕНДЕНЦИЯ* - возвращает значения в соответствии с линейным трендом;
- *УРЕЗСРЕДНЕЕ* - возвращает среднее внутренности множества данных;
- *ТТЕСТ* - возвращает вероятность, соответствующую критерию Стьюдента;
- *ДИСП* - оценивает дисперсию по выборке;
- *ДИСКА* - оценивает дисперсию по выборке, включая числа, текст и логические значения;

- *ДИСПР* - вычисляет дисперсию для генеральной совокупности;
- *ДИСПРА* - вычисляет дисперсию для генеральной совокупности, включая числа, текст и логические значения;

- *ВЕЙБУЛЛ* - возвращает распределение Вейбулла;
- *ЗТЕСТ* - возвращает двустороннее Р-значение z-теста.

Функции баз данных:

- *БСЧЁТ* - подсчитывает количество ячеек, содержащих числа;
- *БСЧЁТА* - подсчитывает количество непустых ячеек;
- *ДМАКС* - ищет максимальное значение;
- *ДМИН* - ищет минимальное значение;
- *БДСУММ* - вычисляет сумму числовых значений;
- *БДПРОИЗВЕД* - перемножает числовые значения;
- *ДСРЗНАЧ* - считает среднее значение;
- *ДСТАНДОТКЛ* - оценивает стандартное отклонение;
- *ДСТАНДОТКЛП* - вычисляет стандартное отклонение по генеральной совокупности;
- *БДДИСП* - оценивает дисперсию;
- *БДДИСПП* - вычисляет дисперсию по генеральной совокупности;
- *БИЗВЛЕЧЬ* - ищет одну запись (если критерию удовлетворяют несколько записей, возвращается ошибка #ЧИСЛО!).

Задание на выполнение лабораторной работы.

Для своего варианта выбрать из таблицы 8 совокупность случайных чисел и рассчитать следующие статистики и параметры:

1 Среднее арифметическое значение выборочной совокупности. (Средняя величина. Среднее значение показателя, при вычислении которого общий объём показателя в совокупности остается неизменным. Среднее слагаемое показателя в выборочной совокупности. Средняя арифметическая выборки (выборочная средняя) является одной из оценок математического ожидания генеральной совокупности. Другими оценками математического ожидания могут быть мода и медиана. Формула для вычисления простой средней по негруппированным данным: $\bar{x} = (1/n) \cdot (x_1 + x_2 + \dots$

x_n); $\bar{x} = (1/n) \cdot \sum x_i$; где \bar{x} - средняя величина, x_i - i -тый элемент выборки, $i = 1, 2, 3, \dots, n$ - размер (объём) выборки).

2 Медиану. (Медиана (лат.: medianus - находящийся посреди, средний, центральный;). Медианой m_e называют такое значение признака в изучаемом ряду, которое делит вариационный ряд на две, равные по числу вариантов, части. Если вариационный ряд имеет нечетное число вариантов $n = 2k+1$, то $m_e = x_{k+1}$. При четном $n = 2k$ в качестве медианы обычно принимают $m_e = (x_k + x_{k+1})/2$. Например, для ряда: 2 3 5 6 7 медиана $m_e = 5$; для ряда: 2 3 5 6 7 9 медиана равна $m_e = (5+6)/2 = 5,5$).

3 Минимальное и максимальное значения элементов выборки.

4 Моду. (Мода - (лат.: modus - мера, величина, размеры, предел, образ жизни, правило, предписание;). Модой M_o называют такое значение признака в изучаемом ряду, которое встречается чаще всего (имеет наибольшую частоту). Например, для ряда, где варианты $X: 1, 4, 7, 9$; частоты $f: 5, 1, 20, 6$; мода равна $M_o = 7$. Мода характеризует структуру, строение вариационного ряда. Чаще встречаются ряды с одним модальным значением признака. В этом случае мода может указывать на положение центра распределения, его симметрию. Бимодальные или мультимодальные ряды могут указывать на неоднородность совокупностей, которые они представляют. Это может быть агрегат нескольких совокупностей с разными модами.

5 Среднее геометрическое значение. (Геометрическое среднее – это "итоговая" статистика, полезная при нелинейной шкале измерений, которая вычисляется как: $G = (x_1 \cdot x_2 \cdot \dots \cdot x_n)^{1/n}$, где n - число наблюдений (объем выборки)).

6 Среднее гармоническое значение. (Гармоническое среднее - это итоговая статистика, используемая в анализе частот, которая вычисляется как:

$$1/\bar{x}_H = n \cdot \sum 1/x_i,$$

где n - число наблюдений (объем выборки))

7 Дисперсию генеральной совокупности. Генеральной совокупностью называют совокупность объектов, из которых производится выборка. Генеральная совокупность может содержать конечное число объектов. Однако если это число достаточно велико, то иногда, в целях упрощения вычислений или для облегчения

теоретических выводов, допускают, что генеральная совокупность состоит из бесчисленного множества объектов. Такое допущение оправдывается тем, что увеличение объёма генеральной совокупности практически не сказывается на результатах обработки данных выборки. Генеральная совокупность характеризуется генеральным распределением с генеральными параметрами, например генеральным математическим ожиданием и генеральной дисперсией. Генеральные параметры могут быть оценены по выборочным данным.

8 Дисперсию выборочной совокупности. (Выборочная совокупность. Часть данных генеральной совокупности, отобранных из нее случайным образом.)

9 Сумму квадратов отклонений. (Сумма квадратов отклонений - показатель "полной" вариации количественной переменной относительно одной или нескольких фиксированных величин. Сумма квадратов отклонений значений измеряемой физической величины вычисляется по формуле 5.

10 Ранг числа ХЗ. (Ранг числа - это показатель его величины относительно других значений в списке. **Стандартные отклонения для выборочной и генеральной совокупностей.** (хс. (Стандартное отклонение (термин был впервые введен Пирсоном, 1894) - это широко используемая мера разброса или вариабельности (изменчивости) данных. Стандартное отклонение популяции определяется формулой: $\sqrt{\sum (x_i - \mu)^2 / N}$ - среднее популяции, N - размер популяции. Выборочное стандартное отклонение или оценка стандартного отклонения вычисляется по формуле: $s\mu$, где $(x\sum = \sum [x_i - \bar{x}]^2 / (n-1))^{1/2}$, где \bar{x} - выборочное среднее, n - число наблюдений в выборке. Генеральная совокупность - вся изучаемая выборочным методом статистическая совокупность объектов и/или явлений общественной жизни, имеющих общие качественные признаки или количественные переменные).

11 Из предложенной совокупности образовать вариационный ряд, расположив элементы в порядке их возрастания.

$$Z = \sum_{i=0}^n [f(x_i) - \varphi(x_i)]^2 \rightarrow \min \quad (5)$$

Варианты заданий на выполнение лабораторной работы показаны в таблице 8.

Таблица 8 – Варианты заданий на выполнение лабораторной работы

Номер вар.	X1	X2	X3	X4	X5	X6	X7
1	0,1517	0,2341	0,4451	0,9813	0,2341	0,1496	0,6308
2	1,2237	1,4317	1,9311	1,2237	1,4613	1,9773	1,1698
3	2,1563	2,2416	2,1563	2,8113	2,9224	2,3348	2,6313
4	3,8317	3,5142	3,0580	3,1703	3,0580	3,6812	3,5556
5	4,1122	4,2211	4,1112	4,1551	4,1238	4,1122	4,1233
6	5,1552	5,1463	5,1272	5,1481	5,1463	5,5336	5,2234
7	6,5530	6,5011	6,3088	6,4057	6,3088	6,9907	6,0306
8	7,0803	7,1193	7,9111	7,9422	7,1193	7,5462	7,9936
9	8,5118	8,6157	8,7188	8,7919	8,8834	8,8463	8,5118
10	9,1161	9,1081	9,1183	9,1156	9,1389	9,1518	9,1183
11	10,1553	10,1148	10,2112	10,2814	10,2814	10,1144	10,5534
12	11,4610	11,4513	11,4895	11,4399	11,9915	11,4895	11,5863
13	12,1134	12,1242	12,1388	12,1516	12,5544	12,1242	12,8194
14	13,4851	13,5182	13,8498	13,9552	13,4851	13,8216	13,9437

Содержание отчёта:

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.9 Лабораторная работа № 9. Многомерная обработка данных с использованием интегрированной системы Statistica

Цель работы. Изучить возможности ППП Statistica при установлении зависимости между переменными и оценке характера зависимости. Провести графический анализ данных.

Краткие теоретические сведения.

Совокупность методов оценки корреляционных характеристик и проверка статистических гипотез о них по выборочным данным называется корреляционным анализом. В корреляционном анализе используются следующие основные приемы:

- построение корреляционного поля (диаграммы рассеяния) для двух экономических показателей или двумерных сечений;
- определение выборочных коэффициентов корреляции или составление корреляционных матриц [6];
- проверка статистических гипотез о значимости связи между показателями.

Коэффициент корреляции является мерой линейной зависимости двух величин. Чем больше коэффициент корреляции по модулю, тем сильнее линейная зависимость. Значение коэффициента корреляции лежит в интервале $[-1; 1]$.

Ориентировочно определить значение коэффициента корреляции можно, анализируя диаграмму рассеяния. Чем теснее расположены точки относительно некоторой прямой (существует линейная тенденция), тем больше по абсолютной величине он стремится к единице, и наоборот, чем более расплывчата диаграмма рассеяния, тем ближе к нулю коэффициент корреляции [6].

Задание на выполнение лабораторной работы.

По 14 предприятиям отрасли исследуется зависимость производительности труда (y), (т/час) от уровня механизации работ (x_1) (%), среднего возраста работников (лет) (x_2) и энерговооруженности труда (x_3) (КВт/100 работающих). Провести корреляционный анализ. Отчет представить преподавателю. Данные по предприятиям приведены в таблице 9.

Таблица 9 - Данные по предприятиям

Варианты заданий	
1	2
Вариант 1	Вариант 2
x1 32 30 36 40 41 47 56 54 60 55 61 67 69 76	x1 55 46 40 39 35 29 31 75 68 66 60 54 59 53
x2 33 31 41 39 46 43 34 38 42 35 39 44 40 41	x2 33 42 45 38 40 30 32 40 39 43 38 34 41 37
x3 30 29 35 40 40 48 50 52 59 54 60 70 70 75	x3 50 45 39 40 34 30 30 74 69 66 59 54 60 52
y 20 24 28 30 31 33 34 37 38 40 41 43 45 48	y 33 32 30 29 27 23 19 47 44 42 40 39 37 36

Продолжение таблицы 9

1	2
<p>Вариант 3</p> <p>x1 48 57 55 61 56 62 68 70 77 42 41 37 31 33</p> <p>x2 44 35 39 43 36 40 45 41 42 47 40 42 32 34</p> <p>x3 47 56 54 62 56 62 67 70 76 42 40 37 30 32</p> <p>y 34 35 38 39 41 42 44 46 49 32 31 29 25 21</p>	<p>Вариант 4</p> <p>x1 52 54 45 39 38 34 28 30 74 67 65 59 53 58</p> <p>x2 36 32 41 44 37 39 29 31 39 38 42 37 33 40</p> <p>x3 52 53 45 38 38 34 28 31 73 66 65 60 52 57</p> <p>y 35 32 31 29 28 26 22 18 46 43 41 39 38 36</p>
<p>Вариант 5</p> <p>x1 43 49 58 56 62 57 63 69 71 78 34 32 38 42</p> <p>x2 48 45 36 40 44 37 41 46 42 43 35 33 43 41</p> <p>x3 42 48 58 55 61 56 62 70 70 78 35 32 38 41</p> <p>y 33 35 36 39 40 42 43 45 47 50 22 26 30 32</p>	<p>Вариант 6</p> <p>x1 52 57 51 53 44 38 37 33 27 29 73 66 64 58</p> <p>x2 32 39 35 31 40 43 36 38 28 30 38 37 41 36</p> <p>x3 52 56 50 53 45 37 37 32 28 30 72 66 64 59</p> <p>y 37 35 34 31 30 28 27 25 21 17 45 42 40 38</p>
<p>Вариант 7</p> <p>x1 39 43 44 50 59 57 63 58 64 70 72 79 35 33</p> <p>x2 44 42 49 46 37 41 45 38 42 47 43 44 36 34</p> <p>x3 45 42 50 46 38 40 45 39 41 48 43 44 35 34</p> <p>y 31 33 34 36 37 40 41 43 44 46 48 51 23 27</p>	<p>Вариант 8</p> <p>x1 63 57 51 56 50 52 43 37 36 32 26 28 72 65</p> <p>x2 40 35 31 38 34 30 39 42 35 37 27 29 37 36</p> <p>x3 39 38 35 35 32 31 28 28 25 25 21 15 45 40</p> <p>y 39 37 36 34 33 30 29 27 26 24 20 16 44 41</p>
<p>Вариант 9</p> <p>x1 64 59 65 71 73 80 36 34 40 44 45 51 60 58</p> <p>x2 46 39 43 48 44 45 37 35 45 43 50 47 38 42</p> <p>x3 50 40 50 55 50 60 35 34 42 41 48 49 50 50</p> <p>y 42 44 45 47 49 52 24 28 32 34 35 37 38 41</p>	<p>Вариант 10</p> <p>x1 46 52 61 59 65 60 66 72 74 81 37 35 41 45</p> <p>x2 51 48 39 43 47 40 44 49 45 46 38 36 46 44</p> <p>x3 46 52 60 58 64 61 65 72 74 80 38 34 40 44</p> <p>y 36 38 39 42 43 45 46 48 50 53 25 29 33 35</p>
<p>Вариант 11</p> <p>x1 64 59 65 71 73 80 36 34 40 44 45 51 60 58</p> <p>x2 44 35 39 43 36 40 45 41 42 47 40 42 32 34</p> <p>x3 47 56 54 62 56 62 67 70 76 42 40 37 30 32</p> <p>y 34 35 38 39 41 42 44 46 49 32 31 29 25 21</p>	<p>Вариант 12</p> <p>x1 52 54 45 39 38 34 28 30 74 67 65 59 53 58</p> <p>x2 36 32 41 44 37 39 29 31 39 38 42 37 33 40</p> <p>x3 47 56 54 62 56 62 67 70 76 42 40 37 30 32</p> <p>y 35 32 31 29 28 26 22 18 46 43 41 39 38 36</p>
<p>Вариант 13</p> <p>x1 43 49 58 56 62 57 63 69 71 78 34 32 38 42</p> <p>x2 48 45 36 40 44 37 41 46 42 43 35 33 43 41</p> <p>x3 64 59 65 71 73 80 36 34 40 44 45 51 60 58</p> <p>y 33 35 36 39 40 42 43 45 47 50 22 26 30 32</p>	<p>Вариант 14</p> <p>x1 52 57 51 53 44 38 37 33 27 29 73 66 64 58</p> <p>x2 64 59 65 71 73 80 36 34 40 44 45 51 60 58</p> <p>x3 52 56 50 53 45 37 37 32 28 30 72 66 64 59</p> <p>y 37 35 34 31 30 28 27 25 21 17 45 42 40 38</p>
<p>Вариант 15</p> <p>x1 32 30 36 40 41 47 56 54 60 55 61 67 69 76</p> <p>x2 46 39 43 48 44 45 37 35 45 43 50 47 38 42</p> <p>x3 30 29 35 40 40 48 50 52 59 54 60 70 70 75</p> <p>y 42 44 45 47 49 52 24 28 32 34 35 37 38 41</p>	<p>Вариант 16</p> <p>x1 55 46 40 39 35 29 31 75 68 66 60 54 59 53</p> <p>x2 33 42 45 38 40 30 32 40 39 43 38 34 41 37</p> <p>x3 46 39 43 48 44 45 37 35 45 43 50 47 38 42</p> <p>y 35 32 31 29 28 26 22 18 46 43 41 39 38 36</p>

Порядок выполнения работы.

Рассмотрим построение корреляционной матрицы на примере анализа показателей хозяйственной деятельности 30 предприятий отрасли. Исследуется зависимость выработки продукции на 1 работника (y), тыс.руб. от ввода в действие основных фондов в % от стоимости фондов на конец года (x₁), от удельного веса (%) рабочих высокой квалификации в общей численности рабочих (x₂), среднего

возраста работников (лет) (x_3). Проведем графический анализ полученной матрицы, где данные по предприятиям приведены в таблице на рисунке 27 [6].

	1 Выработка	2 Фонды	3 Рабочие	4 Средний возраст
1	7	3,9	10	33
2	7	3,9	14	44
3	7	3,7	15	40
4	7	4,0	16	35
5	7	3,8	17	38
6	7	4,8	19	46
7	8	5,4	19	42
8	8	4,4	20	37
9	8	5,3	20	35
10	10	6,8	20	33
11	9	6,0	21	44
12	11	6,4	22	42
13	9	6,8	22	48
14	11	7,2	25	42
15	12	8,0	28	42
16	12	8,2	29	40
17	12	8,1	30	40
18	12	8,5	31	39
19	14	9,6	32	38
20	14	9,0	36	38
21	10	6,4	18	33
22	10	6,1	22	33
23	11	6,3	24	32
24	12	6,5	26	30
25	7	4,5	13	38
26	12	7,8	28	40
27	11	7,8	26	41
28	9	5,8	19	41
29	16	10,2	40	48
30	14	9,8	34	46

Рисунок 27 - данные по предприятиям

Шаг 1. Запустите программу Statistica. Создайте таблицу данных. Переключитесь в модуль **Статистика/Основные статистики**.

Шаг 2. В стартовой панели модуля **Основные статистики** выберите пункт **Correlation matrices** (Корреляционные матрицы). Откроется окно **Product-Moment Correlation** (Корреляция Пирсона), показанное на рисунке 28.

Нажмите на кнопку **Two lists** (Два списка), откроется окно выбора переменных. Выберите переменные 2-4 в **First variables list** (Первый список и переменную 1 в **Second variables list** (Второй список переменных). Подсчитаем корреляции между переменной 1 (результат) и переменными 2-4 (факторами). Нажмите в окне кнопку **ОК**, вернетесь в предыдущее окно.

Шаг 3. В окне **Product-Moment Correlation** нажмите кнопку **Summary**. На

экране вы увидите корреляционную матрицу, показанную на рисунке 29.

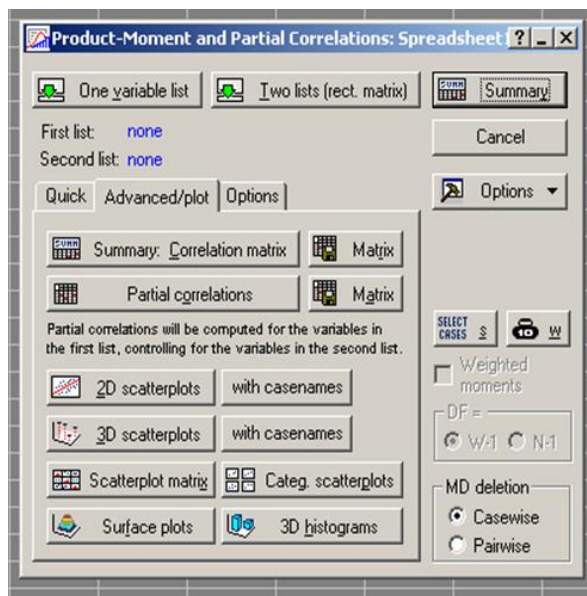


Рисунок 28 - Окно **Product-Moment Correlation** (Корреляция Пирсона)

Variable	Выработка
Фонды	0,96
Рабочие	0,95
Средний возраст	0,15

Рисунок 29 - Корреляционная матрица

В этой матрице имеется только один столбец, так как во втором списке выбрана одна зависимая переменная. Красным цветом автоматически выделены коэффициенты, значимые на уровне $p < 0,05$. Именно на эти коэффициенты следует обратить наибольшее внимание. Грубо говоря, зависимость между переменными с выделенными красным цветом коэффициентами корреляции наиболее значимая. В нашем случае переменная 1 (Выработка) чуть более зависима от переменной 2 (Фонды). Хотя от переменной 3 (Рабочие) установлена также тесная зависимость, чего нельзя сказать о переменной 4 (Средний возраст). Все три коэффициента

корреляции положительны, значит делаем вывод, что при увеличении каждой из переменных 2-4 увеличивается переменная 1. Просмотрим зависимость между переменными 1 и 2, 1 и 3 графически [6].

Шаг 4. В вернитесь в окно **Product-Moment Correlation**. Нажмите на кнопку **Two lists**. Откроется окно выбора переменных. Выберите в первом списке переменную 2, во втором переменную 1 (связь 2-1) и нажмите кнопку **ОК**. Перейдите в окне **Product-Moment Correlation** к вкладке **Advanced/plot**. Нажмите кнопку **2D scatterplot** (2D диаграмма рассеяния). Появится окно диаграммы рассеяния, как на рисунке 30, для выбранных переменных.

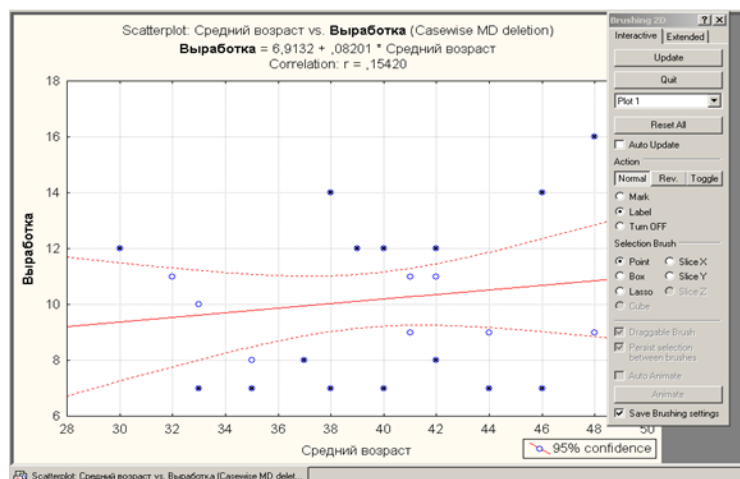


Рисунок 30 - Окно диаграммы рассеяния для выбранных переменных

Из графика отчетливо видно, что зависимость линейная, за исключением некоторых точек, отдаленных от линии. На графике системой дана лучшая прямая. При смене коэффициента наклона подгонка будет хуже. Постройте графики для двух других переменных.

Для случая связи 4-1 график выглядит следующим образом, как показано на рисунке 31.

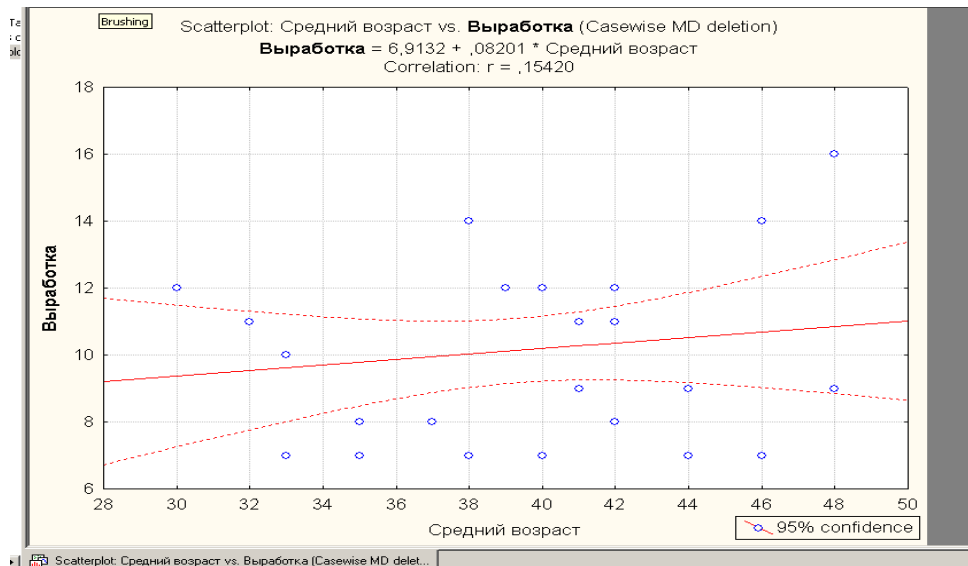


Рисунок 31 – Результат в виде нелинейной зависимости

Содержание отчёта.

- 1 Цель работы.
- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.10 Лабораторная работа № 10. Компонентный анализ

Цель работы. Освоение методов построения регрессионных моделей на основе использования компонентного анализа.

Краткие теоретические сведения.

Компонентный анализ проводится с несколькими частными целями. Как метод снижения размерности он позволяет выявить закономерности, которые непосредственно не наблюдаются. Эта задача решается по матрице нагрузок, как и классификация признаков в пространстве главных компонент. А индивидуальные значения используются для классификации объектов (не по исходным признакам, а

программ статистического анализа, например Statgraphics Plus.

Программа выдает матрицу коэффициентов корреляции A между центрированно-нормированными исходными переменными и ненормированными главными компонентами размерностью $(n \times k)$ [4].

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1k} \\ a_{21} & a_{22} & \dots & a_{2k} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nk} \end{pmatrix} \quad (7)$$

Коэффициенты показывают наличие, силу и направление линейной связи между соответствующими исходными переменными $x_1 - x_n$ и соответствующими главными компонентами $z_1 - z_k$. Уравнение в матричной форме записи принимает вид:

$$Z = X A \quad (8)$$

Программа выдает матрицу наблюдаемых значений главных компонент Z размерностью $(n \times k)$:

$$Z = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1k} \\ z_{21} & z_{22} & \dots & z_{2k} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nk} \end{pmatrix} \quad (9)$$

Используя значения главных компонент, построим модель главных компонент:

$$x_{ji} = a_{1i} z_{j1} + a_{2i} z_{j2} + a_{3i} z_{j3} + \dots + a_{ki} z_{jk}, \quad j=1, 2, \dots, N, \quad i = 1, 2, \dots, n. \quad (10)$$

Для интерпретации используются, как правило, наиболее весомые главные компоненты. Уравнение регрессии на выделенные главные компоненты строится методом множественной регрессии. Оценивается значимость уравнения регрессии в целом и значимость коэффициентов регрессии при главных компонентах

$$y = b_0 + b_1 z_1 + b_2 z_2 + \dots + b_k z_k. \quad (11)$$

Подставляя в полученное уравнение значения главных компонент $z_1 - z_k$,

выраженные через центрированные переменные $x_1 - x_n$, получаем окончательное уравнение регрессии:

$$y = c_0 + c_1x_1 + c_2x_2 + \dots + c_nx_n. \quad (12)$$

Это уравнение отличается более высокой точностью, чем получаемое с использованием классического метода наименьших квадратов [4].

Задание на выполнение лабораторной работы:

- 1 Получить задание от преподавателя на выполнение лабораторной работы.
- 2 Ввести исходные данные в ППП Statgraphics Plus или в другой пакет, например Statistica.
- 3 Оценить мультиколлениарность факторных переменных и сделать вывод о целесообразности построения модели на главных компонентах.
- 4 Рассчитать главные компоненты.
- 5 Построить уравнение в главных компонентах и оценить его адекватность.
- 6 Построить уравнение регрессии в исходных факторных переменных.
- 7 Сравнить точность регрессионной модели с моделью на главных компонентах. Сделать выводы по результатам исследований.

Порядок выполнения работы:

- 1 Подготовить данные для расчетов в ППП Statgraphics Plus либо в другой программе.
- 2 Проверить независимость (немультиколлениарность) факторных переменных.
- 3 Выделить главные компоненты, построить уравнения главных компонент.
- 4 Построить уравнение регрессии на выделенных главных компонентах методом множественной регрессии.
- 5 Оценить значимость уравнения регрессии в целом и коэффициентов регрессии при главных компонентах.
- 6 Сделать выводы по результатам исследований [4].

Содержание отчёта:

- 1 Цель работы.

- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

2.11 Лабораторная работа № 11. Работа со списками в MS Excel

Цель работы. Изучить основные понятия списков и освоить технологию создания списков в Excel, поиска записей, сортировки.

Краткие теоретические сведения.

Понятие списка и базы данных в Excel. Для анализа и обработки данных в Microsoft Excel имеется специальный инструментарий, позволяющий структурировать и хранить данные в связанном виде, а также составлять и получать различные отчеты. При этом таблицы данных должны быть представлены в виде списка или базы данных.

Списком называются таблицы, содержащие уникальные записи (имена полей) в первой строке.

В терминологии Excel понятия «список» и «база данных» являются синонимами.

Строки таблицы называются *записями* базы данных, а столбцы — *полями*.

Начните создавать список, аналогичный представленному на рисунке 32, заполнив только первые 3 записи, № личного дела не заполняйте [5].

	A	B	C	D	E	F	G
1	СВЕДЕНИЯ О СОТРУДНИКАХ ПРЕДПРИЯТИЯ						
2							
3	№ личного дела	Фамилия, имя, отчество	Должность	Пол	Дата рождения	Оклад	Образование
4	1264	Баранов Г.А.	Коммерческий директор	Мужской	27.12.1951	40000	высшее
5	1265	Вилков С.В.	Менеджер по рекламе	Мужской	28.06.1981	15000	высшее
6	1266	Голубева М.С.	Бухгалтер	Женский	02.02.1980	25000	высшее
7	1267	Маркова О.И.	Секретарь-референт	Женский	01.05.1985	14000	незаконченное высшее
8	1268	Петров В.А.	Начальник 1 отдела	Мужской	08.10.1979	25000	высшее
9	1269	Чесноков К.А.	Начальник 2 отдела	Мужской	07.05.1977	25000	высшее
10	1270	Петранов В.А.	Финансовый директор	Мужской	03.04.1956	40000	высшее
11	1271	Волков А.П.	Инженер	Мужской	06.08.1976	17000	высшее
12	1272	Ломов С.В.	Водитель	Мужской	08.05.1984	12000	среднее специальное
13	1273	Петросян А.П.	Библиотекарь	Мужской	03.06.1976	12000	высшее
14	1274	Баннов Р.П.	Программист	Мужской	10.08.1984	38000	высшее
15	1275	Маркова М.И.	Зам. директора	Женский	09.09.1982	28000	высшее
16	1276	Чепалова П.Н.	Старший экономист	Женский	25.06.1980	18000	высшее
17	1277	Барсуков И.А.	Менеджер по продажам	Мужской	29.09.1979	14000	высшее
18	1278	Голубинская А.М.	Менеджер по продажам	Женский	19.10.1978	14000	высшее
19	1279	Иванов С.В.	Дворник	Мужской	14.12.1948	10000	среднее специальное
20	1280	Егопцев Е.А.	Охранник	Мужской	08.01.1961	11000	среднее специальное
21	1281	Попов И.А.	Старший инженер	Мужской	17.11.1956	24000	высшее
22	1282	Зимин О.Н.	Курьер	Мужской	07.10.1986	10000	незаконченное высшее
23	1283	Александров А.А.	Экономист	Мужской	16.05.1968	15000	высшее
24	1284	Губина П.В.	Кассир	Женский	05.05.1955	12000	высшее
25	1285	Путин В.В.	Генеральный директор	Мужской	01.01.1970	50000	высшее

Рисунок 32 - Окно Excel со списком – сведениями о сотрудниках предприятия

Список содержит информацию о каждом сотруднике: *номер личного дела, ФИО, должность, пол, год рождения* и т.д. – это поля списка. Так, в строке 5 показана запись на сотрудника с личным номером 1265, а в строке 8 – на сотрудника с личным номером 1269.

При создании списков следует придерживаться следующих правил:

- на одном рабочем листе следует размещать только один список;
- список отделяется от других данных по крайней мере одной пустой строкой и одним пустым столбцом;
- каждый столбец списка должен содержать однородную информацию;
- первая строка таблицы должна содержать имена полей списка, описывающих назначение соответствующего столбца;
- необходимо избегать пустых строк и столбцов внутри списка.

Над созданными списками Excel можно выполнять следующие операции [5]:

- осуществлять поиск данных по заданным критериям;
- выполнять сортировку записей;

- фильтровать список;
- подводить промежуточные итоги;
- создавать итоговые таблицы данных.

Технология создания списков.

Создание списка начинается с ввода наименований полей списка (строка 3 рисунка 32). Заполнение записей списка — ввод данных в их поля (строки А — 23 рисунка 32) осуществляется двумя способами: *вручную*, как и при подготовке небольших таблиц, и с использованием специальной встроенной *формы данных*.

Форма данных – это наиболее удобный способ для просмотра, изменения, добавления, удаления записей списка, а также для поиска записей, удовлетворяющих заданным условиям.

Форма данных, используемая при создании списков, а также при их обработке, вызывается следующими командами:

Установить курсор в пределах списка»¹ → Данные → Форма

Если выполнен только ввод наименований полей списка, для вызова формы следует выделить эти поля.

В результате появляется диалоговое окно формы с пользовательскими текстовыми полями для ввода значений полей списка, как на рисунке 33. Имя окна формы соответствует имени листа, на котором создается список.

Введите оставшиеся записи списка с помощью формы данных.

Рисунок 33 - Диалоговое окно формы для работы со списками Excel

Он меняется в зависимости от ситуации обработки записей. Краткая информация о кнопках представлена в таблице 10 [5].

Таблица 10 - Назначение кнопок формы работы со списками

Кнопка	Назначение
Добавить	Открытие пустой формы для добавления новой записи
Удалить	Удаление текущей записи из списка
Вернуть	Восстановление измененных значений поля записи
Назад	Переход к предыдущей записи; возврат из режима задания критерия
Далее	Переход к следующей записи
Критерии	Переход в режим задания критерия поиска данных в списке
Заккрыть	Закрытие окна формы
Очистить	Удаление данных поля (в режиме задания критерия)
Вернуть	Восстановление данных поля (в режиме критерия)
Правка	Переход к редактированию содержимого полей записи

Для быстрого заполнения базы можно использовать команду меню Правка → Заполнить.

Более широкие возможности автозаполнения открываются с использованием команд меню. Здесь можно реализовать не только последовательности по правилам арифметической прогрессии, но и геометрической, в том числе с указанием граничного значения. Схема автозаполнения в том случае может быть представлена следующим образом:

Выделить ячейку с начальным значением последовательности → Команда Правка → Заполнить → Прогрессия...¹ → Задать расположение последовательности: по строкам или по столбцам → Задать значение шага → Выбрать тип последовательности → Для последовательности Даты выбрать единицы измерения: день, рабочий день, месяц, год → Задать предельное значение последовательности → ОК

Последующие операции выполняются в окне Прогрессия, изображенном на рисунке 34.

Автозаполнение, реализованное с помощью команд меню, целесообразно применять, например, при формировании графиков дежурств (работников охраны,

специалистов по отделу и т.п.).

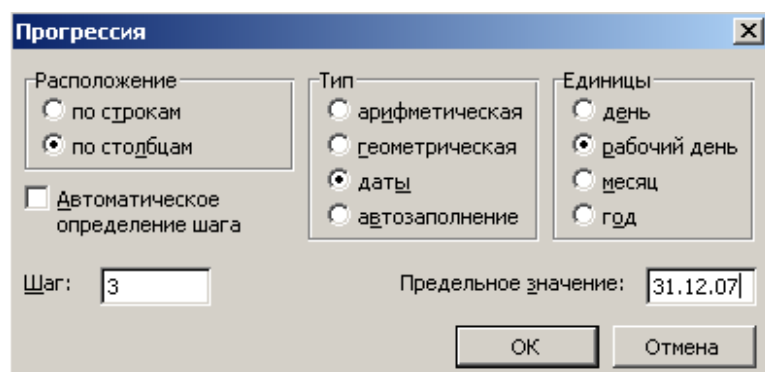


Рисунок 34 - Диалоговое окно задания параметров последовательности

Заполните с помощью вышеописанной команды поле «№ личного дела».

Еще одна возможность автоматизации заполнения списков открывается за счет обеспечения ускоренного ввода требуемых данных, введенных в столбце выше, только по их первым символам.

Возможность достигается, если в окне Параметры, как показано на рисунке 35, установлен флажок Автозавершение значений ячеек (Команда Сервис → Параметры... → Вкладка Правка).

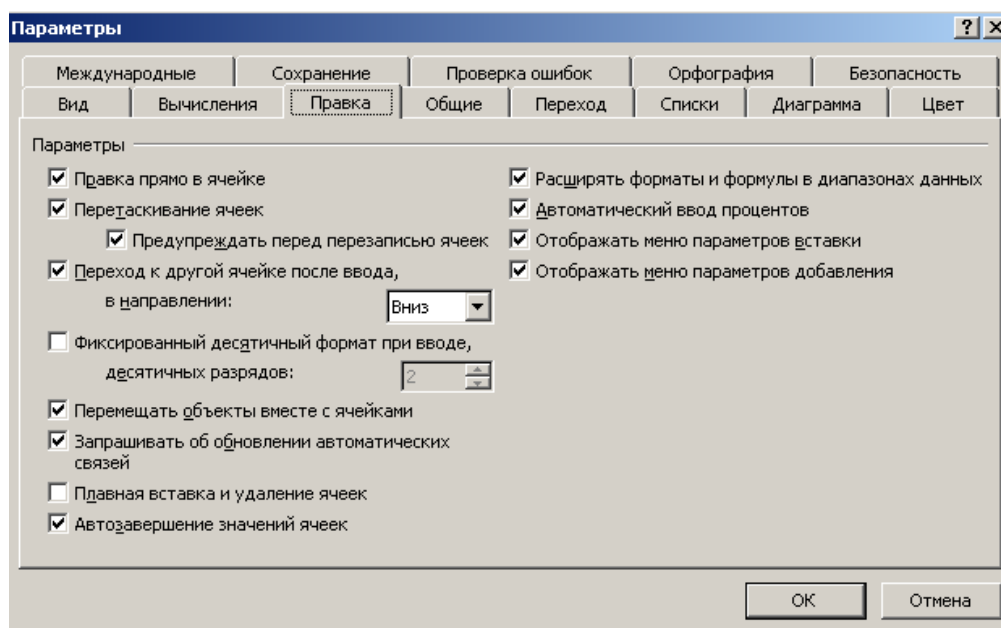


Рисунок 35 - Диалоговое окно задания параметров правки данных

При работе со списками можно задать условия, позволяющие осуществлять проверку данных, вводимых в ячейки, с целью недопущения ошибок. Диалоговое окно задания параметров показано на рисунке 36.

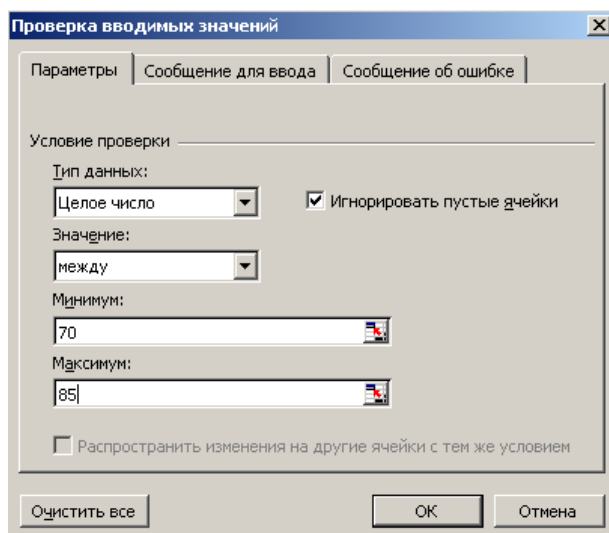


Рисунок 36 - Диалоговое окно задания параметров проверки вводимых значений в ячейки таблицы

Условия для определенного диапазона ячеек, или критерий правильности ввода, реализуются следующей схемой [5]:

Выделить ячейки → Команда Данные → Проверка → Вкладка Параметры → Выбрать из списка тип данных → Задать интервал значений → Вкладка Сообщение для ввода → Задать параметры подсказки → Вкладка Сообщение об ошибке → Выбрать вид действия при попытке ввода неверных данных → Ввести текст сообщения об ошибке → ОК.

Критерий правильности ввода обязательно следует задавать для столбцов с критически важными значениями, для которых данные не должны выходить за установленные пределы.

Задание и выполнение лабораторной работы.

Задайте проверку вводимых значений для оклада сотрудников предприятия от 10000 до 50000 [5].

Следует отметить, что если поле списка содержит формулу, то в форме выводится ее результат. Изменять это поле в форме нельзя. При изменении записи, содержащей формулу, результат формулы не будет вычислен до нажатия клавиши **Enter** или кнопки **Заккрыть**. Диалоговое окно формирования подсказки на рисунке 37.

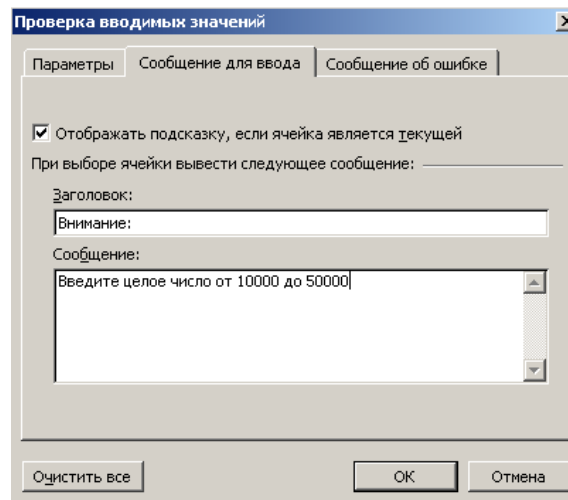


Рисунок 37 - Диалоговое окно формирования подсказки, предваряющей ввод данных в ячейку

Диалоговое окно формирования сообщения об ошибке на рисунке 38.

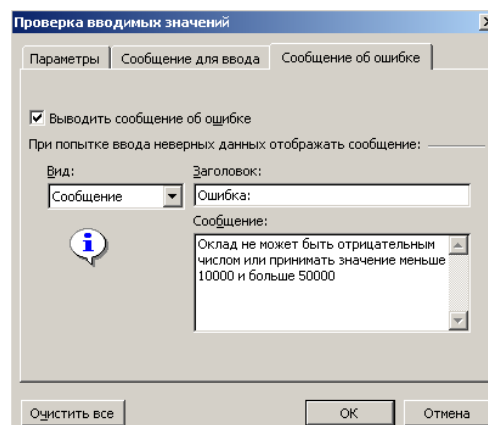


Рисунок 38 - Диалоговое окно формирования сообщения об ошибке при введении данных в ячейку таблицы

Поиск записей.

Перед началом поиска следует обратить внимание на то, чтобы табличный курсор находился в пределах списка, например на его первой записи. Поиск записей осуществляется из диалогового окна формы по нажатию кнопки Критерии. В результате Excel очистит все поля в форме данных и заменит номер записи словом Критерии для того, чтобы осуществить ввод критерия поиска в чистые текстовые поля [5].

Критерий поиска вводится в одно или несколько полей, по которым нужно найти совпадение. Например, из списка требуется найти сотрудника, фамилия которого начинается на букву «М» и оклад более 15 000 руб.

Просмотр результатов поиска осуществляется по кнопке Далее. При этом Excel отобразит форму данных с первой найденной записью.

Просмотр следующей записи также выполняется по кнопке Далее. Для возврата к предыдущей записи используется кнопка Назад. При необходимости выполнить поиск записей списка, удовлетворяющих новому критерию, следует из окна формы выбрать операции: *Кнопка Критерии* → *Кнопка Очистить* → *Ввести новый критерий* → *Кнопка Далее*.

Сортировка записей.

Сортировка — упорядочивание информации в списке в соответствии со значением или с типом информации.

Excel предоставляет многочисленные способы сортировки записей списка. Возможна сортировка записей по возрастанию или убыванию, по любому одному или любым нескольким полям любого интервала ячеек, а также задавая пользовательский порядок сортировки. Таким образом, гибкость упорядочения записей – это то, что требуется для удовлетворения различных нужд при работе с данными. Именно это и позволяет делать команда Сортировка.

Следует различать сортировку, производимую по одному столбцу, и сортировку по нескольким столбцам.

В первом случае записи сортируются на основании выбранного поля по возрастанию (или убыванию) значений в этом поле независимо от места

расположения данного поля. Окно формы в ходе задания критерия выборки на рисунке 39.

Список

№ личного дела:

Фамилия, имя, отчество:

Должность:

Пол:

Дата рождения:

Оклад:

Образование:

Критерии

Добавить

Очистить

Вернуть

Назад

Далее

Правка

Заккрыть

Рисунок 39 - Окно формы в ходе задания критерия выборки записей из списка

Окно формы с записью списка показано на рисунке 40.

Список

№ личного дела:

Фамилия, имя, отчество:

Должность:

Пол:

Дата рождения:

Оклад:

Образование:

4 из 22

Добавить

Удалить

Вернуть

Назад

Далее

Критерии

Заккрыть

Рисунок 40 - Окно формы с записью списка, удовлетворяющей критерию

Во втором случае записи упорядочиваются по более сложному алгоритму, т.е. сначала строки списка группируются по значениям одного столбца, а внутри каждой группы строки, относящиеся к одному значению, располагаются по значениям второго столбца и т.д.

Для того чтобы выполнить сортировку записей заданного списка, сначала необходимо выделить в нем любую ячейку и вызвать команду Сортировка из меню команды Данные. Excel автоматически определит размер списка, выделит ячейки и *откроет диалоговое окно настройки сортировки*, как на рисунке 41, в котором следует указать, какие поля будут определять новый порядок записей, а затем указать тип упорядочения для этих полей [5].

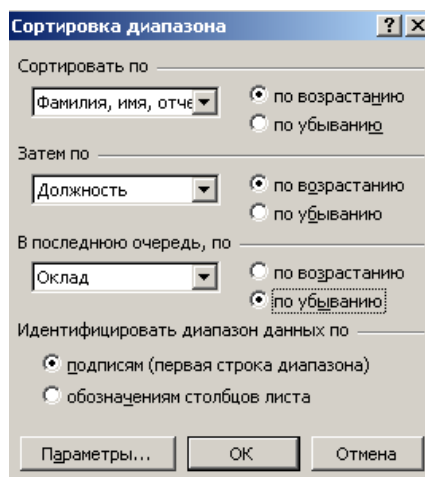


Рисунок 41 - Окно задания параметров сортировки диапазона

Существует два типа упорядочения:

– *по возрастанию* – числа располагаются в порядке увеличения от наименьшего к наибольшему; *текст* сортируется в алфавитном порядке; *даты* – от наиболее ранней даты к самой поздней дате;

– *по убыванию* – числа сортируются от наибольшего к наименьшему; *текст* – в порядке, обратном алфавитному; *даты* – от самой поздней даты к наиболее ранней дате.

Excel дает возможность отсортировать *данные списка* за один раз не более чем по трем полям, установив порядок по возрастанию или убыванию для каждого поля отдельно.

Содержание отчёта:

1 Цель работы.

- 2 Задание.
- 3 Результаты выполнения работы.
- 4 Выводы.

Список использованных источников

- 1 Борздова, Т. В. Основы статистического анализа и обработка данных с применением Microsoft Excel : учеб. пособие / Т. В. Борздова. – Минск : ГИУСТ БГУ,
- 2 Задорина, Н. А. Обработка экспериментальных данных на ЭВМ: Учебное пособие. – Рыбинск: РГАТА, 2009. – 100 с.
- 3 Леонова, Н.Л. Элементы численных методов в Excel: учебно-методическое пособие по выполнению лабораторных работ /СПб ГТУРП.-СПб., 2012. - 47 с.
- 4 Макаров, Р.И. Модели и методы планирования экспериментов, обработки экспериментальных данных : метод. указания к лаб. работам / сост.: Р. И. Макаров, Е. Р. Хорошева; Владим. гос. ун-т имени Александра Григорьевича и Николая Григорьевича Столетовых. – Владимир: Изд-во ВлГУ, 2013. – 61 с.
- 5 Поликарпов, В.М. Современные методы компьютерной обработки экспериментальных данных : учебное пособие / В.М. Поликарпов, И.В. Ушаков, Ю.М. Головин. – Тамбов : Изд-во Тамб. гос. техн. ун-та, 2006. – 84 с. ISBN 5-8265-0498-6.
- 6 Шашков, В.Б. Обработка экспериментальных данных и построение эмпирических формул. Курс лекций. : Учебное пособие.- Оренбург: ГОУ ОГУ, 2005. – 150 с.
- 7 Шевелев, Г. Е. Информатика: лабораторный практикум: Учеб. пособие/ Том. политехн. ун-т. – Томск, 2004. - 118 с.