Лабораторная работа № 1

1.1 Ознакомление с программой Excel 97

Цель работы: Научиться производить вычисления, создавать таблицы и строить графики при помощи программы.

1.2 Теоретические сведения

Excel – одна из самых мощных программ для создания электронных таблиц и работы с ними. Сильная сторона Excel не только в ее способности выполнять различные вычисления, это можно сделать и с помощью простого калькулятора, а то, что Excel позволяет проводить глубокий анализ данных и получать в результате новую полезную информацию.

В каждой новой версии компания Microsoft предлагает дополнительные возможности и совершенствует старые, чтобы облегчить работу пользователей с Excel. Excel 97 существенно усовершенствована по сравнению с Excel 5.0 и Excel 95. среди самых важных усовершенствований можно назвать улучшение работы с графиками и вводом данных; кроме того, добавлены новые инструменты графического редактирования и интеграции с Internet, включая возможности использования Excel 97 при создании Web-публикаций.

Элементы главного окна Excel.

В главное окно Excel 97 добавлено несколько новых элементов (рисунок 1.1). В строке меню и на панелях инструментов есть маркеры для их перемещения. Кнопки панели инструментов и пункты меню стали более анимационными. Когда вы наводите указатель мыши на кнопку, вокруг нее появляется обрамление, выделяющее активный объект в строке других кнопок. В панель формул добавлена новая кнопка *Изменить формулу* для включения режима редактирования формул. Подсказки, которые существовали только для кнопок панели инструментов, теперь используются для определения других элементов окна: поле *Имя*, кнопка *Изменить формулу* или кнопка *Свернуть окно*. Знакомство с различными элементами

окна Excel и умение ими пользоваться поможет в работе с программой.

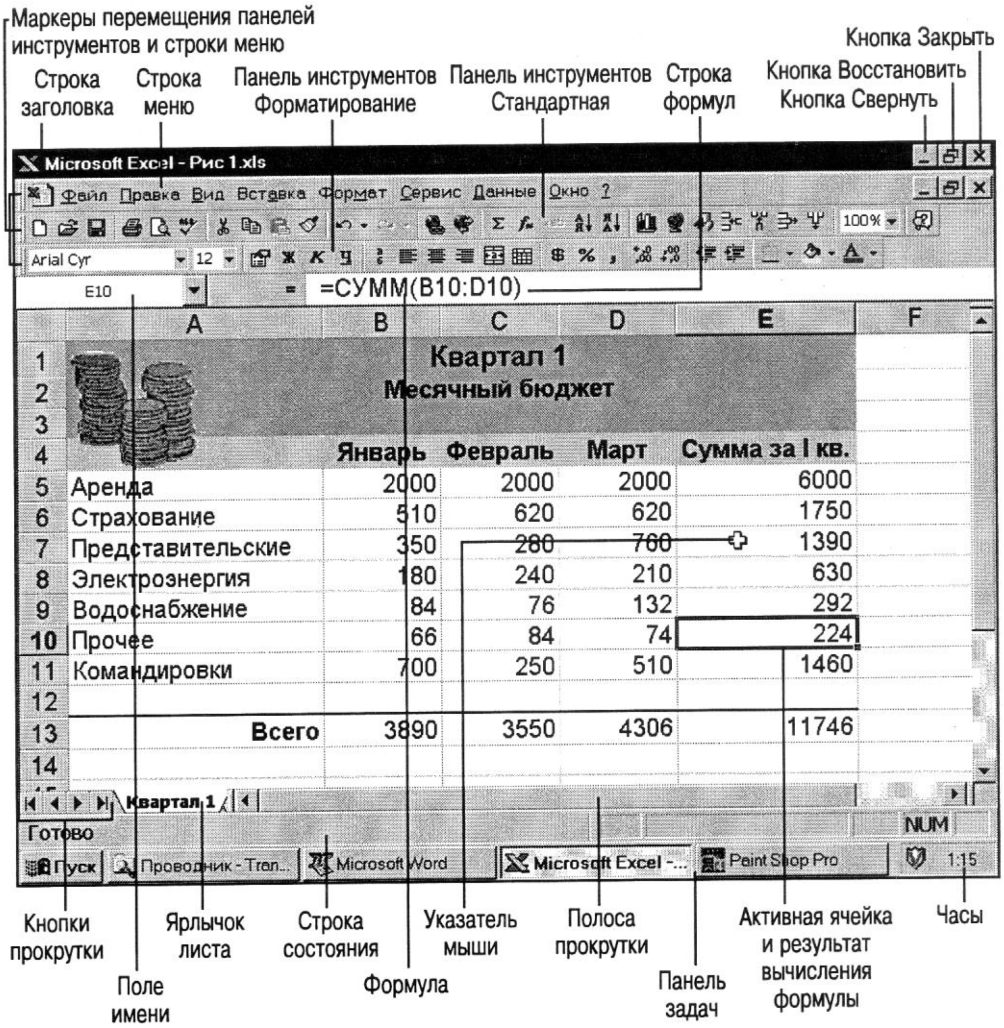


Рисунок. 1.1 - Элементы главного окна Excel

Проектирование эффектных электронных таблиц.

Перед тем как начать ввод данных, продумайте макет своей электронной таблицы, в результате вы сэкономите много сил и времени. При проектировании электронных таблиц обратите внимание на ряд общих моментов. Заголовки в таблице должны быть отделены от данных несколькими пустыми строками. Каждая строка и каждый столбец также имеют свои заголовки. Заголовки столбцов обычно располагаются над данными таблицы, а заголовки строк — слева от данных. На рисунке 1.2 показан оптимальный макет электронной таблицы.

Между данными и формулами нужно оставлять пустую строку, которую следует включать в ссылку формулы наряду с данными таблицы.

Рисунок 1.2 - Оптимальный макет электронной таблицы

Различные способы ввода данных в рабочие листы.

В программе Excel используется два типа данных — константы и формулы. *Константы* включают в себя текст, числа, даты и время (по сути все, кроме формул) и при вводе отображаются в ячейке таблицы. *Формулы* задают алгоритм вычислений, результаты которых отображаются в ячейках; сама формула, по которой происходит вычисление результата, отображается в строке формул (см. рисунок 1.1).

Если необходимо ввести в ячейки рабочего листа последовательность данных, то предварительно выделите эти ячейки. После ввода первого значения активной автоматически станет следующая ячейка в выбранном диапазоне.

Существует ряд приемов, позволяющих выделять несколько ячеек или диапазоны ячеек с помощью клавиатуры или мыши.

• Диапазон ячеек

С помощью мыши. Если необходимо выделить несколько смежных ячеек, поместите указатель мыши на первой ячейке в этой группе. Нажмите и не отпускайте кнопку мыши и переместите указатель с первой ячейки диапазона на последнюю ячейку либо выберите первую ячейку

диапазона, нажмите <Shift> и, удерживая ее, щелкните на последней ячейке диапазона. Независимо от того, каким способом выбраны ячейки, они подсвечиваются черным цветом, а вокруг всего диапазона появится толстая граница. Первая ячейка диапазона не будет подсвечена. Таким образом выделится активная ячейка диапазона, в которую будут вводиться данные.

С помощью клавиатуры. Клавишами со стрелками переведите указатель активной ячейки (активная ячейка выделяется толстой границей) на первую ячейку необходимого вам диапазона. Удерживая клавишу <Shift> и используя клавиши со стрелками, расширьте область выделения.

• Несколько диапазонов ячеек

С помощью мыши. Чтобы выбрать несколько диапазонов, можно использовать описанный выше метод выделения одного диапазона с помощью мыши. Выберите первый диапазон ячеек, а затем, удерживая клавишу <Ctrl>, следующие диапазоны.

С помощью клавиатуры. Выберите первый диапазон методом, описанным выше. Затем нажмите комбинацию клавиш <Shift+F8>, после чего на панели состояния появится надпись ДОБ; этот индикатор говорит о том, что к выбранному диапазону сейчас можно добавить дополнительный диапазон. Далее, с помощью клавиш со стрелками переведите указатель активной ячейки на начало следующего диапазона. Удерживая клавишу <Shift> и применяя клавиши со стрелками, выберите новый диапазон. Выбор каждого нового диапазона следует начинать с комбинации <Shift+F8>.

Выделив ячейки, которые содержат значения, приступайте к вводу данных. После ввода каждого значения в текущую ячейку автоматически активной становится следующая ячейка в диапазоне. После того как данные будут введены во все выбранные ячейки, активной снова становится

первая ячейка диапазона.

После выделения диапазона для перемещения по ячейкам следует использовать только клавиши, представленные в таблице 1.1. Если нажать любую другую клавишу, например со стрелкой, выделение ячеек будет отключено.

Таблица 1.1 - Клавиши для перемещения по ячейкам выделенного диапазона

Перемещение активной ячейки	Клавиши
Вниз	<Enter>
Вверх	<Shift+Enter>
Вправо	<Tab>
Влево	<Shift+Tab>

Работа с формулами.

Обработка чисел — одна из главных задач Excel. Можно создавать простые формулы, где складываются последовательности чисел, вычисляется разность двух чисел или определяется процент, составляющий конкретное значение от общей суммы значений. Кроме того, в Excel существует больше 320-ти функций для выполнения сложных финансовых вычислений, например функция возвращения полного объема вклада при ежемесячных платежах по 100 руб. или функция, автоматически возвращающая стоимость товара по его названию из прейскуранта для заполнения счета или формы заказа.

Существует ряд общих правил, о которых следует помнить при создании формул.

• Формулы начинаются со знака равенства (=).

• В формулу может входить до 1024-х символов.

• Результат вычисления формулы отображается в рабочем листе; формула, по которой происходят вычисления, отображается в строке формул.

• В формулах можно использовать ссылки на ячейки.

Когда необходимо включить в формулу ссылку на ячейку, часто

легче указать на ячейку мышью, чем ввести адрес этой ячейки с клавиатуры. Использование этого метода поможет также избежать ошибок при вводе.

Чтобы добавить в формулу ссылки на ячейки, выполните следующие действия.

1. Выберите ячейку, в которой нужно показать результат вычисления.

2. Введите знак равенства (=), с которого начинается каждая формула.

3. Чтобы ввести ссылку на ячейку, щелкните на ячейке мышью или выберите ее с помощью клавиш со стрелками. Вокруг ячейки появится мерцающая граница из подвижных точек, а в создаваемой формуле появится ссылка.

4. Введите следующую часть формулы, например арифметический оператор, и продолжайте строить формулу.

5. Нажмите клавишу <Enter> или щелкните на кнопке Enter в строке формул, чтобы закончить ввод формулы. Результат ввода формулы будет отображен в рабочем листе, а формула появится в строке формул.

Если необходимо применить одну из функций Excel, вам поможет кнопка *Вставка функции*  (которая в предыдущих версиях Excel называлась мастером функций) на панели инструментов Стандартная.

Порядок операций при вычислениях.

При вычислении формул в Excel применяются известные математические операции. В таблице 1.2 они представлены в порядке убывания их приоритета.

Запись в формуле	**Описание**
( )	Скобки (действия в скобках выполняются первую очередь)
^	Возведение в степень (крыша)
*	Умножение (звездочка)
/	Деление (слеш)

+	Сложение (плюс)
-	Вычитание (дефис)

Копирование формул.

Довольно часто в электронных таблицах в смежных ячейках используются одни и те же формулы для вычисления значений. Например, на рисунке 1.3 вычисляется разность между продажами в 1997 и 1998 годах по каждому напитку. Для каждого напитка схема вычислений одна и та же — продажи в 1998 году минус продажи в 1997 году. Если в смежных ячейках используется однотипная формула, то формулу следует создать только в первой ячейке, а затем просто скопировать ее во все остальные с помощью маркера заполнения (рисунок 1.3).

Маркер заполнения может выполнять две функции: создавать последовательность и копировать содержимое ячеек. При использовании средства *Автозаполнение* формат первой ячейки будет скопирован в новые ячейки вместе с формулой.

При копировании формул важно учитывать разницу между относительными и абсолютными адресами ячеек.

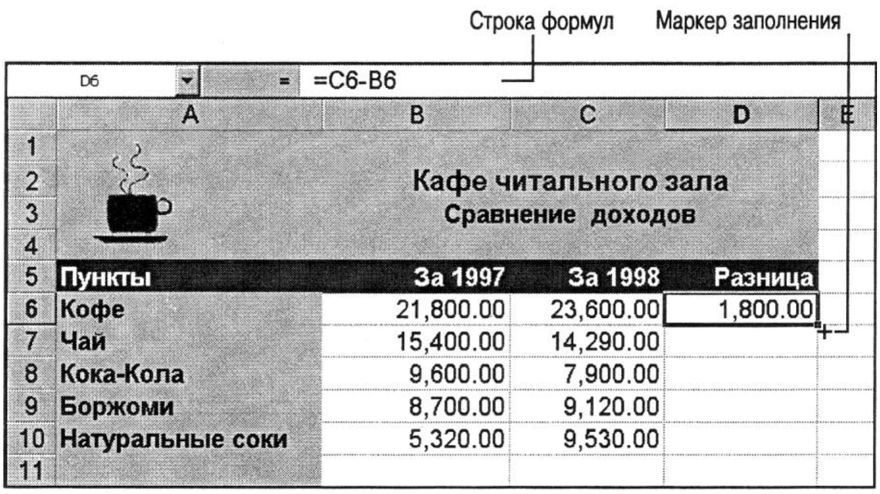


Рисунок 1.3 - Копирование формул с помощью маркера заполнения

Использование диаграмм.

Место размещения диаграммы зависит от того, как вы намереваетесь ее использовать.

• *Внедренные диаграммы.* Эти диаграммы удобны, когда необходи-

мо видеть диаграмму рядом с данными, на основании которых она построена, например в отчете. Если нужно напечатать или увидеть диаграмму вместе с данными, используется внедренная диаграмма. Однако внедренную диаграмму можно распечатать и отдельно от данных.

• *Диаграммные листы.* Диаграммы на отдельных листах в рабочей книге следует размещать в тех случаях, если вы планируете ее использовать самостоятельно без исходных данных. Например, можно вставить диаграмму в книгу, которая размещена в сети и доступна всем сотрудникам, но сами данные, на основании которых была построена эта диаграмма, сделать недоступными для совместного пользования.

Независимо от того, каким образом была создана диаграмма, она всегда связана с данными листа. При изменении данных диаграмма будет автоматически обновляться.

Диаграммы можно создавать двумя способами — с помощью быстрых клавиш и с помощью мастера диаграмм. Первый метод выполняется быстрее и всегда приводит к созданию диаграммного листа; автоматически создается диаграмма того типа, который принят по умолчанию. Обычно по умолчанию создается стандартная двухмерная гистограмма.

Второй способ позволяет создавать как внедренные диаграммы, так и диаграммные листы. Использование мастера диаграмм позволяет модифицировать диаграмму уже в процессе ее создания.

Каждая диаграмма Excel состоит из ряда объектов. Знание их названий и назначений облегчит создание и редактирование диаграмм. Чтобы помочь вам идентифицировать объекты диаграмм, Excel показывает подсказку при наведении указателя мыши на объект. На рисунке 1.4 показаны наиболее распространенные объекты диаграмм, а в таблице 1.3 приведено их описание.

Таблица 1.3 - Объекты диаграммы Excel

Объект диаграммы	**Описание**
Стрелка	Графический объект, созданный с помощью кнопки со стрелкой на панели инструментов *Рисование*. На рисунке 1.4 добавлена стрелка, направленная от надписи на точку “Литвы” ряда данных за 1998 г.
Ось категорий	Ось, на которой отображаются заголовки из листа. На рисунке 1.4 на оси категорий отображаются названия стран. Как правило, ось категорий расположена по горизонтали. Примерами подписей для оси категорий могут быть: Январь. Февраль, ...; Квартал 1, Квартал 2, ...; 1997, 1998, ...
Метка данных	Обычно добавляется в диаграмму для отображения конкретного значения точки данных
Точка данных	Элемент ряда данных, соответствующий значению одной ячейки в листе. Так, объем продаж по Украине за 1998 г. являются точкой данных ряда “1998”
Ряд данных	Строка или столбец данных из листа. Названия всех рядов приводятся в легенде. На рисунке 1.4 имеется два ряда данных – “1997” и “1998”
Таблица данных	Таблица, добавляемая ниже оси категорий, в которой отображаются исходные данные диаграммы
Линии сетки	Линии, начинающиеся с делений шкалы. Хотя линии сетки можно добавлять к любой оси, как правило, их добавляют к оси значений, чтобы идентифицировать значения точек данных
Область построения	Прямоугольная область, ограниченная осями диаграммы
Легенда	Текстовое поле с описанием рядов данных. Имя ряда и цветной маркер помогают найти ряд в диаграмме. Хотя легенда по умолчанию выводится в правой части области построения, ее можно переместить куда угодно
Шкала	Цифровые деления на оси значений. Минимальное значение, как правило, равно нулю. Максимальное превышает самое большое из отображаемых чисел
Маркеры выделения	При щелчке на объекте диаграммы вокруг него появляется несколько черных квадратов, показывающих, что объект выбран. Объект следует выбирать, чтобы переместить его или изменить его свойства. На рисунке 1.4выбрана легенда
Надпись	Текст, не зависимый от данных. Надписи, добавленные к диаграмме, можно свободно перемещать и форматировать. На рисунке 1.4 надпись была добавлена, чтобы подчеркнуть рост объема продаж в Литве

Метки делений	Маленькие линии, отображающие деления шкалы по осям категорий и значений
Подсказка	Форма экранной справки, используемая для идентификации объектов диаграммы
Заголовки	В диаграммах Excel существует три типа заголовков: название диаграммы, заголовок оси значений и заголовок оси категорий. Название диаграммы обычно выводится в верхней части диаграммы вне области построения и используется для описания данных, представленных в диаграмме. На рисунке 1.4 в название диаграммы вынесен текст “Сравнительные значения продажи книг за прошлый и текущий год”. Заголовок оси значений описывает тип значений, представленных по этой оси. Заголовок оси категорий используется для описания категорий, отображаемых в диаграмме. Можно свободно перемещать и форматировать заголовки. Чтобы создать в заголовке новую строку, нажмите <Enter>
Ось значений	Ось, на которой располагаются значения данных из листа. На рисунке 1.4 цифры продаж показаны по оси значений. Обычно ось значений вертикальная. Кроме денежных сумм, на этой оси можно показывать количество или температуру

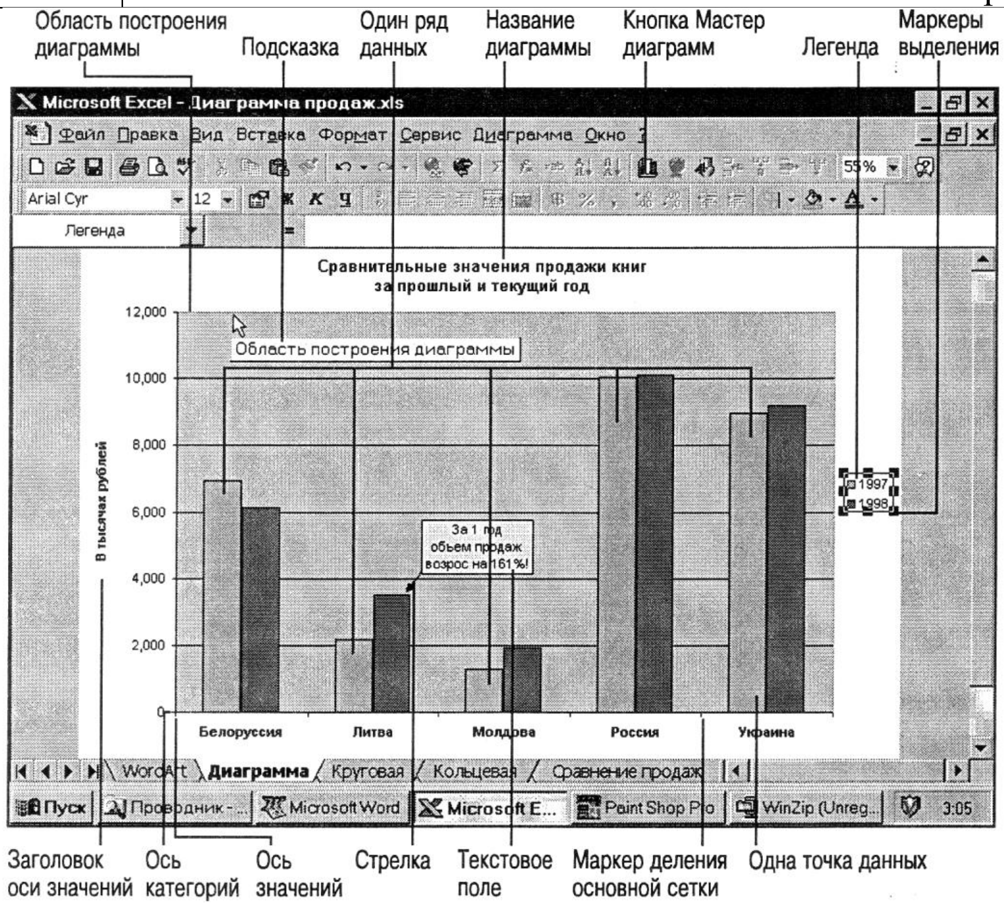


Рисунок 1.4 - Объекты диаграммы Excel

Типы диаграмм.

В Excel имеется 14 стандартных типов диаграмм. Многие из них представлены несколькими видами диаграмм. Распространенными разновидностями являются диаграммы с накоплением, разрезанные диаграммы и диаграммы, нормированные на 100%. *Диаграммы с накоплением* используются для отображения итоговых данных, *разрезанные диаграммы* — для отображения одной точки отдельно от других и являются разновидностями кругового и кольцевого типа диаграмм. *Диаграммы, нормированные на 100%,* используются для отображения процентного содержания каждой точки в общей сумме.

Некоторые типы диаграмм представлены только двухмерными видами, другие имеют как двухмерные, так и объемные варианты. По умолчанию используется стандартная двухмерная гистограмма (см. левую часть рисунка 1.5).

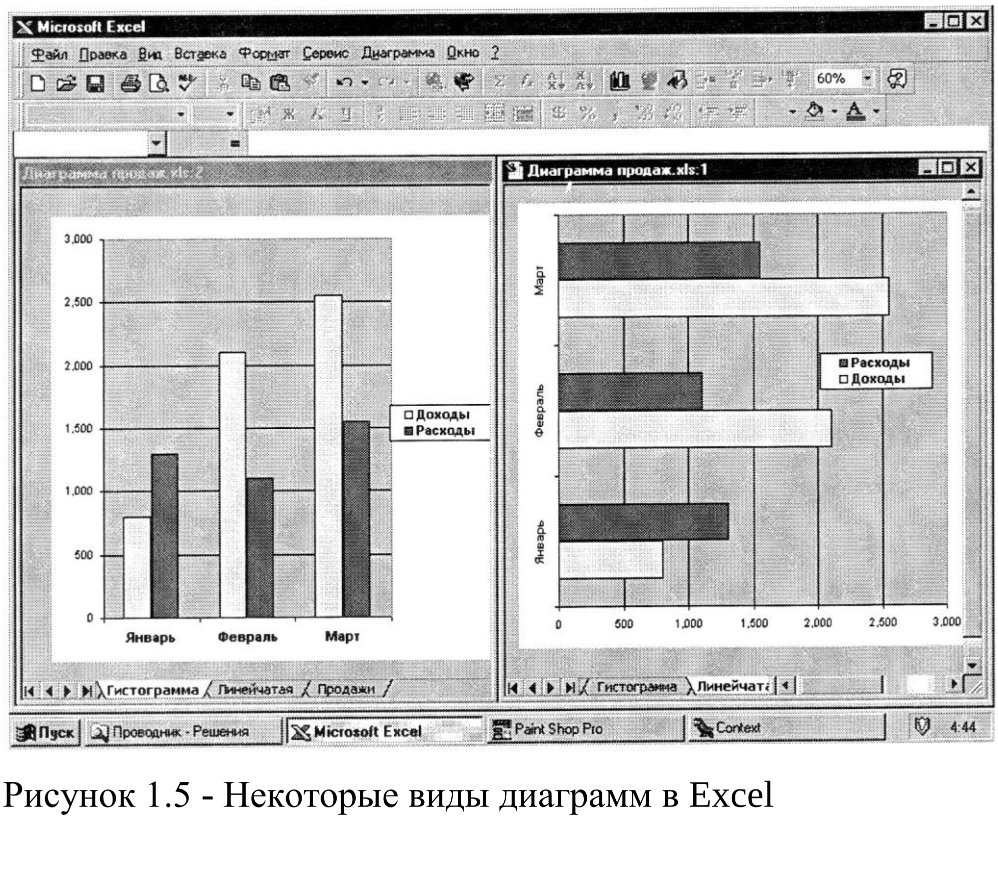


Рисунок 1.5 - Некоторые виды диаграмм в Excel

Таблица 1.4 - Типы диаграмм в Excel

Тип диаграммы	Размерность	Описание
Гистограмма	Двухмерная и трехмерная	Данные отображаются в виде вертикальных полос. Наиболее широко распространенный тип. В Excel используется по умолчанию. Представлена следующими видами: объемная, с накоплением и нормированная на 100%. См. левую часть рисунка 1.5.
Линейчатая	Двухмерная и трехмерная	Данные отображаются в виде горизонтальных полос. Полезна при сравнении величин за один временной период или в тех случаях, когда подписи категорий слишком длинные. Представлена следующими видами: объемная, с накоплением и нормированная на 100%. См. правую часть рисунка 1.5
Коническая, цилиндрическая и пирамидальная	Трехмерная	Привлекательные трехмерные варианты гистограммы и линейчатой диаграммы
График	Двухмерная и трехмерная	Данные отображаются в виде точек, соединенных линиями. Этот тип диаграмм используется для отображения изменений данных во времени. Представлен следующими видами: объемный и с накоплением. См. левую часть рисунка 1.6
С областями	Двухмерная и трехмерная	Похожа на график, только область под линией закрашена. Используется для отображения большого числа точек. Имеет следующие разновидности: объемная и с накоплением. См. правую часть рисунка 1.6
Круговая	Двухмерная и трехмерная	Отображается только один ряд или категория данных. Используется, чтобы показать, сколько процентов составляет каждая точка данных от общей суммы. Этот тип диаграммы представлен следующими видами: объемная, разрезанная, вторичная, круговая и вторичная. Вторичные диаграммы применяются в тех случа-

		ях, когда часть элементов необходимо выделить из общей суммы и представить их в отдельной круговой диаграмме или гистограмме. См. левую часть рисунка 1.7
Кольцевая	Двухмерная	Похожа на круговую. Используется для отображения процента от общей суммы, но для нескольких рядов. Данные отображаются в виде колец. Возможно создание разрезной кольцевой диаграммы
Точечная	Двухмерная	Отображаются маркеры для каждой точки данных. Предназначена для отображения разброса данных, что часто используется в научных работах. Представлена следующими видами: только точки; точки, соединенные сглаживающими линиями; точки, соединенные отрезками
Пузырьковая	Двухмерная	Напоминает точечную диаграмму, в которой для каждой точки отображаются метки. Чем больше значение, тем больше пузырек. Может быть представлена в объемном виде
Биржевая	Двухмерная	Используется, в основном, для отображения изменения курса биржевых цен. Данные для этого типа диаграмм должны быть расположены в определенном порядке. Представлена следующими разновидностями: для набора из трех значений (самый высокий курс, самый низкий и курс закрытия), для набора из четырех значений (курс открытия, самый высокий курс, самый низкий курс, курс закрытия), еще для одного набора из четырех значений (объем, высокий курс, низкий и курс закрытия) и для набора из пяти значений (объем, курс открытия, высокий курс, низкий и курс закрытия)
Поверхность	Трехмерная	Подобно топографическим картам эта диаграмма отображает “возвышения” и “впадины” для совокупностей данных. Разновидность этого типа

		диаграмм — контур
Лепестковая	Двухмерная	Ось значений представлена лучами, исходящими из общего центра, а точки данных соединены отрезками, образуя структуру, напоминающую паутину. Возможен вариант лепестковой диаграммы с полями

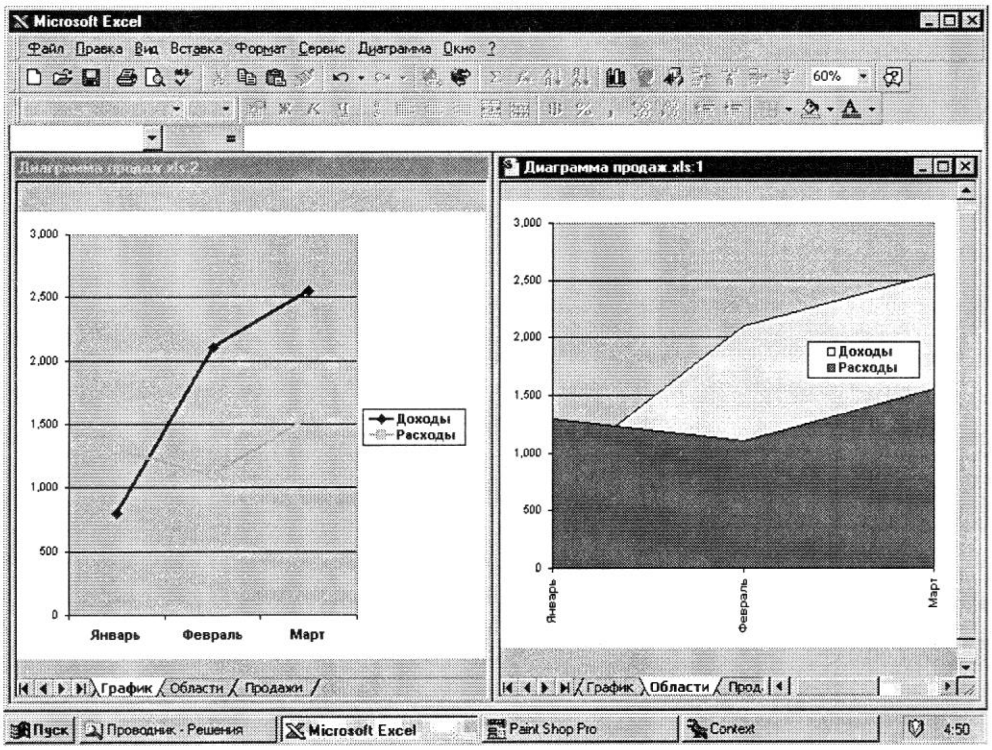


Рисунок 1.6 - Плоский график и плоская диаграмма с областями

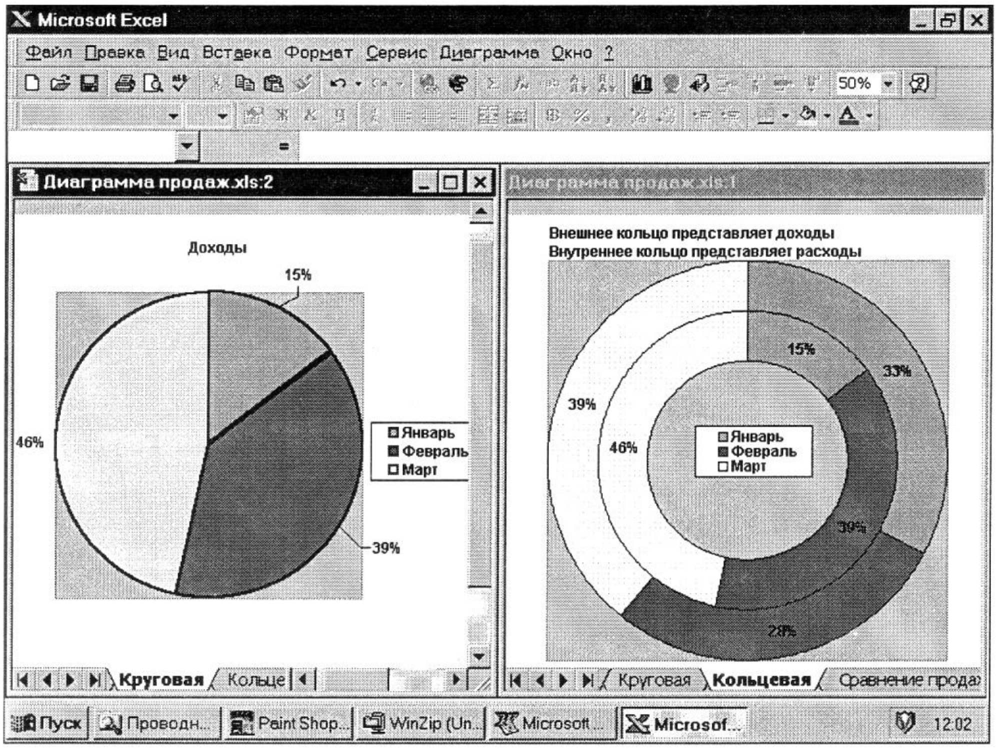


Рисунок 1.7 - Круговая и кольцевая диаграммы

Создание диаграмм с помощью мастера диаграмм.

Мастер диаграмм позволяет легко и быстро создавать профессиональные диаграммы, предлагая на выбор разнообразные опции, представленные на сменяющих друг друга диалоговых окнах. По мере выполнения работы мастер показывает вам пример диаграммы, так что вы можете сразу увидеть влияние выбранных опций. Если вам не нравится диаграмма, вы можете изменить ее. С помощью мастера можно создавать как внедренные диаграммы, так и диаграммные листы.

Если необходимо представить в диаграмме данные из несмежных строк и столбцов, выберите первый диапазон ячеек, включая заголовки и подписи данных, а затем, удерживая клавишу <Ctrl>, выберите дополнительные диапазоны с помощью мыши.

Если в листе имеются данные, которые не следует включать в диаграмму, можно скрыть строки и столбцы, содержащие такие данные. Перейдите к строке или столбцу, которые необходимо скрыть, и выберите из меню *Формат* команду *Строка/Столбец–>Скрыть*.

Если была создана структура документа, скройте уровни, которые не нужно представлять в диаграмме.

Работа с мастером диаграмм состоит из четырех этапов.

1. Выберите ячейки, содержащие данные, которые следует отобразить в диаграмме. Не забудьте включить заголовки столбцов и строк, которые будут представлены по оси категорий и в легенде. Не включайте пустые строки в выбранный диапазон.
2. Щелкните на кнопке *Мастер диаграмм* на панели инструментов *Стандартная*. Появится окно мастера диаграмм.
3. В первом из четырех окон мастера выберите стандартный или нестандартный тип диаграммы (рисунке 1.8). Установите требуемый вид диаграммы. Щелкните на кнопке *Далее*.

Рисунок 1.8 - Первое окно мастера диаграмм

4. Во втором окне мастера проверьте правильность указания исходных данных для диаграммы. Если окажется, что вы неправильно выбрали диапазон ячеек, воспользуйтесь кнопкой свертывания диалогового окна (рисунок 1.9), чтобы временно скрыть окно и выбрать новый диапазон ячеек. Выберите способ отображения рядов данных: по столбцам или по строкам.

На вкладке *Ряды* можно удалить существующий или добавить новый ряд. Посмотрите образец диаграммы и щелкните на кнопке *Далее*.

5. В третьем окне мастера диаграмм установите параметры диаграммы. Эти параметры собраны на вкладках: *Заголовки*, *Оси, Линии сетки, Легенда, Подписи данных и Таблица данных* (рисунок 1.10). Проверьте правильность установок по образцу и щелкните на кнопке *Далее*.

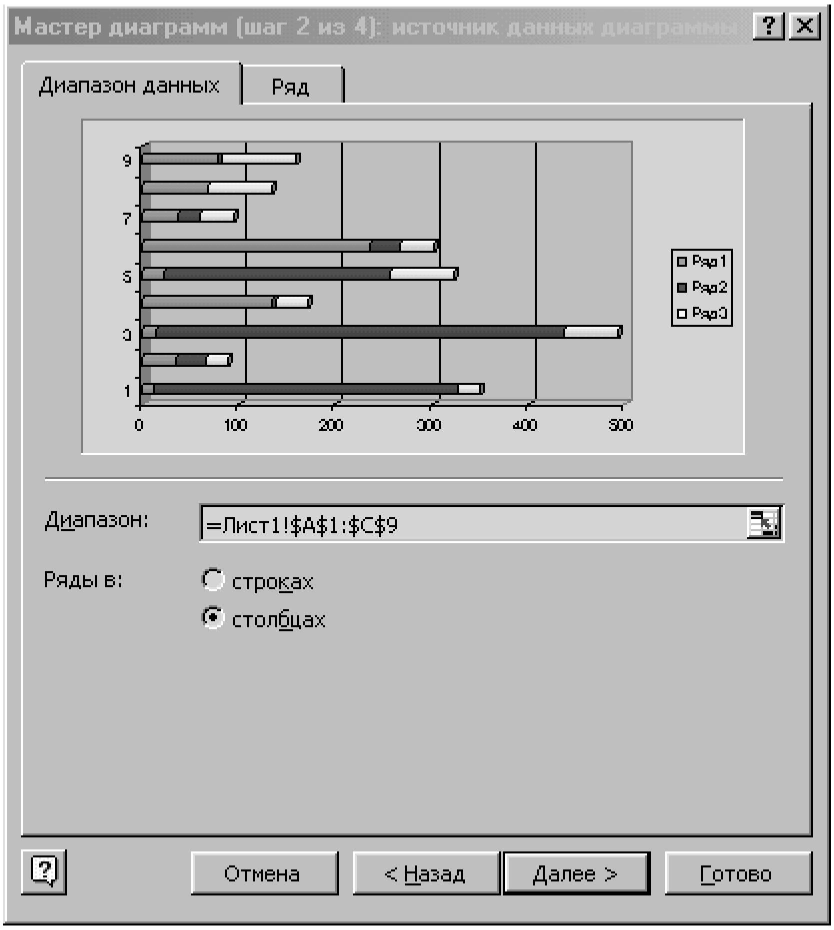


Рисунок 1.9 - Второе окно мастера диаграмм

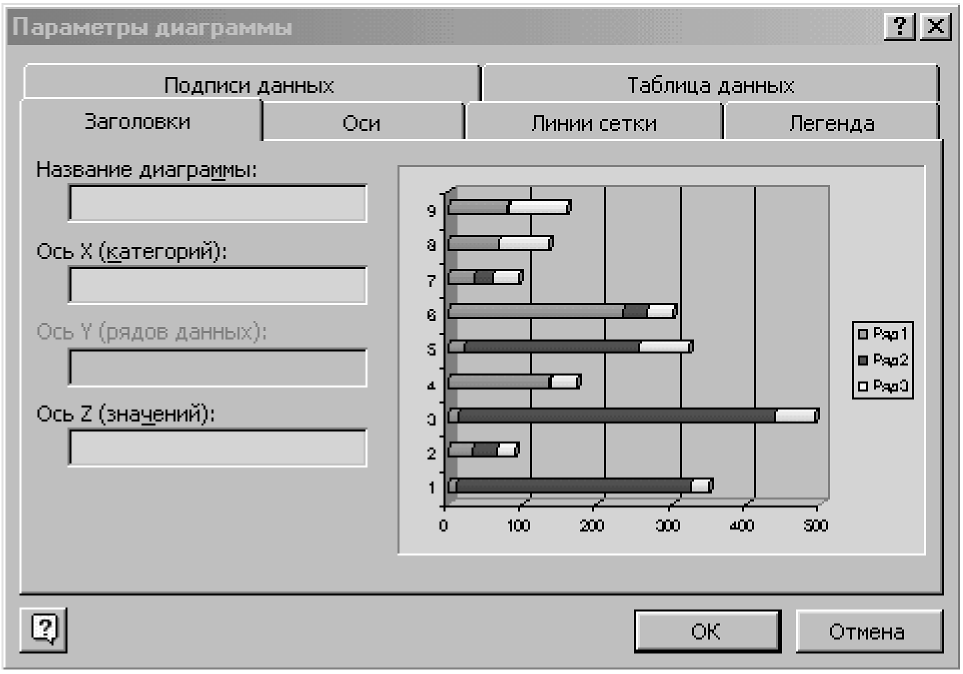


Рисунок 1.10 - Третье окно мастера диаграмм

6. В четвертом окне определите, куда следует поместить диаграмму (рисунок 1.11). Чтобы создать диаграммный лист, выберите пере-

ключатель *отдельном*, а чтобы создать внедренную диаграмму — переключатель *имеющемся*. Внедренную диаграмму можно разместить на любом листе активной книги. Если вы захотите вернуться к предыдущему окну, щелкните на кнопке *Назад*.

7. Чтобы закончить создание диаграммы, щелкните на кнопке *Готово*.

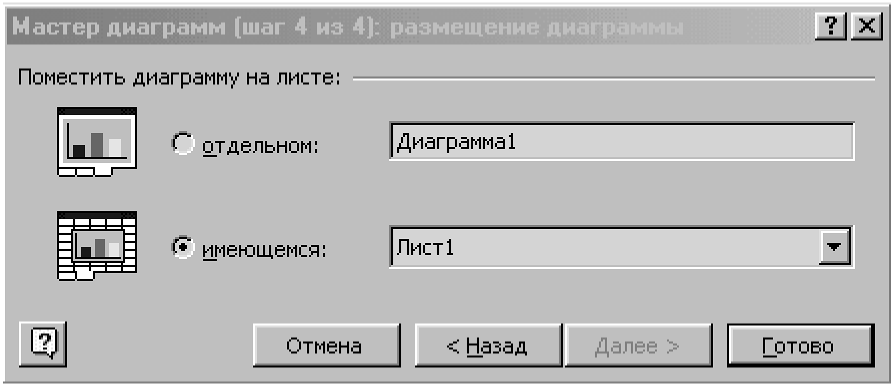


Рисунок 1.11 - Четвертое окно мастера диаграмм

Если при работе с мастером вам потребуется дополнительная информация, щелкните в любом из четырёх окон на кнопке помощника по офису.

Сохранение диаграмм.

При создании внедренной диаграммы или диаграммного листа эти объекты становятся частью книги. При сохранении книги сохраняются и диаграммы.

Чтобы сохранить книгу, щелкните на кнопке *Сохранить*  на панели инструментов *Стандартная*.

Печать диаграмм.

Вывод диаграмм на печать выполняется аналогично печати листов. Перед печатью диаграммы следует просматривать. Для этого на панели инструментов *Стандартная* щелкните на кнопке *Предварительный просмотр*.

Если диаграмма внедрена в лист, то по умолчанию она будет печататься вместе с листом (рисунок 1.12).

Поскольку диаграмма внедрена в лист, параметры страницы остаются неизменными. Диаграмма считается составной частью листа.

Чтобы напечатать внедренную диаграмму отдельно от листа, ее вначале следует выбрать. Тогда при предварительном просмотре вы увидите только диаграмму. Если вы собираетесь напечатать внедренную диаграмму отдельно от листа или напечатать диаграммный лист, диалоговое окно *Параметры страницы* примет немного другой вид (рисунок 1.13).

Вместо вкладки *Лист* в окне *Параметры* страницы появится вкладка *Диаграмма*. В ней можно установить параметры печати диаграммы. Размер диаграммы можно уменьшить или увеличить. При выборе опции *использовать всю страницу*, размер диаграммы будет подобран таким образом, чтобы заполнить весь печатный лист. Это может нарушить соотношение размеров элементов диаграммы. При выборе опции *уместить на странице* размер диаграммы изменится пропорционально.

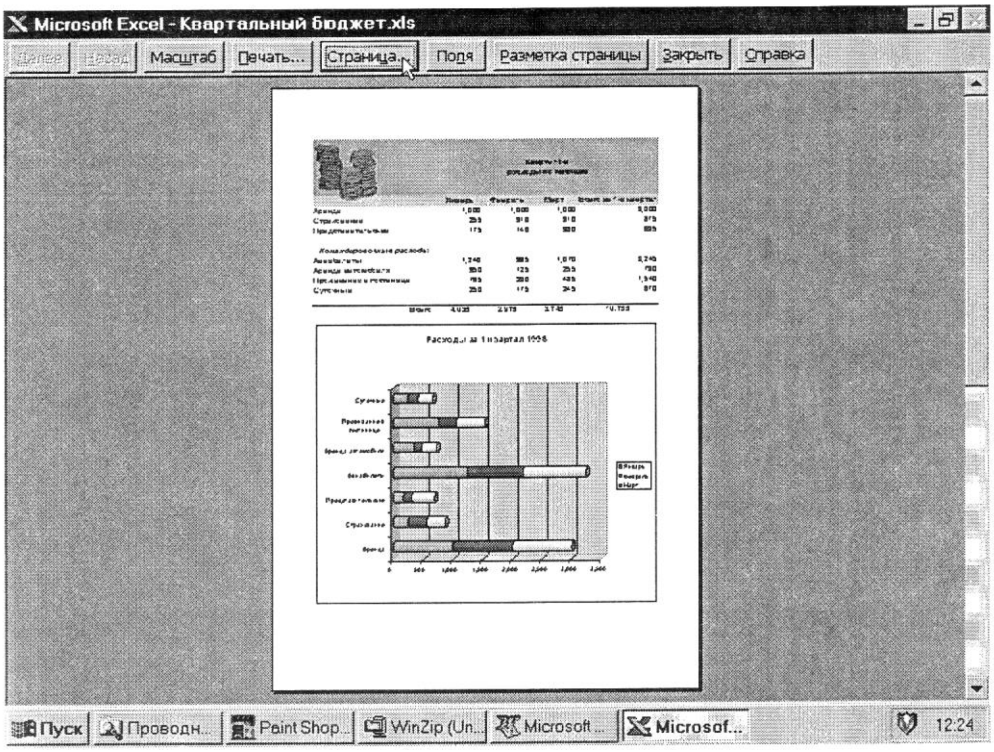


Рисунок 1.12 - Предварительный просмотр листа с внедренной диаграммой

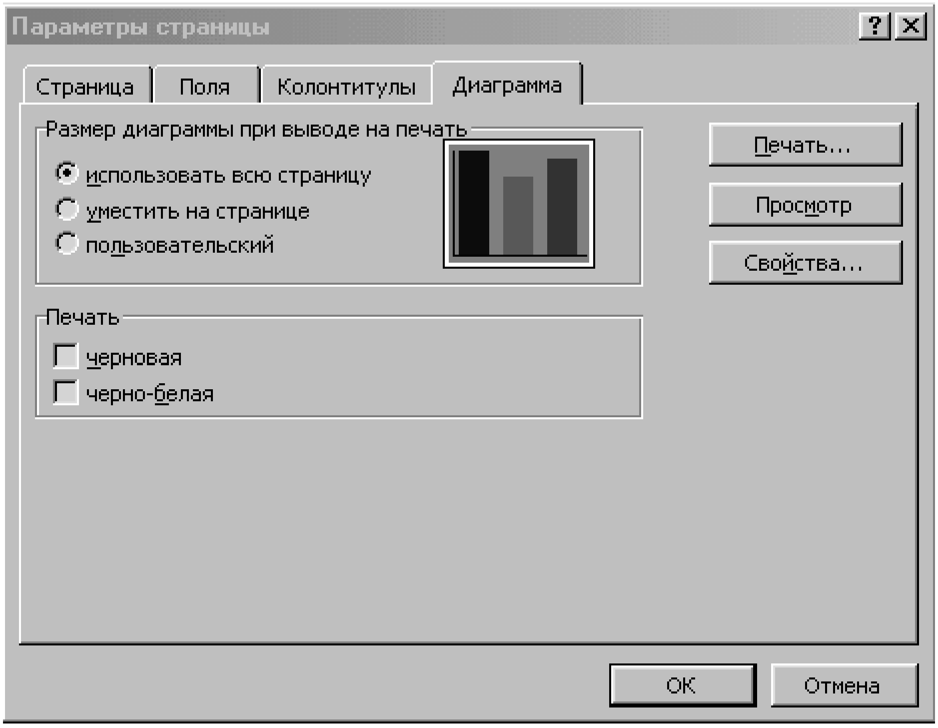


Рисунок 1.13 - Окно *Параметры страницы* при печати диаграммного листа или внедренной диаграммы отдельно от листа

1.3 Порядок выполнения работы

1. Создать таблицу количества выходных и праздничных дней по месяцам года.
2. Построить диаграмму зависимости количества выходных и праздничных дней от месяца года.
3. Составить калькуляцию доходов и расходов семьи за месяц используя математические операции.

1.4 Контрольные вопросы

1. Элементы главного окна.
2. Способы ввода данных.
3. Правила работы с формулами.
4. Математические операции.
5. Порядок построения диаграмм.

Лабораторная работа № 2

2.1 Оценка неизвестных параметров нормального распределения

Когда мы распространяем представления о конечной группе лиц на другие группы или на всю совокупность, мы пользуемся информацией о выборке. Когда врач хочет получить представление о составе и состоянии крови пациента, он проводит анализ небольшой выборки крови. Любое значение искомого параметра, вычисленное на основе ограниченного числа опытов, всегда будет содержать элемент случайности. Работники здравоохранения постоянно имеют дело с информацией, базирующейся на ограниченных выборках. Поэтому они должны хорошо представлять себе границы надежности анализа информации на основе выборочных данных.

Цель работы: Изучить понятия “генеральная совокупность” и “выборка”, научиться вычислять выборочную среднюю, исправленную выборочную дисперсию, исправленное среднеквадратическое отклонение, научиться вычислять доверительный интервал для математического ожидания, соответствующий заданной доверительной вероятности.

2.2 Теоретические сведения

В биологической и медицинской статистике часто приходится исследовать распределение того или иного признака для весьма большой совокупности индивидуумов, образующих статистический коллектив (таким признаком может быть, например, содержание белка в зерне Пшеницы, вес новорожденного ребенка, период колебаний маятника и т.д.). Данный признак является случайной величиной, значение которой от индивидуума к индивидууму меняется. Однако, для того, чтобы составить представление о распределении этой случайной величины или о ее важнейших характеристиках, нет необходимости обследовать каждый объект данной обширной (генеральной) совокупности, а можно обследовать некоторую выборку достаточно большого объема для того, чтобы в ней были выявлены существенные черты изучаемого распределения.

Статистическая совокупность представляет собой множество объектов, однородных относительно признака, характеризующего эти объекты.

Генеральной совокупностью называется совокупность, состоящая из всех объектов, которые могут быть к ней отнесены. Теоретически это бесконечно большая или приближающаяся к бесконечности совокупность. Число объектов генеральной совокупности называют ее объемом и обозначают N.

Выборочной совокупностью или выборкой называется множество объектов, случайно отобранных из генеральной совокупности. Число объектов выборки называют ее объемом и обозначают n.

Для того, чтобы свойства выборки достаточно хорошо отражали свойства генеральной совокупности, выборка должна быть осуществлена случайно, то есть все объекты должны иметь одинаковую вероятность попасть в выборку.

Поскольку на практике приходится иметь дело с ограниченным количеством экспериментальных данных, то результаты наблюдений и их обработки содержат больший или меньший элемент случайности.

Характеристики статистического распределения выборки применяются для оценки неизвестных параметров теоретического распределения вероятностей.

Различают точечные оценки случайной величины (одним числом) и интервальные (оценивание параметра совокупности в виде интервала).

Введем некоторые понятия.

Генеральная средняя $\overline{X}_{\Gamma}$ — среднее арифметическое значение признака $X_1, X_2, \ldots, X_n$ генеральной совокупности, т.е.

$$\overline{X}_{\Gamma} = \frac{1}{N}\sum_{i=1}^{N} X_i .$$

Генеральная средняя равна математическому ожиданию случайной

величины:

$$\overline{X}_{\Gamma} = \mu.$$

Выборочная средняя $\overline{X}_{B}$ — среднее арифметическое значение признака выборочной совокупности $X_1, X_2, \ldots, X_n$ то есть

$$\overline{X}_{B} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

Генеральная дисперсия:

$$D(x) = \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \mu\right)^2 .$$

Выборочная дисперсия:

$$\sigma_B^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}_B\right)^2 .$$

Точечные оценки. За оценку значения μ измеряемой величины принимается выборочная средняя:

$$\mu \approx \overline{X}_{B} = \frac{1}{n}\sum_{i=1}^{n} X_i .$$

За оценку дисперсии D принимается значение исправленной выборочной дисперсии S^2:

$$D \approx S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}_B\right)^2 = \frac{n}{n-1}\sigma_B^2 .$$

Интервальная оценка математического ожидания (доверительный интервал для математического ожидания случайной величины, распределенной по нормальному закону, при неизвестном σ).

Пусть случайная величина А имеет нормальное распределение, причем неизвестны μ и σ.

В ряде задач требуется не только найти для параметра μ подходящее численное значение, но и оценить его точность. Требуется знать, к каким ошибкам может привести замена параметра μ его точечной оценкой

$\overline{X}_B$ и с какой степенью уверенности можно ожидать, что эти ошибки не выйдут за известные пределы?

Такого рода задачи особенно актуальны при малом числе наблюдений, когда точечная оценка в значительной мере случайна и приближенная замена может привести к серьезным ошибкам.

Чтобы дать представление о точности и надежности в математической статистике пользуются так называемыми доверительным интервалом и доверительной вероятностью.

Разные выборки дадут разные оценки. Пусть для параметра μ получена из некоторого опыта точечная оценка $\overline{X}_B$. При этом, заменяя μ на $\overline{X}_B$, мы совершаем некоторую ошибку.

В теории математической статистики показывается, что с заданной вероятностью α неизвестное значение параметра μ попадает в определенный интервал (рисунок 2.1):

$$\left(\overline{X}_B - \Delta X, \overline{X}_B + \Delta X\right)$$

или

$$\overline{X}_B - \Delta X < \mu < \overline{X}_B + \Delta X .$$

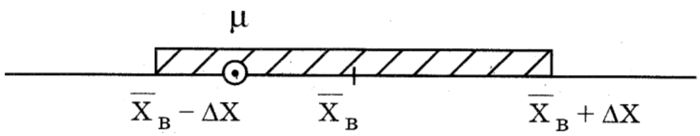


Рисунок 2.1 - Доверительный интервал

Вероятность α принято называть доверительной вероятностью. С такой вероятностью мы “доверяем” результату. Величина α выбирается самим исследователем самостоятельно, например, α = 0,95; 0,98.

С заданной вероятностью α доверительный интервал накрывает точку μ.

Величина ΔX – полуширина доверительного интервала. Точки $\overline{X}_B + \Delta X$ и $\overline{X}_B - \Delta X$ — доверительные границы.

Величины $\overline{X}_B$ и ΔX вычисляются на основе экспериментальных

данных.

Допустим случайная величина А подчиняется нормальному закону распределения.

В эксперименте получены ее значения:

$$X_1, X_2, \ldots, X_n.$$

Если объем выборки невелик, (n<30), то полуширина доверительного интервала в этом случае вычисляется по формуле:

$$\Delta X = t_{\alpha,n} \cdot \frac{S}{\sqrt{n}},$$

где $t_{\alpha,n}$ – коэффициент Стъюдента, значение которого зависит от доверительной вероятности α и от объема выборки n. Его значения приведены в специальной таблице.

Тогда доверительный интервал для μ:

$$\left(\overline{X}_B - t_{\alpha,n} \cdot \frac{S}{\sqrt{n}}; \overline{X}_B + t_{\alpha,n} \cdot \frac{S}{\sqrt{n}}\right).$$

Таким образом, математическое ожидание μ находится в доверительном интервале:

$$\overline{X}_B - t_{\alpha,n} \cdot \frac{S}{\sqrt{n}} < \mu < \overline{X}_B + t_{\alpha,n} \cdot \frac{S}{\sqrt{n}}$$

с заданной доверительной вероятностью α.

Чем выше мы задаем вероятность α, тем шире становится доверительный интервал. И, наоборот, чем меньше α, тем уже интервал.

При увеличении объема выборки ширина интервала уменьшается.

Пример расчета ΔX. При измерении некоторой величины получены следующие значения: $X_1 = 3{,}1$; $X_2 = 3{,}3$; $X_3 == 3{,}2$. С доверительной вероятностью $\alpha = 0{,}95$ оценить истинное значение измеряемой величины.

Решение. Вычисляем среднее выборочное значение

$$\overline{X}_B = \frac{3{,}1 + 3{,}3 + 3{,}2}{3} = 3{,}2.$$

Вычисляем исправленную дисперсию

$$S^2 = \frac{(3{,}2-3{,}1)^2 + (3{,}2-3{,}3)^2 + (3{,}2-3{,}2)^2}{3-1} = 0{,}01.$$

Вычисляем полуширину доверительного интервала. Значение коэффициента Стьюдента находим но соответствующей таблице.

$$t_{0,95;3} \cong 4{,}3 \rightarrow \Delta X \cong 0{,}25 \cong 0{,}3.$$

Следовательно:

$$3{,}2\text{-}0{,}3 < \overline{X}_Г < 3{,}2+0{,}3.$$

Ответ: С доверительной вероятностью α = 0,95 генеральное среднее измеряемой величины находится в доверительном интервале (2,9; 3,5).

2.3 Порядок выполнения работы

Задание 1. При измерении периода колебания математического маятника получены следующие значения:

3,0; 2,8; 3,1; 3,0; 2.9; 3,1; 2,8 с.

Оцените доверительный интервал для математического ожидания периода колебаний. Попадает ли в этот интервал теоретическое значение периода? Длина маятника 0,78 м.

Задание 2. При измерении гемоглобина у двух женщин получены следующие данные:

1 пациент

Содержание гемоглобина в крови, г/л	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
	128	127	126	122	128	126	129	125

2 пациент

Содержание гемоглобина в крови, г/л	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
	85	88	90	85	86	91	88	87

Оцените доверительный интервал для математического ожидания концентрации гемоглобина в крови для данных пациентов. Сравните с

нормой:

$$C_{норма}=130\pm10 \text{ (г/л)}.$$

Задание 3. Проанализируйте изменение полуширины доверительного интервала в зависимости от задаваемой доверительной вероятности (при фиксированном объеме выборки).

Данные измерения концентрации соли в растворе:

с, г/л   10; 10,2; 10,1; 10,4; 10,4; 10,3; 10,2.

Найдите для различных а значения ΔX и занесите в таблицу:

α	0,6	0,8	0,9	0,95	0,99
$t_{\alpha,n}$					
ΔX					

Представьте результаты графически (в виде рисунка 2.1).

Задание 4. Проанализируйте изменение полуширины доверительного интервала в зависимости от объема выборки (при фиксированной доверительной вероятности α = 0,95).

Значения случайной величины приведены в таблице:

X_i	100, 102, 101	99, 103	99, 101	104, 102
n	3	5	7	9
ΔX				
$\bar{X}_B$				

Результат представьте графически.

Приложение. Значения коэффициента Стьюдента $t_{\alpha,n}$ при различных значениях α и n:

n/α	0,6	0,8	0,9	0,95	0,99
2	1,38	3,08	6,31	12,71	63,66
3	1,06	1,89	2,92	4,30	9,93
4	0,98	1,64	2,35	3,18	5,84
5	0,94	1,53	2,13	2,78	4,60
6	0,92	1,48	2,02	2,57	4,03

7	0,91	1,44	1,94	2,45	3,71
8	0,90	1,42	1,90	2,37	3,50
9	0,89	1,40	1,86	2,31	3,36

2.4 Контрольные вопросы

1. Что называется “генеральной совокупностью”? Выборочной совокупностью?
2. Формулы для вычисления генеральной средней, выборочной средней, исправленной выборочной дисперсии.
3. Какая величина является точечной оценкой математического ожидания? Какая величина является точечной оценкой дисперсии?
4. Смысл доверительного интервала, доверительной вероятности.
5. Формулы для их расчетов.

Лабораторная работа № 3

3.1 Использование гистограмм в задачах медицинской статистики

Цель работы: ознакомиться с нормальным законом распределения случайных величин (законом Гаусса); научиться строить график кривых распределения по нормальному закону для различных параметров; научиться проводить статистическую обработку результатов измерений, строить гистограммы и на базе этих данных; обосновывать выводы о результатах проведенных экспериментов.

3.2 Теоретические сведения

Целью любого эксперимента, будь то медицинский, биологический или физический эксперимент, является получение надежных выводов об измеряемых величинах или каких-либо функциях от них. Эта цель еще не достигается с окончанием измерений. Результаты измерений необходимо подвергнуть тщательному анализу и провести необходимую математическую обработку. Только после этого возможно сформулировать выводы относительно величин, представляющих интерес. В данной лабораторной работе вы познакомитесь с одним из методов обработки и графическим представлением экспериментальных данных — методом гистограмм, широко используемым в практике медицинских исследований.

При измерении какой-либо величины несколько раз, экспериментатор получает ряд значений, которые, как правило, Оказываются различными. Этому есть много причин, например, отклонения от начальных условий эксперимента, которые могут быть малы и не поддаваться контролю. Или, например, в клинике врач измеряет один и тот же физиологический показатель пациента несколько раз (температуру, артериальное давление, количество сокращений сердца в минуту и т. д.) и, естественно, получает разные значения этого показания. В этом случае о результатах эксперимента говорят как о случайных величинах.

Случайная величина — это одно из важнейших основных понятий

теории вероятностей. Рассмотрим несколько примеров. Число космических частиц, попадающих на определенный участок земной поверхности в единицу времени подвергается значительным колебаниям в зависимости от многих случайных обстоятельств.

Размер уклонения точки попадания снаряда от центра цели определяется большим количеством разнородных причин, носящих случайных характер. В результате в теории стрельбы вынуждены считаться с явлением рассеивания снарядов около центра цели как со случайными явлениями и рассматривать указанное отклонение как случайную величину.

Скорость молекул газа не остается неизменной, а меняется в зависимости от столкновений с другими молекулами. Этих столкновений очень много даже в течение короткого промежутка времени. Зная скорость молекулы в данный момент, нельзя с полной определенностью указать ее значение, например, через 0.001 с. Изменение скорости молекулы носит случайный характер.

Случайной величиной является и количество эритроцитов в мазке крови в поле зрения микроскопа.

Со случайными величинами приходится иметь дело в самых разнообразных областях науки и техники. Поэтому важна задача создания: и изучения метода исследования случайных величин.

Случайной величиной называется величина, которая в результате опыта может принимать то или иное значение; причем неизвестно заранее какое именно.

Случайная величина может быть дискретной, то есть принимать счетное множество значений, которые можно пронумеровать (например, число клеток в поле зрения микроскопа, число пациентов в отделении, количество показателей состояния больного и т.д.) или непрерывной, которая может принимать все значения из некоторого интервала (бесчисленное множество возможных значений, сплошь заполняющих некоторый проме-

жуток). Непрерывными величинами являются, например, длительность интервалов между зубцами в ЭКГ, значение артериального давления, размер диаметра зрачка и др. Полученное отдельное значение результата измерения какого-либо из указанных параметров А обозначим “x”. Например, А — температура, x— значение температуры: x = 36,9°.

Функция плотности распределения вероятностей. Пусть А — некоторая непрерывная случайная величина, например, вес новорожденного, x — значение случайной величины. Со значением x случайной величины связана функция f(x) — функция плотности распределения вероятностей (ПРВ), такая что произведение f(x)dx пропорционально вероятности события, состоящего в том, что значение x величины А заключено в интервале [x, x + dx].

Функция ПРВ имеет очень важное значение. Происхождение каждого эмпирического распределения (то есть вид функции f(x)) обусловлено совокупностью определенных причин. Совокупность причин, приводящих к тому или иному виду f(x) может быть в каждом случае иной. Задача состоит в том, чтобы представить, себе, за счет каких причин могло получиться найденное распределение, то есть построить подходящую математическую или физическую модель явления. Таким образом, установление вида функции f(x) имеет большое значение для получения информации об изучаемом процессе.

Нормальный закон распределения (закон Гаусса). Значительное число случайных явлений, встречающихся в природе, может быть описано с помощью нормального закона распределения (закона Гаусса).

Закон Гаусса:

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

где x – любое значение изучаемой величины; μ — математическое ожидание; σ — среднее квадратическое отклонение.

$$f_{max} = \frac{1}{\sigma\sqrt{2\pi}} \qquad \text{при } x=\mu.$$

График функции f(x) нормально распределенной случайной величины представляет собой колоколообразную кривую (рисунки 3.1, 3.2), симметричную относительно оси, проходящей через точку $x = \mu$ параллельно ординате. Максимальное значение кривая достигает в точке $x = \mu$.

Функция имеет точки перегиба при $x=\mu\pm\sigma$, ось абсцисс служит для нее асимптотой при $x\to\pm\infty$.

Если изменить значение μ, а σ оставить постоянной, то кривая будет перемещаться вдоль оси OX, сохраняя свою форму (рисунок 3.1).

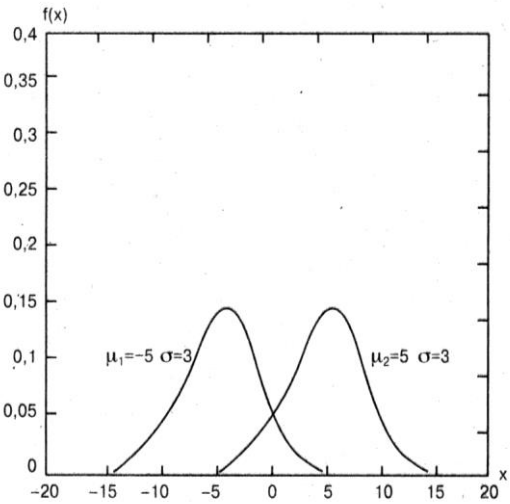


Рисунок 3.1 - Закон Гаусса (различные математические ожидания)

Если изменить σ — среднее квадратическое отклонение, а μ оставить неизменной, то изменяется форма кривой (рисунок 3.2). Параметр σ характеризует не положение, а форму кривой распределения. Это есть характеристика рассеивания. При увеличении σ максимальная ордината уменьшается. Так как площадь под кривой распределения всегда должна оставаться равной единице, то при увеличении о кривая становится более плоской (пологой). Наоборот, при уменьшении σ кривая распределения вытягивается вверх.

Вероятность попадания случайной величины А в интервал значений x, заключенный между числами x_1 и x_2, определяется формулой:

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)dx,$$

т. е. это площадь криволинейной трапеции, ограниченной сверху функцией f(x), снизу — осью x, слева и справа — ординатами, проходящими через точки x_1 и x_2. Раздвинем границы отрезка $[x_1, x_2]$: $x_1 \to -\infty$, $x_2 \to +\infty$, тогда

$$P(-\infty < x < +\infty) = \int_{-\infty}^{+\infty} f(x)dx = 1,$$

то есть площадь под всей кривой f(x) должна оставаться постоянной и равной 1.

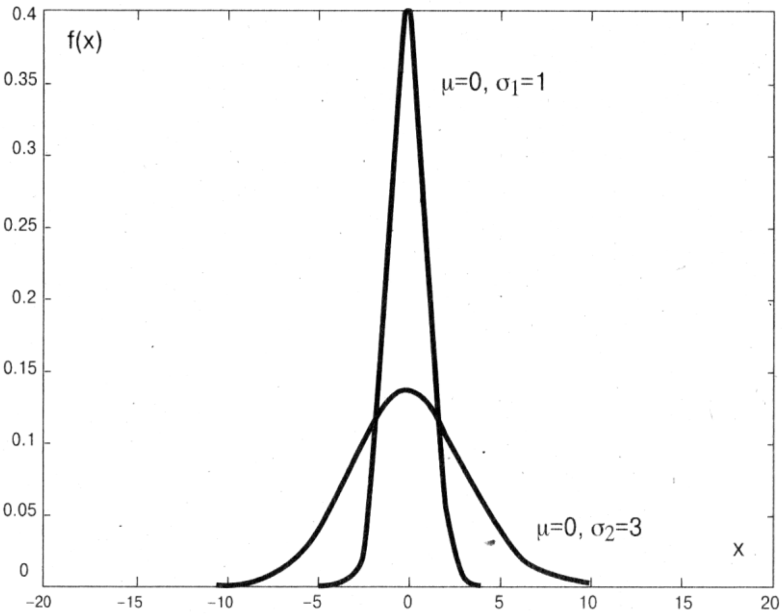


Рисунок 3.2 - Закон Гаусса (различные дисперсий)

Правило 3-х сигм.

Расчетами показано, что вероятность попадания нормально распределенной случайной величины в интервал значений (рисунок 3.3):

I. $P(\mu - \sigma < x < \mu + \sigma)$ $\cong$68,26%.

II. $P((\mu - 2\sigma < x < \mu + 2\sigma)$ $\cong$95,44%.

III. $P(\mu - 3\sigma < x < \mu + 3\sigma)$ $\cong$99,72%,

Таким образом, вероятность того, что отклонение значений нормально распределенной случайной величины превысит *3d,* чрезвычайно мала, а именно 0,0028. Такое событие можно считать практически невоз-

можным. Поэтому границы μ + 3σ и μ — 3σ принимаются за границы практически возможных значений нормально распределенной случайной величины. Это позволяет, зная среднее квадратическое отклонение и математическое ожидание случайной величины, ориентировочно указать интервал ее практически возможных значений. Такой способ оценки диапазона возможных значений случайной величины известен в математической статистике под названием “правило трех сигм”.

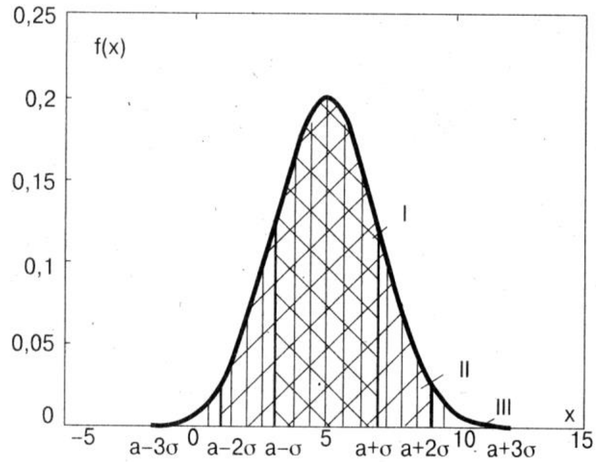


Рисунок 3.3 - Правило 3-х сигм

Пример. На рисунке 3.4 приведены графики нормального закона распределения температуры тела человека в норме и при патологии (например, при заболевании гриппом). Изменяются оба параметра μ и σ.

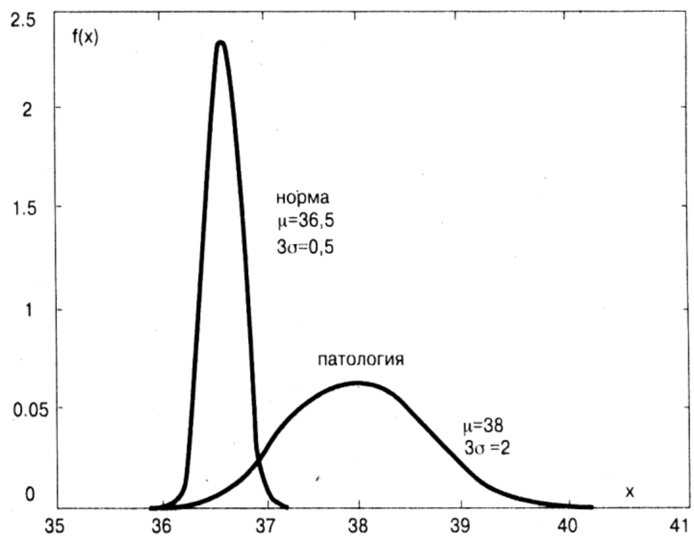


Рисунок 3.4 - Закон Гаусса (изменения μ и σ)

Графическое изображение статистического распределения. Ги-

стограмма.

Для оценки вида функции распределения вероятностей по экспериментальным данным часто используют графический метод, связанный с построением гистограммы. Он состоит в следующем. Пусть проведено n измерений непрерывной случайной величины А. Обозначим минимальное значение случайной величины x_{min}, максимальное —x_{max}.

Разобьем интервал, содержащий полученные значения величины А, на “k” интервалов одинаковой ширины Δx.

Подсчитаем количество значений случайной величины (частоту), попавших в каждый интервал Δx (i = 1, 2, ..., k). Получим частоты m_i (I = 1, 2, ..., k), каждую частоту поделим на ширину интервала Δx.

Величина $\frac{m_i}{\Delta x}$ называется плотностью частоты. Затем на каждом интервале Δx, следует построить прямоугольник с основанием Δx, и высотой $\frac{m_i}{\Delta x}$ (или высотой $\frac{P_i^*}{\Delta x}$ - плотностью относительной частоты $P_i^* = \frac{m_i}{n}$).

Полученную ступенчатую фигуру, состоящую из прямоугольников, называют гистограммой. (Гистограмма — от греческих слов “histos”— столб и “gramma” — запись.)

Задача. В 20 экспериментах непрерывная случайная величина А принимает значения: 21, 11, 17, 23, 28, 14, 19, 22, 24, 33, 16, 21, 18, 29, 23, 22, 31, 24, 27, 26. Построить гистограмму частот и гистограмму относительных частот.

Решение. Находим среди данных минимальное и максимальное значения случайной величины:

x_{min}= 11, x_{max}=33. Самым простым было бы разделить разность x_{min} - x_{max} на равное число частей. Но часто эта разность не делится нацело на требуемое число частей. В таком случае весь интервал несколько расширяется как в сторону меньших, так и в сторону больших значений. В рас-

сматриваемой задаче удобно выбрать $\Delta x = 5$. Тогда логично рассмотреть интервал (10, 35). Получаем, что в первый интервал (10-15) попадают всего два значения переменной x, равные 11, 14, то есть частота $m_1=2$. Во второй интервал (15-20) попадают значения переменной x, равные 17, 19, 16, 18, из чего следует $m_2=4$. Продолжая аналогичные рассуждения, составим таблицу, содержащую последовательность интервалов и соответствующих им частот – статистический интервальный ряд распределения:

X	10-15	15-20	20-25	25-30	30-35
M	2	4	8	4	2

В общем виде статистический интервальный ряд распределения имеет вид таблицы:

Интервалы значений x	(x_0, x_1)	(x_1, x_2)	…	$(x_{к-1}, x_к)$
Частоты m	m_1	m_2	…	m_k

Зная частоты и величину Δx, найдем плотности частот $\frac{m_i}{\Delta x}$ и плотности относительных частот – $\frac{P_i^*}{\Delta x}$.Например, для 1-го интервала плотность частоты $\frac{m_i}{\Delta x}$=0,1, плотность относительной частоты $\frac{P_i^*}{\Delta x}$=0,02.

Данные обработки результатов представлены в таблицах:

X	10-15	15-20	20-25	25-30	30-35
$\frac{m_i}{\Delta x}$	0,4	0,8	1,6	0,8	0,4
$\frac{P_i^*}{\Delta x}$	0,02	0,04	0,08	0,04	0,02

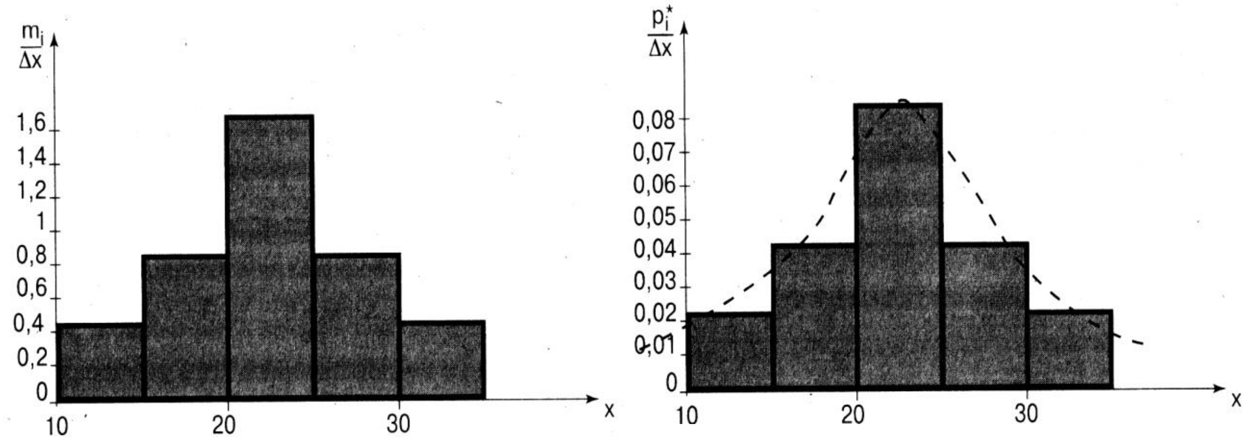


Рисунок 3.5 - Гистограмма плотности частот и ее огибающая

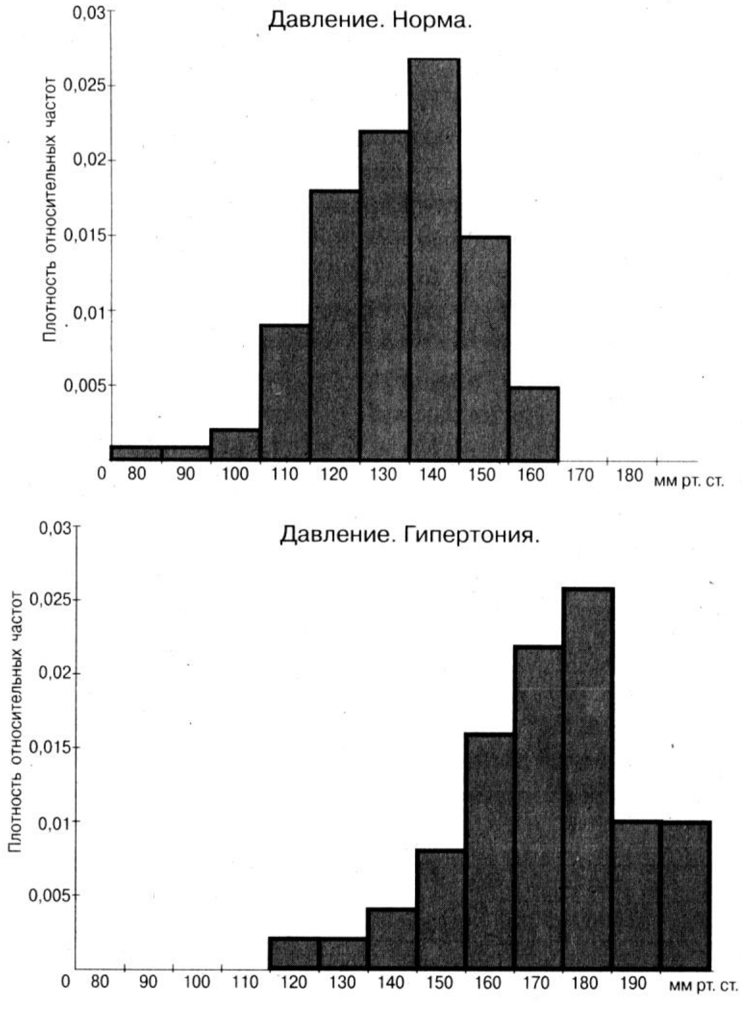


Рисунок 3.6 - Гистограмма плотности относительных частот и ее огибающая

Замечание. Гистограммы на рисунке 3.5 и рисунке 3.6 имеют один и тот же вид, что и следовало ожидать, исходя из метода обработки экспериментальных данных. Поэтому с точки зрения вида гистограммы не име-

ет значения, представлять ли данные в виде гистограммы плотности частот или гистограммы плотности относительных частот. Однако для установления вида функции плотности распределения вероятностей (ПРВ) необходимо пользоваться гистограммой плотности относительных частот. Это можно пояснить, рассматривая предельный случай, когда объем совокупности очень большой, а интервал разбиения Δx — мал. Прямоугольники гистограмм будут узкими, и число их будет велико. Ступенчатая линия гистограммы станет мало отличаться от плавной кривой, которая и будет являться функцией $y = f(x)$, указывающей чему равна ордината y, соответствующая заданной абсциссе x. Приближенно предполагаемый вид функции ПРВ показан на рисунке 3.6 пунктирной линией. Кроме этого представление экспериментальных данных именно в виде гистограммы плотности относительных частот необходимо, если ставится задача, например, о сравнении вида распределений двух или нескольких совокупностей. В этом случае бывает полезным совмещение различных гистограмм, а это возможно только, если рассматриваются плотности относительных частот, что позволяет исключить зависимость от объема выборки и ширины интервала Δx. Так, в клинической практике часто приходится сравнивать разные группы пациентов, например: здоровые и больные, принимающие лекарство и не принимающие и т.п., причем, количество пациентов в сравниваемых группах, как правило, не одинаково (48 здоровых и 21 больной). В этом случае для сравнения можно пользоваться только гистограммой плотности относительных частот. Если же взять гистограммы плотности частот, то высоты столбцов для здоровых (48) и больных (21) будут заведомо не одинаковы.

При построении гистограммы весьма важно правильно выбрать ширину интервала Δx. Если число интервалов “k” будет мало (ширина интервала Δx — велика), следует ожидать, что частично информация о случайной величине может быть потеряна. С другой стороны, если “k” будет

слишком велико (Δx — мало), обработка результатов измерений будет излишне трудоемкой, не давая при этом существенного выигрыша информации. Практика показывает, что рационален выбор числа интервалов “k” в зависимости от объема выборки с помощью таблицы:

Объем выборки (n)	25-40	40-60	60-100	100-200	200
Число интервалов (k)	5-6	6-8	7-10	8-12	10-15

Для более наглядного сравнения нескольких гистограмм (например, при сравнении физиологических данных в норме и при патологии) их необходимо строить одну под другой в одном масштабе как по горизонтальной, так и по вертикальной оси.

На рисунке 3.7 представлены для сравнения гистограммы, построенные на основании измеренных значений артериального давления у женщин в норме и с диагнозом “гипертоническая болезнь”. Видно, что смещается значение μ, в то время как σ почти не изменяется.

Рисунок 3.7 - Изменение параметров гистограммы (давление крови)

На рисунке 3.8 представлены гистограммы, полученные на основе измерения длительности 100 интервалов RR электрокардиограммы у здорового человека и у больного с диагнозом “мерцательная аритмия”. Видно, что μ почти не изменяется, в то время как при аритмии существенно возрастает σ.

Рисунок 3.8 - Изменение параметров гистограммы (ЭКГ-диагностика)

Выравнивание (сглаживание) статистических рядов. При обработке статистического материала часто приходится решать вопрос о том, как подобрать для данного статистического ряда теоретическую кривую распределения, выражающую лишь существенные черты статистического материала, но не случайности, связанные с недостаточным объемом экспериментальных данных. Такая задача называется задачей выравнивания (сглаживания) статистических рядов.

Задача сглаживания заключается в том, чтобы подобрать теоретическую плавную кривую распределения, с той или иной точки зрения наилучшим образом описывающую данное статистическое распределение.

Допустим величина A подчиняется нормальному закону. Тогда задача сглаживания переходит в задачу о рациональном выборе параметров μ и σ в законе Гаусса. Из теории известно, что прежде всего необходимо вычислить среднее выборочное значение ($\overline{x_B}$), определяемое для непрерывной случайной величины по формуле:

$$\overline{x_B} = \frac{m_1 \cdot x_1^* + m_2 \cdot x_2^* + \ldots + m_k \cdot x_k^*}{m_1 + m_2 + \ldots + m_k} = \frac{1}{n}\sum_{i=1}^{k} m_i x_i^* \qquad (3.1)$$

где $m_1 + m_2 + ... + m_k$ – частоты в соответствующих интервалах, $x_1^* + x_2^* + ... + x_k^*$ - середины интервалов, которые вычисляются по формулам:

$$x_1^* = \frac{x_0 + x_1}{2}, x_2^* = \frac{x_1 + x_2}{2}, ..., x_k^* = \frac{x_{k-1} + x_k}{2}$$

Термин “выборочное” означает, что среднее значение вычисляется по данной группе (например, по группе больных включающей 22 пациента), называемой выборкой.

Кроме среднего выборочного значения случайную величину характеризуют параметром, показывающим насколько широко разбросаны отдельные значения случайной величины относительно среднего значения, так называемым выборочным средним квадратическим отклонением σ_B.

$$\sigma_B = \sqrt{\frac{m_1\left(x_1^* - \overline{x_B}\right)^2 + m_2\left(x_2^* - \overline{x_B}\right)^2 + ... + m_k\left(x_k^* - \overline{x_B}\right)^2}{m_1 + m_2 + ... + m_k}} = \sqrt{\frac{\sum_{i=1}^{k} m_i\left(x_i^* - \overline{x_B}\right)^2}{n}} \qquad (3.2)$$

Поясним сказанное примером.

Пользуясь данными вышеуказанной таблицы, вычислить $\overline{x_B}$ и σ_B.

Вычислим середины интервалов: x_1=12,5, …, x_5=32,5.

Подставляя найденные значения и соответствующие частоты в формулы (3.1), (3.2), получаем:

$\overline{x_B}$ =22,5.

σ_B=5,5

Напишем выражение нормального закона в этом случае:

$$f(x) = \frac{1}{5{,}5\sqrt{2\pi}} e^{\frac{(x-22{,}5)^2}{2 \cdot 5.5^2}}.$$

Построим на одном графике и гистограмму, и выравнивающую ее кривую (рисунок 3.6).

3.3 Выполнение работы

Все задания выполняются на ПК с помощью пакета “Excel”.

Задание 1. Проведите анализ кривых распределения случайной величины по нормальному закону.

Для этого:

1) Запишите закон Гаусса для заданных параметров μ и σ,

2) Постройте графики (выберите масштаб, отложите по осям величины и единицы их измерения). Сделайте вывод о влиянии μ и σ на положение и форму кривой распределения.

3) Вычислите интервал 3σ, куда попадают практически все случайные величины. Покажите его на графике.

Варианты изменения μ и σ представлены в таблицах:

а)

μ	25	25	25
σ	2	5	10

б)

μ	10	25	50
σ	2	2	2

в)

	μ, г/л	σ, г/л
Женщины	140	6,6
Мужчины	158	6,8

В таблице в) указано содержание гемоглобина в крови.

Задание 2. Постройте и проанализируйте гистограммы плотности относительных частот. Проведите их выравнивание, считая закон распределения нормальным.

Варианты данных измерения длительности интервалов RR.

Напомним, что ЭКГ представляет зависимость мгновенных значений разности потенциалов между определенными точками на теле челове-

ка (проекции интегрального электрического вектора сердца — ИЭВС) на одно из отведений от времени (рисунок 3.9), на рисунке 3.9 отмечена длительность интервала RR.

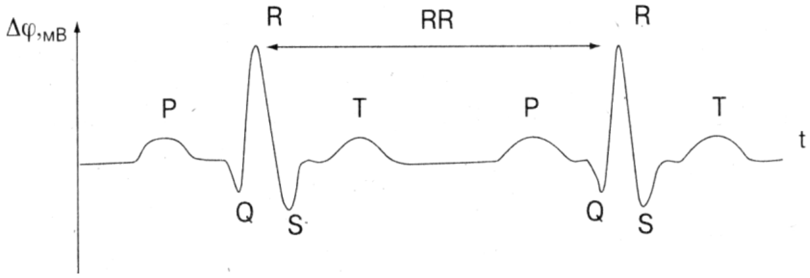


Рисунок 3.9 - Вид ЭКГ

а) Норма

Измерения длительности 100 интервалов RR (в мс) по ЭКГ здорового человека приведены ниже.

RR (мс):

787	801	869	923	872	764	822
943	868	918	881	771	827	907
843	826	826	763	775	873	883
887	896	802	816	925	854	857
764	802	839	822	831	762	755
841	836	799	824	799	773	757
743	819	788	792	752	732	769
864	777	816	734	757	850	805
798	755	807	741	764	799	775
743	780	799	823	757	815	748
778	790	734	788	832	801	728
706	718	744	827	730	753	775
845	791	755	828	773	793	821
730	740	749	808	829	894	824
773	746					

б) Мерцательная аритмия

Измерения длительности 100 интервалов RR (в мс) по ЭКГ больной с диагнозом “Мерцательная аритмия” приведены ниже.

RR (мс):

676	793	827	734	955	730	489
1051	1074	846	757	921	856	785
741	1020	805	875	712	1036	928
861	802	844	715	743	651	1075

902	668	948	727	681	774	698
876	1268	980	861	748	819	637
1085	753	756	773	772	1086	1376
881	950	854	902	718	646	1156
1200	1016	1028	660	715	809	1025
812	946	882	742	911	797	1078
1249	691	697	1047	1034	690	789
1046	760	1075	843	844	681	743
1175	903	856	725	1018	741	1209
1001	723	631	1169	708	739	690
1219	985					

Мерцательная аритмия возникает обычно в результате мерцания предсердий (мерцания желудочков несовместимы с жизнью, и требуются срочные реанимационные меры, включая дефибрилляцию), при котором общая систола предсердий заменена беспорядочным возбуждением и сокращением отдельных групп мышечных волокон. При мерцании предсердий желудочковые сокращения аритмичны. Часть сердечных сокращений, при слабом наполнении желудочков, неэффективны, то есть сопровождаются малым сердечным выбросом и отсутствием пульсовой волны. В результате этого число сердечных сокращений будет больше, чем число пульсовых волн (дефицит пульса). Отрицательное воздействие на кровообращение при мерцательной аритмии обусловлено выпадением сокращений предсердий, что уменьшает наполнение желудочков, а также нарушением желудочкового ритма.

в) Синусовая тахикардия

Измерения длительности 100 интервалов RR (в мс) по ЭКГ больной с диагнозом “Синусовая тахикардия” (больная на 6-е сутки после нейрохирургического вмешательства) приведены ниже:

RR (мс):

500	492	492	496	496	496	500
496	496	492	496	496	492	496
492	492	492	492	492	488	488
492	492	492	492	492	488	492
492	492	488	492	488	488	492
492	488	488	488	488	488	488

488	492	488	492	488	488	488
492	492	492	492	492	492	492
492	496	492	492	496	496	496
496	496	496	496	496	496	492
496	488	492	488	496	492	492
492	492	492	496	496	496	492
492	496	492	496	488	488	488
484	484	484	488	488	488	492
488	484					

Синусовая тахикардия в условиях патологии встречается при интоксикациях, шоке, неврастении и т.д. Синусовая тахикардия сопровождается увеличением числа сокращений сердца от 90 и более в минуту.

г) Желудочковая экстрасистолия

Измерение длительности 100 интервалов RR в мсек по ЭКГ больной с диагнозом “Желудочковая экстрасистолия” (больная на 6-е сутки после нейрохирургического вмешательства) Приведены ниже.

RR (мс):

720	724	720	420	1004	728	724
728	728	440	1044	752	748	740
740	728	1096	776	736	740	740
424	1096	760	728	744	736	744
736	732	736	748	748	748	728
736	732	748	736	736	736	736
736	728	720	404	1028	716	704
700	692	688	676	680	688	696
708	708	704	708	716	724	720
712	724	720	728	416	1028	720
716	724	732	1452	732	716	728
720	724	740	744	748	744	732
744	420	1080	740	724	728	736
736	728	744	748	744	748	752
760	764					

Экстрасистолия характеризуется преждевременным возбуждением и сокращением сердца в результате появления дополнительного очага повышенной возбудимости в сердечной мышце. (Уточнение локализации очага возбуждения (предсердия, желудочки) возможно только при электрокардиографическом, исследовании.) После такого преждевременного

сокращения очередной импульс, возникающий в синусовом узле, не реализуется, и поэтому следует более длинная пауза, которую больные нередко ощущают как “замирания”, перебои в работе сердца. Экстрасистолы могут возникать при всех органических заболеваниях сердца, прежде всего при ишемической болезни, пороках и т.д., но могут наблюдаться и без органической патологии, прежде всего при неврастении.

Все гистограммы плотности относительных частот в заданиях а-г необходимо построить в одном масштабе, друг под другом.

Вычислите среднее выборочное значение $\overline{x_B}$ и выборочное среднее квадратическое отклонение σ_B. Проведите сглаживание гистограмм с помощью функций нормального распределения.

Для сравнения данных некоторых сердечных аритмий с данными в норме заполните таблицу:

Вид сердечной аритмии	Минимальное значение x_{min}	Максимальное значение x_{max}	Величина интервала Δx	Среднее значение $\overline{x_B}$	Среднее квадратическое отклонение σ_B	Частота ритма сердца $N = \frac{1}{\overline{x_B}}$	Закон Гаусса для конкретных параметров
Норма							
Мерцательная аритмия							
Синусовая тахикардия							
Желудочковая экстрасистолия							

Сделайте вывод.

Варианты, данных измерения артериального давления:

Данные о систолическом давлении крови x (мм рт.ст.) у 100 прак-

тически здоровых женщин в возрасте 60—69 лет приведены ниже. Построить гистограмму плотности относительных частот. Вычислить $\overline{x_B}$, σ_B.

x (мм рт.ст.): 121 152 81 101 134 142 73 159 112

113	127	126	95	101	135	137	143	133	98	102	132	133
136	129	115	111	106	109	134	131	118	113	122	150	141
136	136	106	110	124	125	134	139	111	114	138	148	151
120	130	131	141	115	122	128	132	140	143	144	154	131
140	120	123	140	145	146	116	147	148	156	112	125	135
149	105	119	121	142	116	119	128	137	137	144	107	113
123	127	139	124	126	139	139	125	130	127	127		

Указание. Для выполнения задания рекомендуется принять x_0=70, $x_к$= 160, число интервалов k = 9.

б) гипертоническая болезнь

Значения артериального давления крови x (мм рт.ст.) у 50 женщин в возрасте 60-69 лет с диагнозом “гипертоническая болезнь” составляют:

192	185	171	119	172	186	193	126	173
187	194	145	149	151	152	144	173	194
137	187	175	166	156	153	175	196	161
162	165	176	154	164	159	177	182	161
165	179	172	158	148	464	169	171	157
162	178	163	173					

Используя ПК, получить гистограмму плотности относительных частот и рассчитать $\overline{x_B}$, σ_B.

Указание. Рекомендуется выбрать x_{min}=110, x_{max}=200, n=7. Гистограмму (б) следует построить под гистограммой задания (а), используя числовую ось с тем же началом и масштабной единицей, что и в задании (а).

Сравнить результаты, полученные в задании б) с результатами задания а). Вывод записать в отчет.

3.4 Контрольные вопросы

1. Понятие дискретной и непрерывной случайной величины (привести примеры).
2. Понятие частоты, плотности частоты, относительной частоты, плотности относительной частоты, случайной величины.

3. Статистическая вероятность случайного события.
4. Функция плотности распределения вероятностей.
5. Что называется статистическим интервальным рядом распределения?
6. Что называется гистограммой?
7. Формулы для вычисления среднего выборочного значения и выборочного среднего квадратического отклонения непрерывной случайной величины?

Лабораторная работа № 4

4.1 Моделирование изменения численности популяции. Модель естественного роста (модель Мальтуса)

Цель работы: научиться составлять и решать (аналитически и с помощью ПК) кинетические уравнения при моделировании процессов изменения численности популяций, проводить анализ полученных решений, графически представлять результаты.

4.2 Теоретические сведения

Реальная система: имеется некоторая популяция одного вида (микроорганизмы, зайцы и т.п.), в которой происходят жизненные процессы во всем их многообразии.

Постановка задачи. Найти законы изменения численности популяции во времени.

Основные допущения:

1. Существуют только процессы размножения и естественной гибели, скорости которых пропорциональны численности, особей в данный момент времени.

2. Не учитываем биохимические, физиологические процессы.

3. Нет борьбы между особями за место обитания, за пищу (бесконечно большое пространство и количество пищи).

4. Рассматриваем только одну популяцию, нет хищников.

Модель.

Введем величины:

x — численность популяции в момент t;

R — скорость размножения, γ — коэффициент размножения;

S — скорость естественной гибели, σ — коэффициент естественной гибели;

$\frac{dx}{dt}$ — скорость изменения численности популяции;

ε — коэффициент роста.

Тогда R=γx, S=-σx.

Составим дифференциальное уравнение баланса: изменение численности особей в единицу времени определяется количеством рожденных за это время и умерших:

$$\frac{dx}{dt} = (\gamma - \sigma)x\,,$$

или

$$\frac{dx}{dt} = \varepsilon x\,.$$

Начальное условие: при t=0 численность особей $x=x_0$.

Решим уравнение:

$$\int_{x_0}^{x} \frac{dx}{x} = \int_{0}^{t} \varepsilon dt \qquad \ln\frac{x}{x_0} = \varepsilon t\,,$$

отсюда

$$x = x_0 \cdot e^{\varepsilon t}\,.$$

Графики для различных параметров системы приведены на рисунках 4.1-4.4.

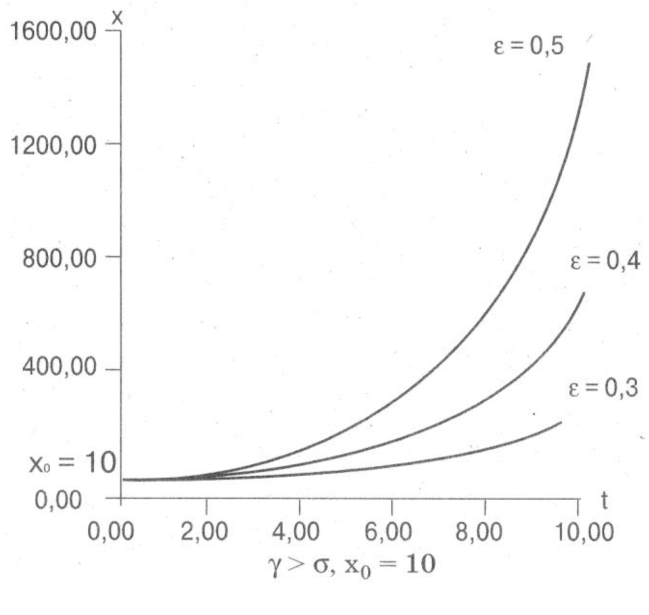


Рисунок 4.1 - Изменение численности особей при ε>0

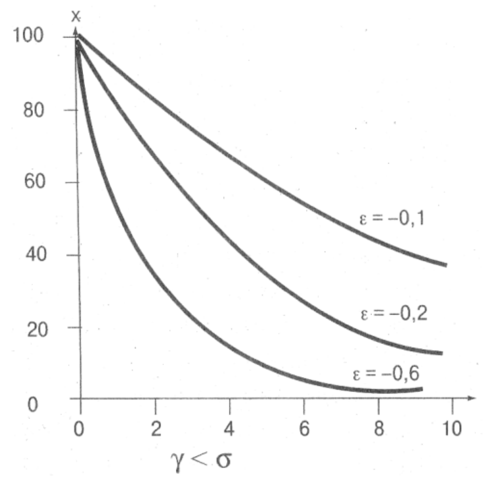


Рисунок 4.2 - Изменение численности особей при ε<0

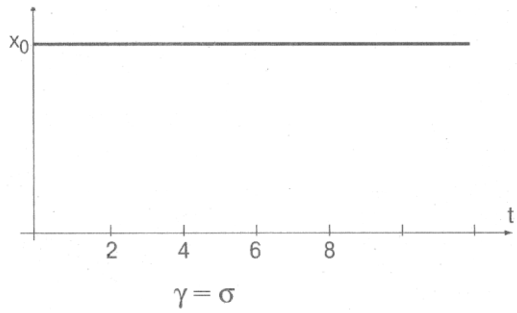
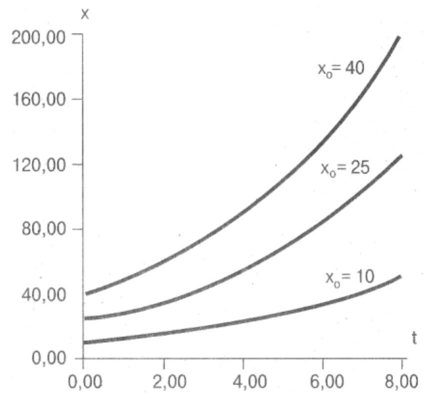


Рисунок 4.3 - Постоянное значение численности особей при ε=0



ε=const=0,2

Рисунок 4.4 - Изменение численности особей вариации x_0

4.3 Порядок выполнения работы

Проанализируйте поведение системы при различных параметрах γ, σ, x_0 .

Для этого:

1. Запишите закон изменения x(t) для заданных параметров.

2. Рассчитайте с помощью компьютера x(t).

3. Постройте графики x(t). Кривые x(t) для разных γ должны быть представлены все на одном рисунке, соответственно для разных σ — на другом, для разных x_0 — на третьем (см. примеры в теоретических сведениях).

4. Оцените из графиков характерные величины процесса:

а) Время $T_{0,5}$, когда численность особей изменится в 2 раза по сравнению с первоначальной. Сопоставьте с расчетными величинами:

$$T_{0,5} = \frac{\ln 2}{|\varepsilon|}.$$

Постройте график зависимости $T_{0,5}$ ($|\varepsilon|$).

Сделайте вывод о влиянии коэффициента роста на характерное время $T_{0,5}$.

б) Скорость изменение численности популяции при t=0. Сопоставьте с расчетными величинами:

$$\left.\frac{dx}{dt}\right|_{t=0} = x_0\varepsilon .$$

Сделайте вывод о влиянии ε на скорость изменения численности популяции в начальный момент времени.

Задание 1. Проанализируйте поведение системы при изменении коэффициента роста $\varepsilon > 0$ ($\gamma>\sigma$). Заполните таблицу:

Параметры	γ, 1/час.	σ, 1/час.	ε, 1/час.	x_0	Закон изменения x(t)
1 система	0,9	0,6	0,3	50	
2 система	1,0	0,6	0,4	50	
3 система	1,1	0,6	0,5	50	

Задание 2. Проанализируйте поведение системы при изменении коэффициента роста $\varepsilon<0$ ($\gamma<0$). Заполните таблицу:

	γ	σ	ε	x_0	Закон изменения x(t)
1 система	0,5	0,6	-0,1	50	

2 система	0,4	0,6	-0,2	50	
3 система	0	0,6	-0,6	50	

Задание 3. Проанализируйте поведение системы при коэффициенте роста ε=0 (γ=0). Заполните таблицу:

Параметры	γ	σ	ε	x_0	Закон изменения x(t)
Система	0,6	0,6	0	50	

Задание 4. Проанализируйте поведение системы при изменениях начальной численности особей x_0. Заполните таблицу:

Параметры	γ	σ	ε	x_0	Закон изменения x(t)
1 система	0,8	0,6	0,2	10	
2 система	0,8	0,6	0,2	50	
3 система	0,8	0,6	0,2	100	

4.4 Контрольные вопросы

1. Математическое моделирование. Основные этапы моделирования. Основные допущения. Понятие об адекватности модели.
2. Кинетические уравнения (общий принцип их построения). Смысл переменных величин входящих в них (аргумент, параметры функции).
3. Составление уравнений для заданной системы.

Лабораторная работа № 5

5.1 Модель изменения численности популяции с учетом внутривидовой конкуренции (модель Ферхюльста)

Цель работы: Научиться составлять и решать (аналитически и с помощью ПК) кинетические уравнения при моделировании процессов изменения численности популяций, проводить анализ полученных решений, графически представлять результаты.

5.2 Теоретические сведения

Усложним рассмотренную в предыдущей работе модель. С целью получения решения, лучше описываемого изучаемый объект, среди допущений, приведенных в предыдущей модели**,** снимем допущение 3. Пусть существует борьба между особями, например, за место обитания, тем самым добавляется дополнительный источник гибели. Считая, что скорость гибели за счет конкуренции между особями пропорциональна вероятности встречи двух особей, можно записать $S=-\delta x\cdot x-\sigma x$ (δ — коэффициент пропорциональности). Тогда уравнение баланса численности особей:

$$\frac{dx}{dt}=\gamma x-\sigma x-\delta x^2$$

или

$$\frac{dx}{dt}=\varepsilon x-\delta x^2.$$

Это нелинейное дифференциальное уравнение. Сделаем замену переменных: $u=(\varepsilon-\delta\cdot x)/x$. Тогда с учетом, что при $t=0$ $x=x_0$ получим:

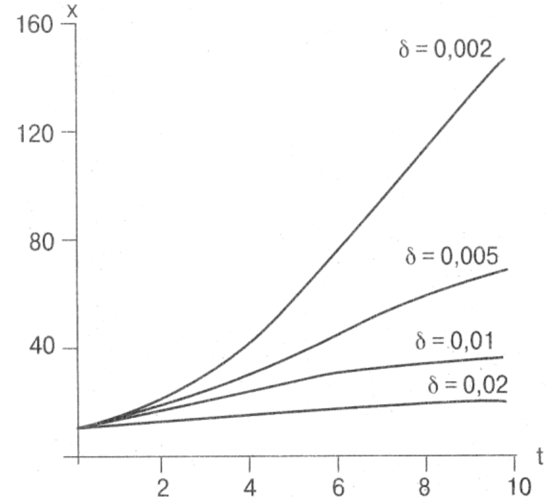
$$\ln\left(\frac{x}{x_0}\right)-\ln\frac{(\varepsilon-\delta x)}{(\varepsilon-\delta x_0)}=\varepsilon\cdot t.$$

Отсюда

$$x(t)=\frac{x_0\cdot\varepsilon}{(\varepsilon-\delta\cdot x_0)\cdot e^{-\varepsilon t}+\delta\cdot x_0}$$

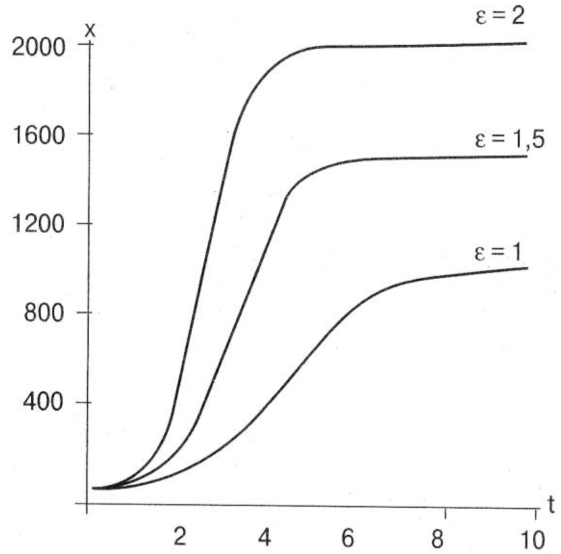
при $t\to\infty$ $x\to X_{ст}=\dfrac{\varepsilon}{\delta}$

Графики для различных параметров системы приведены на рисунках 5.1-5.3.



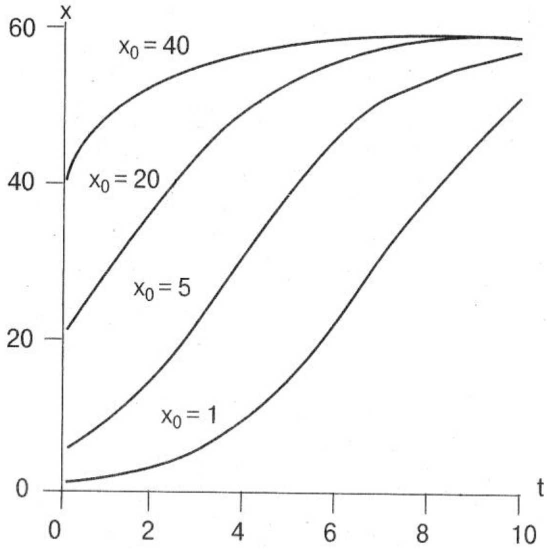
ε=const=0,4 X_0=const=10

Рисунок 5.1 - Влияние δ на x(t)



δ=const=0,001 X_0=const=10

Рисунок 5.2 - Влияние ε на x(t)



δ=const=0,01 ε=const=0,6

Рисунок 5.3 - Влияние x_0 на x(t)

Чем меньше вероятность внутривидовой конкуренции, тем быстрее растет численность особей.

5.3 Порядок выполнения работы

Проанализируйте поведение системы при различных параметрах ε, δ, x_0.

Для этого:

1. Запишите закон изменения x(t) для заданных параметров.

2. Рассчитайте с помощью ПК x(t).

3. Постройте графики x(t). Кривые для каждого вида параметров должны быть представлены на одном рисунке (см. примеры в теоретических сведениях).

4. Оцените характерные величины процесса:

а) Стационарное значение $x_{ст}$, сравните с расчетными данными

$$x_{ст} = \frac{\varepsilon}{\delta}.$$

Постройте графики $x_{ст}(\varepsilon)$, $x_{ст}(\delta)$.

б) Характерное время $T_{0,9}$, когда численность популяции составляет

x=0,9 (т.е. практически выходит на стационарный уровень).

Постройте графики $T_{0,9}(x_0)$, $T_{0,9}(\varepsilon)$, $T_{0,9}(\delta)$.

Сделайте вывод.

в) Параметры точки перегиба – время t_K и численность особей x_K , когда проявляется конкуренция их между собой (рисунок 5.4).

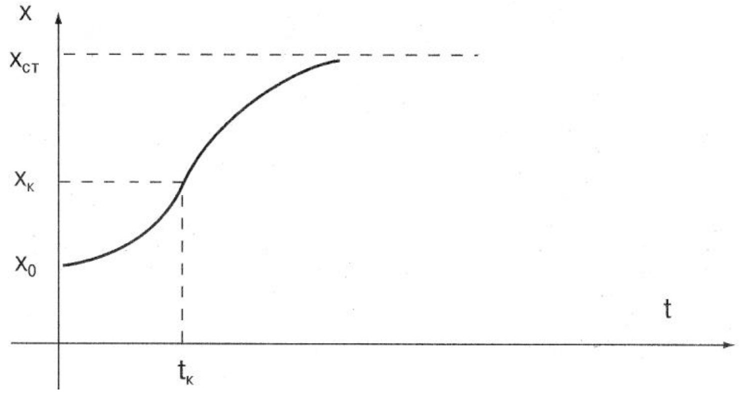


Рисунок 5.4 - Положение точки перегиба

Для этого по графику оцените x_K, сравните с теоретическим значением:

$$x_K = \frac{1}{2} x_{ст} = \frac{\varepsilon}{2\delta}.$$

По графику найдите t_K.

Постройте графики

$x_K(\varepsilon)$, $x_K(\delta)$, $x_K(x_0)$

$t_K(\varepsilon)$, $t_K(\delta)$, $t_K(x_0)$.

Сделайте вывод.

Задание 1. Проанализируйте поведение системы при изменении коэффициента роста ε. Заполните таблицу:

Параметры	ε, 1/час.	δ, 1/час.	x_0	Закон изменения x(t)	$x_{ст}$	$x_к$	t_K
1 система	2,2	0,01	20				
2 система	1,6	0,01	20				
3 система	1	0,01	20				

Задание 2. Проанализируйте поведение системы при изменении коэффициента δ (вероятности конкуренции). Заполните таблицу:

Параметры	ε	δ	x_0	Закон изменения x(t)	$x_{ст}$	$x_к$	t_K
1 система	0,4	0,001	20				
2 система	0,4	0,003	20				
3 система	0,4	0,01	20				
4 система	0,4	0,02	20				

Задание 3. Проанализируйте поведение системы при изменении начальной численности особей x_0. Заполните таблицу:

Параметры	ε	δ	x_0	Закон изменения x(t)	$x_{ст}$	$x_к$	t_K
1 система	0,6	0,01	60				
2 система	0,6	0,01	40				
3 система	0,6	0,01	20				
4 система	0,6	0,01	10				

5.4 Контрольные вопросы

1. Математическое моделирование. Основные этапы моделирования. Основные допущения. Понятие об адекватности модели.
2. Кинетические уравнения (общий принцип их построения). Смысл переменных величин входящих в них (аргумент, параметры функции).
3. Составление кинетических уравнений для заданной системы.
4. Методы решения дифференциальных уравнений первого порядка с разделяющимися переменными.

Лабораторная работа № 6

6.1 Использование методов математической статистики в медицинской диагностике

Цель работы: овладеть знаниями об использовании математической статистики в медицинской диагностике, овладеть навыками практического использования методов математической статистики в медицинской диагностике.

6.2 Теоретические сведения

Математическая статистика – наука о математических методах систематизации и использования статистических данных для решения научных и практических задач.

Предположим, что необходимо изучить множество объектов, по какому-либо признаку. Это, возможно, сделать либо проведя сплошное наблюдение, либо выборочное.

Большая статистическая совокупность, из которой отбирается часть объектов для исследования, называется генеральной совокупностью, а множество объектов, отобранных из нее - выборкой.

Если записать в последовательности измерений все значения величины X в выборке, то получим простой статистический ряд. Например, рост мужчин (см): 171,172,172,168,170,169,… . Такой ряд неудобен для анализа, так как в нем нет последовательности возрастания (или убывания) значений, встречаются и повторяющиеся величины. Поэтому целесообразно ранжировать ряд, например, в возрастающем порядке значений и указать их повторяемость. Тогда статистическое распределение выборки:

$$\begin{matrix} X_1 < X_2 \ldots < X_i \ldots < X_k \\ n_1 \ldots n_2 \ldots\ldots n_i \ldots\ldots n_k \end{matrix},$$

$$p_1^* = \frac{n_1}{n} p_2^* = \frac{n_2}{n} p_i^* = \frac{n_i}{n} p_k^* = \frac{n_k}{n}$$

где X_i - наблюдаемые значения признака; n_i - число наблюдений варианта X_i (частота); p_i - относительная частота.

Общее число объектов в выборке (объем выборки)

$$n = \sum_{i=1}^{k} n_i$$

Здесь k - вариант. Статистическое распределение – это совокупность вариантов и соответствующих им частот.

В медицинской литературе статистическое распределение, состоящее из вариантов и соответствующих им частот, получило название вариационного ряда.

Наряду с дискретным (точечным) статистическим распределением, используют непрерывное (интервальное) распределение:

$$\begin{array}{l} X_0, X_1 \ldots X_1, X_2 \ldots X_{i-1}, X_i \ldots X_{k-1}, X_k \\ n_1 \ldots\ldots\ldots\ldots\ldots n_2 \ldots\ldots\ldots n_i \ldots\ldots\ldots\ldots n_k \\ p_1^* = \dfrac{n_1}{n} \ldots p_2^* = \dfrac{n_2}{n} \ldots p_i^* = \dfrac{n_i}{n} \ldots p_k^* = \dfrac{n_k}{n} \end{array}$$

Здесь X_{i-1}, X_i $-i$-й интервал, в котором заключено количественное значение признака; n_i - сумма частот вариант, попавших в этот интервал; p_i^* - сумма относительных частот.

6.3 Выполнение работы

Задано статистическое распределение массы новорожденных мальчиков (кг) и частоты.

2,7	2,8	2,9	3,0	3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2	4,3	4,4
1	2	3	7	8	12	13	10	7	6	5	6	6	5	3	3	2	1

Общее количество мальчиков

$$n = \sum_{i=1}^{k} n_i = 100$$

Представляем это распределение как непрерывное

2,65-2,75	2,75-2,85	2,85-2,95	2,95-3,05	3,05-3,15	…
1	2	3	7	8	…

Представим статистическое распределение в виде полигона и гистограммы.

Полигон частот – ломаная линия, отрезки которой соединяют точки с координатами $(X_1;n_1)$, $(X_2;n_2)$, или для полигона относительных частот- с координатами $(X_1;p_1^*)$, $(X_2;p_2^*)$ (рисунок 6.1).

Гистограмма частот – совокупность смежных прямоугольников, построенных на одной прямой линии (рисунок 6.2), основания прямоугольников одинаковы и равны *a*, а высоты равны отношению частоты (или относительной частоты) к *a*.

$$\frac{n_i}{a} \text{ или } \frac{n_i}{n_a} = \frac{p_i}{a}.$$

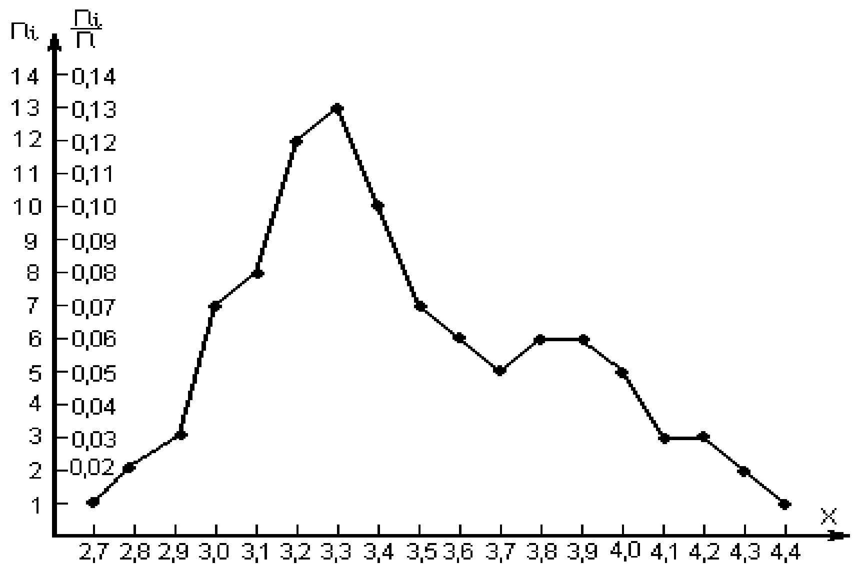


Рисунок 6.1 - Полигон частот

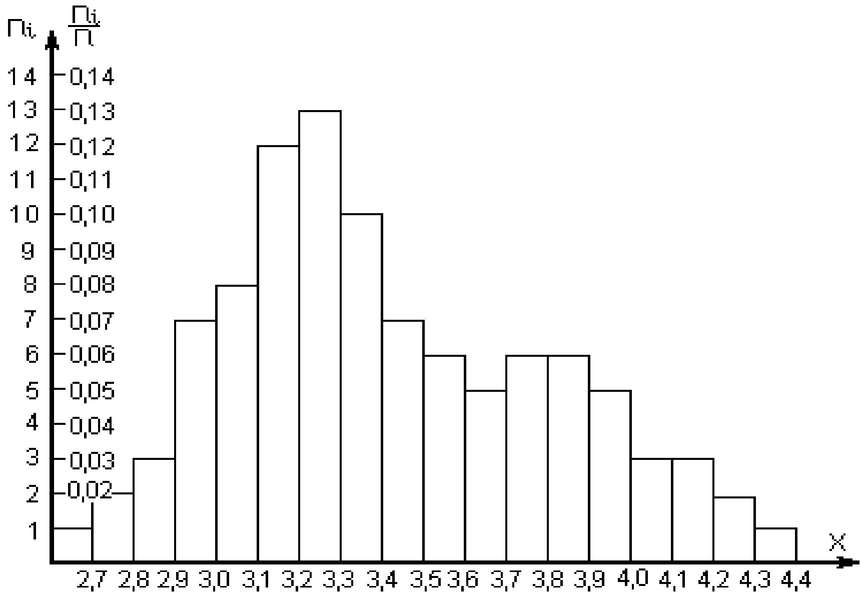


Рисунок 6.2 - Гистограмма частот

Таким образом, площадь каждого прямоугольника равна соответственно

$$\frac{n_i}{a}a = n_i \text{ или } \frac{p_i^*}{a}a = p_i^*.$$

Следовательно, площадь гистограммы частот $n = \sum_{i=1}^{k} n_i$, а площадь гистограммы относительных частот

$$\sum_{i=1}^{k} p_i^* = \sum_{i=1}^{k} \frac{n_i}{n} = \frac{1}{n}\sum_{i=1}^{k} n_i = \frac{n}{n} = 1$$

Рассчитаем моду, медиану и выборочную среднюю.

Мода (M_0) равна варианте, которой соответствует наибольшая частота. В распределении

M_0= 3,3 кг

Медиана (M_e) равна варианте, которая расположена в середине статистического распределения. Она делит статистический ряд на две равные части. При четном числе вариант за медиану принимают среднее значение из двух центральных вариант.

M_e = 3,4 кг

Выборочная средняя $(X_в)$ определяется как среднее арифметическое значение вариант статистического ряда:

$$\overline{X}_в = \frac{1}{n}\sum_{i=1}^{k} X_i n_i$$

или

$$\overline{X}_в = \sum_{i=1}^{k} X_i p_i^*$$

$$X_в = \frac{2.7\cdot 1 + 2.8\cdot 2 + 2.9\cdot 3 + ... + 4.3\cdot 2 + 4.4\cdot 1}{100} = 3.468 кг.$$

Выборочная дисперсия –среднее арифметическое квадратов отклонения вариант от их среднего значения:

$$\sigma_в^2 = \frac{1}{n}\sum_{i=1}^{k}\left(X_i - \overline{X}_в\right)^2$$

Квадратный корень из выборочной дисперсии называют выборочным средним квадратическим отклонением:

$$\sigma_в = \sqrt{\sigma_в{}^2}$$

$$\sigma_в^2 = \frac{(2,7-3,468)^2 + (2,8-3,468)^2 + ... + (4,3-3,468)^2 + (4,4-3,468)^2}{100} = ;$$

$$= 0,1513 кг^2$$

$$\sigma_в = \sqrt{0,1513 кг^2} = 0,3896 кг.$$

Рассчитаем исправленную выборочную дисперсию:

$$s^2 = \frac{n}{n-1}\sigma_в^2$$

$$s^2 = \frac{100}{99}\cdot 0.1513 = 1.5181, \qquad s = 1,2321.$$

Зададим доверительную вероятность $p = 0.95$. Из таблицы 6.1 находим для заданных значений $p = 0.95$ и $n = 100$ коэффициент Стьюдента $t_{0.95;100} = 2$. Находим интервал:

$$\overline{X}_{в} - t\frac{s}{\sqrt{n}} \angle \mu \angle \overline{X}_{в} + t\frac{s}{\sqrt{n}}$$

$$3.47 - 2 \cdot \frac{1.5181}{\sqrt{100}} \angle \mu \angle 3.47 + 2 \cdot \frac{1.5181}{\sqrt{100}}$$

или $3{,}166\text{кг} \angle \mu \angle 3{,}77\text{кг}$.

Сделать выводы по результатам работы.

Таблица 6.1 - Коэффициенты Стьюдента

n	p											
	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,98	0,99
2	0,16	0,33	0,51	0,73	1,00	1,38	2,0	3,1	6,3	12,7	31,8	63,7
3	13	29	45	62	0,82	1,06	1,3	1,9	2,9	4,3	7,0	9,9
4	13	28	42	58	77	0,98	1,3	1,6	2,4	3,2	4,5	5,8
5	13	27	41	57	74	94	1,2	1,5	2,1	2,8	3,7	4,6
6	13	27	41	56	73	92	1,2	1,5	2,0	2,6	3,4	4,0
7	13	27	40	55	72	90	1,1	1,4	1,9	2,4	3,1	3,7
8	13	26	40	55	71	90	1,1	1,4	1,9	2,4	3,0	3,5
9	13	26	40	54	71	90	1,1	1,4	1,9	2,3	2,9	3,4
10	13	26	40	54	70	88	1,1	1,4	1,9	2,3	2,8	3,3
11	13	26	40	54	70	88	1,1	1,4	1,8	2,2	2,8	3,2
12	13	26	40	54	70	87	1,1	1,4	1,8	2,2	2,7	3,1
13	13	26	40	54	70	87	1,1	1,4	1,8	2,2	2,7	3,1
14	13	26	39	54	69	87	1,1	1,4	1,8	2,2	2,7	3,0
15	13	26	39	54	69	87	1,1	1,3	1,8	2,1	2,6	3,0
16	13	26	39	54	69	87	1,1	1,3	1,8	2,1	2,6	2,9
17	13	26	39	54	69	86	1,1	1,3	1,7	2,1	2,6	2,9
18	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,6	2,9
19	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,6	2,9
20	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,5	2,9
21	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,5	2,8
22	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,5	2,8
23	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,5	2,8
24	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,5	2,8

25	13	26	39	53	69	86	1,1	1,3	1,7	2,1	2,5	2,8
26	13	26	39	53	68	86	1,1	1,3	1,7	2,1	2,5	2,8
27	13	26	39	53	68	86	1,1	1,3	1,7	2,1	2,5	2,8
28	13	26	39	53	68	86	1,1	1,3	1,7	2,1	2,5	2,8
29	13	26	39	53	68	86	1,1	1,3	1,7	2,0	2,5	2,8
30	13	26	39	53	68	85	1,1	1,3	1,7	2,0	2,5	2,8
40	13	26	39	53	68	85	1,1	1,3	1,7	2,0	2,4	2,7
60	13	25	39	53	68	85	1,1	1,3	1,7	2,0	2,4	2,7
120	13	25	39	53	68	85	1,1	1,3	1,7	2,0	2,4	2,6
∞	13	25	39	52	67	84	1,1	1,3	1,7	2,0	2,3	2,6

Варианты заданий:

1. Задано статистическое распределение. Доверительная вероятность 0,5.

33	34	35	36	37	38	39	40	41	42	43
1	2	5	7	27	23	16	10	5	2	2

2. Задано статистическое распределение. Доверительная вероятность 0,8.

15	16	17	18	19	20	21	22	23	24	25	26
1	2	2	3	4	5	3	2	1	1	1	1

3. Задано статистическое распределение. Доверительная вероятность 0,95.

52	53	54	55	56	57	58	59	60	61	62	63
1	1	1	2	2	3	4	5	4	2	2	1

4. Задано статистическое распределение. Доверительная вероятность 0,7

10	11	12	13	14	15	16	17	18	19	20	21
1	1	2	2	2	3	3	4	5	5	6	6

5. Задано статистическое распределение. Доверительная вероятность 0,6.

1,9	2,0	2,1	2,2	2,3	2,4	2,5	2,6	2,7	2,8	2,9	3,0
1	1	1	1	2	1	1	1	1	1	1	1

6. Задано статистическое распределение. Доверительная вероятность 0,5.

1,5	1,6	1,7	1,8	1,9	2,0	2,1	2,2	2,3	2,4	2,5	2,6
1	2	2	3	4	5	3	2	1	1	1	1

7. Задано статистическое распределение. Доверительная вероятность 0,4.

0,15	0,16	0,17	0,18	0,19	0,20	0,21	0,22	0,23	0,24	0,25	0,26
1	1	1	2	2	2	3	2	1	1	1	1

8. Задано статистическое распределение. Доверительная вероятность 0,3.

100	101	102	103	104	105	106	107	108	109	110	111
1	3	5	7	18	8	6	4	3	2	2	1

9. Задано статистическое распределение. Доверительная вероятность 0,2.

1,2	1,3	1,4	1,5	1,6	1,7	1,8	1,9	2,0	2,1	2,2	2,3
1	9	11	15	20	20	14	8	8	7	6	1

10. Задано статистическое распределение. Доверительная вероятность 0,1.

0,05	0,06	0,07	0,08	0,09	0,1	0,11	0,12	0,13	0,14
1	1	1	1	2	1	1	1	1	1

11. Задано статистическое распределение. Доверительная вероятность 0,95.

48	49	50	51	52	53	54	55	56	57	58
1	7	23	39	54	61	34	20	11	10	6

12. Задано статистическое распределение. Доверительная вероятность 0,8.

52	53	54	55	56	57	58	59	60	61	62	63
1	2	2	3	4	5	3	2	1	1	1	1

13. Задано статистическое распределение. Доверительная вероятность 0,7.

0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2
1	9	11	15	20	20	14	8	8	7	6	1

14. Задано статистическое распределение. Доверительная вероятность 0,9.

4	5	6	7	8	9	10	11	12	13	14	15
1	2	2	3	3	3	3	2	2	2	2	1

15. Задано статистическое распределение. Доверительная вероятность 0,6.

2	4	6	8	10	12	14	16	18	20	22	24
1	1	1	2	5	8	14	8	8	6	4	2

16. Задано статистическое распределение. Доверительная вероятность 0,4.

89	90	91	92	93	94	95	96	97	98	99	100
1	1	2	2	2	2	4	5	3	2	1	1

17.Задано статистическое распределение. Доверительная вероятность 0,9.

0,11	0,12	0,13	0,14	0,15	0,16	0,17
1	7	10	20	10	10	2

18.Задано статистическое распределение. Доверительная вероятность 0,1.

3,1	3,2	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,0	4,1	4,2
1	1	2	2	2	2	4	5	3	2	1	1

19.Задано статистическое распределение. Доверительная вероятность 0,3.

15	16	17	18	19	20	21	22	23	24	25	26
1	1	2	2	2	2	4	5	3	2	1	1

20.Задано статистическое распределение. Доверительная вероятность 0,4.

52	53	54	55	56	57	58	59	60	61	62	63
1	1	1	2	2	3	4	5	4	2	2	1

21.Задано статистическое распределение. Доверительная вероятность 0,3.

27	28	29	30	31	32	33	34	35	36	37	38
1	3	7	9	10	15	20	20	10	15	9	1

22.Задано статистическое распределение. Доверительная вероятность 0,2.

52	53	54	55	56	57	58	59	60	61	62	63
1	1	1	2	2	3	4	5	4	2	2	1

23.Задано статистическое распределение. Доверительная вероятность 0,1.

5,21	5,22	5,23	5,24	5,25	5,26	5,27	5,28	5,29	5,30	5,31	5,32
1	4	6	14	20	20	20	15	10	5	4	1

24.Задано статистическое распределение. Доверительная вероятность 0,2.

7,0	7,1	7,2	7,3	7,4	7,5	7,6	7,7	7,8	7,9	8,0	8,1
1	2	2	2	4	5	2	1	3	4	2	1

25.Задано статистическое распределение. Доверительная вероятность 0,5.

52	53	54	55	56	57	58	59	60	61	62	63
1	2	2	2	4	5	2	1	3	4	2	1

26.Задано статистическое распределение. Доверительная вероятность 0,95.

32	33	34	35	36	37	38	39	40	41	42	43
1	2	2	4	4	4	5	3	2	1	1	1

27.Задано статистическое распределение. Доверительная вероятность 0,8.

27	28	29	30	31	32	33	34	35	36	37	38
1	2	2	2	4	5	2	1	3	4	2	1

28.Задано статистическое распределение. Доверительная вероятность 0,7.

0,5	0,6	0,7	0,8	0,9	1,0	1,1	1,2	1,3	1,4	1,5	1,6
1	1	2	2	3	3	4	3	3	2	2	1

29.Задано статистическое распределение. Доверительная вероятность 0,4.

27	28	29	30	31	32	33	34	35	36	37	38
1	1	2	3	4	7	10	4	3	2	2	1

30.Задано статистическое распределение. Доверительная вероятность 0,1.

11	12	13	14	15	16	17	18	19	20	21	22
1	1	1	1	1	1	2	2	1	1	1	1

6.4 Контрольные вопросы

1. Статистический ряд. Статистическое распределение выборки, непрерывное распределение.
2. Полигон частот, гистограмма частот, графическое представление.
3. Понятия мода, медиана, выборочная средняя, выборочная дисперсия.
4. Доверительный интервал. Нахождение доверительного интервала.

ЛИТЕРАТУРА

1. *Пригожий И.* От существующего к возникающему / М.: Мир, 1985.
2. *Романовский Ю.М., Степанова* Я.В., *ЧернавскийД.С.* Математическая биофизика. М.: Наука, 1984.
3. *Хакен Г.* Синергетика / М.: Мир, 1980.
4. *К. Каро* и др. Механика кровообращения / Пер. с англ. М.: Мир, 1981
5. Физиология человека / Под ред. Шмидта Р. и Тевса Г., т. 2. М.: Мир, 1996.
6. *6.Кудряшов Ю.Б., Беренфельд Б.С.* Основы радиационной биофизики. М.: МГУ, 1982.
7. Радиация. Дозы, эффекты, риск // М.: Мир, 1988.
8. Физический энциклопедический словарь. - М., Советская энциклопедия, - 1984. .
9. *Годик Э.Э., Гуляев ЮЯ.* Физические поля человека и животных // В мире науки. - 1990. - № 5. - С. 75-83.
10. *Гуляев Ю.В., Годик Э.Э., Петров А.В., Тараторим. А.М.* О возможностях дистантной функциональной диагностики биологических объектов по их собственному инфракрасному излучению // Докл. АН СССР. - 1984. - Т. 277, - № 6. - С. 1486-1491.
11. *Шевелев И А., Кузнецова Г Д., Цыкалов ЕЛ., Горбач АМ., Будко КЛ., ШараевГА.* Термоэнцефалоскопия. - М.: Наука, - 1989. -224 с.