

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

**САНКТ-ПЕТЕРБУРГСКИЙ НАЦИОНАЛЬНЫЙ
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ**

И.С. Осетрова, Н.А. Осипов

Microsoft Excel 2010 для аналитиков

Учебное пособие



Санкт-Петербург

2013

УДК 004.655, 004.657, 004.62

И.С. Осетрова, Н.А. Осипов

Microsoft Excel 2010 для аналитиков – СПб: НИУ ИТМО, 2013. – 65 с.

В пособии излагаются основы наиболее распространенных методов обработки результатов наблюдений и анализа данных и методические указания к выполнению лабораторных работ по дисциплине «Моделирование инфокоммуникационных систем».

Предназначено для студентов, обучающихся по всем профилям подготовки бакалавров направления: 210700 Инфокоммуникационные технологии и системы связи.

Рекомендовано к печати Ученым советом факультета Инфокоммуникационных технологий, протокол № 3 от 19 марта 2013 г.



В 2009 году Университет стал победителем многоэтапного конкурса, в результате которого определены 12 ведущих университетов России, которым присвоена категория «Национальный исследовательский университет». Министерством образования и науки Российской Федерации была утверждена программа его развития на 2009–2018 годы. В 2011 году Университет получил наименование «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики».

© Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики, 2013

© И.С. Осетрова, Н.А. Осипов, 2013.

Оглавление

Введение.....	4
1. Анализ данных с помощью сценариев “ЧТО-ЕСЛИ”	5
2. Расчет характеристик марковских процессов	11
3. Решение задач линейной оптимизации.....	17
4. Изучение законов распределения случайных величин	21
5. Генерация случайных чисел и анализ выборки данных.....	28
6. Анализ временных рядов	37
7. Регрессионный анализ	44
8. Работа со сводными таблицами.....	51
Список литературы	60

Введение

Экспериментальные исследования являются основным источником объективной информации о характеристиках процессов, протекающих в реальных объектах, в том числе в средствах и комплексах телекоммуникаций. Целью обработки экспериментальных данных является выявление закономерностей в характеристиках исследуемых объектов и процессов. Результаты обработки ЭД позволяют оценить качество объекта, они необходимы для оперативного управления процессами, решения задач адаптации объекта к изменившимся условиям или формирования требований ко вновь создаваемым системам.

Практические знания из области проведения и организации эксперимента, умения и навыки в работе с измерительными приборами, владение аппаратом статистического анализа результатов требуются и в деятельности инженера-практика, и в деятельности инженера-исследователя.

Данное пособие посвящено расширенным возможностям MS Excel 2010 для анализа экспериментальных данных.

Для качественного усвоения материала, обучаемые должны знать интерфейс программы Excel 2010, уметь вводить данные, форматировать ячейки, выполнять элементарные математические операции с данными, работать с ярлыками листов, иметь представление о формате данных.

После изучения данного курса Вы расширите свои знания о возможностях MS Excel, получите практические навыки применения функций программы, анализа данных, сможете использовать инструменты MS Excel для статистического и регрессионного анализа.

Для отработки практических навыков используются готовые рабочие книги, которые располагаются в сетевой папке. Скопируйте их на локальную машину перед выполнением упражнений.

1. Анализ данных с помощью сценариев “ЧТО-ЕСЛИ”

Цель работы

Изучить средства программы Microsoft Excel для анализа данных с помощью сценариев и таблиц подстановки.

Теоретические основы

Во многих случаях возникает необходимость просматривать различные варианты значений в ячейках на листе. Для выполнения анализа “что-если” Excel предоставляет следующие возможности:

- диспетчер сценариев;
- таблицы подстановки.

С помощью механизма сценариев можно создать и сохранить на листе различные группы значений ячеек, а затем переключаться на любой из них для просмотра различных результатов.

Сценарий представляет собой версию рабочего листа, отражающую результаты различных условий и предположений в виде набора значений ячеек, которые могут автоматически подставляться в соответствующие ячейки на листе.

Диспетчер сценариев открывается следующим образом.

Перейдите на вкладку **Данные**, в группе **Работа с данными**, нажмите кнопку **Анализ “что-если”** и выберите команду **Диспетчер сценариев...**

В результате откроется окно диалога (см. рис.1.1):

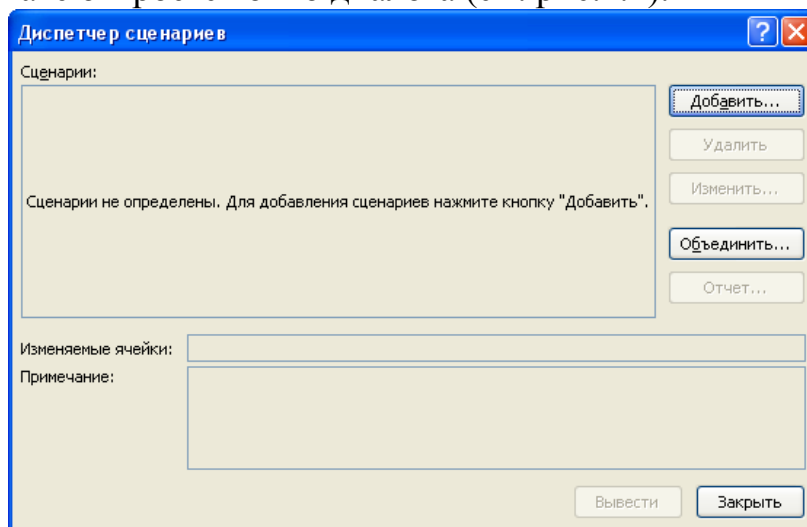


Рис.1.1 Диспетчер сценариев

В окне **Диспетчер сценариев (Scenario Manager)** следует нажать кнопку **Добавить (Add)**. Отобразится окно **Добавление сценария (Add Scenario)**, изображенное на рис. 1.2.

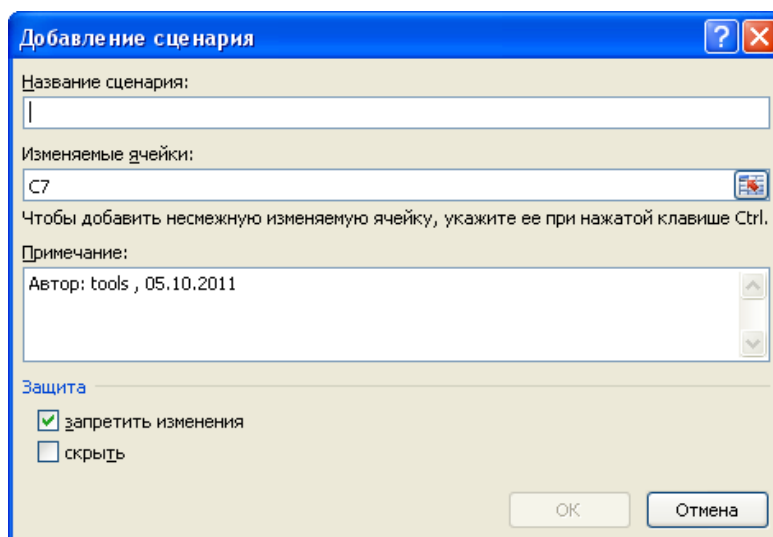


Рис. 1.2 Добавление сценариев

Далее указывается имя сценария, ячейки, в которых записаны изменяемые данные.

Таблицы подстановки (данных) помогают создавать сценарии типа «что-если». Например, изменяя значения одного или двух параметров, можно проследить их влияние на результирующее значение.

Таблица подстановки представляет собой диапазон ячеек, содержащий результаты подстановки различных значений в одну или несколько формул. Существует два типа таблиц данных: таблицы с одним входом и таблицы с двумя входами.

В обоих случаях используется одно диалоговое окно (см. рис. 1.3), для открытия которого следует выполнить следующие действия:

Вкладка **Данные** → группа **Работа с данными** → кнопка **Анализ “что-если”** → команда **Таблица данных...**

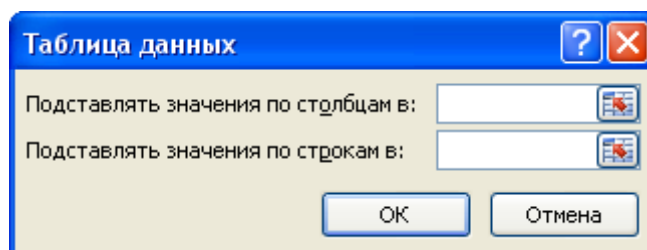


Рис.1.3 Окно диалога для применения таблиц подстановки

В случае одного входа данные подставляются либо по столбцам (если данные располагаются в строках), либо по строкам (если данные располагаются в столбце).

В случае использования двух входных параметров данные подставляются и по строкам и по столбцам.

Содержание работы

Основная часть работы состоит в проведении анализа данных с помощью диспетчера сценариев и таблиц подстановки.

Порядок выполнения

Упражнение 1. Создание и просмотр сценария

1. Откройте книгу ArrayFormulas.xlsx.
2. Скопируйте лист *Товарный чек* в новую книгу, сохраните ее под именем *Мой_Сценарий*. Книгу ArrayFormulas.xlsx можно закрыть.
3. Подсчитайте общую сумму заказа (ячейка C7).
4. Перед созданием сценария рекомендуется выделить ячейки, данные в которых могут изменяться. В таблице могут изменяться значения в ячейках B2:C5, выделите этот диапазон.
 5. Выполните команду: вкладка **Данные** → группа **Работа с данными** → кнопка **Анализ “что-если”** → команда **Диспетчер сценариев...**
6. В окне **Диспетчер сценариев (Scenario Manager)** нажмите кнопку **Добавить (Add)**. Отобразится окно **Добавление сценария (Add Scenario)**.
7. В поле *Название сценария (Scenario name)* введите название создаваемого сценария – «Вариант исходный».
8. Проверьте, что изменяемые ячейки, выделенные до начала создания сценария автоматически указаны в поле *Изменяемые ячейки (Changing cells)*.
9. В поле *Примечание (Comment)* просмотрите автоматически вставляемые имя автора сценария и дата создания. При желании дополните примечание необходимой информацией (или очистите поле). Содержание примечания будет отображаться при выборе сценария в поле *Примечание* окна **Диспетчер сценариев (Scenario Manager)**.
10. Нажмите кнопку **ОК**, в результате чего отобразится окно **Значения ячеек сценария (Scenario Values)**.
11. В полях с именами изменяемых ячеек отображены текущие значения в указанных ячейках.
12. Для создания первого сценария, отражающего текущие (исходные) значения и перехода к созданию нового сценария нажмите кнопку **Добавить (Add)**. Снова отобразится окно **Добавление сценария (Add Scenario)**.
13. Создайте еще два сценария с именами «Вариант один» и «Вариант два». Значения ячеек, влияние которых необходимо просмотреть укажите по своему усмотрению.

14.Нажмите кнопку **ОК**. Отобразится окно **Диспетчер сценариев (Scenario Manager)**, в котором будет показан список доступных сценариев.

15.Выберите сценарий «Вариант один» и нажмите кнопку **Вывести (Show)**.

16.Просмотрите сценарий «Вариант два».

17.Нажмите кнопку **Заккрыть (Close)**.

В дальнейшем для отображения требуемого сценария необходимо в окне **Диспетчер сценариев (Scenario Manager)** выбрать нужный сценарий и нажать кнопку **Вывести (Show)**.

Любой сценарий можно изменить или дополнить. Для этого в окне **Диспетчер сценариев (Scenario Manager)** следует выбрать нужный сценарий и нажать кнопку **Изменить (Edit)**, после чего отобразится окно **Изменение сценария (Edit Scenario)**. После внесенных изменений следует нажать кнопку **ОК**.

Упражнение 2. Объединение сценариев

После создания рабочего листа, содержащего сценарий, можно разослать его копии коллегам, чтобы они могли внести в него коррективы или создать свой собственный сценарий. Когда они вернут свои сценарии их можно объединить на одном листе.

1. Создайте новую книгу и скопируйте в нее данные листа *Товарный чек*.
2. Создайте новый сценарий с именем «Вариант Омега». Изменяемые ячейки те же, значения их укажите по своему усмотрению.
3. Откройте исходную книгу *Мой_Сценарий.xlsx* лист *Товарный чек*.
4. Для объединения исходных сценариев с новым выберите: вкладка **Данные** → группа **Работа с данными** → кнопка **Анализ “что-если”** → команда **Диспетчер сценариев...**
5. В окне **Объединение сценариев (Merge Scenarios)** в поле *Книга (Book)* выберите книгу, в поле *Лист (Sheet)* укажите лист с новым сценарием. Нажмите **ОК**.
6. В окне **Диспетчер сценариев (Scenario Manager)** выберите новый сценарий и выведите его на экран, нажав кнопку **Вывести (Show)**.

Упражнение 3. Создание отчета по сценарию

После завершения построения сценариев можно создать отчет.

Для создания отчета выполните следующие действия:

1. Откройте книгу *Мой_Сценарий.xlsx* лист *Товарный чек*.
2. Выполните команду: вкладка **Данные** → группа **Работа с данными** → кнопка **Анализ “что-если”** → команда **Диспетчер сценариев...**

3. В окне **Диспетчер сценариев (Scenario Manager)** нажмите кнопку **Отчет (Summary)**.
4. Проверьте ячейку результата (C7) и нажмите ОК.
5. В открывшемся новом листе *Структура сценария (Scenario Summary)* просмотрите отчет.

Упражнение 4. Создание таблицы подстановки с одним входом

Таблица подстановки с одним входом позволяет выполнить произвольное количество расчетов, изменяя значения в одной ячейке.

Формировать таблицу подстановки с одной переменной следует так, чтобы введенные значения были расположены либо в столбце, либо в строке. Формулы, используемые в таблицах подстановки с одной переменной, должны ссылаться на ячейку ввода – ячейку, в которую подставляется каждое значения из таблицы данных.

1. Откройте книгу Таблицы_подстановки.xlsx лист 1.
2. Изучите функцию ПЛТ(РМТ) в ячейке В7, рассчитывающую сумму периодического платежа в зависимости от общего числа периодов выплат (ячейка В6), годовой процентной ставки (ячейка В5) и общей суммы (В4).
3. Создайте таблицу, показывающую зависимость размеров выплат от количества периодов при постоянной процентной ставке.
 - a. В отдельную строку (в ячейки с Е6 по I6) введите значения, которые следует подставлять в ячейку ввода – 12, 24, 36, 48, 60.
 - b. Проверьте, что введена формула в ячейку, расположенную на один столбец левее и на одну строку ниже первого значения (ячейка D7).
 - c. Выделите диапазон ячеек, содержащий формулы и значения подстановки (D6:I7).
 - d. Выполните команду: вкладка **Данные** → группа **Работа с данными** → кнопка **Анализ “что-если”** → команда **Таблица данных...**
 - e. Введите в поле *Подставлять значения по столбцам в: (Row input cell)* ссылку на ячейку ввода значения периода – В6. Второе поле оставьте пустым. Нажмите ОК.
4. Просмотрите ячейки E7:I7, в которых создалась формула массива, использующая функцию ТАБЛИЦА(Table) с одним аргументом.

Упражнение 5. Создание таблицы подстановки с двумя входами

Таблица подстановки с двумя входами (переменными) содержит результаты вычислений по одной формуле при изменении двух входных параметров.

Создайте таблицу, показывающую зависимость размеров выплат от количества периодов при изменении процентной ставки:

1. Откройте книгу Таблицы_подстановки.xlsx лист 1.
2. Проверьте, что формула введена в ячейку D10.
3. В том же столбце ниже формулы (D11:D16) введите шесть значений подстановки для величины процентной ставки от 6,0% до 8,5%.
4. В строке правее формулы (E10:I10) введите пять значений периода выплат – 12, 24, 36, 48, 60.
5. Выделите диапазон ячеек, содержащий формулу и оба набора данных подстановки (D10:I16).
6. Выполните команду: вкладка **Данные** → группа **Работа с данными** → кнопка **Анализ “что-если”** → команда **Таблица данных...**
7. Введите в поле *Подставлять значения по столбцам в: (Row input cell)* ссылку на ячейку ввода значения периода – B6.
8. Введите в поле *Подставлять значения по строкам в: (Column input cell)* ссылку на ячейку ввода величины процентной ставки – B5.
9. Нажмите ОК. Просмотрите ячейки E11:I16, в которых создалась формула массива, использующая функцию ТАБЛИЦА с двумя аргументами.

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.
2. Результаты выполнения упражнений.
3. Выводы по результатам работы.

2. Расчет характеристик марковских процессов

Цель работы

Изучить средства программы Microsoft Excel для расчета основных характеристик марковских процессов.

Теоретические основы

В настоящее время имеется большое количество компьютерных программ для использования различных итерационных методов. Наиболее простые и удобные программы оформлены в виде функций MS Excel **Подбор параметра** и **Поиск решения**.

Перед тем как рассмотреть основные средства MS Excel полезно изучить возможности программы для решения менее сложных задач, например, задачи исследования функций.

Многие инженерные задачи сводятся к исследованию функций одной или нескольких переменных вида $Y = f(X)$ или $Y = f(x_1, x_2, \dots, x_N)$.

Исследовать функцию, значит установить область ее существования (значения X при которых возможно вычислить Y), определить области значений X , при которых Y принимает положительные, отрицательные и аномально большие значения ("уходит в бесконечность"), найти максимумы, минимумы, иногда точки перегиба графика функции, а также корни уравнения $Y = f(x)$ – значения x , при которых Y обращается в 0 (график функции пересекает ось абсцисс).

Наиболее простые методы исследования функциональных зависимостей с помощью компьютера – итерационные, основанные на многократном выполнении сравнительно простых операций.

Широко известен итерационный метод – табулирование функции, включающий следующие этапы:

- расчет значений Y при заданных X в большом диапазоне значений X с большим шагом;
- табулирование с небольшим шагом в наиболее близких диапазонах – вблизи корней, максимумов и минимумов;
- сужение диапазонов X и уменьшение шага для получения все более точных значений экстремумов и корней. Получаемые решения зависят от того, в каких диапазонах X и Y ведется их поиск, т.е. от их начальных значений.

Пример 1. Решение уравнений

Пример показывает использование функции **Подбор параметра** для решения уравнений.

Требуется: решить уравнение $Y = 0,1 \cdot x^2 - x - 11$.

Решение.

1. С помощью график функции определить количество действительных корней уравнения. Для этого следует:

- 1.1. Задать область определения (x) от -20 до $+20$: занести в соседние ячейки (например, A5 и A6) -20 и -19 , выделить обе ячейки, поставить курсор на черный квадратик в правом нижнем углу, нажать левую клавишу мыши и потащить вниз до появления числа 20.
- 1.2. Создать область значений: в ячейку рядом с -20 вставить формулу $=0,1*A5^2-A5-11$, скопировать ее вниз.
- 1.3. Выделить область значений и на вкладке **Вставка** в группе **Диаграммы** выбрать кнопку **График**. Обратите внимание, что на оси X указываются не значения аргумента, а порядковые номера.
2. Найти первый корень уравнения:
 - 2.1. Сделать активной ячейку в диапазоне Y вблизи одного из корней (первое пересечение графика функции с осью абсцисс).
 - 2.2. Вызвать **Подбор параметра**: на вкладке **Данные**, в группе **Работа с данными** выбрать кнопку **Анализ «что-если»** и указать команду **Подбор параметра**.
 - 2.3. В окне *Значение* установить 0, в нижнем окне *Изменяя значение ячейки* указать адрес ячейки X , соответствующей активной ячейке Y , после чего кликнуть по кнопке **ОК**.
3. Найти второй корень, выбрав значения Y и X вблизи него.

Пример 2. Решение системы уравнений

Расчет установившихся значений вероятностей состояний системы (финальных вероятностей), описанной марковским процессом, сводится к решению системы алгебраических уравнений. Для изучения принципа решения подобных систем рассмотрим решение системы из трех уравнений с тремя неизвестными вида $a_iX + b_iY + c_iZ = d_i$ ($i = 1,2,3$) с помощью команды **Поиск решения**.

Поиск решения – это надстройка, входящая в поставку Excel и предназначенная для решения задач линейной и нелинейной оптимизации. Для этого в ней используются методы и алгоритмы математического программирования, которые позволяют находить оптимальные решения задач оптимизации, представленные в виде табличных моделей. Пример подобной задачи рассматривается в следующей работе. В данном примере с помощью этой надстройки решается система уравнений.

Для решения системы уравнений следует:

1. Выполнить подготовительные действия (см. табл. 2.1):
 - 1.1 Записать начальные значения неизвестных X , Y , Z , например, нулевые значения в ячейки A10, B10, C10 соответственно.
 - 1.2 Составить таблицу, содержащую значения коэффициентов при этих неизвестных a_i , b_i , c_i ($i = 1,2,3$), и значения соответствующих свободных членов d_i , например, в ячейках A12 : D14.

- 1.3 Перемножить начальные значения X , Y , Z на соответствующие коэффициенты (ячейки A16:C18) и просуммировать произведения по строкам (ячейки D16:D18).
2. Запустить **Поиск решения**: на вкладке **Данные**, в группе **Анализ** выбрать команду **Поиск решения**
3. Заполнить окно диалога **Поиск решения** данными решаемой задачи (см. рис. 2.1):
 - 3.1 В качестве целевой ячейки установить первую сумму: задать “Установить целевую ячейку (в примере это ячейка D16) равной значению” первого свободного члена d_1 .
 - 3.2 На две другие суммы наложить ограничения: равенство двум другим свободным членам d_2 и d_3 (используя кнопку “Добавить”).
 - 3.3 В окне “Изменяя ячейки” указать ячейки с начальными значениями неизвестных X , Y , Z (в этих ячейках окажется результат решения системы).
 - 3.4 Нажать кнопку “*Параметры*” и ознакомиться с параметрами и методами, используемыми при оптимизационных расчетах (при необходимости можно параметры изменить), чтобы закрыть окно “*Параметры*”, следует нажать кнопку ОК.
4. Запустить выполнение программы нажатием кнопки “*Выполнить*”. В результате появится сообщение о нахождении или не нахождении решения (см. рис. 2.2).

Таблица 2.1

	A	B	C	D	Комментарии
10	X	Y	Z		Начальные значения (числа), после решения задачи на их месте появятся результирующие значения
11					
12	a_1	b_1	c_1	d_1	Коэффициенты (числа) в уравнениях и значения свободных членов
13	a_2	b_2	c_2	d_2	
14	a_3	b_3	c_3	d_3	
15					
16	$a_1 \cdot X$	$b_1 \cdot Y$	$c_1 \cdot Z$	=Сумм(A16:C16)	Начальный вид системы уравнений (числа)
17	$a_2 \cdot X$	$b_2 \cdot Y$	$c_2 \cdot Z$	=Сумм(A17:C17)	
18	$a_3 \cdot X$	$b_3 \cdot Y$	$c_3 \cdot Z$	=Сумм(A18:C18)	

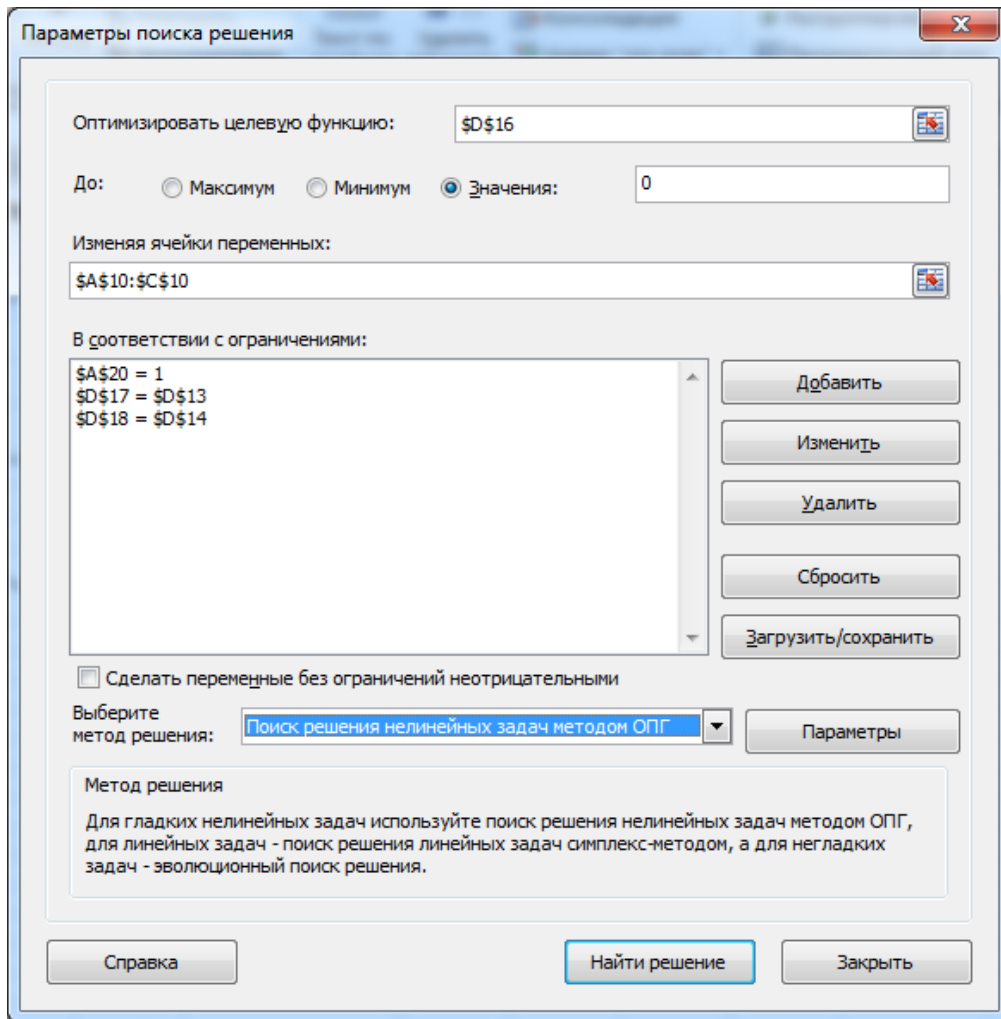


Рис. 2.1 Окно “Поиск решения”

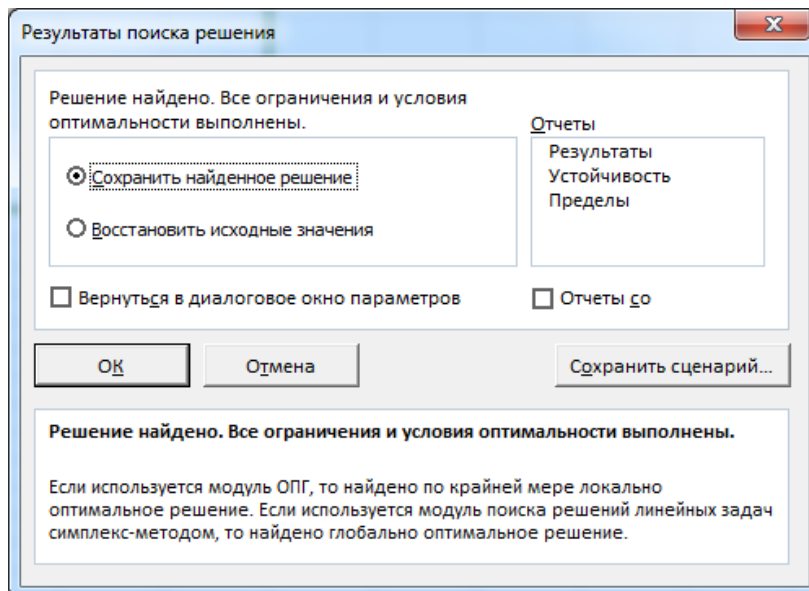


Рис. 2.2 Результаты поиска решения

Содержание работы

Основная часть работы состоит в расчете установившихся значений вероятностей состояний системы (финальных вероятностей), описанной марковским процессом.

Порядок выполнения

1. Изучить и решить примеры, описанные в теоретической части работы.
2. Выполнить задания для самостоятельного решения:

2.1 Найти корни уравнения третьего порядка. Рекомендации к решению: протабулируйте функцию (согласно номеру варианта) на достаточно большом интервале, постройте график, определите, сколько корней и где они примерно находятся, найдите корни с помощью средства

Подбор параметра.

Варианты заданий:

- | | |
|--------------------------|---------------------------|
| 1. $Y=x^3-4x^2-5x+6=0$ | 5. $Y=x^3-5x^2-4x+8=0$ |
| 2. $Y=1,5x^3-x^2-4x+4=0$ | 6. $Y=1,5x^3-3x^2-6x+5=0$ |
| 3. $Y=1,2x^3-2x^2-x+4=0$ | 7. $Y=0,5x^3-2x^2-4x+6=0$ |
| 4. $Y=x^3-3x^2-4x+4=0$ | 8. $Y=1,1x^3-5x^2-3x+7=0$ |

2.2 Решить систему уравнений, используя надстройку **Поиск решения**.

Варианты заданий:

- 1) $x_1 + 2x_2 - x_3 = -15$
 $-x_1 + 7x_2 - 9x_3 = 4$
 $x_1 + 2x_2 + 4x_3 = 18$
- 2) $2x_1 + x_2 - 3x_3 = 5$
 $-2x_1 + 4x_2 - 2x_3 = 2$
 $x_1 + x_2 + 4x_3 = 8$
- 3) $x_1 + x_2 - x_3 = 2$
 $x_1 + 3x_2 - 5x_3 = 5$
 $x_1 + 4x_2 - 3x_3 = 10$

2.3 Определить финальные вероятности состояний процесса, заданного (см. рис. 2.3) в виде графа. Интенсивности переходов считать равными ($1/4$):

$$\lambda_{12} = 0.01, \lambda_{23} = 0.1, \lambda_{34} = 0.5, \lambda_{45} = 0.05, \lambda_{51} = 2, \lambda_{53} = 1.$$

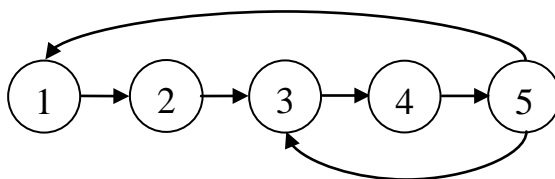


Рис. 2.3 Граф марковского процесса

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.
2. Результаты решений примеров, описанных в теоретической части.
3. Результаты решения контрольных заданий.
4. Выводы по результатам работы.

Контрольные вопросы

1. Объясните назначение и принцип работы средства MS Excel **Подбор параметра**.
2. Какой процесс можно считать марковским?
3. Как можно определить установившиеся значения вероятностей состояний системы?
4. Сформулируйте правило, по которому составляется система уравнений для определения вероятностей состояний системы.

3. Решение задач линейной оптимизации

Цель работы

Изучить средства программы Microsoft Excel для решения задач линейной оптимизации.

Теоретические основы

В общем виде задачу оптимизации формулируется следующим образом:

Пусть $X = (x_1, x_2, \dots, x_n)$ – вектор действительных переменных. Необходимо минимизировать или максимизировать целевую функцию $z = f(X)$ при выполнении нескольких ограничений $g_j(X) \leq b_j$, ($j = 1 \dots m$), которые задаются в виде неравенств или равенств. Могут быть также добавлены условия неотрицательности переменных ($x_i \geq 0$), которые включаются в указанные ограничения.

Если все функции $f(X)$ и $g_j(X)$ линейны относительно переменных x_i то имеем задачу линейной оптимизации, если хотя бы одна из функций нелинейная, то получаем задачу нелинейной оптимизации.

Таким образом, задача оптимизации включает три элемента:

- переменные x_1, x_2, \dots, x_n (в средстве **Поиск решения** ячейки, содержащие значения этих переменных, называются *изменяемыми* ячейками);
- целевая функция (ячейка, содержащая значение этой функции называется целевой ячейкой);
- ограничения (для применения средства **Поиск решения** ограничения могут быть записаны на рабочем листе и затем указаны в диалоговом окне либо заданы непосредственно в этом окне без записи на рабочем листе). При задании ограничений отдельно указываются функции ограничений $g_j(X)$ и вектор правых частей ограничений b_j .

После формулирования математической задачи оптимизации на рабочем листе Excel создается ее табличная модель, в которой в отдельных ячейках содержатся переменные решения, в отдельные ячейки записываются формулы, по которым будут вычисляться целевая функция и функции ограничений (левые части ограничений), также в отдельных ячейках указываются значения правых частей ограничений.

После создания табличной модели задачи оптимизации для нахождения оптимального решения применяют средство **Поиск решения**.

Пример. Минимизирование линейной функции

Дана функция: $z = 2x_1 + 3x_2 + 5x_3$.

Требуется минимизировать эту функцию при следующих ограничениях:

$$x_1 + x_2 - x_3 \geq -5,$$

$$-6x_1 + 7x_2 - 9x_3 \leq 4,$$

$$x_1 + x_2 + 4x_3 = 10,$$

и на переменные наложены условия неотрицательности.

Для решения данной задачи средство **Поиск решения** используется следующим образом:

1. Создать на рабочем листе Excel табличную модель решаемой задачи, например, как на рис. 3.1:

	A	B	C	D	E	F	G	H
1	Линейная задача оптимизации							
2		x1	x2	x3				
3	Переменные	0	0	0	Значение целевой функции			
4	Кoeffициенты целевой функции	2	3	5	0			
5	Ограничения	Кoeffициенты ограничений			Значение ф. ограничений		Правая часть огр	Разности
6	o1	1	1	-1	0		-5	-5
7	o2	-6	7	-9	0		4	4
8	o3	1	1	4	0		10	10
9								

Рис. 3.1 Табличная модель задачи линейной оптимизации

2. Открыть средство **Поиск решения**.

3. В открывшемся окне диалога **Поиск решения** указать данные, требуемые для процесса оптимизации (см. рис. 3.2):

3.1 В поле “*Оптимизировать целевую ячейку*” ввести адрес ячейки, содержащей значение целевой функции. Для этого примера в это поле следует ввести E4, или щелкнуть указателем мыши по этой ячейке и адрес введется автоматически.

3.2 В области “*До*” выбрать переключатель *Минимум*.

3.3 В поле “*Изменяя ячейки переменных*” указать ячейки, в которых содержатся переменные модели (в данном случае это диапазон B3:D3).

4. Задать ограничения – щелкнуть по кнопке “*Добавить*”, откроется окно диалога **Добавление ограничения**. Ввести ограничения поочередно, нажимая кнопку “*Добавить*” для каждого ограничения. После ввода всех ограничений нажать *ОК* для возврата в диалоговое окно **Поиск решения**.

Следует обратить особое внимание на флажок «*Сделать переменные без ограничений неотрицательными*». При необходимости он может быть установлен. В рассматриваемой задаче он должен быть снят.

5. Выбрать метод решения. Для решаемой задачи – *Поиск решения линейных задач симплекс-методом*.

6. После задания всех необходимых данных нажать кнопку “*Найти решение*”. Средство **Поиск решения** выполнит оптимизацию. В процессе вычислений в строке состояния отображаются число итераций и значения целевой функции при переборе множества допустимых решений задачи.

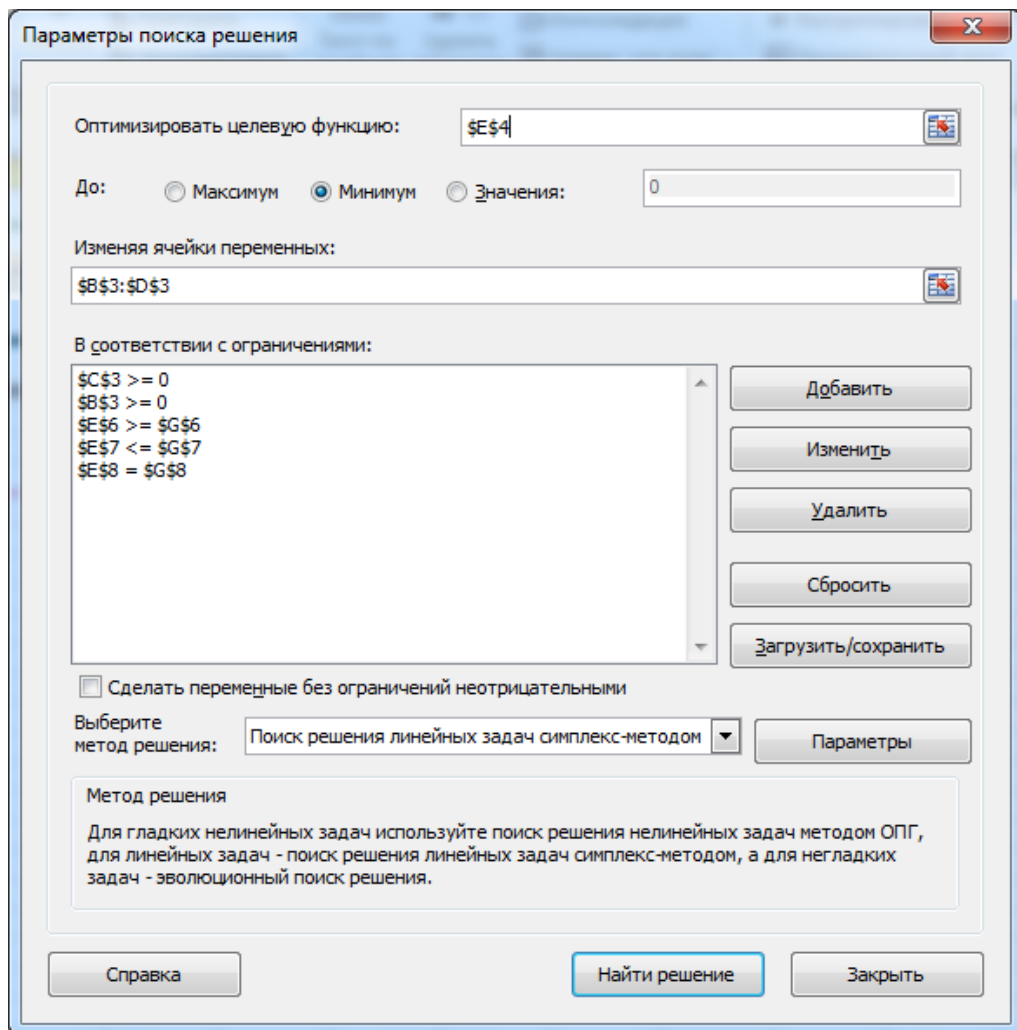


Рис.3.2 Окно Поиск решения для задачи оптимизации

7. Если в табличной модели нет ошибок, то будет выведено на экран диалоговое окно **Результаты поиска решения** (см. рис. 2.2), которое сообщает о завершении поиска. При нахождении оптимального решения в диалоговом окне должны присутствовать два ключевых предложения: *Решение найдено. Все ограничения и условия оптимальности выполнены.* Если хотя бы одного из предложений нет, программе не удалось оптимизировать модель. В этом случае следует сначала проверить правильность ввода данных, затем табличную модель и, наконец, пересмотреть исходную формулировку задачи.

8. В случае оптимального решения выбрать переключатель *Сохранить найденное решение* и нажать *ОК*. Кроме того есть возможность выбрать *Восстановить исходные значения*, что приведет к отказу от решения и восстановлению исходных значений в изменяемых ячейках, а также получить три типа отчетов (выбрав тип в списке *Тип отчета*) с результатами поиска решения.

Типы отчетов:

1. В отчете *Результаты* выводятся исходные и полученные в результате поиска решения значения изменяемых ячеек и целевой функции, а также сведения об ограничениях задачи.

2. Отчет *Устойчивость* дает основную информацию для анализа чувствительности моделей. Этот анализ показывает, насколько чувствительно найденное оптимальное решение к небольшим изменениям параметров модели.
3. Отчет *Пределы* показывает наименьшее и наибольшее значения, которые может принимать каждая переменная решения при удовлетворении ограничений и при постоянстве значений всех остальных переменных.

Содержание работы

Основная часть работы состоит в решении наиболее распространенных практических задач оптимизационных моделей.

Порядок выполнения

1. Изучить и решить пример, описанные в теоретической части работы.
2. Выполнить задания для самостоятельного решения:
 - 2.1 Решить с помощью надстройки **Поиск решения** задачи, описанные в книге *Оптимизация_Поиск_Решения.xlsx* на листах *Оптимизация* и *Транспортная*.
 - 2.2 Получить и проанализировать отчеты всех трех типов, описанные в теоретической части.

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.
2. Результаты решения примера, описанного в теоретической части.
3. Результаты решения контрольных заданий.
4. Выводы по результатам работы.

Контрольные вопросы

1. Объясните назначение и принцип работы средства MS Excel **Поиск решения**.
2. Какие элементы включает в себя задача оптимизации?
3. Что такое целевая функция?
4. В каком случае модель является оптимизационной?
5. Из каких соображений формулируются ограничения?

4. Изучение законов распределения случайных величин

Цель работы

Изучить законы распределения случайных величин с помощью программы Microsoft Excel.

Теоретические основы

При сборе и анализе данных аналитики часто имеют дело с разнообразными случайными величинами.

Под случайной величиной понимается переменная величина, принимающая те или иные значения с определенными вероятностями. Например, объем продаж, количество покупателей товара, число единиц выпущенной продукции за отчетный период, количество баллов студента, полученное в процессе обучения, число вызовов в единицу времени на телефонной станции.

Каждая случайная величина подчиняется определенному закону распределения, т.е. правилу, по которому задается соответствие между значением случайной величины и вероятностью ее появления. Закон распределения может быть задан таблично или в виде функции распределения.

Таким образом, функция распределения задает вероятность появления определенных значений случайной величины.

В данной работе рассматриваются наиболее применяемые на практике законы распределения случайных величин для моделирования случайных событий. Например, распределение времени отказа системы (экспоненциальный закон распределения) используют как модель надежности системы.

Биномиальное распределение

Случайная величина ξ имеет биномиальное распределение с параметрами n и p , где $0 \leq p \leq 1$, если ξ принимает значения $0, 1, 2, \dots, n$ с вероятностями $P(\xi = k) = C_n^k p^k (1-p)^{n-k}$. Случайная величина с таким распределением определяет число успехов в n испытаниях с вероятностью успешного исхода p . Такое распределение, например, используется при определении числа отказов при испытаниях определенного объема изделий, количества заказов, которые будут получены в результате следующих звонков в магазин, а также количества людей среди опрошенных, которые выразят желание купить данный товар и т.д.

Пример. Необходимо определить вероятность того, что трое из пяти человек, зашедших в магазин, купят товар марки «А», причем известно, что 85% покупателей предпочитают именно товар этой марки.

Решение. Для определения вероятности $P(\xi = k)$ следует использовать формулу $=\text{БИНОМРАСП}(k;n;p;0)$, для определения вероятности того, что результат окажется меньше или равным k ($P(\xi \leq k)$) формулу $=\text{БИНОМРАСП}(k;n;p;1)$.

Для вызова требуемой функции следует выполнить следующие действия:

- на вкладке **Формулы** выбрать команду **Вставить функцию**;
- в окне диалога **Мастер функций Шаг 1 из 2** в списке **Категория: Статистические** выбрать функцию **БИНОМРАСП** и нажать кнопку **ОК**.

В диалоговом окне **Аргументы функции** (см. рис. 4.1) задать следующие параметры:

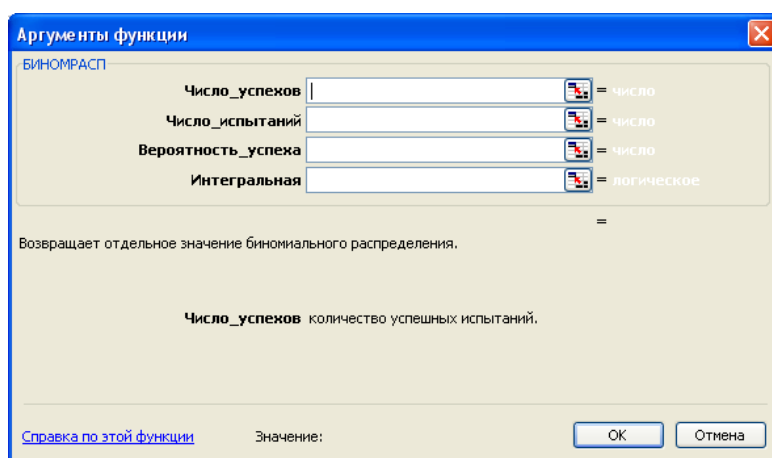


Рис. 4.1 Параметры биномиального распределения

1. *Число_успехов* – количество успешных событий.
2. *Число_испытаний* – количество независимых испытаний.
3. *Вероятность_успеха* – вероятность наступления успешного события.
4. *Интегральная* – логическое значение. Если 1, то функция рассчитывает интегральную (кумулятивную) функцию распределения, т.е. вероятность того, что число успешных событий не больше значения аргумента *Число_успехов*. Если 0, то рассчитывается дифференциальная функция распределения, т.е. вероятность того, что число успешных событий равно значению аргумента *Число_успехов*.

Для рассматриваемого примера значение функции $=\text{БИНОМРАСП}(3;5;0,85;1)$ равно 0,16479, т.е. вероятность того, что из пяти покупателей число купивших товар марки «А» будет не больше трех равно 0,16479.

Значение функции $=\text{БИНОМРАСП}(3;5;0,85;0)$ равно 0,138178, это значит, что трое из пяти человек, зашедших в магазин, купят товар именно марки «А».

Гипергеометрическое распределение

Гипергеометрическое распределение применяется при решении задач контроля качества продукции. Вероятность того, что из n изделий, выбранных случайным образом из партии объемом N , ровно k изделий с дефектом, имеет гипергеометрическое распределение.

Случайная величина ξ имеет гипергеометрическое распределение с параметрами n, N, K , где $K \leq N$, $n \leq N$, если принимает целые значения от максимума $(0, N-K-n)$ до минимума (n, N) с вероятностями

$$P(\xi = k) = \frac{C_K^k C_{N-K}^{n-k}}{C_N^n},$$

где k – число изделий с дефектом в выборке,

K – число изделий с дефектом в генеральной совокупности,

n – объем выборки,

N – объем генеральной совокупности.

Пример. Из партии, содержащей 40 изделий, случайным образом выбираются 5 и подвергаются проверке на качество. Если в результате контроля только одно изделие оказывается бракованным, то партия принимается. В противном случае вся партия бракуется. Требуется определить вероятность того, что партия будет принята, если из 40 изделий 6 имеют дефекты.

Вероятность получить k из выборки объемом n определяется с помощью функции MS Excel =ГИПЕРГЕОМЕТ() со следующими аргументами:

1. Число_успехов_в_выборке (k) – количество успешных событий в выборке.
2. Размер_выборки (n).
3. Число_успехов_в_совокупности (K).
4. Размер_совокупности (N).

Для примера искомая вероятность будет равна сумме двух функций:
=ГИПЕРГЕОМЕТ(0,5,6,40)+ГИПЕРГЕОМЕТ(1,5,6,40).

В итоге вероятность того, что партия будет принята, равна 0,846.

Распределение Пуассона

Данное распределение описывает число событий, происходящих в одинаковых промежутках времени при условии, что события происходят независимо друг от друга с постоянной интенсивностью λ , например, число отказавших изделий за определенный интервал, число требований выплаты страховых сумм в год, число вызовов в ремонтную службу за сутки.

Случайная величина ξ имеет распределение Пуассона с параметром λ , если она принимает значения $0, 1, 2, \dots$ с вероятностями $P(\xi = k) = \frac{\lambda^k}{k!} e^{-\lambda}$.

Пример. Фирма выпускает оборудование высокого качества, благодаря чему каждый день ожидается возврат в среднем только 1,5 единиц товара на гарантийный ремонт. Требуется рассчитать, с какой вероятностью завтра на

гарантийный ремонт не поступит ни одного изделия или, например, поступит не более одного.

Для того чтобы получить вероятность $P(\xi = k)$ следует использовать функцию $=\text{ПУАССОН}(k, \lambda, 0)$, а для вычисления вероятности $P(\xi \leq k)$ - $=\text{ПУАССОН}(k, \lambda, 1)$ со следующими параметрами:

1. *Количество событий* - X .
2. *Среднее* – интенсивность появления событий.
3. *Интегральная* – логическое значение. Если 1, то функция рассчитывает интегральную (кумулятивную) функцию распределения, т.е. вероятность того, что число успешных событий не больше значения аргумента X . Если 0, то рассчитывается дифференциальная функция распределения, т.е. вероятность того, что число успешных событий равно значению X .

Таким образом, вероятность, что завтра на гарантийный ремонт не поступит ни одного изделия $=\text{ПУАССОН}(0, 1,5,0)$ равна 0,22, а вероятность, что не более одного - $=\text{ПУАССОН}(1, 1,5,1) = 0,56$.

Нормальное распределение

Нормальное распределение часто применяется в тех случаях, когда возможные значения лежат вокруг некоего среднего значения, например, при стрельбе по мишени, измерении какого-либо параметра и т.д.

Пример. Компания по производству верхней одежды в результате измерения роста группы людей получила среднее значение, равное 172 см. и стандартное отклонение – 5 см. Требуется определить, какой процент общего числа производимых пальто должны составлять пальто 3-го роста (158 - 164 см.).

Для решения задачи следует использовать функцию $=\text{НОРМРАСП}()$ со следующими параметрами:

1. X – значение, для которого рассчитывается вероятность.
2. *Среднее* – среднее значение.
3. *Стандартное откл* – стандартное отклонение.
4. *Интегральная* – логическое значение. Если 1, то функция рассчитывает интегральную (кумулятивную) функцию распределения, т.е. вероятность того, результат не больше значения аргумента X . Если 0, то рассчитывается дифференциальная функция распределения, т.е. результат равен значению X .

Итоговое значение будет определяться как разность: $=\text{НОРМРАСП}(164;172;5;1) - \text{НОРМРАСП}(158;172;5;1) = 0,0548 - 0,0025 = 0,052$, т.е. пальто 3-го роста должны составлять 5,2% от общего числа.

χ^2 -распределение (Пирсона)

Предположим, что $\xi_1, \xi_2, \dots, \xi_n$ – независимые случайные величины, каждая из которых имеет нормальное распределение. Закон распределения суммы $\xi_1^2 + \xi_2^2 + \dots + \xi_n^2$ называется χ^2 -распределением с n – степенями свободы.

Вероятность $p(\chi^2)$ можно определить с помощью функции =ХИ2РАСП() при следующих аргументах:

1. X – значение, для которого рассчитывается χ^2 -распределение.
2. *Степени_свободы* – число степеней свободы.

Данное распределение используется при проверке статистических гипотез с помощью критерия согласия Пирсона.

Для определения квантиля χ^2 -распределения следует использовать функцию

=ХИ2ОБР(*уровень значимости; число степеней свободы*), возвращающее обратное значение вероятности.

F -распределение (распределение Фишера)

Распределение Фишера используется, например, при проверке адекватности уравнений регрессии. Для расчета значений вероятности используется функция =ФРАСП() со следующими параметрами:

1. X – значение, для которого рассчитывается функция.
2. *Степени_свободы1* – число степеней свободы первого набора данных.
3. *Степени_свободы2* – число степеней свободы второго набора данных.

Для определения квантиля F -распределения следует использовать функцию

=ФРАСПОБР(*уровень значимости; число степеней свободы1; число степеней свободы1*), возвращающее обратное значение вероятности.

Распределение Вейбулла

Распределение Вейбулла часто применяют в теории надежности при описании этапа старения аппаратуры. Для вычисления вероятности применяется функция =ВЕЙБУЛЛ() с параметрами:

1. X – значение, для которого вычисляется функция.
2. *Альфа* – параметр распределения.
3. *Бета* – параметр распределения.
4. *Интегральная* – логическое значение. Если 1, то функция рассчитывает интегральную (кумулятивную) функцию распределения, т.е. вероятность того, результат не больше значения аргумента X . Если 0, то рассчитывается дифференциальная функция распределения, т.е. результат равен значению X .

Экспоненциальное распределение

Экспоненциальное распределение применяют в теории надежности при описании режима нормальной эксплуатации техники. Для вычисления вероятности применяется функция =ЭКСПРАСП() с параметрами:

1. X – значение, для которого вычисляется функция.
2. *Лямбда* – параметр распределения.
3. *Интегральная* – логическое значение. Если 1, то функция рассчитывает интегральную (кумулятивную) функцию распределения, т.е. вероятность того, результат не больше значения аргумента X . Если 0, то рассчитывается дифференциальная функция распределения, т.е. результат равен значению X .

Содержание работы

Основная часть работы состоит в решении наиболее распространенных практических задач, использующих законы распределения случайных величин.

Порядок выполнения

1. Изучить и решить примеры, описанные в теоретической части работы.
2. Выполнить задания для самостоятельного решения:
 - 2.1 Количество клиентов, которыми располагает фирма, равно 8, а вероятность, с которой может позвонить каждый из них, 0,25. Чему равна вероятность, что завтра позвонят не менее 3-х клиентов?
 - 2.2 Чему равна вероятность того, что завтра на ремонт поступит 4 детали, если в среднем возвращается 1,8 единиц.
 - 2.3 Результаты опроса общественного мнения говорят о том, что наибольшей популярностью покупателей пользуются мужские пальто серого цвета (30% опрошенных), затем пальто черного цвета (26% опрошенных), синего (22%) и коричневого (22%). Имеется ли существенная разница между цветами пальто с точки зрения покупателей, если считать, что результаты получены при $n = 100$. Для проверки гипотезы о принадлежности данных опроса к равномерному распределению используйте критерий χ^2 -квадрат при уровне значимости $\alpha = 0.05$.
 - 2.4 Решите задачу из предыдущего пункта при $n = 1000$.

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.
2. Результаты решения примеров, описанных в теоретической части.
3. Результаты решения контрольных заданий.
4. Выводы по результатам работы.

Контрольные вопросы

1. Сформулируйте определение понятия “закон распределения случайной величины”.
2. Приведите примеры применения экспоненциального закона распределения для описания случайных событий.
3. Покажите порядок ввода формул в Excel.
4. Перечислите параметры нормального распределения.
5. Какие случайные события можно описать с помощью нормального закона распределения?
6. Что такое дисперсия?
7. Раскройте понятие квантиля.
8. Каким образом определяется степень свободы?
9. Перечислите действия при применении критерия согласия.
10. В чем причина отличия результатов решения задач из пунктов 2.3 и 2.4?

5. Генерация случайных чисел и анализ выборки данных

Цель работы

Изучить средства программы Microsoft Excel для генерации случайных чисел с требуемыми законами распределения, для построения и анализа выборок данных.

Теоретические основы

Имитационное моделирование представляет собой численный метод проведения на ЭВМ вычислительных экспериментов с математическими моделями, имитирующими поведение реальных объектов, процессов и систем во времени в течении заданного периода. При этом их функционирование разбивается на элементарные явления, подсистемы и модули. Функционирование этих элементарных явлений, подсистем и модулей описывается набором алгоритмов, которые имитируют элементарные явления с сохранением их логической структуры и последовательности протекания во времени.

Одним из видов имитационного моделирования является статистическое имитационное моделирование, позволяющее воспроизводить на ЭВМ функционирование сложных случайных процессов.

В вероятностном имитационном моделировании оперируют не с характеристиками случайных процессов, а с конкретными случайными числовыми значениями параметров системы. При этом результаты, полученные при воспроизведении на имитационной модели рассматриваемого процесса, являются случайными реализациями. Поэтому для нахождения объективных и устойчивых характеристик процесса требуется его многократное воспроизведение, с последующей статистической обработкой полученных данных. Именно поэтому исследование сложных процессов и систем, подверженных случайным возмущениям, с помощью имитационного моделирования принято называть статистическим моделированием.

Статистическая модель случайного процесса – это алгоритм, с помощью которого имитируют работу сложной системы, подверженной случайным возмущениям; имитируют взаимодействие элементов системы, носящих вероятностный характер.

Применение статистического моделирования широко распространено в задачах анализа и проектирования информационно-вычислительных систем и других сложных организационно-технических объектов. При реализации статистического имитационного моделирования возникает задача получения случайных числовых последовательностей с заданными вероятностными характеристиками.

Таким образом, статистическое моделирование – это метод решения вероятностных и детерминированных задач, основанный на эффективном

использовании случайных чисел и законов теории вероятностей. Этот метод использует способность современных компьютеров порождать и обрабатывать за короткие промежутки времени значительное количество случайных чисел.

Подавая последовательность случайных чисел на вход исследуемой модели, на её выходе получают преобразованную последовательность случайных величин – выборку. При правильной организации подобного статистического эксперимента выборка содержит ценную информацию об исследуемой модели, которую трудно или практически невозможно получить другими способами. Информация извлекается из выборки методами математической статистики.

Метод статистического моделирования (метод Монте-Карло) позволяет, опираясь на законы теории вероятностей, свести широкий класс сложных задач к относительно простым арифметико-логическим преобразованиям выборок. Поэтому такой метод и получил весьма широкое распространение. В частности, он почти всегда используется при имитационном моделировании реальных сложных систем.

При реализации на ЭВМ статистического моделирования возникает задача получения (генерирования) на ЭВМ случайных числовых последовательностей с заданными вероятностными характеристиками, в которых каждое число – это имитация случайного значения какого-либо параметра реального процесса или системы, подверженного случайным возмущениям.

Таким образом, при имитационном моделировании стохастических систем приходится имитировать случайное событие с заданными вероятностями, случайные величины с заданными распределениями, случайные процессы с заданными характеристиками (математическое ожидание, дисперсия, корреляционная функция). Для имитации случайной величины с заданным распределением нужно иметь генератор (датчик) случайных чисел, генерирующий числа с заданным законом распределения.

Генерация случайных чисел

Для того чтобы сформировать массив случайных чисел, распределенных по требуемому закону распределения можно воспользоваться функцией **Генерация случайных чисел**.

Данная функция входит в состав надстройки **Анализ данных**. Для вызова этой надстройки необходимо на вкладке **Данные**, в группе **Анализ** выбрать команду **Анализ данных**.

В окне диалога **Анализ данных** необходимо указать функцию **Генерация случайных чисел** и нажать **ОК**. Появится соответствующее диалоговое окно (см. рис. 5.1).

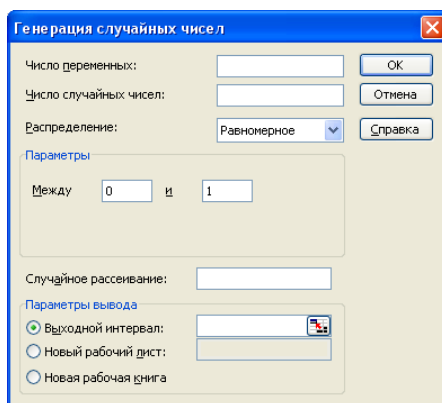


Рис. 5.1. Генерация случайных чисел

В диалоговом окне (см. рис. 5.1) необходимо задать следующие параметры:

1. *Число переменных* – вводится число столбцов значений, которые необходимо разместить в выходном диапазоне.
2. *Число случайных чисел* – вводится число значений, которое необходимо вывести в каждом столбце выходного диапазона.
3. *Распределение* – в выпадающем списке выбирается тип распределения, необходимый для генерации случайных чисел.
4. *Параметры* – вводятся параметры выбранного распределения.
5. *Случайное рассеивание* – вводится произвольное значение, для которого необходимо генерировать случайные числа. Впоследствии можно снова использовать это значение для получения тех же самых случайных чисел.
6. *Параметры вывода* – указывается место, где будет сгенерирован массив.

Построение выборки данных

На практике при проведении исследований часто ограничиваются только выборочным набором объектов из генеральной совокупности, т.е. **выборкой**. Работа с выборкой существенно экономит время и средства.

Случайная выборка строится таким образом, чтобы каждый объект генеральной совокупности имел одинаковую вероятность быть выбранным, и при этом объекты отбираются независимо друг от друга.

Существует два основных типа случайной выборки: выборка без возврата (любой объект не может попасть в выборку более одного раза) и выборка с возвратом (объект может попасть в выборку более одного раза, т.е. после выбора объект снова возвращается в генеральную совокупность).

Для получения систематической выборки в генеральной совокупности определяют случайную начальную точку и отбирают элементы, начиная с этой точки через постоянный интервал (с постоянным шагом отбора). Для того чтобы из совокупности N получить систематическую выборку размером n , необходимо выбирать элементы с шагом N/n . Данный метод используется в том случае, когда есть предварительно пронумерованные списки,

платежные поручения, приходные и расходные чеки и т.д., в подобных случаях он быстрее и проще, чем метод простой случайной выборки.

В Excel реализованы случайная выборка с возвратом и систематическая выборка с помощью функции **Выборка**. Вызвать эту функцию можно в окне диалога **Анализ данных**. В этом случае появится диалоговое окно **Выборка** (см. рис. 5.2).

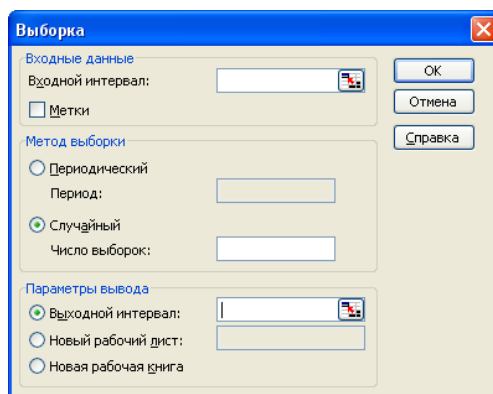


Рис. 5.2. Моделирование выборки

В диалоговом окне необходимо задать следующие параметры:

1. *Входной интервал* – вводится диапазон ячеек, содержащих анализируемые данные.

2. *Переключатель* в группе **Метод выборки** для случайной выборки с возвратом должен быть установлен в положение **Случайный**. В поле **Число выборок** вводится количество размещаемых в выходном столбце случайных значений (размер выборки). Для формирования систематической выборки следует выбрать переключатель **Периодический** и в поле **Период** указать шаг отбора, элементы будут отбираться через постоянный интервал.

3. *Параметры вывода* – указывается место, где будет сгенерирована выборка.

Анализ данных

Для того чтобы обнаружить общие свойства совокупности данных, выявить закономерности, тенденции развития процесса и в результате прийти к правильным выводам, необходимы обобщающие количественные показатели. Эти показатели можно условно разделить на четыре группы:

1. Показатели, которые описывают закон распределения данных: таблицы частот, полигоны, гистограммы.

2. Показатели уровня – описывают положение данных на числовой оси: минимальный и максимальный элементы выборки, верхний и нижний квартили, перцентиль, различные средние и другие характеристики.

3. Показатели рассеивания – описывают степень разброса данных относительно своего центра: дисперсия, среднеквадратичное отклонение, размах выборки.

4. Показатели асимметрии – характеризуют симметрию распределения данных около своего центра: коэффициент асимметрии, эксцесс.

Гистограмма

Для графического отображения данных, представляющих вариационный ряд можно построить **гистограмму** – столбчатую диаграмму частот.

По оси абсцисс откладываются значения интервалов, а по оси ординат – частоты в виде столбиков, высота которых соответствует частоте попадания случайной величины в интервал.

В Excel для построения гистограммы применяется функция **Гистограмма**. Вызвать эту функции можно в окне диалога **Анализ данных**. Появится диалоговое окно **Гистограмма** (рис. 5.3).

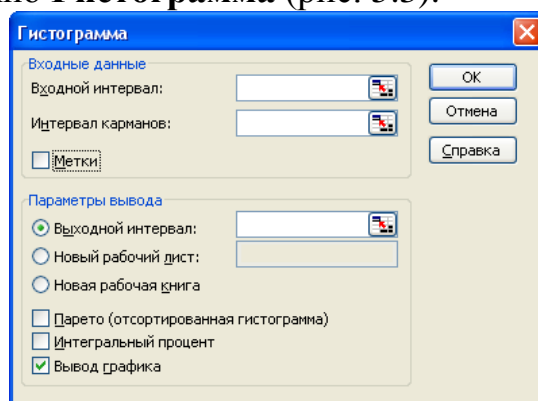


Рис. 5.3. Построение гистограммы

В диалоговом окне **Гистограмма** задаются следующие параметры:

1. *Входной интервал* – вводится диапазон ячеек, содержащих анализируемые данные.

2. В поле *Интервал карманов* вводится диапазон ячеек, содержащих значения границ интервалов (параметр является необязательным, в этом случае набор интервалов создается автоматически).

3. *Параметры вывода* – указывается место, где будет указана таблица частот.

4. Для *вывода гистограммы* следует установить флажок опции **Вывод графика**. Флажки опций **Парето (отсортированная гистограмма)** и **Интегральный процент (накопленные) частоты** следует оставить сброшенными.

Показатели уровня

Средняя арифметическая: для расчета используется функция $=СРЗНАЧ(\text{число1}; \text{число2}; \dots)$ из категории **Статистические**.

Также применяется функция, которая возвращает среднее значение (среднее арифметическое) всех ячеек в диапазоне, которые соответствуют данному условию:

$=СРЗНАЧЕСЛИ(\text{диапазон}; \text{условие}; \text{диапазон усреднения})$

При необходимости отбросить аномальные значения, так называемые «статистические выбросы», применяется функция:

$=УРЕЗСРЕДНЕЕ(\text{массив}; \text{доля}),$

где

- *массив* – массив или интервал усредняемых значений.
- *доля* – доля точек (значений) данных, исключаемых из вычислений.

Например, если доля равна 0,2, то из множества данных, содержащих 20 точек (20 x 0,2), исключаются четыре точки: две точки с наибольшими значениями и две точки с наименьшими значениями на множестве данных.

Данная функция полезна тогда, когда в диапазон, для которого рассчитывают среднее значение, попадают данные, существенно отличающиеся от остальных, и расчет простого среднего арифметического может привести к неправильным выводам.

Средняя геометрическая: $\bar{x} = \sqrt[n]{\prod_{i=1}^n x_i}$, для расчета используется функция =СРГЕОМ(*число1*; *число2*; ...) из категории Статистические.

Медиана: величина, которая соответствует варианту, находящемуся в середине ранжированного ряда.

Пример. Для ряда 1,2,3,3,6,7,8,8,10 медианой будет величина, которая расположена в центре ряда, т.е. пятая величина, действительно значение функции =МЕДИАНА(1;2;3;3;6;7;8;8;10) равно 6.

Мода: значение признака, которое встречается наиболее часто среди элементов совокупности, для расчета используется функция =МОДА(*число1*; *число2*; ...) из категории Статистические.

Минимум: для расчета используется функция =МИН(*число1*; *число2*; ...) из категории Статистические.

Максимум: для расчета используется функция =МАКС(*число1*; *число2*; ...) из категории Статистические.

Наибольший: для расчета используется функция =НАИБОЛЬШИЙ(*массив*; *k*) из категории Статистические. Функция возвращает k-е по величине значение из множества данных и позволяет выбрать значение по его относительному местоположению.

Наименьший: для расчета используется функция =НАИМЕНЬШИЙ(*массив*; *k*) из категории Статистические. Функция возвращает k-е наименьшее значение на множестве данных и используется для определения значения, занимающего заданное относительное положение на множестве данных.

Ранг: определяет номер (порядковое место) значения случайной величины в наборе данных:

=РАНГ(*число*, *ссылка*, [*порядок*]).

Функция РАНГ присваивает повторяющимся числам одинаковые значения ранга. Следует учитывать, что данная функция оставлена для обеспечения совместимости с предыдущими версиями Excel. Однако если обратная совместимость не требуется, целесообразнее перейти к использованию новых функций, так как их названия более точно соответствуют выполняемым действиям:

=РАНГ.СР(*число*, *ссылка*, [*порядок*]) – если несколько значений имеют одинаковый ранг, возвращается среднее.

=РАНГ.РВ(*число*, *ссылка*, [*порядок*]) – если несколько значений имеют одинаковый ранг, возвращается наивысший ранг этого набора значений.

Перцентиль: обобщает информацию о рангах, характеризуя значение, достигаемое заданным процентом общего количества данных, т.е. являются характеристиками набора данных, которые выражают ранги элементов в виде процентов.

Для расчета рангов и перцентилей используется функция **Ранг и перцентиль**.

Для вызова этой функции необходимо в окне диалога **Анализ данных** выбрать соответствующую функцию. Появится диалоговое окно **Ранг и перцентиль** (рис. 5.4).

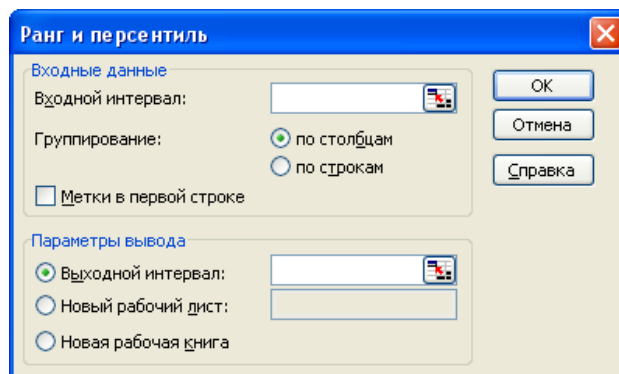


Рис.5.4. Ранг и перцентиль

В диалоговом окне (рис. 5.4) задаются следующие параметры:

1. *Входной интервал* – вводится диапазон ячеек, содержащих анализируемые данные.
2. *Группирование* – по строкам и столбцам в зависимости от расположения данных во входном диапазоне.
3. *Метки в первой строке* – флажок ставится, если первая строка содержит заголовок, в противном случае будут созданы стандартные заголовки автоматически.
4. *Параметры вывода* – указывается место, где будет указана таблица рангов и перцентилей.

Показатели рассеяния

Размах выборки – разности между максимальным и минимальным значением.

Среднее линейное отклонение – среднее арифметическое из абсолютных значений отклонений от средней: $(\sum |x_i - \bar{x}|) / n$. Для расчета среднего линейного отклонения используется функция =СРОТКЛ(*число1*; *число2*; ...) из категории Статистические.

Среднее квадратическое отклонение (ско). Для расчета используется функция =СТАНДОТКЛОН(*число1*; *число2*; ...) из категории Статистические.

Дисперсия. Для расчета используется функция =ДИСП(*число1*; *число2*; ...) из категории Статистические.

Показатели асимметрии

Экссесс – характеризует островершинность (плосковершинность) распределения. Если эксцесс больше нуля, то распределение островершинное, если меньше нуля – плосковершинное. Для расчета

используется функция =ЭКЦЕСС(*число1;число2;...*) из категории Статистические.

Асимметрия – характеризует меру несимметричности (скошенности) распределения. Если коэффициент асимметрии больше нуля, то асимметрия правосторонняя, если меньше нуля – левосторонняя. Для расчета используется функция =СКОС(*число1;число2;...*) из категории Статистические.

Описательная статистика

Для получения статистического отчета одновременно по всем основным показателям используют специальный инструмент анализа: **Описательная статистика**, который находится в надстройке **Анализ данных**.

При вызове функции **Описательная статистика** появится соответствующее диалоговое окно (рис. 5.5).

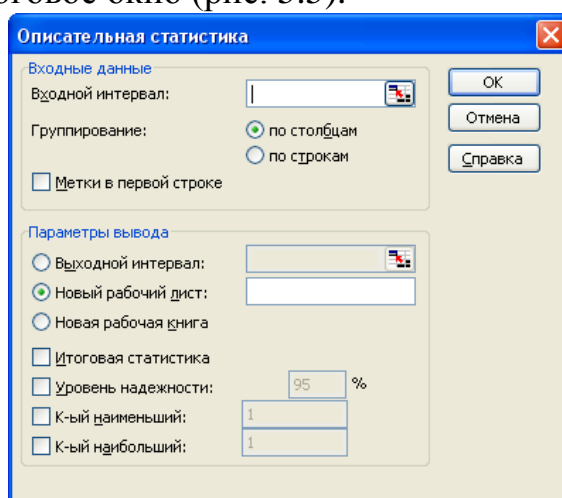


Рис. 5.5. Описательная статистика

В диалоговом окне (рис. 5.5) задаются следующие параметры:

1. *Входной интервал* – вводится диапазон ячеек, содержащих анализируемые данные.

2. *Группирование* – по строкам и столбцам в зависимости от расположения данных во входном диапазоне.

3. *Метки в первой строке* – флажок ставится, если первая строка содержит заголовок, в противном случае будут созданы стандартные заголовки автоматически.

4. *Параметры вывода* – указывается место, где будет указана таблица результатов анализа.

5. *Итоговая таблица* – флажок ставится, если необходимо получить результаты по каждому показателю.

6. *Уровень надежности* – флажок устанавливается, если требуется вывести значение ошибки выборки при установленном уровне надежности.

7. *K-ый наименьший и K-ый наибольший* – флажки устанавливаются, если требуется получить определенный наименьший элемент (начиная с минимального значения среди элементов выборки) или наибольший (начиная с максимального значения выборки).

Содержание работы

Основная часть работы состоит в решении практических задач, использующих возможности программы Excel для генерации случайных чисел с требуемыми законами распределения и анализа выборки данных.

Порядок выполнения

1. Сформировать в столбцах массивы случайных чисел, распределенных по требуемому закону распределения, число переменных = 1, число случайных чисел = 100:
 - 1.1. Распределение Бернулли. Исходные данные: $p = 0,3$.
 - 1.2. Биномиальное распределение. Исходные данные: $p = 0,85$; число испытаний – 25.
 - 1.3. Нормальное распределение. Исходные данные: среднее – 100; стандартное отклонение – 20.
2. На основе построенных в п.1.1-1.3 массивов получить случайную и систематическую выборки размером 20.
3. Построить гистограммы для нормального распределения (п.1.3) и случайной и систематической выборок, полученных в п.2 для нормального распределения. Сравнить вид гистограмм.
4. Для случайных выборок, полученных в п.2 рассчитать показатели анализа данных, описанные в теоретической части с помощью функций Excel.
5. Для случайных выборок, полученных в п.2 рассчитать показатели анализа данных, описанные в теоретической части с помощью инструмента «Описательная статистика». Сделать выводы по результатам анализа данных.

Контрольные вопросы

1. Сформулируйте определение понятия “закон распределения случайной величины”.
2. Приведите примеры применения нормального закона распределения для описания случайных событий.
3. Что показывают показатели асимметрии и эксцесса?
4. Перечислите этапы процесса построения гистограммы.
5. В чем отличие систематической выборки от случайной?

6. Анализ временных рядов

Цель работы

Изучить средства программы Microsoft Excel для анализа временных рядов.

Теоретические основы

Временной ряд представляет собой последовательность данных, описывающих объект в последовательные моменты времени.

Примеры временных рядов:

в экономике – курсы акций, валют, объемы продаж, объем производства;
в метеорологии – температура, интенсивность осадков, сила ветра, давление, влажность и т.д.;

в гидрологии – уровни воды в водоемах;

в технике – параметры сигналов.

При изучении временных рядов, как правило, ставятся следующие цели:

- выявление особенностей временного ряда;

- подбор модели, позволяющей решать задачи прогноза и управления.

При построении модели ряд разделяют на детерминированную (неслучайную) и случайную части.

Детерминированная (закономерная) часть – это составляющая временного ряда, элементы которой вычисляются по определенному правилу как функция от времени,

Случайный компонент временного ряда придает хаотичность и непредсказуемость. Для описания и анализа данного компонента используют понятия и методы теории вероятностей и математической статистики.

В большинстве приложений в детерминированной части чаще всего выделяют три составляющие: тренд, сезонная и циклическая составляющие.

Трендом (trend – тенденция, направление) временного ряда называют изменяющийся, нециклический компонент, описывающий влияние долговременных факторов, эффект которых сказывается постепенно. К таким факторам относятся изменение демографических характеристик, рост потребления, технологическое и экономическое развитие.

Сезонный компонент временного ряда описывает поведение, изменяющееся регулярно в течение заданного периода (года, месяца, недели, дня и т.п.). Состоит из почти повторяющихся циклов.

Примеры: объем продаж перед праздниками, пики грузоперевозок в течение суток, пики перевоза пассажиров в городском транспорте, изменение нагрузки на электрическую и телефонную сети в течение суток и года и т.п.

Циклический компонент описывает длительные периоды относительного подъема и спада и состоит из циклов, меняющихся по амплитуде и протяженности.

Пример: взаимодействие спроса и предложения, рост и истощение ресурсов, изменение налоговой политики, кадровая политика.

Метод скользящего среднего

Данный метод формирует значения, которые в каждой точке определения равны среднему значению исходной функции за предыдущий период.

Скользящие средние обычно используются с данными временных рядов для сглаживания краткосрочных колебаний и выделения основных тенденций или циклов. Наиболее часто используются простое скользящее среднее (среднее арифметическое), экспоненциальное скользящее среднее и взвешенное скользящее среднее.

Для сглаживания на основе простого скользящего среднего используется функция **Скользящее среднее**. Её следует вызвать в окне диалога **Анализ данных**, указав имя соответствующей функции. Появится диалоговое окно (рис. 6.1).

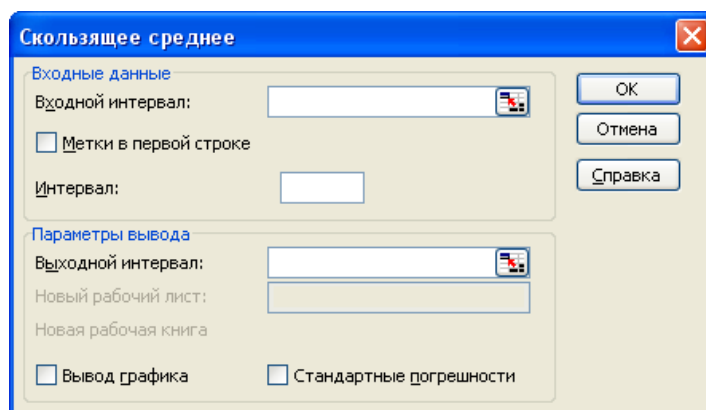


Рис. 6.1. Скользящее среднее

В диалоговом окне (рис. 6.1) задаются следующие параметры:

1. *Входной интервал* – вводится диапазон ячеек, содержащих анализируемые данные.
2. *Метки в первой строке* – флажок ставится, если первая строка содержит заголовок, в противном случае будут созданы стандартные заголовки автоматически.
3. *Интервал* – в поле вводится число уровней, входящих в интервал сглаживания. По умолчанию значение равно 3.
4. *Параметры вывода* – указывается место, где будет указана таблица результатов анализа.
5. Для *вывода графика* и стандартных погрешностей необходимо установить соответствующие опции.

Метод экспоненциального сглаживания

На практике часто возникает ситуация, когда последние данные временного ряда являются более значимыми, чем предыдущие и, следовательно, при вычислении скользящего среднего наблюдениям

необходимо приписывать различные веса. При экспоненциальном сглаживании более старым наблюдениям приписываются экспоненциально убывающие веса, при этом, в отличие от скользящего среднего, учитываются все предшествующие наблюдения ряда.

Формула экспоненциального сглаживания имеет вид:

$S_t = \alpha \cdot y_t + (1 - \alpha) \cdot S_{t-1}$ и применяется рекурсивно, т.е. каждое новое сглаженное значение S_t вычисляется как взвешенное среднее текущего наблюдения y_t и сглаженного ряда S_{t-1} (предыдущие наблюдения). Очевидно, что если $\alpha = 1$, то предыдущие наблюдения полностью игнорируются, а если $\alpha = 0$, то игнорируются текущие наблюдения.

Для экспоненциального сглаживания используется функция **Экспоненциальное сглаживание**. Для ее вызова необходимо выбрать соответствующее имя в окне диалога **Анализ данных**. Появится диалоговое окно (рис. 6.2).

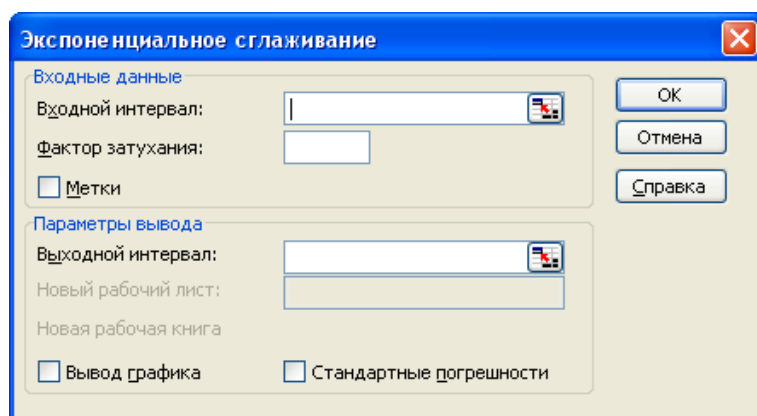


Рис.6.2. Экспоненциальное сглаживание

В диалоговом окне (рис. 6.2) задаются следующие параметры:

1. *Входной интервал* – вводится диапазон ячеек, содержащих анализируемые данные.

2. *Фактор затухания* – в поле вводится значение коэффициента экспоненциального сглаживания. По умолчанию значение принимается равным 0,3.

3. *Метки в первой строке* – флажок ставится, если первая строка содержит заголовки, в противном случае будут созданы стандартные заголовки автоматически.

4. *Параметры вывода* – указывается место, где будет указана таблица результатов анализа.

5. Для *вывода графика* и стандартных погрешностей необходимо установить соответствующие опции.

Метод аналитического выравнивания

Основным содержанием метода аналитического выравнивания временных рядов является расчет общей тенденции развития (тренда) как

функции времени на основе математической модели, которая наилучшим образом отображает основную тенденцию развития временного ряда.

В качестве моделей обычно используют линейную, показательную, степенную и логарифмическую функции в зависимости от особенностей временного ряда.

Для оценки точности трендовой модели используется коэффициент детерминации, равный отношению дисперсии теоретических данных, полученных по трендовой модели к дисперсии эмпирических данных. Считается, что трендовая модель адекватна изучаемому процессу и отражает тенденцию его развития при значениях коэффициента детерминации близких к единице.

Порядок построения трендовой модели с помощью Excel 2010 следующий:

1. Эмпирический временной ряд следует представить в виде диаграммы одного из следующих типов: гистограмма, линейная диаграмма, график, точечная диаграмма, диаграмма с областями, а затем щелкнуть правой кнопкой мыши на одном из маркеров данных и в открывшемся контекстном меню выбрать команду **Добавить линию тренда**. Откроется окно диалога (рис.6.3).

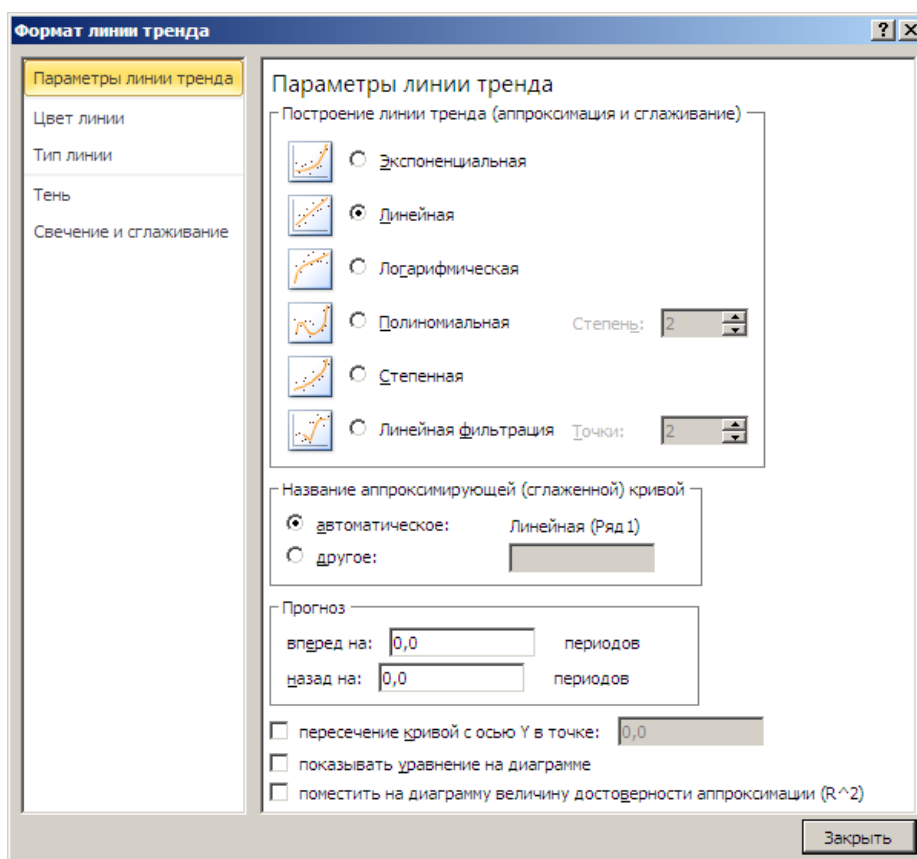


Рис. 6.3. Окно выбора модели тренда

2. Выбрать требуемый вид трендовой модели.
3. Задать дополнительные параметры тренда:

- *Название сглаженной кривой* – автоматическое по типу модели или собственное название функции.
- *Прогноз* – можно указать, на сколько периодов вперед (назад) требуется спроектировать линию тренда в будущее (прошлое).
- Если необходимо указать точку пересечения кривой с осью ординат, показать уравнение аппроксимирующей функции и вывести значение коэффициента детерминации следует установить соответствующие флажки.

Использование статистических функций

Спрогнозировать данные без использования диаграмм, соответствующие линейным и экспоненциальным линиям, можно с использованием статистических функций ТЕНДЕНЦИЯ() и РОСТ().

Функция ТЕНДЕНЦИЯ возвращает значения в соответствии с линейным трендом. Аппроксимирует прямой линией (по методу наименьших квадратов) массивы *известные_значения_y* и *известные_значения_x*. Возвращает значения *y*, в соответствии с этой прямой для заданного массива *новые_значения_x*.

Синтаксис функции: ТЕНДЕНЦИЯ(*y*; *x*; *n_x*; конст):

y - *известные_значения_y* – множество значений *y*, для которых уже известна линейная зависимость;

x - *известные_значения_x* - множество значений *x*, для которых уже известна линейная зависимость;

n_x - *новые_значения_x* – новые значения *x*, для которых функция возвращает соответствующие значения *y*.

конст – логическое значение, если оно равно 0, то свободный член равен нулю, в противном случае свободный член вычисляется обычным образом.

Функция РОСТ рассчитывает прогнозируемый экспоненциальный рост на основании имеющихся данных. Синтаксис функции: РОСТ(*y*; *x*; *n_x*; конст):

y - *известные_значения_y* – множество значений *y*, для которых уже известна экспоненциальная зависимость;

x - *известные_значения_x* - множество значений *x*, для которых уже известна экспоненциальная зависимость;

n_x - *новые_значения_x* – новые значения *x*, для которых функция возвращает соответствующие значения *y*.

конст – логическое значение, если оно равно 0 или отсутствует, то константа равна единице, в противном случае вычисляется обычным образом.

Перед вводом формул ТЕНДЕНЦИЯ() и РОСТ() необходимо выделить результирующие ячейки, ввести ссылки на известные и новые значения, при необходимости ввести единицу в строку *Константа* и нажать комбинацию клавиш <Ctrl+Shift+Enter> для ввода данных в массив.

Содержание работы

Основная часть работы состоит в решении практических задач, использующих возможности программы Excel для анализа временных рядов.

Порядок выполнения

1. Для ряда, изображенного на графике (рис. 6.4) применить методы скользящего среднего и экспоненциального сглаживания средствами Excel, сравнить результаты и сделать выводы.

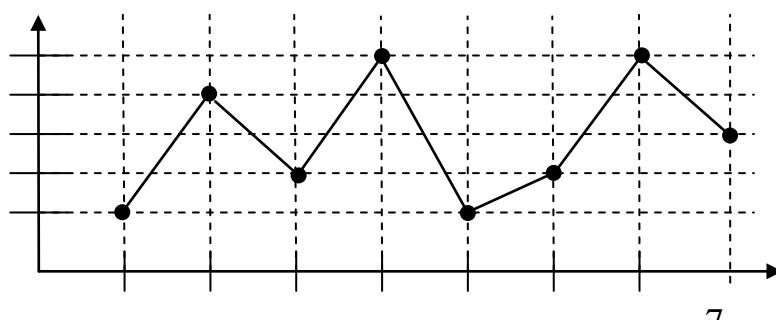


Рис. 6.4 Временной ряд

2. Сформировать в столбцах массивы случайных чисел, распределенных по требуемому закону распределения, число переменных = 1, число случайных чисел = 100:
 - 2.1. Биномиальное распределение. Исходные данные: $p = 0,8$; число испытаний – 25.
 - 2.2. Нормальное распределение. Исходные данные: среднее – 100; стандартное отклонение – 15.
3. На основе построенных массивов п.2 получить случайную выборку размером 14.
4. Для полученных исходных данных п.3 построить несколько вариантов типов моделей и выбрать наиболее адекватную модель тренда.
5. Для имеющихся данных продаж за первое полугодие (см. табл. 6.1) спрогнозируйте объем продаж во втором полугодии, используя линейную и экспоненциальную зависимости с помощью функций ТЕНДЕНЦИЯ() и РОСТ() соответственно. Сравните результаты.

Таблица 6.1

Месяц (x)	1	2	3	4	5	6	7	8	9	10	11	12
тыс. р.(y)	380	250	285	420	440	480						

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.

2. Результаты решения примеров, описанных в теоретической части.
3. Результаты решения контрольных заданий.
4. Выводы по результатам работы.

Контрольные вопросы

1. Дать определение временного ряда. Привести примеры.
2. Перечислить задачи, решаемые при изучении временных рядов.
3. На какие части можно разделить модель временного ряда?
4. Что такое тренд? Какими моделями он может быть описан?
5. Какими способами можно выделить тренд?
6. Раскрыть принцип скользящего среднего.

7. Регрессионный анализ

Цель работы

Изучить средства программы Microsoft Excel для регрессионного анализа данных.

Теоретические основы

На практике встречается достаточно большое количество взаимосвязей, например, между заработной платой работника и его образованием, объемом выпускаемой продукции и затратами на производство, между производительностью работника и его отношением к труду и т.д. В связи с этим возникает необходимость в анализе двухмерных данных. Для этого используют два базовых инструмента: корреляционный анализ, позволяющий оценить степень взаимосвязи между факторами и регрессионный анализ, который показывает, как можно предсказать поведение одной из двух переменных или управлять ею с помощью другой.

Коэффициент корреляции

Наиболее важным для практического использования является случай, когда связь между признаками линейная. Мера силы линейной связи признаков называется коэффициентом корреляции. Величина коэффициента, близкая к 1, указывает, что зависимость между данными случайными величинами почти линейная. Значения, близкие к нулю, означают, что связь между величинами либо слабая, либо не носит линейного характера.

Для расчета коэффициента корреляции можно использовать функцию Excel =КОРРЕЛ(*массив1*; *массив2*) из категории статистические.

Для вывода результатов в виде таблицы применяют функцию **Корреляция** из пакета **Анализ данных**. После ее вызова из окна диалога **Анализ данных** откроется соответствующее диалоговое окно (рис. 7.1).

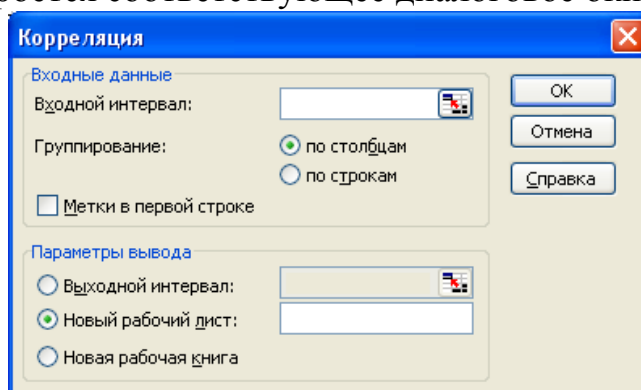


Рис. 7.1. Корреляция

В диалоговом окне (рис. 7.1) задаются следующие параметры:

1. *Входной интервал* - вводится диапазон ячеек, содержащих анализируемые данные.

2. *Группирование* – переключатель устанавливается в требуемое положение в зависимости от расположения исходных данных.

3. *Метки в первой строке* – флажок ставится, если первая строка содержит заголовок, в противном случае будут созданы стандартные заголовки автоматически.

4. *Параметры вывода* – указывается место, где будет указана таблица результатов анализа.

Линейная регрессия

Форма связи результативного признака Y с факторами X_1, X_2, \dots, X_m называется уравнением регрессии. В зависимости от типа выбранного уравнения различают линейную и нелинейную регрессию, а в зависимости от количества факторов – парную (простую, $m = 1$) и множественную (многочленную, $m > 1$).

На этапе регрессионного анализа решаются следующие задачи:

1. Выбор общего вида уравнения регрессии и определение параметров регрессии.
2. Определение степени взаимосвязи результативного признака и факторов, проверка общего качества уравнения регрессии.
3. Проверка статистической значимости каждого коэффициента уравнения регрессии и определение их доверительных интервалов.

Уравнение простой линейной регрессии имеет вид: $y = b_0 + b_1x$, множественная линейная регрессия описывается следующим уравнением: $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_mx_m$.

Параметры уравнений парной и множественной регрессий могут быть определены с помощью метода наименьших квадратов, который реализован в Excel. Для этого используется функция **Регрессия**. Для ее вызова необходимо выбрать требуемое имя в окне диалога **Анализ данных**. В результате появится диалоговое окно (рис. 7.2).

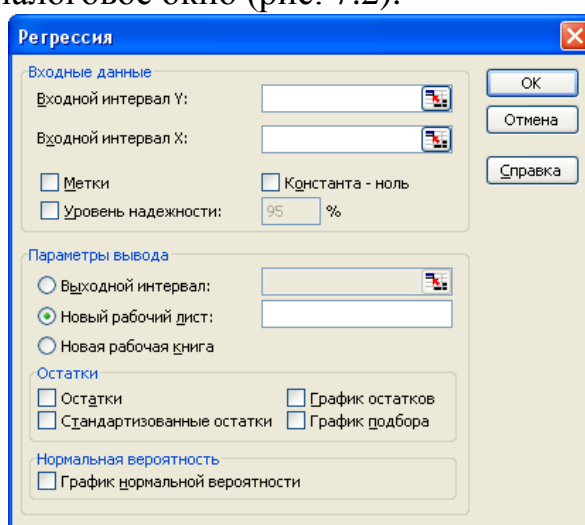


Рис. 7.2 Регрессия

В диалоговом окне (рис. 7.2) задаются следующие параметры:

1. *Входной интервал Y* – вводится диапазон ячеек (один столбец), содержащих исходные данные по результирующему признаку.

2. *Входной интервал X* – вводится диапазон ячеек (число столбцов равно количеству признаков), содержащих исходные данные факторного признака.

3. *Метки* – флажок ставится, если первая строка содержит заголовок, в противном случае будут созданы стандартные заголовки автоматически.

4. *Уровень надежности* – флажок устанавливается, если требуется ввести значение уровня отличное от 95%. При выключенном флажке уровень надежности принимается равным 95%.

5. *Константа-ноль* - флажок устанавливается в том случае, когда требуется, чтобы линия регрессии прошла через начало координат, т.е. $b_0 = 0$.

6. *Параметры вывода* – указывается место, где будут указаны таблицы результатов анализа.

7. *Остатки* – при необходимости вывода столбцов остатков и графиков остатков и подбора необходимо включить соответствующие флажки.

8. *Нормальная вероятность* – флажок устанавливается, если не требуется вывести график зависимости наблюдаемых значений от автоматически формируемых интервалов перцентилей.

Пример. Для исходных данных (табл. 7.1) построить регрессионную линейную однофакторную модель зависимости затрат на ТО от срока службы с помощью функции **Регрессия**.

Таблица 7.1

X	срок службы	2	5	9	3	8
Y	затраты на ТО	6	13	23	5	22

Результаты решения с помощью функции **Регрессия** выводятся в виде нескольких отдельных таблиц.

Результаты расчета регрессионной статистики выводятся в следующей таблице (табл. 7.2):

Таблица 7.2

<i>Регрессионная статистика</i>	
Множественный R	0,984535285
R-квадрат	0,969309728
Нормированный R-квадрат	0,959079637
Стандартная ошибка	1,724792855
Наблюдения	5

В таблице 7.2 указаны следующие элементы:

Множественный R – коэффициент корреляции.

R-квадрат – коэффициент детерминации.

Нормированный R-квадрат – нормированное значение коэффициента корреляции.

Стандартная ошибка - стандартное отклонение для остатков.

Наблюдения - количество исходных наблюдений.

В следующей таблице (табл. 7.3) представлены результаты дисперсионного анализа, которые используются для проверки значимости коэффициента детерминации.

Таблица 7.3

<i>Дисперсионный анализ</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	281,8752688	281,8752688	94,75084337	0,002303227
Остаток	3	8,924731183	2,974910394		
Итого	4	290,8			

В таблице 7.3 указаны следующие элементы:

df – число степеней свободы. Для строки *Регрессия* это количество факторных признаков, для строки *Остаток* – число наблюдений минус количество переменных в уравнении регрессии, для строки *Итого* – сумма степеней свободы для строк *Регрессия* и *Остаток*.

SS – сумма квадратов отклонений. Для строки *Регрессия* это значение определяется как сумма квадратов отклонений теоретических данных от среднего, для строки *Остаток* это сумма квадратов отклонений эмпирических данных от теоретических, для строки *Итого* это сумма квадратов отклонений эмпирических данных от среднего.

MS – дисперсии. Для строки *Регрессия* это факторная дисперсия, для строки *Остаток* это остаточная дисперсия.

F – расчетное значение *F*-критерия Фишера, определяемое как отношение факторной дисперсии к остаточной.

Значимость F – значение уровня значимости, соответствующее вычисленному значению *F*.

Полученные значения коэффициентов регрессии и их статистические оценки сводятся в следующую таблицу (табл. 7.4):

Таблица 7.4

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	-1,064516129	1,710826692	-0,622223241	0,577881718
срок службы	2,752688172	0,282790929	9,734004488	0,002303227

В таблице 7.4 указаны следующие элементы:

Коэффициенты – значения коэффициентов модели.

Стандартная ошибка – стандартные ошибки коэффициентов.

t-статистика – расчетные значения *t*-критерия, вычисляемого как отношение значений коэффициентов к соответствующим стандартным ошибкам.

P-Значение – значения уровней значимости, соответствующие вычисленным значениям t_p .

В экранной таблице Excel также указываются нижние и верхние границы доверительных интервалов для коэффициентов регрессии - *Нижние 95%*, *Верхние 95%* (ввиду ограниченности места в таблице 7.4 они опущены).

На основе данных из полученных таблиц можно сделать следующие выводы:

1. Уравнение регрессии имеет вид: $y = -1,06 + 2,75 \cdot x$.
2. Значение коэффициента детерминации, равного 0,97 показывает, что срок службы существенно влияют на затраты на ТО, что подтверждает правильность включения его в построенную модель.
3. Рассчитанный уровень значимости *Значимость* $F = 0,002$ меньше 0,05 подтверждает значимость величины коэффициента детерминации.
4. *P-Значение* для срока службы, равное 0,002 и меньше 0,05 подтверждает значимость коэффициента b_1 .
5. *P-Значение* для коэффициента b_0 превышает 0,05, это означает, что данный коэффициент для модели не является значимым и его можно опустить, т.е. график модели будет проходить через точку начала координат.

Для получения новой модели без коэффициента b_0 необходимо еще раз запустить функцию **Регрессия**, в окне рис. 7.2 поставить флажок **Константа-ноль**. В результате построятся новые таблицы.

Приведем таблицу для значений коэффициентов регрессии (табл. 7.5).

Таблица 7.5

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	0	-	-	-
срок службы	2,595628415	0,11732738	22,12295556	2,47108E-05

Анализ новых полученных таблиц, показывает значимость коэффициента модели и коэффициента детерминации, что подтверждает адекватность полученного уравнения.

В итоге модель получится следующего вида: $y = 2,6 \cdot x$.

Если в результате анализа незначимыми окажутся коэффициенты b_1, b_2, \dots , то следует пересчитать результаты регрессии, не указывая в поле *Входной интервал X* (см. рис. 7.2) диапазон ячеек с данными соответствующего фактора.

Содержание работы

Основная часть работы состоит в решении практических задач, использующих возможности программы Excel для регрессионного анализа данных.

Порядок выполнения

1. Для исходных данных (табл. 7.1) вычислить корреляцию попарно между признаками, применяя функцию =КОРРЕЛ(*массив1;массив2*) и с помощью функции **Корреляция**. Сделать выводы.

2. Чему равен коэффициент корреляции двух случайных величин, представленных в таблице 7.6.

Таблица 7.6

X	12,1	14,7	20,5	16,6	19,0
Y	53,2	44,2	51,4	45,5	34,0

3. На основе данных (табл. 7.7) построить линейную модель и провести ее анализ.

Таблица 7.7

Количество работников (X)	Объем производства (Y)
7	483
6	489
7	486
8	563
9	570
9	559
9	594
9	575
6	464
9	647

4. Построить модель зависимости величины заработной платы от стажа работы и пола сотрудника (табл. 7.8). Проверить адекватность модели.

Таблица 7.8

Заработная плата, Y	Стаж работы, X ₁	Пол (0-муж., 1-жен) X ₂
38900	15	0
28700	2	1
31600	4	1
33800	13	0
31890	16	0
45000	35	0
24313	10	1
22700	8	1
36300	20	0
32350	7	0
31800	5	0

5. Определить по данным (табл.7.9) параметры уравнения линейной регрессии и провести его анализ.

Таблица 7.9

Предприятие	Прибыль, Y	Оборотные средства, X ₁	Основные фонды, X ₂
1	188	129	510
2	78	64	190
3	93	69	240

Предприятие	Прибыль, Y	Оборотные средства, X_1	Основные фонды, X_2
4	152	87	470
5	55	47	110
6	161	102	420

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.
2. Результаты решения примеров, описанных в теоретической части.
3. Результаты решения контрольных заданий.
4. Выводы по результатам работы.

Контрольные вопросы

1. Что называется регрессионной моделью?
2. Привести общий вид регрессионной модели.
3. Каким образом можно проверить значимость коэффициента регрессии?
4. Какой метод обычно используется при определении коэффициентов регрессионной модели?

8. Работа со сводными таблицами

Цель работы

Изучить возможности сводных таблиц Microsoft Excel для анализа данных.

Теоретические основы

Создание сводной таблицы на основе списка Excel

В процессе работы большинства предприятий накапливаются различные данные, например, для торгового предприятия это могут быть данные о продажах товаров, для предприятий сотовой связи — статистика нагрузки на базовые станции. Очень часто менеджменту предприятия необходима аналитическая информация, которая генерируется на основе необработанных данных, например, посчитать вклад каждого вида товара в доходы предприятия или качество обслуживания в зоне данной станции. Из необработанной информации такие сведения извлечь очень тяжело: нужно выполнять сложные SQL-запросы, которые обрабатываются долго и часто мешают текущей работе. Поэтому в настоящее время необработанные данные сводятся вначале в хранилище архивных данных Data Warehouse, а затем — в кубы OLAP, которые очень удобны для интерактивного анализа. Можно представить кубы OLAP как многомерные таблицы, в которых вместо стандартных двух измерений (столбцы и строки, как в обычных таблицах) измерений может быть очень много. Обычно для описания измерений в кубе используется термин "в разрезе". Например, отделу маркетинга может быть нужна информация во временном разрезе, в региональном разрезе, в разрезе типов продукта, в разрезе каналов продаж и т. п.

При помощи кубов, в отличие от стандартных SQL-запросов, очень просто получать ответы на такие вопросы, как "сколько товаров такого-то типа было продано в четвертом квартале прошлого года в Северо-Западном регионе через региональных дистрибьюторов".

Для работы с такими кубами в Excel встроен специальный клиент — **Pivot Table (Сводная таблица)**, который доступен на вкладке **Вставка** в группе **Таблицы**.

Таким образом, сводная таблица представляет собой интерактивную таблицу, с помощью которой можно быстро объединять и анализировать большие объемы данных. Она позволяет получать динамические представления данных, объединять большие массивы, переупорядочить их, подводить итоги по различным полям.

Отображая разные страницы, можно выполнять следующие действия:

- осуществлять фильтрацию данных,
- детализировать информацию,
- автоматически получать промежуточные или общие итоги,

– добавлять формулы в вычисляемые поля или элементы полей.

Независимо от назначения и оформления все сводные таблицы имеют элементы (области): область строк; область столбцов; фильтр отчета и область данных.

В областях строк, столбцов и фильтре отчета находятся поля, для элементов которых производятся вычисления над элементами полей, помещенных в область данных. Обычно область данных содержит значения, которые обрабатываются определенной функцией.

Поля представляют собой категорию (группу) данных, извлекаемых из одного столбца исходной таблицы. Название поля извлекается из верхней ячейки столбца исходной таблицы. Области сводной таблицы могут иметь любое количество полей.

Для того чтобы сводная таблица выполняла свои функции, область данных должна иметь хотя бы одно поле. Области строк, столбцов и фильтр отчета могут не иметь полей.

Вычисление значений в отчете сводной таблицы

Для вычисления полей значений используются следующие итоговые функции:

Функция	Вычисляемое значение
СУММ	Сумма значений. Это функция для числовых данных по умолчанию.
СРЗНАЧ	Среднее набора значений.
МАКС	Наибольшее значение.
МИН	Наименьшее значение.
ПРОИЗВЕД	Произведение значений.
Кол-во чисел	Количество числовых значений. Итоговая функция «Кол-во чисел» работает так же, как функция СЧЁТ.
СТАНДОТКЛОН	Оценка стандартного отклонения совокупности, в которой примером является подмножество всей совокупности.
СТАНДОТКЛОНП	Стандартное отклонение совокупности, где совокупностью являются все складываемые данные.
ДИСП	Оценка изменчивости совокупности, в которой примером является подмножество всей совокупности.
ДИСПР	Изменчивость совокупности, где совокупностью являются все складываемые данные.

Новые возможности сводных таблиц в Excel 2010

Кроме вычисления суммы или количества в сводной таблице можно подсчитывать различные итоги, включая различные виды долей (от общей суммы, от суммы по строке или столбцу, относительно заданного элемента), нарастающие итоги, динамику изменений (отличие и приведенное отличие) и т.д.

В 2010 версии Excel к этому набору добавились несколько новых функций, доступных через контекстное меню (для их использования

щелкните правой кнопкой мыши по полю в таблице и выберите **Дополнительные вычисления**):

- **% от суммы по родительской строке (столбцу)** - позволяет посчитать долю относительно промежуточного итога по строке или столбцу.

В прошлых версиях можно было вычислять долю только относительно общего итога.

- **% от суммы нарастающим итогом** - работает аналогично функции суммирования нарастающим итогом, но отображает результат в виде доли, т.е. в процентах. Удобно считать, например, процент выполнения плана или исполнения бюджета.
- **сортировка от минимального к максимальному и наоборот** - функция ранжирования, вычисляющая порядковый номер (позицию) элемента в общем списке значений. Например, с ее помощью удобно ранжировать менеджеров по их суммарной выручке, определяя кто на каком месте в общем зачете.

Срезы

Работая с большими сводными таблицами, часто приходится их принудительно упрощать, фильтруя часть информации. Самый простой способ для этого – поместить некоторые поля в область фильтра (в версиях до 2007 она называлась область страниц) и выбирать из выпадающих списков только нужные значения.

Неудобства такого способа очевидны:

- при выборе нескольких элементов - их не видно, а видно текст "(несколько элементов)", что затрудняет просмотр;
- один фильтр отчета жестко привязан к одной сводной таблице. Если у нас несколько сводных таблиц, то для каждой придется создавать свой фильтр и для каждой придется его раскрывать, отмечать галочками нужные элементы.

В Excel 2010 отбор требуемых элементов это можно делать с помощью **срезов (slicers)**.

Срезы – это графическое представление интерактивных фильтров для сводной таблицы или диаграммы.

Если используются несколько срезов, то это позволит быстро и наглядно отобразить взаимосвязи между элементами данных (см. рис. 8.1).

магазин	название столбцов	стол	тумба	Общий итог
Кв-п1	105			105
2007 Итого	105			105
Кв-п1	363			363
Кв-п2	77			77
Кв-п3	270			270
2008 Итого	270	440		710
Общий итог	375	440		815

Срез мебели: диван, стол, тумба, шкаф.

Срез магазина: Адамант, Азия, Европа, Кент, Парнас, Феликс.

Рис. 8.1 Использование срезов для отображения данных

Срез выглядит как отдельный графический объект (как диаграмма или картинка), не связан с ячейками и отображается на листе, что позволяет легко его перемещать.

Чтобы создать срез для текущей сводной таблицы нужно перейти на вкладку **Параметры (Options)** и в группе **Сортировка и фильтр (Sort and filter)** нажать кнопку **Вставить срез (Insert slicer)**, как показано на рис. 8.2:

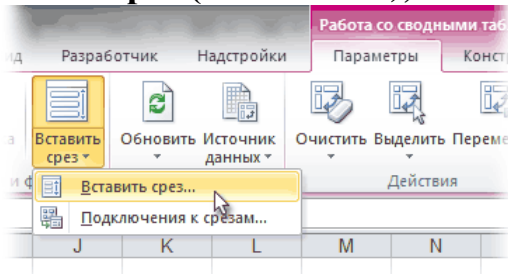


Рис. 8.2 Вставить срез

При выделении или снятии выделения с элементов среза можно использовать клавиши **Ctrl** и **Shift**, а также протягивание с нажатой левой кнопкой мыши для выделения оптом в сводной таблице будут отображаться только отфильтрованные данные по отобранным элементам.

Дополнительная особенность в том, что срез разными цветами отображает не только выделенные, но еще и пустые элементы, для которых нет ни одного значения в исходной таблице.

Вставка вычисляемого поля или вычисляемого элемента

Поскольку сводная таблица является особым типом диапазона данных, в нее нельзя вставлять новые строки или столбцы. Это означает, что нельзя вставить в сводную таблицу формулы, которые будут выполнять операции над данными.

Если итоговые функции и настраиваемые вычисления не дают нужный результат, можно создать собственные формулы в вычисляемых полях и вычисляемых элементах. Например, можно добавить вычисляемый элемент с формулой для комиссионных сборов. В отчет сводной таблицы будут автоматически включены комиссионные сборы в промежуточные и общие итоги.

Особенности применения:

- Вычисляемое поле используется для подстановки в формулу данных из другого поля.
- Вычисляемый элемент применяется для подстановки в формулу данных из одного или нескольких специальных элементов внутри поля.
- Для вычисляемых элементов можно вводить различные формулы в ячейки по очереди. Например, если для всех месяцев в вычисляемом элементе **Продажи** содержится формула $=\text{Продажи} * 0,25$, ее можно заменить формулой $=\text{Продажи} * 0,5$ для июня, июля и августа месяца.

- При наличии нескольких вычисляемых элементов или формул можно изменить порядок их вычисления.

Изменение формулы сводной таблицы

Перед изменением формулы определите, где она находится: в вычисляемом поле или в вычисляемом элементе. Если формула находится в вычисляемом элементе, определите, является ли она единственной для этого элемента. Учитывайте, что в вычисляемых элементах можно изменять отдельные формулы в конкретных ячейках.

Для определения, где находится формула: в вычисляемом поле или вычисляемом элементе можно использовать команду **Вывести формулы** на вкладке **Параметры** в группе **Вычисления** нажмите кнопку **Поля, элементы и наборы**.

Для изменения формулы вычисляемого поля или элемента следует:

- открыть соответствующее окно **Вставка вычисляемого поля (или объекта)**,
- выбрать вычисляемое поле (или элемент), для которого необходимо изменить формулу, в поле с раскрывающимся списком **Имя**,
- нажать кнопку **Изменить** и внести правку в формулу.

Удаление формулы сводной таблицы

Удаление формулы сводной таблицы является необратимой операцией. Если формулу не требуется удалять окончательно, можно скрыть поле или элемент, перетащив его за пределы отчета.

Определите, где хранится формула: в вычисляемом поле или вычисляемом элементе.

Вычисляемые поля отображаются в списке полей сводной таблицы. Вычисляемые элементы отображаются как элементы в других полях.

Выполните одну из следующих процедур:

- чтобы удалить вычисляемое поле, щелкните в любом месте отчета сводной таблицы.
- чтобы удалить вычисляемое поле, в отчете сводной таблицы щелкните поле, содержащее элемент, который требуется удалить.

При этом на экране появится элемент «Работа со сводными таблицами» с добавленными вкладками **Параметры** и **Проектирование**.

На вкладке **Параметры** в группе **Вычисления** нажмите кнопку **Поля, элементы и наборы** и выберите команду **Вычисляемое поле** или **Вычисляемый объект**.

В поле **Имя** выберите требуемое поле или элемент и нажмите кнопку **Удалить**.

Создание сводной таблицы на основе данных, находящихся в нескольких диапазонах

Если данные, которые вы намерены проанализировать имеют одинаковые заголовки строк и столбцов, но размещены в разных списках, на разных листах или в разных файлах, можно использовать метод создания сводной таблицы на основе данных, находящихся в нескольких диапазонах консолидации.

В версиях до 2003 включительно в мастере сводных таблиц была опция "*построить сводную по нескольким диапазонам консолидации*". В версиях Excel 2007/2010 эту функцию убрали из стандартного диалога при создании отчетов сводных таблиц, и она доступна только через настраиваемую кнопку "**Мастер сводных таблиц**" (Pivot Table Wizard), которую можно добавить на панель быстрого доступа.

Содержание работы

Основная часть работы состоит в решении практических задач, использующих возможности сводных таблиц Excel для анализа данных.

Порядок выполнения

Построение сводной таблицы

1. Постройте сводную таблицу, показанную в книге Список мебели.xlsx на листе *Сводная таблица*.
2. Добавьте промежуточные итоги по каждому году. Для этого следует на вкладке **Конструктор** в группе **Макет** воспользоваться кнопкой **Промежуточные итоги** и выбрать команду **Показывать все итоги в нижней части группы** (пример показан на листе *Сводная таблица_Итоги*).

Создание вычисляемого поля в сводной таблице

Откройте книгу Список мебели_Сводная таблица_вычисляемое поле.xlsx.

На листе *Сводная_поле* изучите построенную сводную таблицу.

Предположим, что необходимо постоянно отслеживать остаток продукции на складе. Для этого необходимо вычесть из поля **кол-во на складе** поле **кол-во продано**. Результат будет храниться в новом (вычисляемом) поле сводной таблицы.

Чтобы создать такое поле, выполните следующие действия:

1. Поместите курсор в область сводной таблицы. При этом на ленте появится вкладка **Работа со сводными таблицами** с добавленными вкладками **Параметры** и **Конструктор**.
2. На вкладке **Параметры** в группе **Вычисления** нажмите кнопку **Поля, элементы и наборы** и выберите команду **Вычисляемое поле**

(Calculated Field). Появится диалоговое окно **Вставка вычисляемого поля** (Insert Calculated Field in...).

3. В поле **Имя** введите имя данного поля – **Остаток**.
4. В поле **Формула** введите выражение, с использованием других полей: **= кол-во на складе – кол-во продано**. Чтобы использовать в формуле данные из другого поля, выберите это поле в списке **Поля**, а затем нажмите кнопку **Добавить поле**.
5. Щелкните на кнопке **Добавить** (Add) для добавления нового поля, затем на кнопке **ОК**, чтобы закрыть диалоговое окно.
6. Если итоги по новому полю добавились в область *Названия столбцов*, то для лучшей наглядности перенесите поле в область *Названия строк*.
7. Примерный вид итоговой таблицы можно посмотреть на листе *Сводная_поле* (ответ).

Вставка вычисляемого элемента

Откройте книгу *Список мебели_Сводная таблица_вычисляемый элемент.xlsx* и перейдите на лист *Сводная_элемент*.

Сводная таблица содержит текстовое поле с именем *Мебель*. В эту таблицу можно добавить вычисляемые элементы (назовем их, например, *Мягко* и *Жестко*), которые будут содержать сумму объемов продаж мебели различного вида.

Чтобы просуммировать данные в строках с помощью вычисляемого элемента, выполните следующие действия:

1. Поместите курсор в область сводной таблицы. В нашем случае установите курсор в любой элемент мебели. При этом на ленте появится вкладка **Работа со сводными таблицами** с добавленными вкладками **Параметры** и **Конструктор**.
2. На вкладке **Параметры** в группе **Вычисления** нажмите кнопку **Поля, элементы и наборы** и выберите команду **Вычисляемый объект (Calculated Item...)**. Появится диалоговое окно **Вставка вычисляемого элемента (Insert Calculated Item in...)**.
3. Введите имя вычисляемого элемента в поле с раскрывающимся списком **Имя** – **Мягко**.
4. Введите формулу для элемента **Мягко** в поле **Формула**: **=диван + стул**.
Чтобы использовать в формуле данные элемента поля, выберите этот элемент в списке **Элементы**, а затем нажмите кнопку **Добавить элемент** (элемент должен находиться в одном поле с вычисляемым элементом). Нажмите кнопку **Добавить**.
5. Введите формулу для элемента **Жестко** в поле **Формула**: **=стол + тумба + шкаф**. Нажмите кнопку **Добавить** и закройте окно.
6. Примерный вид итоговой таблицы можно посмотреть на листе *Сводная_элемент* (ответ).

Все созданные вычисляемые элементы будут автоматически добавлены в сводную таблицу, причем новые элементы отобразились после элементов поля *Мебель*. Чтобы вставить вычисляемые поля между исходными элементами, можно перетащить их на требуемое место.

Вычисляемый элемент можно создать только на основе не сгруппированных элементов.

Для вычисляемых элементов можно вводить различные формулы в разные ячейки. Для этого выберите ячейку, в которой требуется изменить формулу, и измените ее в строке формул.

При наличии нескольких вычисляемых элементов или формул можно изменить порядок их вычисления. Для этого на вкладке **Параметры** в группе **Вычисления** нажмите кнопку **Поля, элементы и наборы** и выберите команду **Порядок вычислений**. Выберите формулу и нажмите кнопку **Вверх** или **Вниз**. Продолжайте выполнять предыдущее действие, пока формулы не окажутся на нужных местах.

Можно просмотреть все формулы, используемые в текущем отчете сводной таблицы. Чтобы отобразить список всех формул на вкладке **Параметры** в группе **Вычисления** нажмите кнопку **Поля, элементы и наборы** и выберите команду **Вывести формулы**. В результате добавиться новый лист с отчетом.

Объединение данных с помощью сводной таблицы

Откройте книгу Автосалон_сводная консолидация.xlsx.

Требуется на листе *Summary_Tab* построить сводную таблицу на основе данных, расположенных на листах *Jan*, *Feb* и *Mar*.

Для построения сводной таблицы выполните следующие действия:

1. Добавьте кнопку "Мастер сводных таблиц" (Pivot Table Wizard) на панель быстрого доступа.
2. Перейдите на лист *Summary_Tab* и установите курсор в ячейку E10.
3. Щелкните кнопку "Мастер сводных таблиц", откроется мастер сводных таблиц и диаграмм.
4. На первом шаге работы мастера сводных таблиц выберите опцию *В нескольких диапазонах консолидации* (Multiple consolidation ranges).
5. На шаге 2а укажите, должны ли создаваться в сводной таблице страничные поля. Выберите вариант *Создать поля страницы* (I will create the page fields).
6. На шаге 2б введите адреса диапазонов консолидации. Для этого укажите адрес диапазона (A6:C11) на листе *Jan* и щелкните кнопку **Добавить (Add)**. Затем введите адреса остальных диапазонов.
7. Определите количество полей (How many page fields) равное 1 и для каждого диапазона в поле для ввода меток элементов *Первое поле* (Field one) укажите имена Январь, Февраль и Март соответственно. Нажмите **Далее**.
8. Укажите местоположение таблицы – на существующем листе. При необходимости измените адрес ячейки вывода таблицы. Нажмите

Готово (Finish). На листе в указанном месте построится сводная таблица. Примерный вид итоговой таблицы можно посмотреть на листе Summary_Tab_Res.

Отчётность по работе

После выполнения работы обучаемый представляет отчет. Отчёт должен содержать:

1. Название и цель работы.
2. Результаты решения примеров, описанных в теоретической части.
3. Результаты решения контрольных заданий.
4. Выводы по результатам работы.

Контрольные вопросы

1. Перечислите основные задачи, решаемые с помощью сводных таблиц.
2. Какую роль играют при построении сводных таблиц срезы?
3. Каким образом можно добавить новое поле в сводную таблицу?
4. Каким образом можно добавить вычисляемый элемент в сводную таблицу?
5. Какие методы обычно используются при анализе данных в сводной таблице?

Список литературы

1. Афонин А.Ю. Макарычев П. П. Оперативный и интеллектуальный анализ данных – Пермь: ПГУ, 2010 – 142 с.
2. Билл Джелен, Майкл Александер. Сводные таблицы в Microsoft Excel 2010. – М.: Вильямс, 2011 – 464 с.
3. Джон Уокенбах. Microsoft Excel 2010. Библия пользователя – М.: Диалектика, 2013 – 912 с.
4. Джон Уокенбах. Формулы Microsoft Excel 2010. – М.: Диалектика, 2011 – 704 с.
5. Конрад Карлберг. Бизнес-анализ с использованием Excel. – М.: Вильямс, 2012 – 576 с.
6. Шумейко А.А. Интеллектуальный анализ данных (Введение в Data Mining)/ А.А. Шумейко, С.Л. Сотник. – Днепропетровск: Белая Е.А., 2012. – 212 с.



В 2009 году Университет стал победителем многоэтапного конкурса, в результате которого определены 12 ведущих университетов России, которым присвоена категория «Национальный исследовательский университет». Министерством образования и науки Российской Федерации была утверждена программа его развития на 2009–2018 годы. В 2011 году Университет получил наименование «Санкт-Петербургский национальный исследовательский университет информационных технологий, механики и оптики»

КАФЕДРА ПРОГРАММНЫХ СИСТЕМ

Кафедра **Программных систем** входит в состав нового факультета **Инфокоммуникационные технологии**, созданного решением Ученого совета университета 17 декабря 2010 г. по предложению инициативной группы сотрудников, имеющих большой опыт в реализации инфокоммуникационных проектов федерального и регионального значения.

На кафедре ведется подготовка бакалавров и магистров по направлению **210700 «Инфокоммуникационные технологии и системы связи»**:

**210700.62.10 – ИНТЕЛЛЕКТУАЛЬНЫЕ
ИНФОКОММУНИКАЦИОННЫЕ СИСТЕМЫ (Бакалавр)**

**210700.68.05 – ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ В
ИНФОКОММУНИКАЦИЯХ (Магистр)**

Выпускники кафедры получают фундаментальную подготовку по: математике, физике, электронике, моделированию и проектированию инфокоммуникационных систем (ИКС), информатике и программированию, теории связи и теории информации.

В рамках профессионального цикла изучаются дисциплины: архитектура ИКС, технологии программирования, ИКС в Интернете, сетевые технологии, администрирование сетей Windows и UNIX, создание программного обеспечения ИКС, Web программирование, создание клиент-серверных приложений.

Область профессиональной деятельности бакалавров и магистров включает:

- сервисно-эксплуатационная в сфере современных ИКС;
- расчетно-проектная при создании и поддержке сетевых услуг и сервисов;

- экспериментально-исследовательская;
- организационно-управленческая – в сфере информационного менеджмента ИКС.

Знания выпускников востребованы:

- в технических и программных системах;
- в системах и устройствах звукового вещания, электроакустики, речевой, и мультимедийной информатики;
- в средствах и методах защиты информации;
- в методах проектирования и моделирования сложных систем;
- в вопросах передачи и распределения информации в телекоммуникационных системах и сетях;
- в методах управления телекоммуникационными сетями и системами;
- в вопросах создания программного обеспечения ИКС.

Выпускники кафедры Программных систем обладают компетенциями:

- проектировщика и разработчика структур ИКС;
- специалиста по моделированию процессов сложных систем;
- разработчика алгоритмов решения задач ИКС;
- специалиста по безопасности жизнедеятельности ИКС;
- разработчика сетевых услуг и сервисов в ИКС;
- администратора сетей: UNIX и Windows;
- разработчика клиентских и клиент-серверных приложений;
- разработчика Web – приложений;
- специалиста по информационному менеджменту;
- менеджера проектов планирования развития ИКС.

Трудоустройство выпускников:

1. ОАО «Петербургская телефонная сеть»;
2. АО «ЛЕНГИПРОТРАНС»;
3. Акционерный коммерческий Сберегательный банк Российской Федерации;
4. ОАО «РИВЦ-Пулково»;
5. СПб ГУП «Петербургский метрополитен»;
6. ООО «СоюзБалтКомплект»;
7. ООО «ОТИС Лифт»;
8. ОАО «Новые Информационные Технологии в Авиации»;
9. ООО «Т-Системс СиАйЭс» и др.

Кафедра сегодня имеет в своем составе высококвалифицированный преподавательский состав, в том числе:

- 5 кандидатов технических наук, имеющих ученые звания профессора и доцента;
- 4 старших преподавателя;

- 6 штатных совместителей, в том числе кандидатов наук, профессиональных IT - специалистов;
- 15 Сертифицированных тренеров, имеющих Западные Сертификаты фирм: Microsoft, Oracle, Cisco, Novell.

Современная техническая база; лицензионное программное обеспечение; специализированные лаборатории, оснащенные необходимым оборудованием и ПО; качественная методическая поддержка образовательных программ; широкие Партнерские связи существенно влияют на конкурентные преимущества подготовки специалистов.

Авторитет специализаций кафедры в области компьютерных технологий подтверждается Сертификатами на право проведения обучения по методикам ведущих Западных фирм - поставщиков аппаратного и программного обеспечения.

Заслуженной популярностью пользуются специализации кафедры ПС по подготовке и переподготовке профессиональных компьютерных специалистов с выдачей Государственного Диплома о профессиональной переподготовке по направлениям: **"Информационные технологии (инженер-программист)"** и **"Системный инженер"**, а также Диплома о дополнительном (к высшему) образовании с присвоением квалификации: **"Разработчик профессионально-ориентированных компьютерных технологий "**. В рамках этих специализаций высокопрофессиональные преподаватели готовят компетентных компьютерных специалистов по современным в России и за рубежом операционным системам, базам данных и языкам программирования ведущих фирм: Microsoft, Cisco, IBM, Intel, Oracle, Novell и др.

Профессионализм, компетентность, опыт, и качество программ подготовки и переподготовки IT- специалистов на кафедре ПС неоднократно были удостоены **высокими наградами «Компьютерная Элита» в номинации лучший учебный центр России.**

Партнеры:

1. **Microsoft** Certified Learning Solutions;
2. **Novell** Authorized Education Center;
3. **Cisco** Networking Academy;
4. **Oracle** Academy;
5. **Sun Java** Academy и др;
6. **Prometric**;
7. **VUE**.

Мы готовим квалифицированных инженеров в области инфокоммуникационных технологий с новыми знаниями, образом мышления и способностями быстрой адаптации к современным условиям труда.

Ирина Станиславовна Осетрова
Никита Алексеевич Осипов

Microsoft Excel 2010 для аналитиков

УЧЕБНОЕ ПОСОБИЕ

В авторской редакции
Редакционно-издательский отдел НИУ ИТМО
Зав. РИО
Лицензия ИД № 00408 от 05.11.99
Подписано к печати
Заказ №
Тираж 100 экз.
Отпечатано на ризографе

Н.Ф. Гусарова

Редакционно-издательский отдел
Санкт-Петербургского национального
исследовательского университета
информационных технологий, механики
оптики
197101, Санкт-Петербург, Кронверкский пр., 49

