

КАЗАНСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ
ИНСТИТУТ УПРАВЛЕНИЯ, ЭКОНОМИКИ И ФИНАНСОВ
Кафедра экономико-математического моделирования

И.И.Исмагилов, Г.И. Лисогор, С.Ф. Хасанова

РЕШЕНИЕ ЭКОНОМИЧЕСКИХ ЗАДАЧ
СРЕДСТВАМИ MS EXCEL

Учебно-методическое пособие

Казань 2015

УДК 681.3

ББК 32.97

Публикуется по решению заседания кафедры экономико-математического моделирования

Протокол № 2 от 15 октября 2015 года

Рецензент:

доктор технических наук

доцент кафедры экономико-математического моделирования

В.А. Качалкин

Исмагилов И.И., Лисогор Г.И., Хасанова С.Ф.

Решение экономических задач средствами MS EXCEL: учебно-методическое пособие для студентов, обучающихся по направлению 38.03.01 «Экономика»

/ И.И. Исмагилов, Г.И. Лисогор, С.Ф. Хасанова – 36 с.

Данное учебно-методическое пособие предназначено для проведения практических занятий по дисциплине «Экономико-математические модели» при подготовке студентов по направлению 38.03.01 «Экономика». Цель учебно-методического пособия – развить методические и практические умения и навыки построения и проверки качества оптимизационных и регрессионных моделей в среде MS Excel.

Оглавление

ВВЕДЕНИЕ	4
1. РЕШЕНИЕ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ С ПОМОЩЬЮ MS EXCEL	4
1.1. Задача распределения ресурсов.....	5
1.2. Экономико-математическая модель задачи и технология ее решения	5
Контрольные задания по разделу 1.	10
Тесты по разделу 1.	12
2. РЕШЕНИЕ ЗАДАЧ СТАТИСТИЧЕСКОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ДАННЫХ В СРЕДЕ MS EXCEL	13
2.1. Описательные статистики	15
2.2. Корреляционный и регрессионный анализ.....	20
2.3. Анализ и прогнозирование временных рядов	28
Контрольные задания по разделу 2.	32
Тесты по разделу 2.	33
Список литературы.....	35

ВВЕДЕНИЕ

Наиболее важной задачей экономистов является поддержка принятий управленческих решений. Современные компьютерные технологии позволяют существенно ускорить и расширить спектр решаемых задач. Электронные таблицы предоставляют пользователю широкие возможности для выполнения экономических расчетов. В настоящее время мировое первенство на рынке электронных таблиц занимает MS Excel, входящий в состав пакета Microsoft Office. Использование специальных средств *Надстроек* расширяет возможности табличного процессора. Из основных надстроек MS Excel отметим две.

Пакет анализа – обеспечивает дополнительные возможности анализа наборов данных при решении сложных статистических, экономических и инженерных задач. Для проведения анализа данных с помощью этого инструмента следует указать входные данные и выбрать параметры; анализ будет проведен с помощью подходящей статистической или инженерной макрофункции, а результат будет помещен в выходной диапазон. Дополнительные инструменты позволяют представить результаты анализа в графическом виде.

Поиск решения – используется для решения задач оптимизации. С помощью этого инструмента возможны постановка, решение и анализ всех классов задач оптимизации: линейного, целочисленного, нелинейного и стохастического программирования в области менеджмента, консалтинга и при оптимальном проектировании технических систем.

1. РЕШЕНИЕ ЗАДАЧ ЛИНЕЙНОГО ПРОГРАММИРОВАНИЯ С ПОМОЩЬЮ MS EXCEL

В экономике оптимизационные задачи возникают в связи с многочисленностью возможных вариантов функционирования конкретного экономического объекта, когда возникает ситуация выбора варианта, наилучшего по некоторому правилу, критерию, характеризующему соответствующей целевой функцией (например, иметь минимум затрат, максимум продукции).

Продемонстрируем возможности надстройки – *Поиск решения* на примерах решения задач линейного программирования.

Перед началом работы необходимо убедиться в наличии установки надстройки на вашем компьютере. Если в *Меню Данные* отсутствует команда *Поиск решения*, значит, необходимо загрузить эту надстройку. Выберите команду *Файл – Параметры – Надстройки – Перейти* - и активизируйте надстройку *Поиск решения*. Если же надстройки нет в диалоговом окне *Надстройки*, то вам необходимо обратиться к панели управления Windows, щелкнуть на пиктограмме *Установка и удаление программ* и с помощью программы установки MS Excel (или MS Office) установить надстройку *Поиск решения*.

1.1. Задача распределения ресурсов

Если финансы, оборудование, сырье и даже людей полагать ресурсами, то значительное число задач в экономике можно рассматривать как задачи распределения ресурсов. Достаточно часто математической моделью таких задач является задача линейного программирования.

Рассмотрим следующий пример.

Требуется определить, в каком количестве надо выпускать продукцию четырех типов Прод1, Прод2, Прод3, Прод4, для изготовления которой требуются ресурсы трех видов: трудовые, сырье, финансы. Количество ресурса каждого вида, необходимое для выпуска единицы продукции данного типа, называется нормой расхода. Нормы расхода, а также прибыль, получаемая от реализации единицы каждого типа продукции, приведены в табл. 1. Там же приведено наличие располагаемого ресурса.

Таблица 1

Ресурсы	Нормы расхода ресурсов на единицу продукции				знак	Наличие ресурсов
	Прод1	Прод2	Прод3	Прод4		
Трудовые	1	1	1	1	\leq	16
Сырье	6	5	4	3	\leq	110
Финансы	4	6	10	13	\leq	100
Прибыль	60	70	120	130	max	-

Требуется найти такой план выпуска продукции, при котором будет максимальной общая прибыль.

1.2. Экономико-математическая модель задачи и технология ее решения

Обозначим через X_1, X_2, X_3, X_4 количество выпускаемой продукции каждого типа. Тогда экономико-математическая модель задачи, включающая требование максимизации целевой функции F и ограничения по ресурсам, будет иметь вид:

$$\text{целевая функция } F = 60X_1 + 70X_2 + 120X_3 + 130X_4 \rightarrow \max \quad (1)$$

при ограничениях:

$$\begin{cases} X_1 + X_2 + X_3 + X_4 \leq 16, \\ 6X_1 + 5X_2 + 4X_3 + 3X_4 \leq 110, \\ 4X_1 + 6X_2 + 10X_3 + 13X_4 \leq 100, \\ X_1, X_2, X_3, X_4 \geq 0. \end{cases} \quad (2)$$

Для решения задачи с помощью надстройки *Поиск решения* необходимо выполнить действия в ряд шагов:

Шаг 1. Создать формулу на рабочем листе MS Excel для ввода условий задачи.

Шаг 2. Указать адреса ячеек, в которые будет помещен результат решения (изменяемые ячейки).

Шаг 3. Ввести исходные данные.

Шаг 4. Ввести зависимость для целевой функции.

Шаг 5. Ввести зависимости для ограничений.

Шаг 6. Указать назначение целевой функции (установить целевую ячейку).

Шаг 7. Ввести ограничения.

Шаг 8. Ввести параметры для решения задачи линейного программирования.

Рассмотрим подробно все перечисленные выше этапы.

Шаг 1. Для задачи (1) - (2) подготовим форму, представленную на рис. 1, для ввода условий.

	А	В	С	Д	Е	Ф	Г	Н
1				Переменные				
2	имя	прод1	прод2	прод3	прод4			
3	значение							
4	нижн. гр.							
5	верх. гр.					ЦФ		
6	коэф. в ЦФ						макс	
7				Ограничения				
8	вид					левая часть	знак	правая часть
9	трудовые							
10	сырье							
11	финансы							

Рисунок 1

Весь текст на рис. 1 (и в дальнейшем) является комментарием и на решение задачи не влияет.

Шаг 2. В нашей задаче оптимальные значения переменных X_1, X_2, X_3, X_4 будут помещены в ячейках В3:Е3, оптимальное значение целевой функции – в ячейке F6.

Шаг 3. Введем исходные данные в созданную форму. Получим результат, показанный на рис. 2.

	А	В	С	Д	Е	Ф	Г	Н
1				Переменные				
2	имя	прод1	прод2	прод3	прод4			
3	значение							
4	нижн. гр.							
5	верх. гр.					ЦФ		
6	коэф. в ЦФ	60	70	120	130		макс	
7				Ограничения				
8	вид					левая часть	знак	правая часть
9	трудовые	1	1	1	1		<=	16
10	сырье	6	5	4	3		<=	110
11	финансы	4	6	10	13		<=	100

Рисунок 2.

Шаг 4. Введем зависимость для целевой функции (1).

При данной форме ввода исходных данных на рабочий лист необходимо в ячейку F6 поместить формулу $=B6*B3+C6*C3+D6*D3+E6*E3$. Однако для формул подобной структуры удобнее использовать встроенную функцию СУММПРОИЗВ(), которая возвращает сумму произведений соответствующих элементов массивов (в данном случае массивов В6:Е6 и В3:Е3).

- Поместите курсор в ячейку F6.

- Курсор на кнопку *Мастер функций*.
- М1 (обозначим через М1 действие – «один щелчок левой кнопкой мыши»).
На экране: диалоговое окно *Мастер функций шаг 1 из 2*.
- Курсор в окно *Категория* на категорию *Математические*.
- М1.
- Курсор в окно *Функция* на СУММПРОИЗВ.
- М1.
- ОК.

На экране появляется диалоговое окно. Сместите его перетаскиванием мышью в нижнюю часть экрана как на рис. 3.

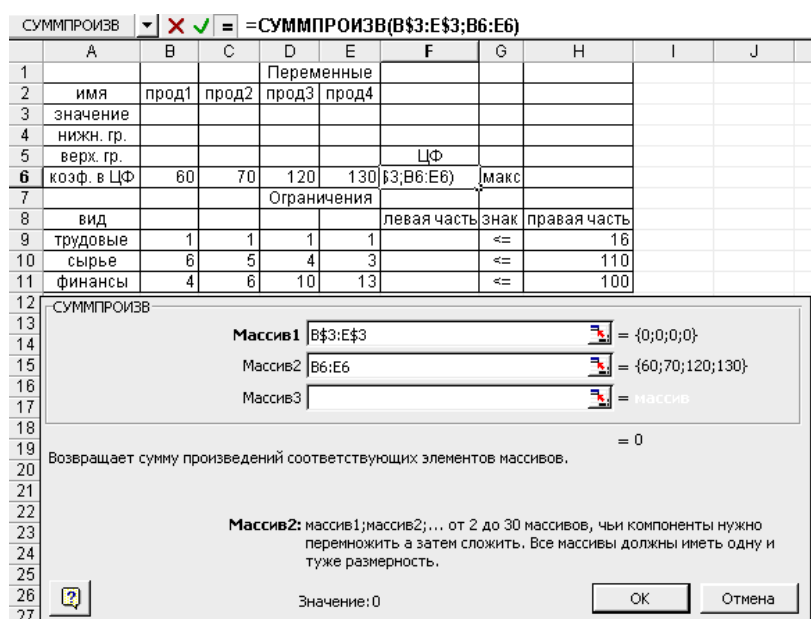


Рисунок 3.

- В массив 1 ввести B\$3:E\$3. Использование здесь адресации диапазона как для *смешанных ссылок* в дальнейшем упростит ввод зависимостей для левых частей ограничений (2).

Заметим, что во все диалоговые окна адреса ячеек удобно вводить не с клавиатуры, а протаскивая мышью по ячейкам, чьи адреса следует ввести.

- В массив 2 ввести B6:E6.
- ОК.

Шаг 5. Введем зависимости для левых частей ограничений (2).

- Курсор в F6.
- Копировать в буфер обмена
- Курсор в F9.
- Вставить из буфера
- Скопировать F9 в F10:F11.

На этом ввод зависимостей закончен.

Запуск надстройки *Поиск решения* осуществляется командой *Данные → Поиск решения*, после применения которой появляется окно *Поиск решения*, представленное на рис. 4.

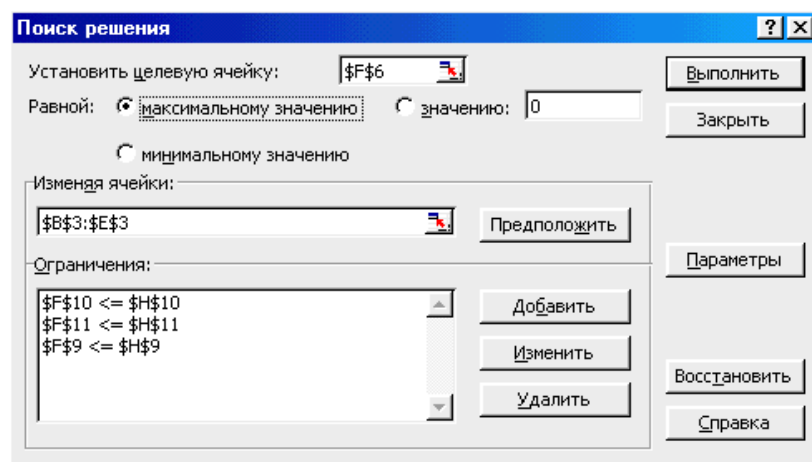


Рисунок 4.

В диалоговом окне *Поиск решения* есть три основных параметра:

- Установить целевую ячейку.
- Изменяя ячейки.
- Ограничения.

Сначала нужно заполнить поле *Установить целевую ячейку*. Во всех задачах для средства *Поиск решения* оптимизируется результат в одной из ячеек рабочего листа. Целевая ячейка связана с другими ячейками этого рабочего листа с помощью формул. Средство *Поиск решения* использует формулы, которые дают результат в целевой ячейке, для проверки возможных решений. Можно выбрать поиск наименьшего или наибольшего значения для целевой ячейки или же установить конкретное значение.

Второй важный параметр средства *Поиск решения* – это параметр *Изменяя ячейки*. Изменяемые ячейки – это те ячейки, значения в которых будут изменяться для того, чтобы оптимизировать результат в целевой ячейке. Для поиска решения можно указать до 200 изменяемых ячеек. К изменяемым ячейкам предъявляются два основных требования: они не должны содержать формул, и изменение их значений должно отражаться на изменении результата в целевой ячейке. Другими словами, целевая ячейка зависима от изменяемых ячеек.

Третий параметр, который нужно вводить для *Поиска решения* – это *Ограничения*.

Шаг 6. Назначение целевой функции (установить целевую ячейку).

- Курсор в поле *Установить целевую ячейку*.
- Ввести адрес $\$F\6 (щелчком мыши по ячейке F6).
- Ввести направление изменения целевой функции: *максимальному значению*.

Ввести адреса искомых переменных:

- Курсор в поле *Изменяя ячейки*.
- Ввести адреса $\$B\$3:\$E\3 (протаскиванием мыши по диапазону ячеек).

Шаг 7. Ввод ограничений.

- Курсор на кнопку *Добавить*.

- M1.
- Появится диалоговое окно *Добавление ограничения* (рис. 5).

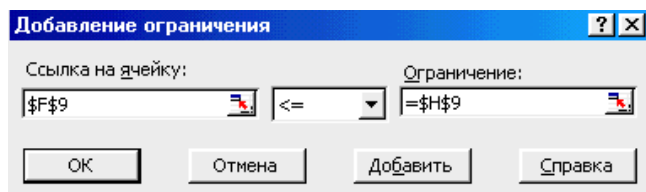


Рисунок 5.

- В поле *Ссылка на ячейку* ввести адрес \$F\$9.
- С помощью раскрывающегося списка в центре окна ввести знак ограничения \leq .
- Курсор в правое поле *Ограничение*.
- Ввести адрес \$H\$9.
- Кнопка *Добавить*. На экране опять диалоговое окно *Добавление ограничения*.
- Аналогично ввести остальные ограничения.
- После ввода последнего ограничения применить кнопку *ОК*.

Если при вводе условий задачи возникает необходимость в изменении или удалении внесенных ограничений, то это делается с помощью команд *Изменить*, *Удалить*.

- На экране появится диалоговое окно *Поиск решения* с введенными условиями (2) (см. рис. 4).

Шаг 8. Ввод параметров для решения задачи линейного программирования.

- С помощью кнопки *Параметры* открыть окно *Параметры поиска решения*, представленного на рис. 6.

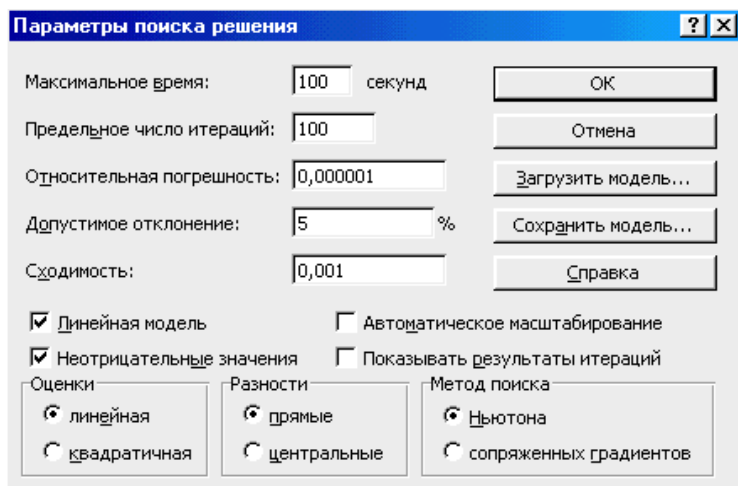


Рисунок 6.

С помощью команд, находящихся в этом диалоговом окне, можно вводить условия для решения задач оптимизации всех классов.

Поле *Максимальное время* служит для назначения времени в секундах, выделяемого на поиск решения задачи. В поле можно ввести время, не превышающее 32767 с (более 9 часов).

Поле *Предельное число итераций* служит для назначения числа итераций.

Команды, используемые по умолчанию, подходят для решения большей части практических задач.

- Установить флажок *Линейная модель*, что обеспечивает применение симплекс-метода.
- Установить флажок *Неотрицательные значения*, что позволит выполнить условия последней строки системы неравенств (2).
- *ОК*.

На экране появится диалоговое окно Поиск решения.

- *Выполнить*.

На экране появятся диалоговое окно *Результаты поиска решения*, а в соответствующих ячейках рабочего листа численные результаты оптимального решения задачи, которые представлены на рис. 7.

	A	B	C	D	E	F	G	H
1				Переменные				
2	имя	прод1	прод2	прод3	прод4			
3	значение	10	0	6	0			
4	нижн. гр.							
5	верх. гр.					ЦФ		
6	коэф. в ЦФ	60	70	120	130	1320	макс	
7				Ограничения				
8	вид					левая часть	знак	правая часть
9	трудовые	1	1	1	1	16	<=	16
10	сырье	6	5	4	3	84	<=	110
11	финансы	4	6	10	13	100	<=	100
12	Результаты поиска решения							
13	Решение найдено. Все ограничения и условия оптимальности выполнены.							
14					Тип отчета			
15	<input checked="" type="radio"/> Сохранить найденное решение				Результаты			
16	<input type="radio"/> Восстановить исходные значения				Устойчивость			
17					Пределы			
18	ОК				Отмена			
19					Сохранить сценарий...			
20					Справка			

Рисунок 7.

Полученное решение означает, что максимальная общая прибыль $F_6 = 1320$ может быть получена при следующем плане выпуска продукции: $Прод1 = B_3 = 10$, $Прод2 = C_3 = 0$, $Прод3 = D_3 = 6$, $Прод4 = E_3 = 0$. При этом ресурсы трудовые и финансы будут использованы полностью, а из 110 единиц сырья будет использовано 84.

Таково оптимальное решение рассматриваемой задачи распределения ресурсов. Отметим, что подобная технология используется для решения широкого класса задач, например, определения оптимального плана размещения предприятий, организации оптимального снабжения, управления запасами, транспортной задачи, задачи о назначениях и др.

Контрольные задания по разделу 1.

Задание 1 . Решите задачу о размещении средств.

Пусть собственные средства банка вместе с депозитами в сумме составляют 100 млн. долл. Часть этих средств, но не менее 35 млн. долл., должна быть размещена в кредитах. Кредиты являются неликвидными активами банка, так как в случае непредвиденной потребности в наличности обратит кредиты в деньги без существенных потерь невозможно.

Другое дело ценные бумаги, особенно государственные. Их можно в любой момент продать, получив некоторую прибыль или, во всяком случае, без большого убытка. Поэтому существует правило, согласно которому коммерческие банки должны покупать в определенной пропорции ликвидные активы – ценные бумаги, чтобы компенсировать неликвидность кредитов. В нашем примере ликвидное ограничение таково: ценные бумаги должны составлять не менее 30% средств, размещенных в кредитах и ценных бумагах.

Обозначим через X_1 средства (млн. долл.), размещенные в кредитах, через X_2 – средства, вложенные в ценные бумаги. Цель банка состоит в том, чтобы получить максимальную прибыль от кредитов и ценных бумаг: $F = C_1X_1 + C_2X_2$, где C_1 – доходность кредитов, C_2 – доходность ценных бумаг.

Целевая функция – это выражение, которое необходимо максимизировать: $F = 9X_1 + 6X_2$.

Имеем следующую систему линейных ограничений:

1. $X_1 + X_2 \leq 100$ – балансовое ограничение;
2. $X_1 \geq 35$ – кредитное ограничение;
3. $X_2 \geq 0,3(X_1 + X_2)$ – ликвидное ограничение;
4. $X_1 \geq 0, X_2 \geq 0$.

Используя предложенную в разделе технологию решения задачи линейного программирования, найдите оптимальные значения искомым величин X_1, X_2 .

Ответ: $X_1 = 70$ млн. долл., $X_2 = 30$ млн. долл., $F = 810$ млн. долл.

Задание 2 . Для производства двух видов изделий типа А и типа В предприятие использует 3 вида сырья. Нормы расхода каждого вида сырья на изготовление единицы продукции каждого типа, общее количество сырья, имеющегося в наличии, а также прибыль от реализации одного изделия каждого типа приведены в таблице:

Таблица 2

Виды сырья	Нормы расхода сырья		Общее количество сырья (кг)
	А	В	
I	12	10	300
II	4	5	120
III	3	2	252
Прибыль от реализации одного изделия (руб.)	40	30	

Найти план выпуска продукции, обеспечивающий в рамках имеющихся ресурсов наибольшую прибыль.

Ответ: $X_A = 25, X_B = 0, F = 1000$ руб.

Задание 3. Повторите решение задачи из задания 6.3 при условии $X_B \geq 6$ на нижнюю границу выпуска продукции типа В.

Ответ: $X_A = 20$, $X_B = 6$, $F = 980$ руб.

Задание 4 . Шоколадный цех выпускает шоколадные батончики трех видов. Известна норма расхода сырья на каждый батончик:

Таблица 3

Виды сырья	Нормы расхода сырья (гр.)			Общее количество сырья (кг)
	«Белка»	«Зайка»	«Рыбка»	
Сахар	30	30	20	100
Шоколад	10	15	10	50
Нуга	15	10	15	25
Орех	10	0	0	10
Вафли	0	10	5	5

Один батончик «Белка» стоит 7 руб., «Зайка» - 6 руб., «Рыбка» - 5 руб.

Как спланировать производство шоколадных батончиков, чтобы прибыль от их производства была наибольшей?

Ответ: $X_1 = 1000$, $X_2 = 250$, $X_3 = 500$, $F = 11000$ руб.

Задание 5. Решить следующие задачи:

1. $F(X) = -6X_1 - 4X_2 + 4X_3 \rightarrow \max$ **3.** $F(X) = 2X_1 + 4X_2 + 6X_3 \rightarrow \min$

при ограничениях:

$$\begin{cases} X_1 + X_2 + X_3 \geq -1, \\ -2X_1 - X_2 + X_3 \leq 1, \\ X_1 \geq 0, X_2 \geq 0, X_3 \geq 0. \end{cases} \quad \begin{cases} -X_1 + X_2 + X_3 \geq 1, \\ X_1 + X_2 + 2X_3 \geq 1, \\ X_1 \geq 0, X_2 \geq 0, X_3 \geq 0. \end{cases}$$

2. $F(X) = -2X_1 - 6X_2 + 6X_3 + 6X_4 \rightarrow \max$ **4.** $F(X) = -3X_1 - 4X_2 + X_3 \rightarrow \min$

при ограничениях:

$$\begin{cases} -X_1 + 2X_2 + 2X_3 + X_4 \leq 1, \\ 2X_1 - X_2 + X_3 + 2X_4 \leq 1, \\ X_1 \geq 0, X_2 \geq 0, X_3 \geq 0, X_4 \geq 0. \end{cases} \quad \begin{cases} -2X_1 - X_2 + X_3 \geq -10, \\ X_1 - X_2 - 2X_3 \geq 5, \\ X_1 \geq 0, X_2 \geq 0, X_3 \geq 0. \end{cases}$$

Тесты по разделу 1.

1. К основным надстройкам MS Excel относятся:

- а) *Анализ пакетов*;
 в) *Пакет анализа*;
- б) *Решение поиска*;
 г) *Поиск решения*.
2. Надстройка *Поиск решения* используется для решения задач оптимизации:
 а) линейного программирования;
 б) нелинейного программирования;
 в) стохастического программирования;
 г) целочисленного программирования.
3. Установить надстройку на ПК можно с помощью команды:
 а) *Данные - Надстройки*;
 в) *Параметры-Надстройки*;
- б) *Сервис - Поиск решения*;
 г) *Данные - Поиск решения*.
4. В диалоговом окне *Поиск решения* есть параметры:
 а) *Ограничения*;
 б) *Изменяя ячейки*;
 в) *Ссылка на ячейку*;
 г) *Установить целевую ячейку*.
5. Запуск надстройки *Поиск решения* осуществляется командой:
 а) *Данные → Надстройки*;
 в) *Сервис → Надстройки*;
- б) *Сервис → Поиск решения*;
 г) *Данные → Поиск решения*.
6. Выберите правильную последовательность установки параметров в диалоговом окне надстройки *Поиск решения*:
 а) *Данные → Поиск решения → Ограничения*;
 б) *Установить целевую ячейку → Изменяя ячейки → Ограничения*;
 в) *Поиск решения → Параметры → Изменяя ячейки*;
 г) *Изменяя ячейки → Ограничения → Данные*.
7. В надстройке *Поиск решения* можно указать число изменяемых ячеек до:
 а) 20;
 б) 200;
 в) 2000.
8. Окно *Параметры поиска решения* позволяет управлять временем, выделяемым на поиск решения задачи:
 а) максимальным;
 б) минимальным;
 в) средним.
9. При использовании надстройки *Поиск решения* время поиска решения задачи ограничено:
 а) 9 часами;
 б) 5 часами;
 в) 1 часом;
 г) не ограничено.
10. К изменяемым ячейкам в надстройке *Поиск решения* предъявляются требования:
 а) они не должны содержать формул; б) они должны содержать формулы;
 в) изменение их значений не должно отражаться на изменении результата в целевой ячейке;
 г) изменение их значений должно отражаться на изменении результата в целевой ячейке.

2. РЕШЕНИЕ ЗАДАЧ СТАТИСТИЧЕСКОГО АНАЛИЗА И ПРОГНОЗИРОВАНИЯ ДАННЫХ В СРЕДЕ MS EXCEL

Методы статистического анализа позволяют изучать совокупность массовых явлений с целью выявления закономерностей и их характеристик при помощи обобщенных показателей. При этом должно выявляться типичное, характерное, массовое.

Моделирование процессов с помощью статистических методов выражается в конечном итоге в построении статистических моделей. Важную роль при этом играют правильность проведения статистического наблюдения, умение оперировать относительными и средними величинами, проводить группировку статистических материалов, использовать основные виды выборок.

Методы математической статистики в приложении к экономическим процессам могут применяться для решения задач:

- анализа с использованием описательных статистик;
- дисперсионного анализа;
- корреляционного анализа;
- регрессионного анализа;
- многомерного статистического анализа;
- факторного анализа;
- прогнозирования.

Очевидно, что этот перечень задач не полон и может быть расширен.

Современное математико-статистическое программное обеспечение представлено значительным количеством пакетов прикладных программ (ППП) соответствующего функционального назначения. Отметим, что мировыми лидерами являются пакеты STATISTICA и SPSS. Представлены статистические методы также в ряде приложений офисного назначения, например, в табличном процессоре MS Excel.

Отметим, что при использовании компьютерных технологий целесообразно придерживаться графически-ориентированного подхода к анализу данных. Этот подход состоит в том, чтобы получить всестороннее визуальное представление данных на всех этапах статистической обработки.

Процесс статистической обработки данных с помощью ППП, как правило, включает следующие этапы:

1. Экономическая постановка задачи и подготовка данных.
2. Ввод данных в систему.
3. Преобразование данных, адекватное выбранным статистическим методам.
4. Визуализация данных с помощью различных типов графиков и диаграмм.
5. Реализация статистического алгоритма.
6. Выбор результатов анализа в виде графиков и таблиц с числовой и текстовой информацией.
7. Интерпретация полученных результатов.

При этом пользователь освобождается от рутинной, трудоемкой работы (по проведению расчетов, построению таблиц, графиков и т.д.). На его долю

остается исследовательская работа по постановке задачи, выбора метода обработки и оценки качества моделей интерпретации, которая носит творческий характер.

Для статистического анализа и прогнозирования MS Excel предоставляет большое число встроенных функций категории *Статистические* и специализированное средство – *Пакет анализа*. Пакет анализа – это надстройка, содержащая набор функций и инструментов для расширения встроенных аналитических возможностей MS Excel. Следует отметить, что весьма значительная мощь рассматриваемого табличного процессора заключена в библиотеке встроенных функций и дополнительных программных надстройках.

При необходимости использования средства *Пакет анализа* предварительно следует проверить доступ к нему командой *Файл - Параметры - Надстройки - Перейти* (должен быть установлен флажок *Пакет анализа*). Если *Пакет анализа* отсутствует в списке надстроек, необходимо его установить, запустив программу установки MS Excel. Для выбора требуемого инструмента анализа используется команда *Данные → Анализ данных*.

При использовании инструментов *Пакета анализа* следует помнить, что выходные данные являются таблицами констант. Следовательно, при изменениях в исходных данных необходимо запустить соответствующую процедуру повторно, так как автоматического перерасчета выходных данных не производится. В случае же встроенных статистических функций возможность автоматического перерасчета предусмотрена.

Рассмотрим решение отдельных задач статистического анализа и прогнозирования с использованием средств MS Excel. При этом ограничимся рассмотрением кратких теоретических сведений и технологий реализации методов описательной статистики, корреляционного и регрессионного анализа, анализа и прогнозирования временных рядов. За более полными теоретическими сведениями следует обратиться к соответствующей учебной литературе по статистике и эконометрике.

2.1. Описательные статистики

Методы описательной статистики служат для выявления самых общих свойств наблюдаемых величин в компактной форме. К числу основных описательных статистик относятся: среднее значение, дисперсия, стандартное отклонение, медиана, мода, максимальное и минимальное значения, размах, квантили. На практике чаще всего используются выборочное среднее, медиана и дисперсия (или стандартное отклонение).

Следует отметить, что в основе математической статистики лежат понятия генеральной совокупности и выборки (выборочной совокупности). Все множество изучаемых элементов называется генеральной совокупностью, а выбранная для исследования группа элементов – выборкой. Таким образом, выборка – это лишь часть генеральной совокупности. Выборку называют репрезентативной

(представительной), если она достаточно полно представляет изучаемые признаки и параметры генеральной совокупности. Отметим, что на практике обычно располагают лишь выборочной совокупностью.

Очевидно, что в зависимости от вида совокупности показатели описательной статистики могут быть определены либо для генеральной совокупности, либо для выборки. Ограничимся приведением формул для вычисления основных выборочных характеристик:

1. Выборочная средняя: $\bar{y} = \frac{1}{N} \cdot \sum_{i=1}^N y_i$.

2. Выборочная дисперсия: $d = \frac{1}{N} \cdot \sum_{i=1}^N (y_i - \bar{y})^2$.

В статистике чаще выборочную дисперсию определяют по формуле:

$$d = \frac{1}{N-1} \cdot \sum_{i=1}^N (y_i - \bar{y})^2.$$

Именно такое определение выборочной дисперсии используется в статистических функциях MS Excel.

3. Стандартное отклонение (среднее квадратическое отклонение): $\sigma = \sqrt{d}$.

4. Размах вариации (интервал): $R = y_{\max} - y_{\min}$.

5. Коэффициент вариации: $v = \frac{\sigma}{\bar{y}} \cdot 100\%$.

При $v > 40\%$ в выборке наблюдается большая вариация данных.

К другим важным показателям, характеризующим взаимное расположение данных в выборке, относятся мода и медиана.

Мода – это наиболее часто встречающееся значение в выборке. Если значения не повторяются, мода отсутствует.

Медиана – срединное значение вариационного ряда при нечетном N и полусумма двух срединных значений при четном N . Вариационный ряд – это результат упорядочения исходной выборки данных по возрастанию их численных значений. Выборочная медиана разбивает выборку пополам: слева и справа от неё оказывается одинаковое число элементов выборки.

Форма распределения случайной величины характеризуется значениями асимметрии и эксцесса. Принято считать, что асимметрия характеризует несимметричность распределения случайной величины, а эксцесс – степень выраженности остроконечности или сглаженности распределения. Асимметрия принимает положительное (отрицательное) значение, если более “длинная” часть кривой плотности распределения лежит правее (левее) моды. Эксцесс равен нулю для нормального распределения, положителен для остроконечных и отрицателен для сглаженных по сравнению с нормальной плотностью распределения вероятностей.

В MS Excel достаточно полно представлены методы описательной статистики. Для определения статистик можно использовать следующие инструменты:

- встроенные статистические функции;

- инструмент **Описательная Статистика** надстройки **Пакет Анализа**.

Рассмотрим наиболее часто используемые встроенные статистические функции. Аргументом всех этих функций может быть диапазон ячеек или список чисел, разделенных точкой с запятой. Ниже аргумент *диапазон* задает диапазон ячеек.

Функция **СРЗНАЧ** (*диапазон*) находит среднее значение чисел из указанного диапазона ячеек.

Функция **МЕДИАНА** (*диапазон*) вычисляет медиану чисел из указанного диапазона ячеек.

Функция **МОДА** (*диапазон*) определяет значение, которое чаще других встречается во множестве чисел. Если диапазон не содержит повторяющихся чисел, функция возвращает ошибочное значение #Н/Д.

Функция **МАКС** (*диапазон*) находит наибольшее значение из набора данных.

Функция **МИН** (*диапазон*) находит наименьшее значение из набора данных.

Четыре статистические функции **ДИСП** (*диапазон*), **ДИСПР** (*диапазон*), **СТАНДОТКЛОН** (*диапазон*) и **СТАНДОТКЛОНП** (*диапазон*) вычисляют дисперсию и стандартное отклонение для чисел, расположенных в диапазоне ячеек. Отметим, что функции **ДИСП** и **СТАНДОТКЛОН** используются в случае, когда множество чисел является выборкой из генеральной совокупности. Функции **ДИСПР** и **СТАНДОТКЛОНП** предполагают, что значения образуют конечную генеральную совокупность.

Показатели формы распределения случайной величины асимметрия и эксцесс вычисляют статистические функции **СКОС** и **ЭКСЦЕСС**.

Часто целесообразно провести анализ распределения исходного множества чисел с использованием функции **ЧАСТОТА**. Функция **ЧАСТОТА** (*массив данных; массив границ*) для исходного множества значений (*массив данных*) определяет число значений (частот), попадающих в интервалы разбиения (*массив границ*).

Функция **ЧАСТОТА** является функцией массива. Для использования функции выполните следующие действия:

1. Создайте столбец границ интервалов. Интервалы разбиения не обязательно должны быть одной ширины, но они должны идти по возрастанию.
2. Выделите диапазон для выходных данных (только один столбец ячеек).
3. Введите формулу, задав входной диапазон.
4. Введите диапазон границ интервалов в качестве аргументов. Нажмите одновременно клавиши [Ctrl] +[Shift]+[Enter] (для создания формулы массива).

Сводную таблицу основных статистических характеристик для одного или нескольких множеств данных можно получить с помощью инструмента анализа данных **Описательная Статистика**. Диалоговое окно инструмента представлено на рис. 8.

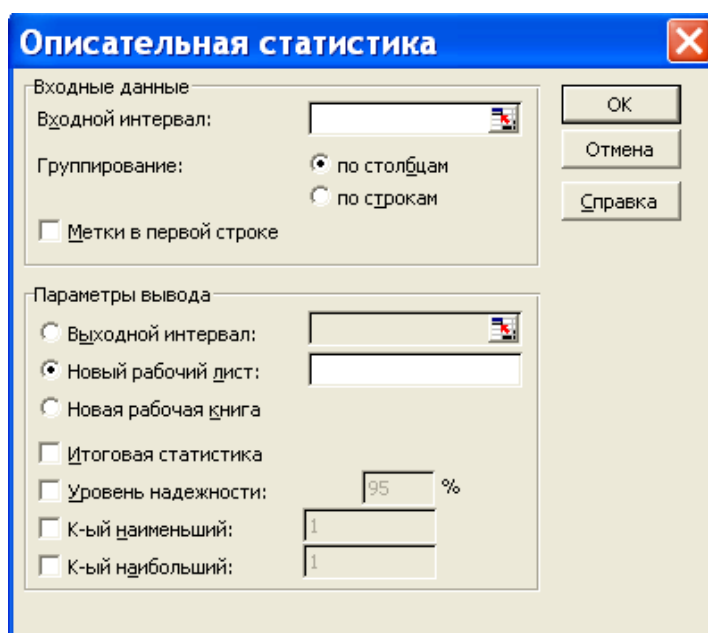


Рисунок 8. Диалоговое окно инструмента *Описательная Статистика*

Для использования инструмента **Описательная Статистика** выполните следующие действия:

1. В главном меню выберите пункты *Данные* → *Анализ данных* → *Описательная статистика* и нажмите кнопку *OK*.

2. Заполните диалоговое окно ввода данных и параметров вывода и щелкните на кнопке *OK*.

Инструмент **Описательная Статистика** требует задания следующей информации:

- Входной интервал – диапазон, содержащий исходные данные;
- Группирование – по столбцам или строкам;
- Метки в первой строке – флажок (указывает, содержит ли первая строка названия столбцов или нет);
- Выходной интервал – диапазон для статистик (достаточно указать левую верхнюю ячейку диапазона).

Если необходимо получить дополнительную информацию ***Итоговая статистика***, ***Уровень надежности***, ***K-ые наибольшее и наименьшее значения***, установите соответствующие флажки в диалоговом окне.

Пример вычисления описательных статистик приведен на рис. 9. В выходной таблице содержатся показатели итоговой статистики. Эта таблица является таблицей констант. В случае необходимости большинство показателей можно получить с использованием соответствующих встроенных статистических функций MS Excel.

Национальное богатство ряда стран в конце XX века			
Страна	На душу населения, тыс. долларов	<i>Столбец1</i>	
США	461,5	Среднее	248,24
Россия	400	Стандартная ошибка	49,27
Япония	423,4	Медиана	324,40
Китай	28,5	Мода	#Н/Д
Германия	375,1	Стандартное отклонение	170,66
Франция	359,6	Дисперсия выборки	29125,80
Великобритания	353	Эксцесс	-1,92
Италия	295,8	Асимметричность	-0,27
Бразилия	89	Интервал	441,50
Индонезия	60	Минимум	20,00
Индия	20	Максимум	461,50
Мексика	113	Сумма	2978,90
		Счет	12,00
		Наибольший (2)	423,40
		Наименьший (2)	28,50
		Уровень надежности (95,0%)	108,43

Рисунок 9. Исходные данные и результат применения инструмента *Описательная статистика*

Исходные данные рассмотренного примера представляют национальное богатство ряда стран в конце XX века на душу населения (оценки по методологии Всемирного банка)*. Данные приведены по странам, общий объем национальных богатств которых превышает 10 трлн. долларов.

Полученные показатели описательной статистики свидетельствуют о значительной неравномерности распределения национальных богатств стран на душу населения (коэффициент вариации $v = \frac{170,66}{248,24} \cdot 100\% = 69\%$). Наблюдается заметное отличие выборочной медианы от выборочной средней. Следовательно, данные не подчиняются нормальному распределению. По этой причине при анализе невозможно применение показателя уровня надежности (доверительный интервал для средней), так как он определяется исходя из предположения о нормальности выборки.

В заключение отметим следующее. Описательные статистики сильно искажаются при наличии в выборке данных выбросов - грубых (ошибочных), сильно отличающихся от основной массы наблюдений. Поэтому необходимо провести хотя бы визуальный анализ данных с целью обнаружения выбросов. В случае

* Нестеров Л.И. Новый этап расчетов показателей национального богатства в мире // Вопросы статистики. – 2001. – № 3. – С. 12.

выявления аномального выброса необходимо определить причину его возникновения (ошибки технического характера или аномальные уровни из-за воздействия объективных факторов). В случае точного установления, что выброс обусловлен ошибками технического характера необходимо аномальное наблюдение либо устранить, либо заменить, по крайней мере, средней арифметической двух соседних уровней ряда данных. При построении трендовой модели целесообразно также редактирование выбросов, обусловленных объективными факторами.

2.2. Корреляционный и регрессионный анализ

Задачей экономического исследования является уяснение природы экономического объекта, раскрытие механизма взаимосвязи между важнейшими его переменными. Решение этой задачи позволяет разработать и осуществить экономическую политику, необходимую для управления данным объектом. Следует отметить, что любая экономическая политика заключается в регулировании экономических переменных, и она должна базироваться на знании взаимосвязи ключевых переменных.

Изучение взаимосвязи экономических показателей проводится с использованием методов корреляционного и регрессионного анализа. На основе этих методов строятся однофакторные или многофакторные регрессионные модели экономических объектов.

В общем случае моделирование производится с использованием уравнений множественной регрессии. Множественная регрессия – это уравнение связи с несколькими независимыми переменными:

$$y = f(x_1, x_2, \dots, x_k),$$

где y – зависимая переменная (результативный признак);

x_1, x_2, \dots, x_k – независимые переменные (факторы).

Основными этапами построения регрессионной модели являются:

1. Отбор факторов для построения модели.
2. Корреляционный анализ данных и отбор существенных факторов.
3. Выбор вида модели и численная оценка ее параметров.
4. Проверка качества модели.

Отметим, что в ряде случаев по результатам проверки качества модели возможны повторные выполнения этапов 3, 4 с использованием уточненных наборов существенных факторов.

Выбор факторов, влияющих на исследуемый показатель, производится на основе качественного теоретико-экономического анализа. Для получения надежных оценок в модель не следует включать слишком много факторов. Обычно при оценке множественной регрессии требуется, чтобы число наблюдений, по крайней мере, в 3 раза превосходило число оцениваемых параметров. В противном случае параметры окажутся статистически незначимыми.

Однако теоретический анализ часто не позволяет однозначно ответить на вопрос о количественной взаимосвязи рассматриваемых признаков и целесообразности включения фактора в модель. Поэтому производится сравнительная оценка и отсев части факторов на основе коэффициентов парной корреляции.

Для этого составляется матрица коэффициентов парной корреляции, оценивающих тесноту связи каждого из факторов с результативным признаком и между собой. Коэффициенты парной корреляции вычисляются по формуле:

$$r_{y,x} = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \cdot \sum_{i=1}^N (x_i - \bar{x})^2}},$$

где \bar{y}, \bar{x} – соответствующие выборочные средние.

Связь считается достаточно сильной, если $|r_{y,x}| \geq 0,7$. Отметим, что этот коэффициент дает объективную оценку тесноты связи лишь при линейной зависимости переменных.

Для включения в модель отбираются те факторы, связь которых с зависимой переменной наиболее сильная. При этом анализируются также коэффициенты интеркорреляции (т.е. корреляция между факторами), что позволяет исключить из модели дублирующие факторы. Считается, что два фактора коллинеарны (т.е. находятся между собой в линейной зависимости), если коэффициент корреляции больше 0,7.

Одним из условий построения множественной регрессии является независимость действия факторов. Коллинеарность факторов нарушает это условие. Если факторы коллинеарны, то они дублируют друг друга. Поэтому один фактор необходимо исключить. Предпочтение при этом отдается не тому фактору, который более тесно связан с результативным признаком, а тому фактору, который при достаточно тесной связи с результирующим имеет наименьшую тесноту связи с другими факторами.

Парные коэффициенты корреляции обнаруживают лишь коллинеарность между парами факторов. Возможна также мультиколлинеарность факторов, когда более чем два фактора связаны между собой линейной зависимостью, т.е. имеет совокупное воздействие факторов друг на друга. В результате нельзя оценить воздействие на результативный признак каждого фактора в отдельности.

Включение в модель мультиколлинеарных факторов нежелательно, так как:

- затрудняется интерпретация модели, параметры теряют экономический смысл;
- оценки параметров ненадежны (в ряде случаев даже невозможны, например, при использовании для оценивания метода наименьших квадратов).

Модель оказывается не пригодной для анализа и прогнозирования. Поэтому необходимо устранение мультиколлинеарности, которое достигается либо исключением одного или нескольких факторов, либо использованием методов преобразования факторов, уменьшающих или исключаящих межфакторную корреляцию.

При выборе вида модели часто используются следующие уравнения регрессии:

1) линейная: $y = a_0 + a_1x_1 + a_2x_2 + \dots + a_kx_k$;

2) степенная $y = a_0x_1^{a_1}x_2^{a_2}\dots x_k^{a_k}$.

На практике наибольшее применение находит линейная модель регрессии.

Для оценивания параметров уравнения регрессии обычно используется метод наименьших квадратов, при котором минимизируется сумма квадратов остатков регрессии $s = \sum_{i=1}^N e_i^2$, где $e_i = y_i - \hat{y}_i$. Здесь \hat{y}_i - это оценка значения y_i по модели.

Качество модели оценивается на основе анализа остатков регрессии. Целесообразно провести хотя бы визуальный анализ графика остатков на наличие какой-то зависимости, не учтенной в модели, и аномальных наблюдений – выбросов.

Оценка качества построенной модели проводится также с использованием стандартной ошибки оценки и коэффициента детерминации.

Стандартная ошибка оценки определяется по формуле

$$S_e = \sqrt{\frac{\sum_{i=1}^N e_i^2}{N - k - 1}}.$$

где k – число факторов.

Очевидно, чем ниже стандартная ошибка оценки, тем лучше аппроксимирует регрессионная зависимость исходные данные.

Коэффициент детерминации R^2 вычисляется по формуле

$$R^2 = 1 - \frac{\sum_{i=1}^N e_i^2}{\sum_{i=1}^N (y_i - \bar{y})^2} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad 0 \leq R^2 \leq 1.$$

Можно также оперировать коэффициентом множественной корреляции (индексом корреляции) R , являющимся корнем квадратным из коэффициента детерминации.

Коэффициент детерминации является весьма важной характеристикой качества построенной регрессии, он показывает долю общего разброса (относительно выборочного среднего зависимой переменной), которая объясняется ею.

В множественной регрессии добавление дополнительных факторов увеличивает коэффициент детерминации. Поэтому целесообразно использование скорректированного (нормированного) коэффициента детерминации, определяемого по формуле:

$$\bar{R}^2 = 1 - (1 - R^2) \frac{N - 1}{N - k - 1}.$$

Для проверки значимости модели регрессии в целом и показателя тесноты связи R используется F -критерий Фишера

$$F_r = \frac{R^2}{1 - R^2} \cdot \frac{N - k - 1}{k}.$$

Проверка производится с использованием табличного (критического) значения критерия $F_{кр}$ при значениях степеней свободы $\nu_1 = k$, $\nu_2 = N - k - 1$ и уровне значимости α . Табличное значение F -критерия Фишера – это максимально возможное значение критерия под влиянием случайных факторов при данных степенях свободы и уровне значимости. Здесь уровень значимости α – это вероятность отвергнуть гипотезу о статистической незначимости уравнения регрессии и показателя тесноты связи при условии, что она верна. Обычно α принимается равной 0,05 или 0,01.

Если $F_r > F_{кр}$, то уравнение в целом и величина R признается статистически значимой. Отметим, что обычно на практике модель считается пригодной для прогнозирования при $R > 0,7$.

Значимость отдельных коэффициентов регрессии определяется с использованием t -критерия Стьюдента, определяемого как отношение коэффициента к его стандартной ошибке. По этому статистическому критерию проводится проверка гипотезы о равенстве нулю коэффициента уравнения. Если расчетное значение t -критерия с $(N - k - 1)$ степенями свободы превосходит его табличное значение при заданном уровне значимости, коэффициент считается значимым. В противном случае фактор, соответствующий этому коэффициенту, следует исключить из модели. Таким образом, производится окончательный отбор факторов, включаемых в модель.

Рассмотрим технологию решения задач корреляционного и регрессионного анализа с помощью *Пакета анализа MS Excel*. Основными средствами решения этих задач являются инструменты **Корреляция** и **Регрессия**. Инструмент **Корреляция** позволяет получить матрицу парных корреляций. С помощью инструмента **Регрессия**, помимо результатов регрессионной статистики, дисперсионного анализа и доверительных интервалов, можно получить остатки и графики подбора линии регрессии, остатков и нормальной вероятности.

Порядок действий при решении задач многофакторного регрессионного анализа следующий:

1. Активизируйте инструмент **Корреляция** командой *Данные* → *Анализ данных* → *Корреляция*. Диалоговое окно инструмента представлено на рис. 10.

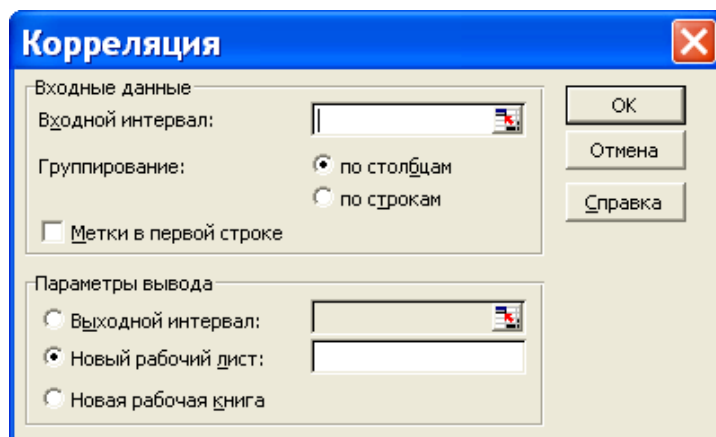


Рисунок 10. Диалоговое окно инструмента **Корреляция**

2. В диалоговом окне **Корреляция** введите следующую информацию:
- Входной интервал - диапазон ячеек, содержащих исходные данные;
 - Группирование - по столбцам или строкам;
 - Метки в первой строке – флажок (в случае выделения заголовков столбцов);
 - Параметры вывода – установить переключатель в выбранном варианте (целесообразно новый рабочий лист).

Щелкните на кнопке *OK* диалогового окна.

3. Выполните анализ матрицы коэффициентов парной корреляции и проведите на основе его результатов отбор значимых независимых переменных.

4. Сформируйте новую таблицу, включающую значения зависимой переменной и значимых независимых переменных на новом рабочем листе путем копирования данных из исходной таблицы.

5. Активизируйте инструмент **Регрессия** командой *Данные → Анализ данных → Регрессия*. Диалоговое окно инструмента представлено на рис. 11.

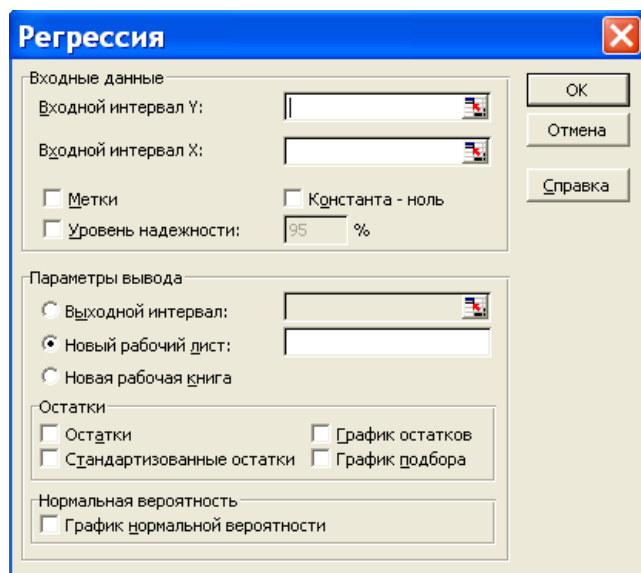


Рисунок 11. Диалоговое окно инструмента Регрессия

6. Заполните диалоговое окно ввода данных и параметров вывода:
- Входной интервал Y – диапазон, содержащий значения зависимой переменной;
 - Входной интервал X – диапазон, содержащий значения независимых переменных;
 - Метки – флажок (первая строка содержит название столбцов);
 - Константа – ноль – флажок (отсутствие свободного члена в уравнении);
 - Выходной интервал – диапазон для вывода результатов (достаточно указать левую верхнюю ячейку);
 - Новый рабочий лист – произвольное имя нового листа.

7. Если необходимо получить дополнительную информацию и графики остатков, установите соответствующие флажки в диалоговом окне. Щелкните по кнопке *OK*.

8. Проверьте качество модели с помощью вычисленного значения F -критерия Фишера, используя его минимальный уровень значимости для проверки гипотезы о значимости регрессии. Если полученный уровень значимости (**Значимость F**) меньше уровня $\alpha = 0,05$, то модель статистически значима.

9. Проверьте адекватность уравнения регрессии по величине коэффициента детерминации R^2 . Чем ближе значение R^2 к 1, тем лучше модель описывает исходные данные.

10. Проанализируйте коэффициенты регрессии на значимость, сравнивая **P -значения** (они соответствуют минимальным уровням значимости полученных значений t -критериев Стьюдента для проверки гипотезы о равенстве значений коэффициентов нулю) и выбранную величину уровня значимости $\alpha = 0,05$. Если полученный уровень значимости (**P -значение**) для коэффициента меньше этого уровня, то он статистически значим.

11. Проведите анализ результатов исследований качества модели. В случае статистической незначимости модели в целом и ее отдельных коэффициентов сформируйте новый набор существенных факторов и повторно выполните этапы 4-11.

12. Сравните расчетные (предсказанные) значения зависимой переменной с исходными значениями путем анализа остатков.

13. Выполните визуальный анализ качества модели с использованием графиков остатков, подбора регрессии и примете решение об ее пригодности для моделирования и прогнозирования изучаемого экономического процесса.

В качестве примера рассмотрим решение задачи корреляционного и регрессионного анализа данных о деятельности крупнейших компаний США в 1996 г.

Исходные данные задачи* и результат корреляционного анализа (часть матрицы коэффициентов парной корреляции) приведены на рис. 12.

Линейное уравнение множественной регрессии, включающее все факторы, имеет следующий вид:

$$y = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4.$$

Анализ матрицы коэффициентов парной корреляции показывает, что наиболее тесную связь с результирующим признаком y имеют следующие факторы (в порядке убывания тесноты связи):

- чистый доход x_1 ;
- численность служащих x_3 ;
- оборот капитала x_2 .

Фактор x_4 имеет слабую связь с результирующим признаком y и не включается в набор существенных факторов.

Наблюдается также тесная межфакторная связь между факторами x_1 , x_2 и x_3 (факторы коллинеарны между собой). Поэтому целесообразно исключить из уравнения множественной регрессии факторы x_2 и x_3 как малоинформативные, недостаточно статистически надежные.

* Практикум по эконометрике/ Под ред. И.И. Елисеевой. – М.: Финансы и статистика, 2002. – С. 92.

В результате корреляционного анализа приходим к следующей структуре регрессионной модели:

$$y = a_0 + a_1x_1.$$

№ п/п	Чистый доход, млрд. долл.	Оборот капитала, млрд. долл.	Использованный капитал, млрд. долл.	Численность служащих, тыс. чел.	Рыночная капитализация, млрд. долл.
	y	x1	x2	x3	x4
1	0,9	31,3	18,9	43	40,9
2	1,7	13,4	13,7	64,7	40,5
3	0,7	4,5	18,5	24	38,9
4	1,7	10	4,8	50,2	38,5
5	2,6	20	21,8	106	37,3
6	1,3	15	5,8	96,6	26,5
7	4,1	137,1	99	347	37
8	1,6	17,9	20,1	85,6	36,8
9	6,9	165,4	60,6	745	36,3
10	0,4	2	1,4	4,1	35,3
11	1,3	6,8	8	26,8	35,3
12	1,9	27,1	18,9	42,7	35
13	1,9	13,4	13,2	61,8	26,2
14	1,4	9,8	12,6	212	33,1
15	0,4	19,5	12,2	105	32,7
16	0,8	6,8	3,2	33,5	32,1
17	1,8	27	13	142	30,5
18	0,9	12,4	6,9	96	29,8
19	1,1	17,7	15	140	25,4
20	1,9	12,7	11,9	59,3	29,3
21	-0,9	21,4	1,6	131	29,2
22	1,3	13,5	8,6	70,7	29,2
23	2	13,4	11,5	65,4	29,1
24	0,6	4,2	1,9	23,1	27,9
25	0,7	15,5	5,8	80,8	27,2

Матрица коэффициентов парной корреляции

	y	x1	x2	x3	x4
y	1				
x1	0,8480292	1			
x2	0,763298	0,8977439	1		
x3	0,8295679	0,9115047	0,712513648	1	
x4	0,2689771	0,2486806	0,348493848	0,11519554	1

Рисунок 12. Исходные данные и результат применения инструмента Корреляция

Вычисление параметров этого линейного уравнения парной регрессии проводится с помощью инструмента **Регрессия**. Часть результатов вычислений представлена на рис. 13.

Регрессионная статистика	
Множественный R	0,848
R-квадрат	0,7192
Нормированный R-квадрат	0,7069
Стандартная ошибка	0,7792
Наблюдения	25

Дисперсионный анализ					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>
Регрессия	1	35,7563	35,7563	58,8953	8,68526E-08
Остаток	23	13,9637	0,60712		
Итого	24	49,72			

Рисунок 13. Результаты регрессионной статистики и дисперсионного анализа

Параметры уравнения регрессии представлены в табл.4.

Таблица 4

	Y-пересечение	x1
Коэффициенты	0,756294811	0,031503026
Стандартная ошибка	0,187755993	0,00410499
t-статистика	4,028072807	7,674325418
P-Значение	0,000524651	8,68526E-08
Нижние 95%	0,367892475	0,023011219
Верхние 95%	1,144697146	0,039994832

В результате вычислений имеем следующее уравнение регрессии:

$$y = 0,756294811 + 0,031503026x_1.$$

Коэффициенты уравнения статистически значимы и надежны, так как их **P-значения** меньше 0,05. Отметим, что при включении в уравнение регрессии всех факторов коэффициенты оказываются незначимыми (**P-значения** всех коэффициентов оказываются выше даже уровня 0,1).

Оценка надежности уравнения регрессии в целом оказывается достаточно высокой, так как значение параметра **Значимость F** меньше 0,05. На это указывают также значения скорректированного и нескорректированного коэффициентов детерминации в рамках регрессионной статистики.

В итоге решения задачи можно сделать следующий вывод: многофакторная линейная регрессионная модель содержит лишь один существенный фактор x_1 ,

поэтому можно ограничиться линейным уравнением парной регрессии. Это относительно простое уравнение оказывается пригодным для анализа и прогнозирования.

В заключение отметим, что углубленный анализ адекватности регрессионной модели требует более детального анализа остатков. Согласно общим предположениям регрессионного анализа, остатки должны вести себя как независимые (в действительности почти независимые), одинаково распределенные случайные величины. В классических методах регрессионного анализа предполагается также нормальный закон распределения остатков. Используемые в инструменте **Регрессия** статистические критерии оценки качества модели базируются именно на этом предположении.

На практике в большинстве случаев целесообразна проверка остатков на независимость с использованием соответствующих статистических критериев, например критерия Дарбина-Уотсона. Эффективные методы имеются также для проверки нормальности распределения. Однако детальная проверка гипотезы о нормальности распределения требует довольно значительных объемов выборки (как минимум, порядка сотни наблюдений), что часто ограничивает возможность проведения такого исследования.

2.3. Анализ и прогнозирование временных рядов

Временной ряд (ВР) – это последовательность упорядоченных по времени числовых показателей, характеризующих уровень состояния и изменения изучаемых явлений.

Данные типа ВР широко распространены в экономике. В качестве примера можно привести ежедневные цены на акции, курсы валют, ежедневные, месячные и годовые объемы продаж и т.п.

Исследованием ВР занимается самостоятельная и весьма обширная область статистики – анализ ВР. Круг задач, решаемых в рамках этой науки, весьма обширен. Ограничимся рассмотрением решения задачи прогнозирования ВР на основе экстраполяции тренда, которая часто встречается в прикладных задачах в области экономики.

Экстраполяция – это распространение выявленной при анализе ВР закономерности развития изучаемого явления на будущее. При этом часто используется закономерность изменения основной тенденции (тренда).

Тренд ВР – это плавно изменяющаяся компонента, описывающая влияние долговременных факторов, эффект которых сказывается постепенно. В качестве примеров таких факторов в экономике можно привести рост потребления и изменения его структуры, изменение демографических характеристик популяции и т.д.

Тренд предпочитают описывать с помощью гладких кривых, задающихся в достаточно простом аналитическом виде. Следует отметить, что анализ ВР обычно начинается с выделения именно этой компоненты. Выделение тренда позволяет перейти к дальнейшей идентификации других компонент ряда, например, сезонных и циклических.

Решение задач прогнозирования по тренду выполняется в следующей последовательности:

1. Предварительный анализ ВР и формирование набора моделей тренда для прогнозирования.
2. Оценка погрешностей моделей.
3. Получение прогнозов и графический анализ результатов.
4. Выбор лучшей модели и соответствующего прогноза.
5. Содержательный анализ полученного прогноза.

Относительно просто решение этой задачи в среде MS Excel достигается с использованием средства *Мастер диаграмм*. Типы линий тренда, используемые в этом средстве MS Excel, приведены в табл. 5.

Таблица 5

<i>Тип зависимости</i>	Уравнение
Линейная	$y = a_0 + a_1x$
Полиномиальная	$y = a_0 + a_1x + a_2x^2 + \dots + a_kx^k, k = \overline{2,6}$
Логарифмическая	$y = a_0 + a_1 \ln x$
Экспоненциальная	$y = a_0 e^{a_1x}$
Степенная	$y = a_0 x^{a_1}$

Технология решения рассматриваемой задачи в среде MS Excel (без детализации шагов построения графика ВР) может быть представлена следующей последовательностью действий:

1. Постройте график ВР с помощью *Мастера диаграмм*.
2. Проведите визуальный анализ графика. Добавьте *линию тренда* в диаграмму следующей последовательностью действий:
 - выделите линию графика;
 - выполните команду *Макет - Линия тренда – Дополнительные параметры линии тренда*;
 - в диалоговом окне выберите вкладку *Параметры линии тренда*, щелкните на требуемом типе тренда (при выборе типа *Полиномиальная* введите также значение степени от 2 до 6 в поле *Степень*);
 - установите флажки *Показывать уравнение на диаграмме* и *Величина достоверности аппроксимации*, в соответствующих полях введите количество периодов прогнозирования, которые будут добавлены к линии тренда впереди;
 - щелкните на кнопке *ОК* для завершения процесса создания линии тренда.
3. Выполните шаг 2 для других выбранных типов тренда.
4. Выберите лучшие модели тренда на основе анализа величин достоверности аппроксимации (модели с наибольшими значениями R^2).
5. Проведите визуальный анализ результатов прогноза по лучшим моделям и выберете наилучшую модель для прогноза.
6. Занесите параметры уравнения наилучшей модели тренда в диапазон ячеек соответствующей размерности рабочего листа.

7. Вычислите теоретические значения ВР по трендовой модели на анализируемом периоде и периоде прогнозирования.

8. Проведите содержательный анализ прогнозных значений ВР с целью выявления возможных противоречий известным фактам и сложившимся к настоящему времени представлениям о характере развития в периоде прогнозирования.

Остановимся подробнее на отдельных шагах представленного укрупненного алгоритма решения задачи.

При выборе типа тренда не следует выбирать *Скользящее среднее*, так как он не позволяет получить аналитическое описание, необходимое для прогнозирования. Это тип тренда можно использовать для визуального анализа ВР с целью предварительного выбора вида трендовой модели.

Вычисление значений уровней тренда проводится по полученной модели тренда $f(t)$, поставляя в него в качестве аргумента значения соответствующего условного времени:

$$\hat{y}_t = f(t), \quad t = \overline{1, N},$$

$$\hat{y}_{N+l}^* = f(N+l), \quad l = \overline{1, L},$$

где: \hat{y}_t – значение уровня тренда;

\hat{y}_{N+l}^* – прогнозируемый уровень;

N – количество уровней ВР;

l – период упреждения.

Отметим, что оценки параметров уравнения тренда, вычисляемые средствами *Мастера диаграмм*, являются оценками метода наименьших квадратов. В качестве показателя достоверности аппроксимации используется величина R^2 – коэффициент детерминации. При сравнении моделей с разным количеством параметров целесообразно оперировать скорректированным коэффициентом детерминации.

В качестве примера рассмотрим задачу прогноза объема продаж товара фирмы на 2003 и 2004 годы, используя метод экстраполяции ВР по тренду. Данные о продажах за последние 7 лет (в тыс. долларах США) приведены в табл. 6.

Таблица 6

Годы	Объем продаж, тыс. долларов США	
	исходное значение	трендовое значение
1996	212	219
1997	280	274
1998	317	307
1999	332	330
2000	347	348
2001	356	363
2002	375	375
2003		386
2004		395

Отметим, что трендовые значения в таблице округлены до точности представления исходных данных.

Для решения задачи с помощью *Мастера диаграмм MS Excel* построим график исходного ВР. Визуальный анализ графика показывает отсутствие выбросов, что позволяет провести построение трендовой модели без редактирования данных.

Исследуем различные модели тренда и выберем из них лучшую для построения прогноза по наибольшей величине коэффициента детерминации R^2 . В рассматриваемом примере по разным уравнениям тренда получены следующие результаты:

Таблица 7

Вид уравнения	R^2	Скорректированное R^2
линейный	0,8737	0,8484
логарифмический	0,9878	0,9854
степенной	0,9673	0,9608
экспоненциальный	0,8167	0,78

Использованные модели тренда характеризуются двумя параметрами, поэтому требование, как минимум, трехкратного превышения количества уровней ВР над числом оцениваемых параметров выполняется. Отметим также, что здесь нет необходимости в дополнительном анализе значений скорректированных коэффициентов детерминации, так как выбранные регрессионные модели характеризуются одинаковым количеством оцениваемых параметров.

Анализ результатов показывает, что исходные данные лучше всего описывает логарифмическая модель тренда, описываемая уравнением

$$f(t) = 219,11 + 80,378 \ln(t), \quad t = \overline{1,7}.$$

Следовательно, расчет прогнозных значений следует вести по этому уравнению. Вычисленные трендовые значения ВР также приведены в табл. 6.

Графики ВР и его логарифмического тренда с прогнозными значениями представлены на рис. 14.

Визуальный анализ графиков показывает, что выбранный тип тренда хорошо описывает исходный ряд и прогнозные значения не противоречат общей тенденции продаж. Полученной кривой и ее уравнением можно воспользоваться для построения прогноза.

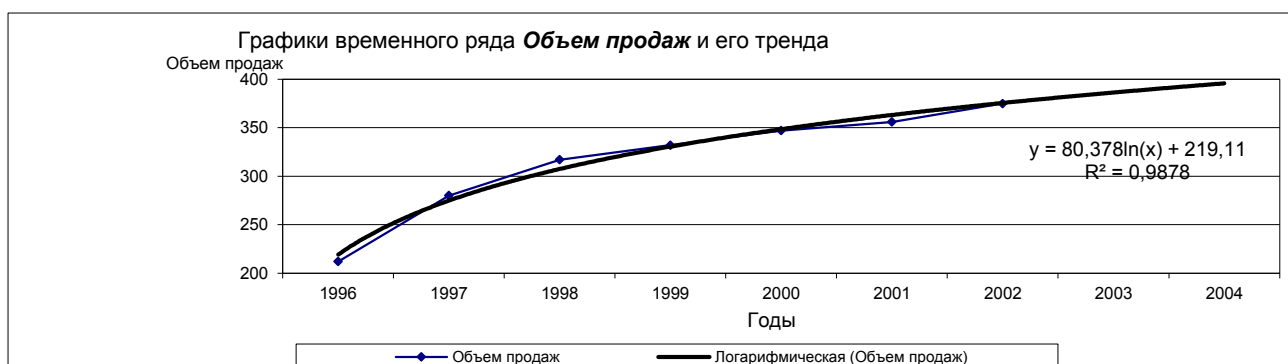


Рисунок 14.

В итоге получены следующие результаты. Прогноз объема продаж товара фирмы может составить:

- 386 тыс. долларов США на 2003 г.;
- 395 тыс. долларов США на 2004 г.

Рассмотренный метод прогнозирования ВР по тренду достаточно прост в реализации средствами MS Excel и часто используется в прикладных экономических задачах для получения точечных прогнозов.

В заключение отметим следующее. Прогноз экономических показателей на базе трендовых моделей основывается на допущении, что закономерности их изменения будут действовать на определенном отрезке времени в будущем. Однако такое условие в реальности часто нарушается. Поэтому прогнозирование по тренду в большинстве случаев можно применять с упреждением на один, максимум на два интервала ВР. В случае среднесрочного прогнозирования целесообразнее использование многофакторных регрессионных моделей. Многофакторные модели позволяют точнее отразить процесс формирования экономических ВР, чем трендовые однофакторные модели.

Контрольные задания по разделу 2.

Задание 2.1. Для статистических рядов данных **Y**, **X2** и **X4**, приведенных в табл. 8 получить сводную таблицу основных статистических характеристик (описательные статистики). Провести анализ результатов.

Таблица 8

Y	X1	X2	X3	X4	X5
126	0	4	15	17	100
137	1	4,8	14,8	17,3	98,4
148	2	3,8	15,2	16,8	101,2
191	3	8,7	15,5	16,2	103,5
274	4	8,2	15,5	16	104,1
370	5	9,7	16	18	107
432	6	14,7	18,1	20,2	107,4
445	7	18,7	13	15,8	108,5
367	8	19,8	15,8	18,2	108,3
367	9	10,6	16,9	16,8	109,2
321	10	8,6	16,3	17	110,1
307	11	6,5	16,1	18,3	110,7
331	12	12,6	15,4	16,4	110,3
345	13	6,5	15,7	16,2	111,8
364	14	5,8	16	17,7	112,3
384	15	5,7	15,1	16,2	112,9

Задание 1.2. На основе статистических данных, приведенных в табл. 8, построить линейную многофакторную регрессионную модель и проверить ее качество. Провести интерпретацию результатов.

Задание 1.3. По данным о недельных объемах продаж торгового предприятия по одной группе товаров (\$ США), представленных в табл. 9, составить прогноз на две недели по тренду.

Таблица 9

Неделя	1	2	3	4	5	6	7	8	9	10	11	12
Объем продаж	175	263	326	297	247	298	366	297	420	441	453	399

Тесты по разделу 2.

- Статистическую обработку данных можно проводить с использованием:
 - табличных процессов;
 - математических пакетов общего назначения;
 - статистических пакетов;
 - пакетов для решения задач математического программирования.
- При статистической обработке данных с использованием компьютерных технологий целесообразно придерживаться:
 - режима пакетной обработки данных;
 - режима автоматической обработки данных;
 - графически-ориентированного подхода;
 - мультикластерного подхода.
- Решение задач статистического анализа и прогнозирования в среде MS Excel поддерживается следующими средствами:
 - встроенными функциями категории *Статистические*;
 - надстройкой *Пакет анализа*;
 - надстройкой *Мастером диаграмм*;
 - надстройкой *Анализ решения*.
- Окно выбора требуемого инструмента надстройки *Пакет анализа* открывается командой:
 - Данные* → *Анализ данных*;
 - Сервис* → *Пакет анализа*;
 - Вставка* → *Функция* → *Инструменты*;
 - Вид* → *Панели инструментов* → *Анализ данных*.
- Если статистическая функция является формулой массива, то ввод формулы необходимо завершить нажатием:
 - на клавишу F4;
 - на клавишу *Ctrl*;
 - на комбинацию клавиш *Ctrl+Shift+Enter*;
 - на комбинацию клавиш *Ctrl+Enter*.
- Инструменты надстройки *Пакет анализа* возвращают результаты в виде:
 - массива констант;
 - массива формул;
 - массива статистических функций;
 - в представленных вариантах нет правильного ответа.
- При изменении исходных данных, являющихся аргументами встроенных статистических функций MS Excel, результат пересчитывается:
 - в автоматическом режиме;
 - после нажатия на клавишу F9;

- в) после нажатия на клавиши *Ctrl+F9*; г) в режиме отложенного времени.
8. Автоматический перерасчет результата при изменении исходных данных отсутствует при использовании:
- а) статистических функций MS Excel;
 - б) инструментов *Пакета анализа*;
 - в) векторно-матричных функций MS Excel;
 - г) формулы массива.
9. Мастер диаграмм MS Excel позволяет получить аналитическое описание:
- а) линейной трендовой модели ряда данных;
 - б) нелинейной трендовой модели ряда данных;
 - в) остаточного ряда;
 - г) статистических характеристик ряда данных.
10. Статистические характеристики рядов данных можно получить с использованием:
- а) статистических функций MS Excel;
 - б) инструмента *Описательная статистика* Пакета анализа;
 - в) таблицы дисперсионного анализа инструмента *Регрессия* Пакета анализа;
 - г) надстройки *Статистика*.

Список литературы

1. Марков, Ю. Н. Табличный процессор MS Excel: Основы работы и применения в экономике : Учебное пособие / И. И. Исмаилов, Ю. Н. Марков .— 2010
2. Математика и экономико-математические модели: Учебник/С.В.Юдин - М.: ИЦ РИОР, НИЦ ИНФРА-М, 2016. - 374 с.: 60x90 1/16. - (Высшее образование: Бакалавриат) (Переплёт) ISBN 978-5-369-01409-7, 300 экз.
3. Сидорова, М.И. Экономико-математические модели в управленческом учете и анализе [Электронный ресурс] : Монография / М. И. Сидорова, А. И. Мастеров. — М.: Издательско-торговая корпорация «Дашков и К°», 2013. — 229 с. - ISBN 978-5-394-02330-9 - Режим доступа: <http://znanium.com/catalog.php?bookinfo=514585>
4. Хуснутдинов Р. Ш. Экономико-математические методы и модели: Учебное пособие / Р.Ш. Хуснутдинов. - М.: НИЦ Инфра-М, 2013. - 224 с.: 60x90 1/16. - (Высшее образование). (переплет) ISBN 978-5-16-005313-4.
5. Гетманчук А. В. Экономико-математические методы и модели [Электронный ресурс] : Учебное пособие для бакалавров / А. В. Гетманчук, М. М. Ермилов. - М. : Издательско-торговая корпорация «Дашков и К», 2013. - 188 с. - ISBN 978-5-394-01575-5.
6. Орлова И. В. Экономико-математическое моделирование: Практическое пособие по решению задач / И.В. Орлова. - 2-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ ИНФРА-М, 2014. - 140 с.: 60x88 1/16. (обложка) ISBN 978-5-9558-0107-0
7. Алексеев, Г. В. Численное экономико-математическое моделирование и оптимизация [Электронный ресурс] : учебное пособие / Г. В. Алексеев, И. И. Холявин. - СПб., 2011. - 209 с. - Режим доступа: <http://www.znanium.com/>
8. Машунин, Ю. К. Теория управления. Математический аппарат управления в экономике [Электронный ресурс] : учеб. пособие / Ю. К. Машунин. - М.: Логос, 2013. - 448 с. - (Новая университетская библиотека). - ISBN 978-5-98704-736-1.
9. Каплан, А. В. Решение экономических задач на компьютере [Электронный ресурс] / А. В. Каплан, В. Е. Каплан, М. В. Мащенко, Е. В. Овечкина. - М. : ДМК Пресс, 2008. - 600 с. : ил. - ISBN 5-94074-243-2.
10. Методы и модели принятия управленческих решений: Учебное пособие / Е.В. Бережная, В.И. Бережной. - М.: НИЦ ИНФРА-М, 2014. - 384 с.: 60x90 1/16 + (Доп. мат. znanium.com). - (Высшее образование: Бакалавриат). (п) ISBN 978-5-16-006914-2
11. Оптимизационные модели управления финансовыми ресурсами предприятия: Моногр. / А.В.Мищенко, Е.В.Виноградова - М.: ИЦ РИОР: НИЦ ИНФРА-М, 2013. - 337 с. + Vc.: 60x88 1/16. (Доп. мат. znanium.com). - (Науч. мысль). (о) ISBN 978-5-369-01152-2
12. Головицына М.В. Информационные технологии в экономике // Интернет-Университет Информационных Технологий - 2012

13. Киселев Г.М., Бочкова Р.В., Сафонов В.И. Информационные технологии в экономике и управлении (эффективная работа в MS Office 2007) // Дашков и К - 2013
14. Информатизация бизнес-процессов в Microsoft Excel 2010 // Саратовский государственный аграрный университет - 2014
15. Анеликова Л.А. Лабораторные работы по Excel // СОЛОН-ПРЕСС - 2010
16. Горбунова Т.Н., Журавлева Т.Ю. Автоматизированный лабораторный практикум по информатике. Освоение работы в MS Excel 2007 // Вузовское образование - 2014
17. Журавлева Т.Ю. Практикум по дисциплине «Бизнес-анализ с помощью Microsoft Excel» // Вузовское образование - 2014
18. Катаргин Н.В. Экономико-математическое моделирование в Excel // Вузовское образование - 2013
19. Пакулин В.Н. Решение задач оптимизации управления с помощью MS Excel 2010 // Интернет-Университет Информационных Технологий - 2012
20. Левина Н.С., Харджиева С.Б., Цветкова А.Л. MS Excel и MS Project в решении экономических задач // СОЛОН-ПРЕСС – 2010
21. Машунин, Ю. К. Теория управления : математический аппарат управления в экономике [Текст] : учеб. пособие / Ю. К. Машунин. – М.: Логос, 2013. – 448 с. – (Новая университетская библиотека). – ISBN 978-5-98704-736-1.
22. Седышев, В. В. Информационные технологии в профессиональной деятельности [Текст] : учеб. пособие / В. В. Седышев. — М.: ФГБОУ «Учебно-методический центр по образованию на железнодорожном транспорте, 2013. — 262 с. — ISBN 978-5-89035-660-4.