

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования**

«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»

Кафедра статистики и прикладной математики

**КОМПЬЮТЕРНЫЙ АНАЛИЗ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ДАННЫХ
(MS Excel, MS PROJECT)**

Методические указания
для студентов очной и заочной формы обучения

Краснодар 2011

УДК – 311:004.9(076)

ББК 60.6

К. 63

Рецензент: заведующий кафедрой информационных систем
доктор экономических наук, профессор
Е.В. Попова

Бондаренко П.С. Компьютерный анализ социально-экономических данных. Методические указания для студентов очной и заочной формы обучения / П.С.Бондаренко, А.М. Ляховецкий, Д.А. Крепышев, С.А. Кацко, А.Е. Сенникова, А.Е. Жминько – Краснодар: ФГОУ ВПО «КубГАУ». - 2011. - 81 с.

Методические указания и задания рассмотрены и рекомендованы к изданию кафедрой статистики и прикладной математики КубГАУ, протокол № 2 от «10» октября 2011 г.

Одобрены методической комиссией учетно-финансового факультета КубГАУ, протокол № ____ от «__» _____ 2011 г.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	4
ЛАБОРАТОРНАЯ РАБОТА № 1	5
1.1 <i>Анализ вариационных рядов в Excel</i>	5
1.2 Парная регрессия и корреляция	15
ЛАБОРАТОРНАЯ РАБОТА № 2	19
2.1 <i>Множественный корреляционно-регрессионный анализ в Excel 2007</i>	19
2.2 <i>Группировка в Excel 2007 (Сводные таблицы)</i>	24
ЛАБОРАТОРНАЯ РАБОТА №3	
Пакет анализа	30
ЛАБОРАТОРНАЯ РАБОТА № 4	
<i>Анализ временных рядов в Excel 2007</i>	49
ЛАБОРАТОРНАЯ РАБОТА № 5	
Финансовые вычисления в Excel 2007	60
ЛАБОРАТОРНАЯ РАБОТА № 6	
Управление проектами в MS Project	66
ПРИЛОЖЕНИЕ Основные показатели производства в сельскохозяйственных предприятиях Краснодарского края	79

ВВЕДЕНИЕ

Методические указания предназначены для выполнения лабораторных заданий и самостоятельной работы студентами очной и заочной форм обучения.

Цель заданий – оказать помощь студентам в овладении приемами и методами статистического анализа данных с использованием компьютера. Содержатся начальные сведения работы с часто употребляющимися средствами (MS EXCEL, Statistica, MS PROJECT – 2007).

Предлагаемые лабораторные работы охватывают стандартные разделы многомерного статистического анализа: метод визуализации, поиска зависимостей (дисперсионный анализ, корреляционно-регрессионный анализ и анализ временных рядов), классификации и методы снижения размерности изучаемого признакового пространства (факторный анализ).

Задания могут быть использованы при самостоятельном изучении курса. Для систематизации и закрепления изучаемого материала по каждой теме по каждой теме даны контрольные вопросы.

ЛАБОРАТОРНАЯ РАБОТА № 1

1.1 Анализ вариационных рядов в Excel

Цель работы: ознакомиться с возможностями методик анализа временных рядов, парного корреляционно-регрессионного анализа, получить навыки анализа данных в Excel.

Теоретические сведения

В реальных социально-экономических системах нельзя проводить активные эксперименты, поэтому данные обычно представляют собой наблюдения за происходящим процессом, например: курс валюты на бирже в течение месяца, урожайность пшеницы в хозяйстве за 30 лет, производительность труда рабочих за смену и т.д. Результаты наблюдений – это, в общем случае, ряд чисел, расположенных в беспорядке, который для изучения необходимо упорядочить (проранжировать).

Операция, заключенная в расположении значений признака по возрастанию, называется ранжированием опытных данных.

После операции ранжирования опытные данные можно сгруппировать так, чтобы в каждой группе признак принимал одно и то же значение, которое называется вариантом (X_i). Число элементов в каждой группе называется частотой варианта (n_i).

Размахом вариации называется число $W = X_{\max} - X_{\min}$,

где, X_{\max} – наибольший вариант;

X_{\min} – наименьший вариант.

Сумма всех частот равна определенному числу n , которое называется объемом совокупности:

$$\sum_{i=1}^k n_i = n_1 + n_2 + \dots + n_k = n$$

Отношение частоты данного варианта к объему совокупности называется относительной частотой или частостью этого варианта:

$$\hat{p} = \frac{n_i}{n},$$
$$\sum_{i=1}^k \hat{p}_i = \sum_{i=1}^k \frac{n_i}{n} = \frac{\sum_{i=1}^k n_i}{n} = \frac{n}{n} = 1$$

Последовательность вариант, расположенных в возрастающем порядке, называется вариационным рядом (вариация - изменение).

Вариационные ряды бывают дискретными и непрерывными. Дискретным вариационным рядом называется ранжированная

последовательность вариант с соответствующими частотами и (или) частотами.


Пример 1 В результате тестирования группа из 24 человек набрала баллы: 4, 0, 3, 4, 1, 0, 3, 1, 0, 4, 0, 0, 3, 1, 0, 1, 1, 3, 2, 3, 1, 2, 1, 2. Построить дискретный вариационный ряд.

Проранжируем исходный ряд, подсчитаем частоту и частость вариант: 0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4.

В результате получим дискретный вариационный ряд (табл.1).

Таблица 1 – Ранжированный ряд успеваемости студентов

Балл, x_i	Число студентов, n_i	Относительная частота, \hat{p}_i
0	6	6/24
1	7	7/24
2	3	3/24
3	5	5/24
4	3	3/24
Σ	24	1,000

В Excel. Проранжируем исходный ряд. Для этого введём все данные в диапазон A1:A24 и воспользуемся кнопкой  (Сортировка по возрастанию).

Подсчитаем частоту и частость вариант. Построим таблицу в диапазоне D2:G7 (рис.1). Рассмотрим два варианта подсчёта частот:

1) Выделим диапазон A1:A6 – в котором находятся нули. Щёлкнем в нижней правой части окна Excel правой кнопкой мыши и выберем в контекстном меню вид итога, который по умолчанию будет появляться в итоговой строке при выделении произвольного диапазона (см. рис.1) – количество. Таким образом, последовательно выделяя диапазоны с одинаковыми значениями вариант мы получим все частоты.

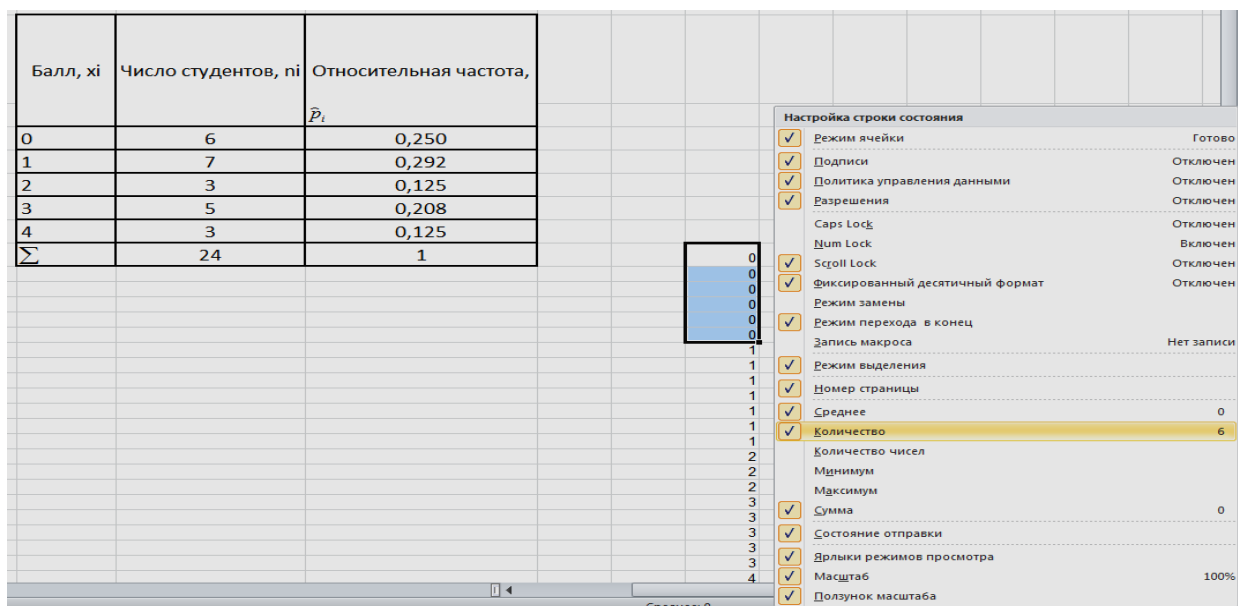


Рисунок 1 – Контекстное меню строки состояния

2) Выполним команду Данные – Анализ данных – Гистограмма. Заполним диалоговое окно в соответствии с рисунком 2.

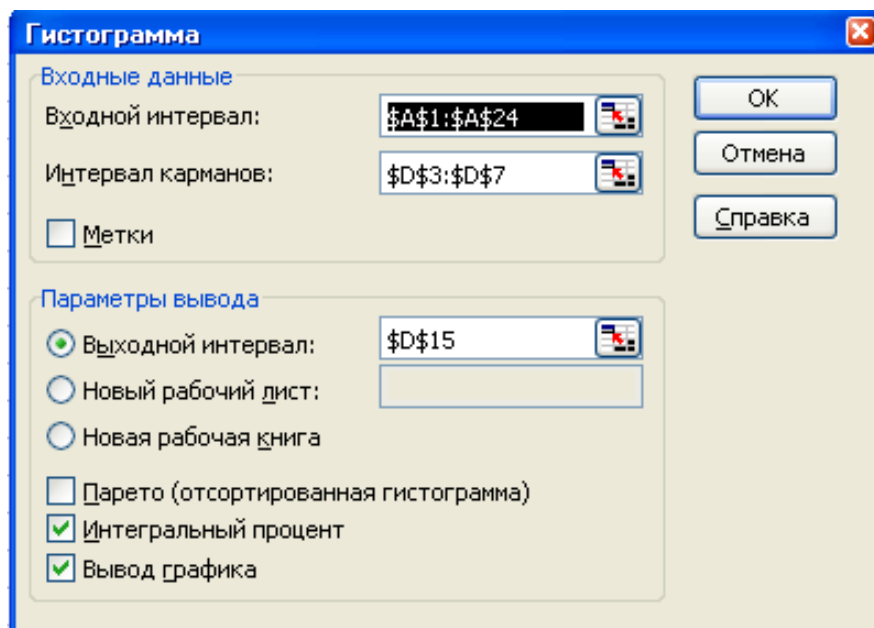


Рисунок 2 – Диалоговое окно инструмента пакета анализа Гистограмма

В результате получим таблицу с частотами вариантов и соответствующий график (рис.3).

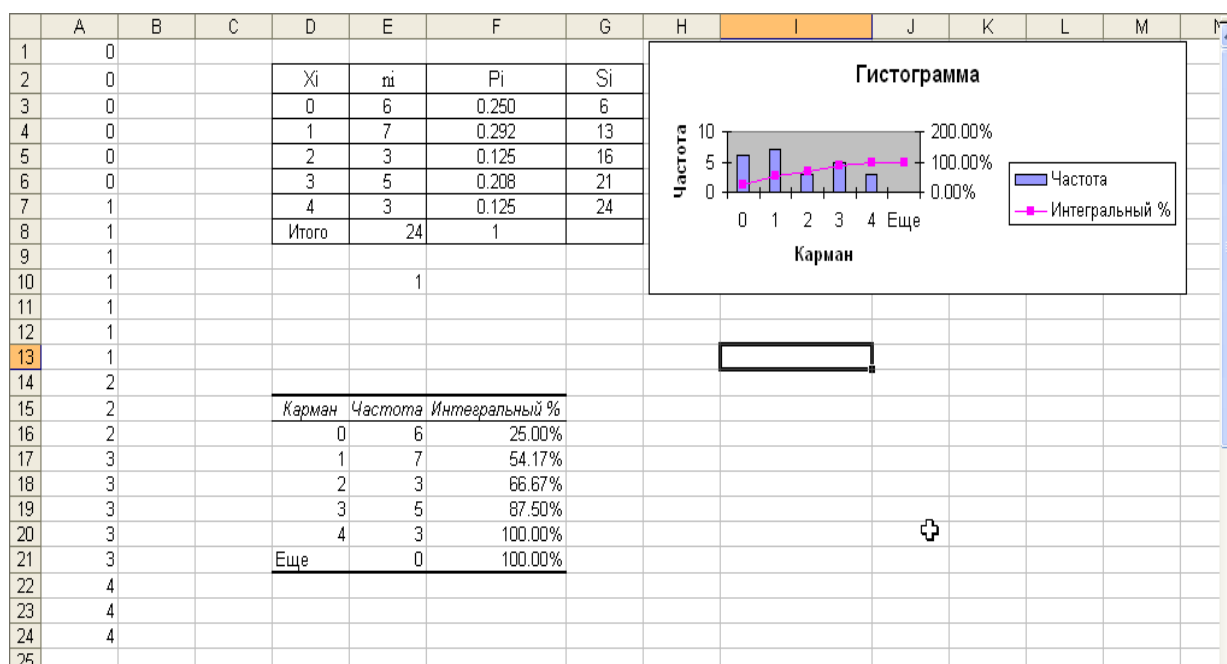


Рисунок 3 – Результаты применения инструмента Гистограмма

Найдём объём выборки, заполнив все частоты вариант в диапазоне E3:E7, выделим в его левой кнопкой мыши и щёлкнем по кнопке Σ (автосумма).

В ячейку F3 введём формулу «=E3/\$E\$8», за маркер заполнения (крест в правом нижнем углу ячейки) с помощью мыши скопируем до F7 и выберем кнопку автосумма, в результате мы получим частоты вариантов и их сумму (1). В ячейку G3 введём частоту варианта 0 – цифру 6 (или ссылку на ячейку её содержащую – E3), в ячейку G4 введём формулу «=G3+E4» и скопируем её до ячейки G7, в результате получим накопленные частоты. Таким образом, мы получили дискретный вариационный ряд.

Частоты необходимо округлить, но так, чтобы их сумма равнялась 1,000. Для этого выделим левой кнопкой мыши диапазон частот (F3:F7), щёлкнув по правой кнопке, откроем контекстное меню и выполним команду Формат ячеек – Числовой – Число знаков 3 – ОК. Преобразовав обозначения, получим дискретный вариационный ряд, изображённый в таблице 2.

Таблица 2 – Дискретный вариационный ряд

Балл, x_i	Число студентов, n_i	Относительная частота, \hat{p}_i	Сумма накопленных частот, S_i
0	6	0,250	6
1	7	0,292	13
2	3	0,125	16
3	5	0,208	21
4	3	0,125	24
Итого	24	1,000	-

Вариационные ряды изображают графически с помощью полигона и гистограммы.

Полигон частот - это ломаная, отрезки которой соединяют точки

$$(x_1; n_1), (x_2; n_2), \dots, (x_k; n_k).$$

Полигон относительных частот - это ломаная, отрезки которой соединяют точки:

$$(x_1; \frac{n_1}{n}), (x_2; \frac{n_2}{n}), \dots, (x_k; \frac{n_k}{n}).$$

Изобразим ряд графически. Для этого построим полигон частот с помощью мастера диаграмм - .

Выделим диапазон D2:E7, щёлкнем по кнопке мастера диаграмм и выберем диаграмму Точечная – Далее – Ряды в столбцах (рис.4).

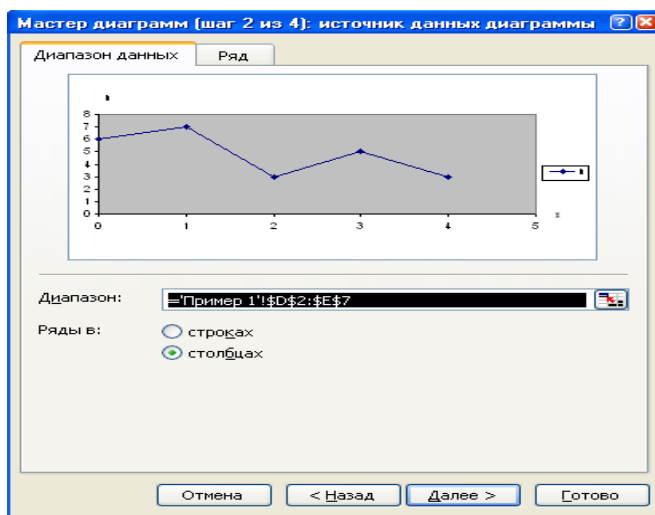
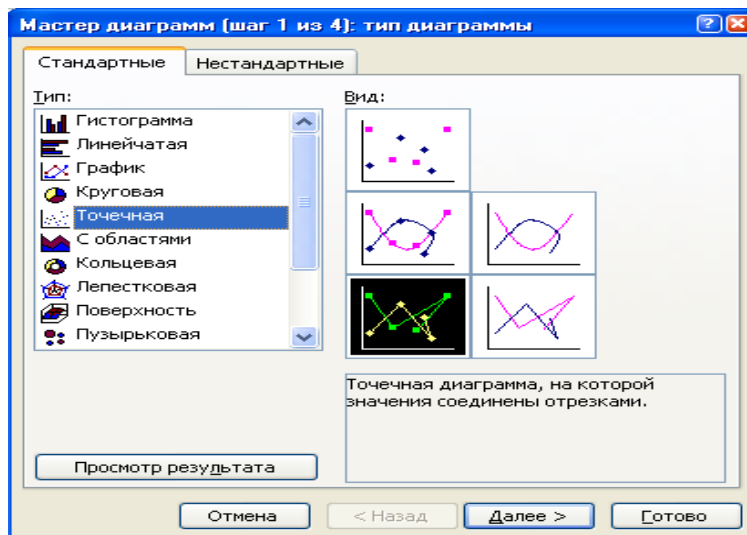


Рисунок 4 – Построение полигона – шаг 1 и 2

На третьем шаге, открыв вкладку «Заголовки» заполним её согласно рисунку 5, на вкладке Линии сетки отменим все линии. Перейдём к шагу 4 и

выберем Готово.

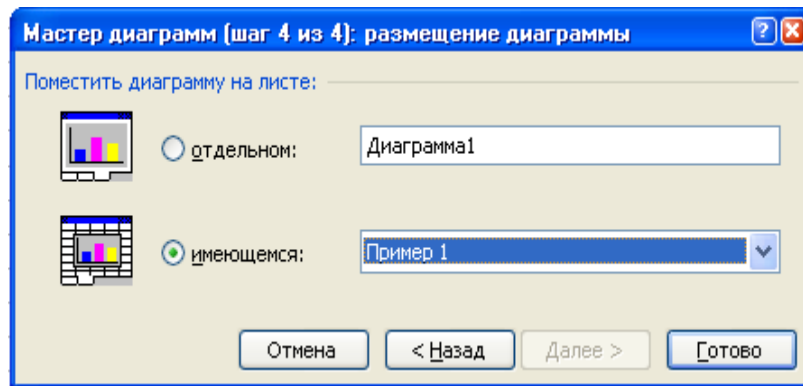
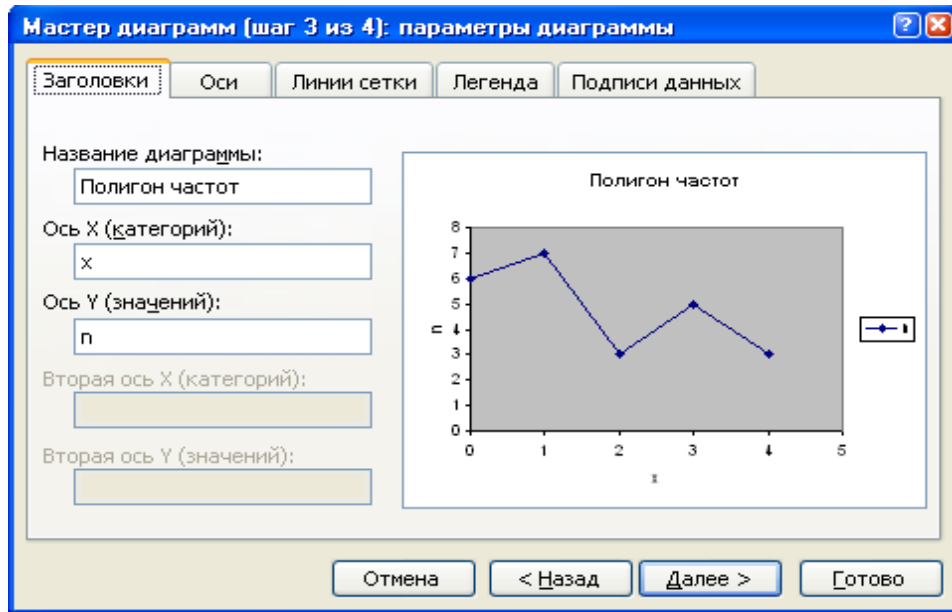


Рисунок 5 – Построение полигона – шаги 3 и 4

Полигон частот

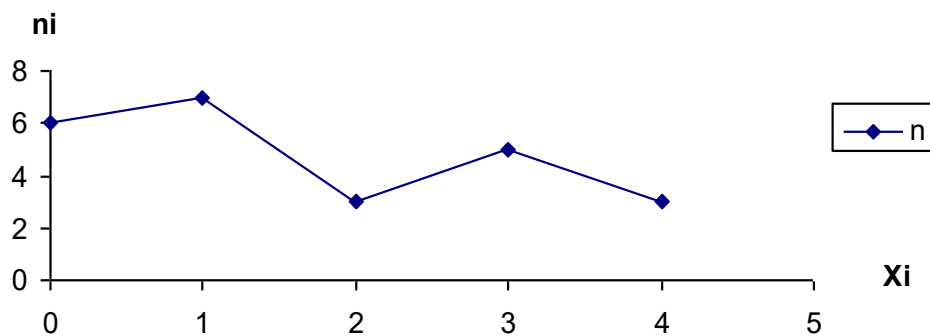


Рисунок 6 - Полигон частот

Получившийся график полигона частот желательно преобразовать, открыв с помощью контекстного меню, Формат области построения и Формат области диаграммы (уберите границы и заливку диаграммы). В результате получится полигон частот, изображённый на рисунке 6.

Постройте самостоятельно полигон относительных частот и кумуляту (рис. 7, 8). Для этого можно выделить в Excel всю таблицу (диапазон D2:G7, см. рис.3) и уже на графике удалить лишнее.

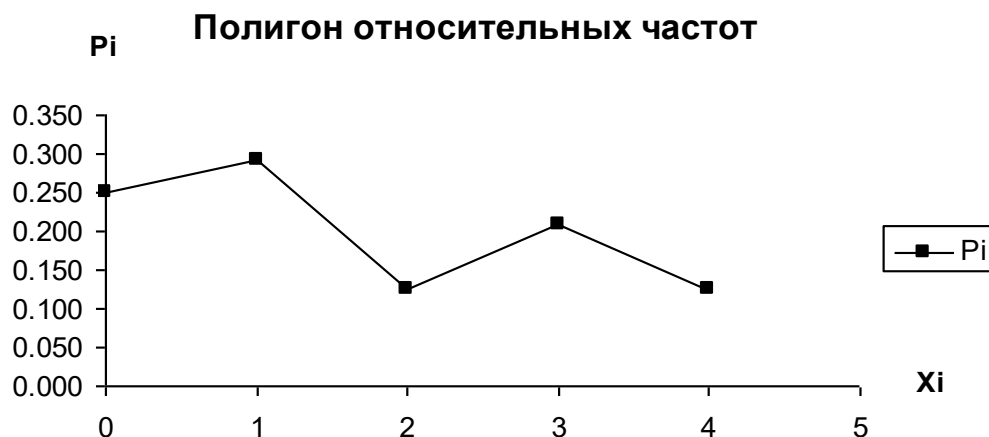


Рисунок 7 - Полигон относительных частот

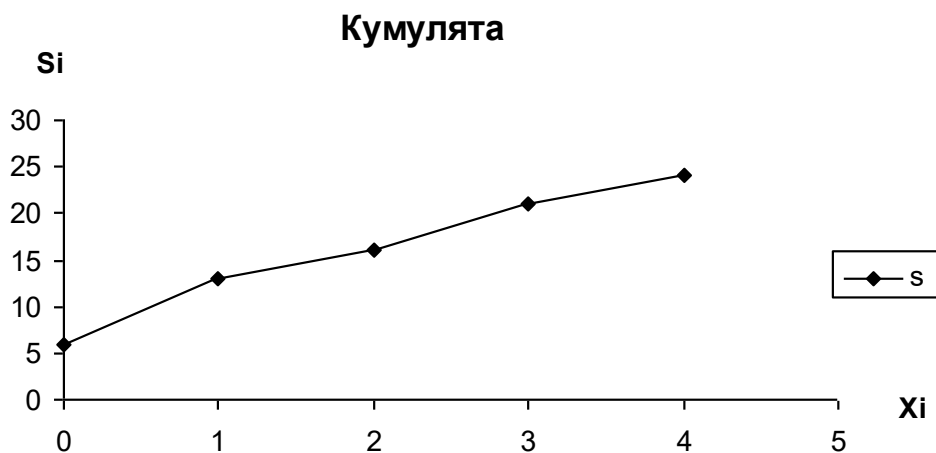


Рисунок 8 - Кумулята

Построение дискретного вариационного ряда нецелесообразно, если число значений признака велико. В этом случае следует построить интервальный вариационный ряд. Для построения такого ряда промежутки изменения признака разбиваются на ряд отдельных интервалов и подсчитывается количество значений величины в каждом из них.

Будем считать, что отдельные (частичные) интервалы имеют одну и ту же длину. Число интервалов (k), в случае нормально распределённой совокупности, можно определить по формуле Стерджесса:

$$k = 1 + 3,322\lg(n)$$

или приближённо: $k \in [6,12]$.

Пример 2 Пусть дан ряд распределения хозяйств по количеству рабочих на 100 га с/х угодий ($n=60$):

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

Построить интервальный вариационный ряд.

Для определения числа групп подставим значение $n=60$ в формулу Стерджесса:

$$k = 1 + 3,322\lg 60 \approx 6,907; \quad k = 7.$$

Длина частичного интервала определяется по формуле:

$$h = \frac{W}{k} = \frac{x_{\max} - x_{\min}}{k} = \frac{15 - 4}{7} \approx 1,6.$$

Построим интервальный вариационный ряд, для этого в качестве начального значения используем x_{\min} .

Разобьем интервал вариации признака X на $k=7$ частичных интервалов (табл. 3) с шагом $h=1,6$ (4,0; 5,6; 7,2; 8,8; 10,4; 12,0; 13,6; 15,2).

Таблица 3 – Группировка хозяйств по численности работников на 100 га сельхозугодий

Группы хозяйств по численности работников на 100га с/х угодий	Число хозяйств в группе (n_i)	Накопленное число хозяйств (S_i)	Относительная частота (\hat{P}_i)
4,00 - 5,60	5	5	5/60
5,61 - 7,20	17	22	17/60
7,21 - 8,80	9	31	9/60
8,81 - 10,40	15	46	15/60
10,41 - 12,00	10	56	10/60
12,01 - 13,60	1	57	1/60
13,61 - 15,20	3	60	3/60

Итого:	60	-	1,000
--------	----	---	-------

Далее подсчитаем количество рабочих на 100 га сельскохозяйственных угодий в каждом интервале с использованием инструмента Гистограмма пакета анализа (рис. 9, 10).

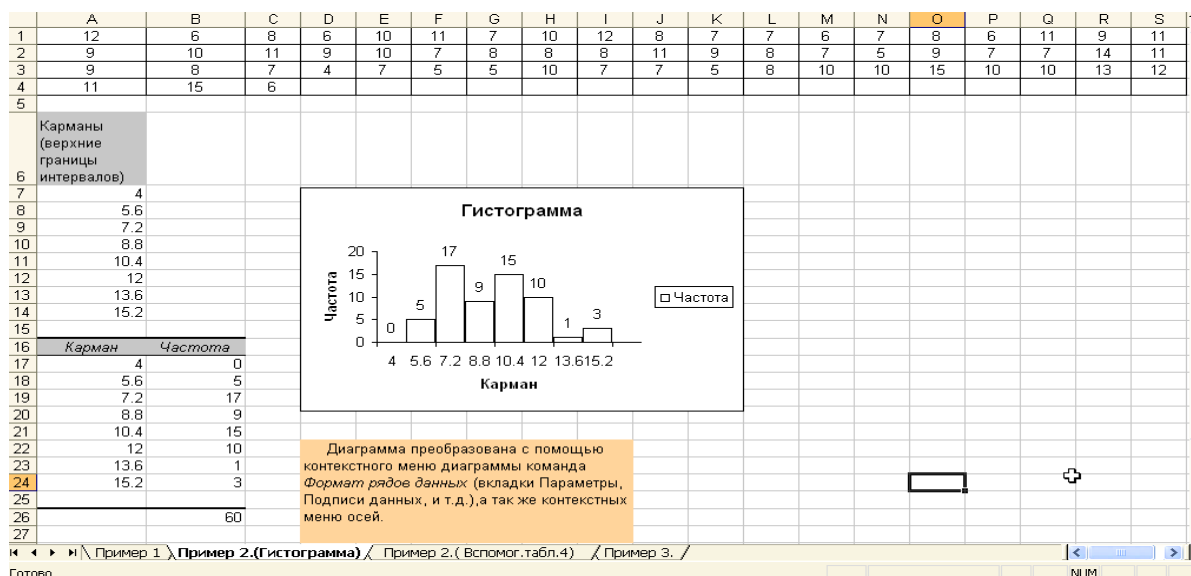


Рисунок 9 – Подсчёт частот и построение гистограммы для интервального вариационного ряда

Гистограммой частот называется фигура, состоящая из прямоугольников с основанием h и высотами n_i . Для гистограммы относительных частот в качестве высоты рассматривают n_i/n . Гистограмма относительных частот является аналогом дифференциальной функции случайной величины.

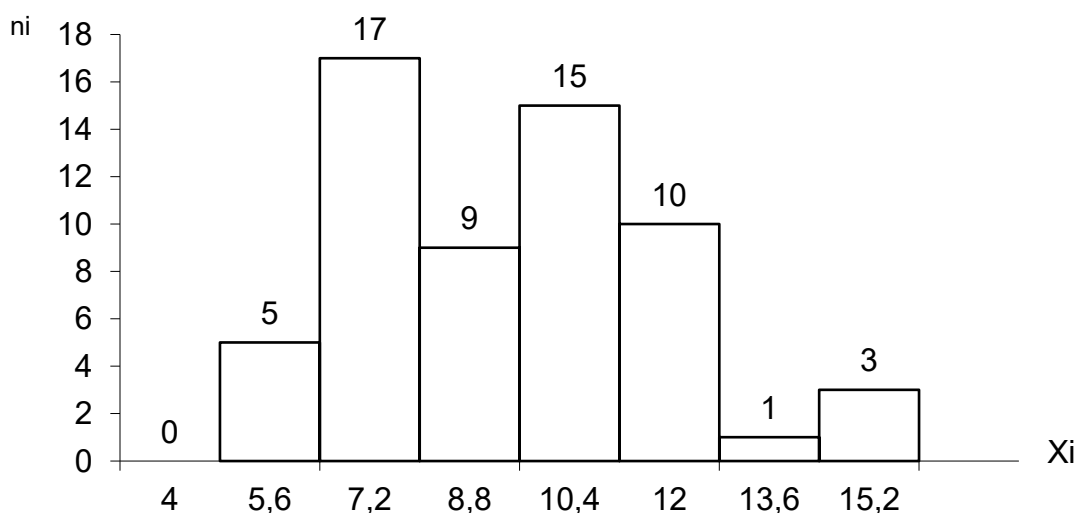


Рисунок 10 – Гистограмма частот

Построим гистограмму частот для примера 2. (рис.10). Гистограмма

построена в Excel (Пакет анализа - инструмент Гистограмма) подпись под прямоугольником означает верхнюю границу интервала, над прямоугольником – соответствующую частоту.

График гистограммы относительных частот можно получить из графика рисунка 10 сжатием в 60 раз вдоль оси ординат.

Для примера 2 можно подобно примеру 1, построить полигон частот и полигон относительных частот.

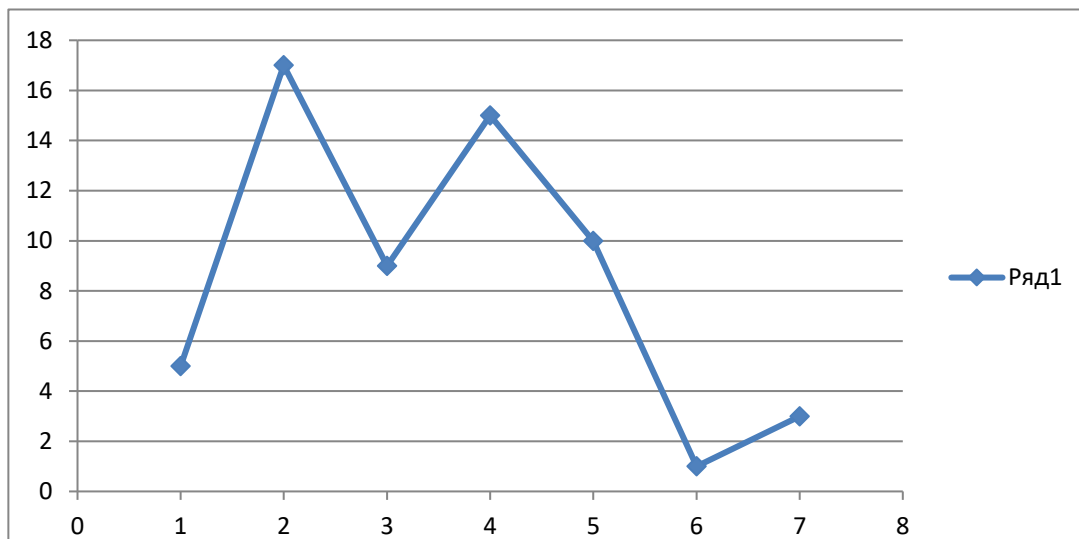


Рисунок 11 – Полигон частот

Кумулята будет иметь следующий вид (рисунок 12).

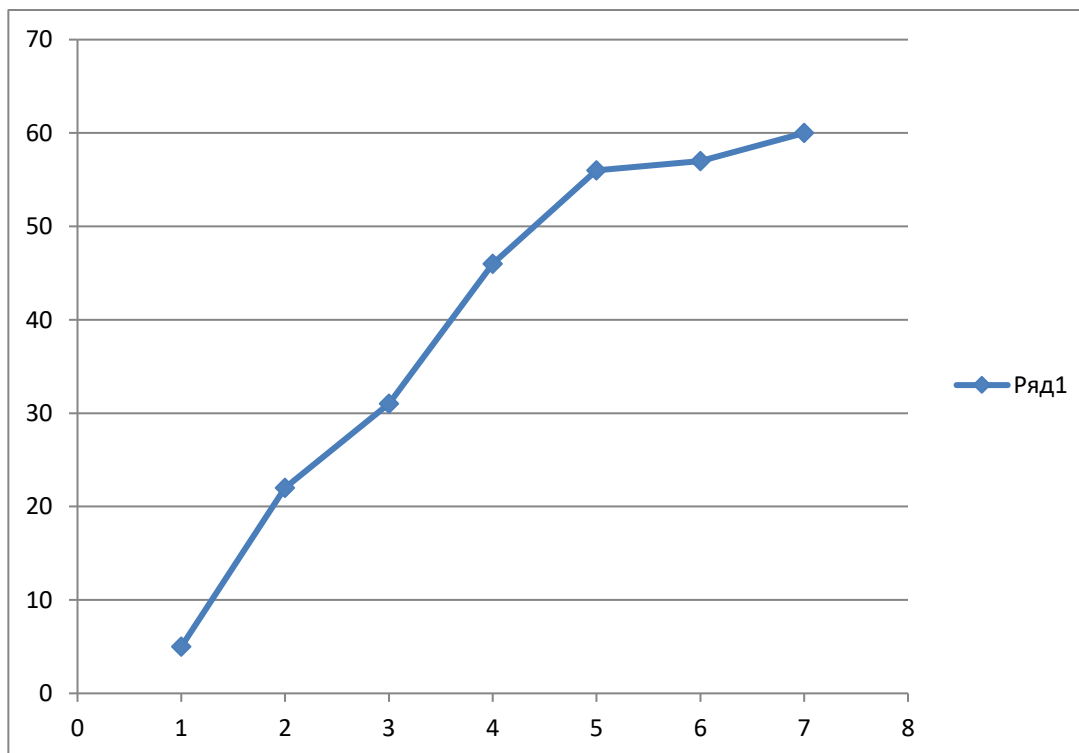


Рисунок 12 - Кумулята частот

1.2 Парная регрессия и корреляция

Пример 3 По имеющимся данным требуется построить график зависимости между переменными, по которому необходимо подобрать модель уравнения регрессии. Используя следующие функции:

- линейную;
- степенную;
- экспоненциальную;
- показательную.

Замечание. Важнейшим методом анализа данных является визуализация (представление данных в виде таблиц, диаграмм, кросс-таблиц, кросс-диаграмм, графиков).

Таблица 4 - Фондообеспеченность и производство продукции

№	Фондообеспеченность на 1 га сельхозугодий, тыс. руб.; (x)	Стоимость валовой продукции на 1 га сельхозугодий, тыс. руб.; (y)
1	38,4	62,3
2	24,2	30,1
3	29,2	47,3
4	23,0	29,9
5	18,2	37,2
6	33,2	46,1
7	14,1	22,3
8	26,2	43,0
9	20,1	34,1
10	35,0	49,2
11	31,7	41,4
12	24,4	37,4
13	18,9	28,2
14	27,1	37,0
15	17,0	26,1

Рассмотрим применение диаграммы рассеяния. Выделим в Excel диапазон В2:С16 (рисунок 2), выполним команду 1: Вставка - Точечная – Точечная с маркерами. В результате получим рисунок 13.

¹ Словосочетание *Выполнить команду (выбрать команду)* означает, что необходимо установить на вкладку Ленты (Excel 2007) указатель и щёлкнуть левой кнопкой мыши

Важность графического представления данных заключается в возможности увидеть возможные ошибки, допущенные при вводе данных (артефакты – объекты созданные человеком) или неоднородные значения признаков - выбросы – явно не принадлежащие изучаемой совокупности.

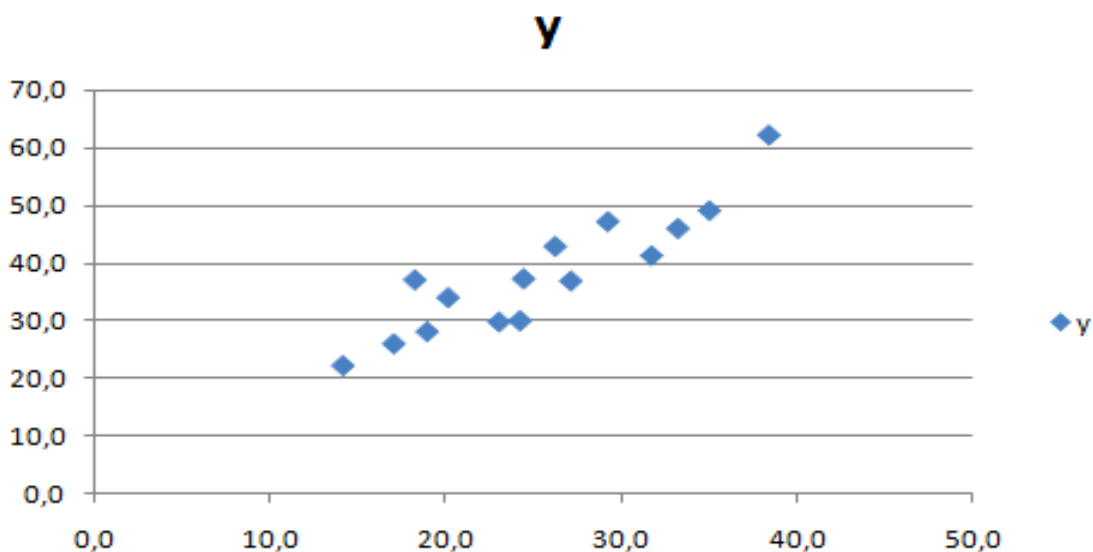


Рисунок 13 - Диаграмма рассеяния

Например, при вводе исходных данных мы вместо 62,3 ввели 623. Построим соответствующую диаграмму рассеяния (рисунок 14) из которой видно, что есть наблюдение, отличающееся от других данных.

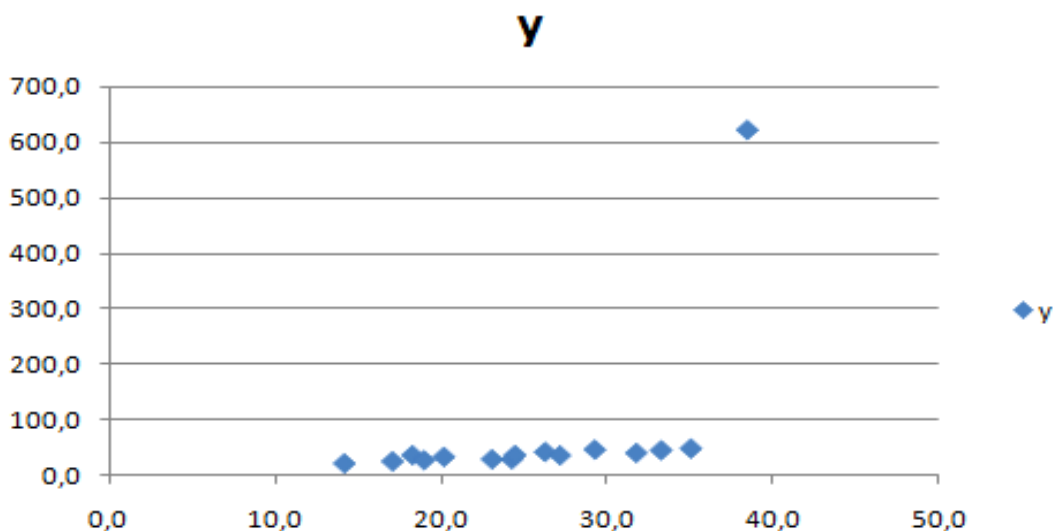


Рисунок 14 – Диаграмма рассеяния с артефактом (или выбросом)

Важным методом анализа данных в Excel являются диаграммы. Выделим на рисунке 33 щелчком левой клавиши мыши маркеры наблюдений; с помощью правой клавиши откроем контекстное меню (рисунок 15) и выберем одну из перечисленных линий трендов (рисунок 16):

- Линейная;
- Логарифмическая;
- Полиномиальная;
- Степенная;
- Экспоненциальная;
- Линейная фильтрация (Скользящая средняя).

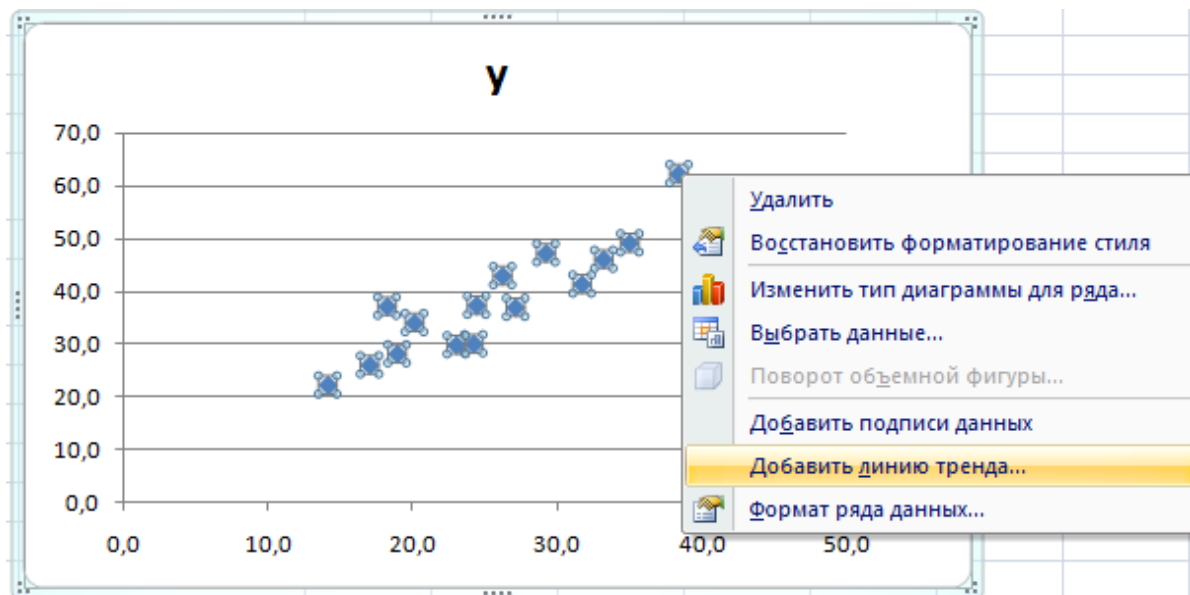


Рисунок 15 – Контекстное меню выделенных точек наблюдений

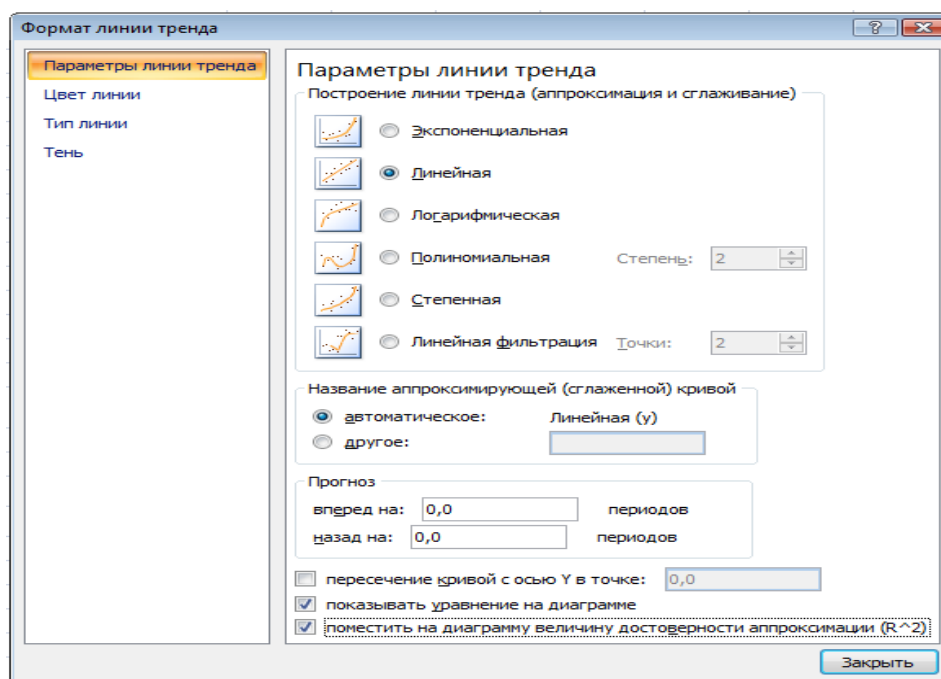


Рисунок 16 – Диалоговое окно выбора линии тренда

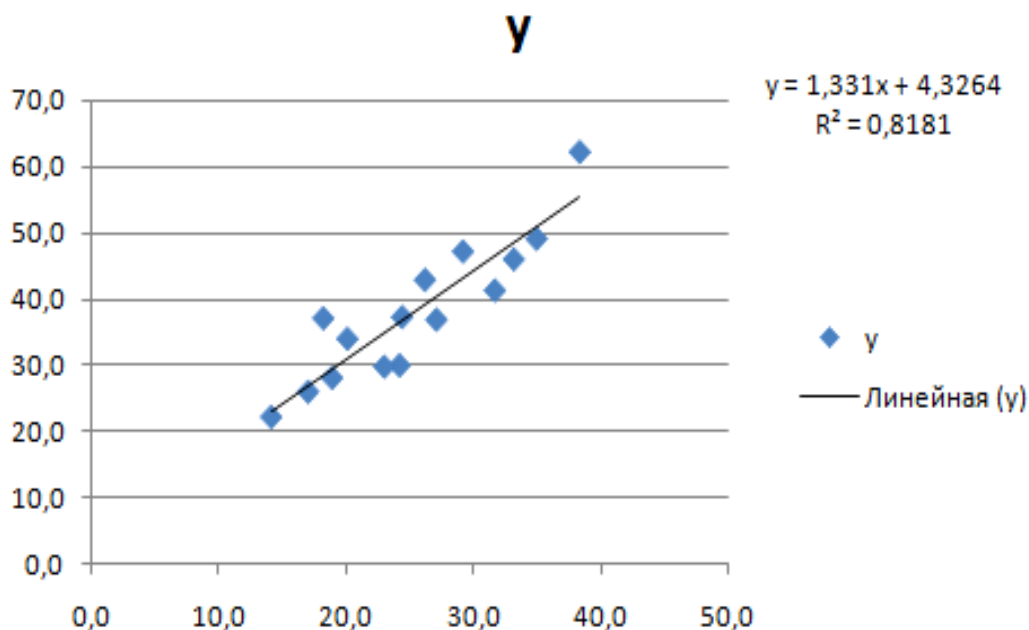


Рисунок 17- График линейного уравнения

После выбора одного из трендов, например, линейного - выберем и заполним вкладку Параметры диалогового окна (рисунок 16). Можно выбрать название (назвать тренд самостоятельно) или оставить автоматически предлагаемое Excel; для прогноза согласно выбранной линии тренда на 5 лет вперед выберем соответствующее значение в диалоговом окне; для отображения на диаграмме уравнения тренда и коэффициента детерминации отметим соответствующие элементы. Далее выберем ОК.

Задания

1. Построить вариационный ряд по заданию преподавателя.
2. Построить график точечной зависимости $y=2x+3$ для x от 1 до 50.

Наложить на переменную y «шум» (например, случайную величину, подчиняющуюся равномерному закону распределения), рассмотреть несколько вариантов с учетом уравнения тренда и коэффициента детерминации. Сделать выводы.

Вопросы для самоконтроля

1. Определение и виды вариационных рядов.
2. Способы графического изображения вариационных рядов.
3. Как определяется число групп и величина интервала при построении интервального вариационного ряда.
4. Основные этапы корреляционно-регрессионного анализа.
5. Расчет показателей тесноты связи между двумя признаками.

Уравнение множественной регрессии может быть построено в стандартизованном масштабе, когда единицей измерения признаков принимается их среднее квадратическое отклонение:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + \dots + \beta_p t_{x_p},$$

где,

$$t_y = \frac{y - \bar{y}}{\sigma_y}, t_{x_i} = \frac{x_i - \bar{x}_i}{\sigma_{x_i}}$$

β_i - стандартизованные коэффициенты регрессии,
 σ – среднее квадратическое отклонение.

Для оценки тесноты связи между признаками применяются парные, частные и множественные коэффициенты (индексы) корреляции и детерминации. Если изучение зависимости проводится по выборочным данным, то оценивается значимость коэффициентов регрессии, корреляции и всего уравнения множественной регрессии в целом.

Пример 4 Рассмотрим пример построения линейного уравнения регрессии между валовой продукцией на 1 га с/х угодий (Y), энергообеспеченностью на 100 га с/х угодий, л.с. (X1), фондообеспеченностью на 100 га с/х угодий, тыс. руб. (X2), затратами на производство продукции на 1 га с/х угодий, тыс. руб. (X3) по данным приложения.

В книге Excel **Группировка** вставим новый лист «Регрессия» и введём исходные данные (рис. 18).

	A	B	C	D	E	F	G	I	K	M	
	Валовая продукция на 1 га с/х угодий, тыс.руб.	Энергообеспеченность на 100 га с/х угодий, л.с.	Фондообеспеченность на 100 га с/х угодий, тыс. руб.	Затраты на производство продукции на 1 га с/х угодий, тыс.руб.							
1	Y	X1	X2	X3							
3	9,923057913	284,2326386	611,0140868	9,166817695							
4	4,630444412	217,0087111	950,5684335	4,571091037							
5	5,791561278	178,2708669	801,5750747	4,375258668							
6	6,096240136	254,4097168	408,0457992	4,89494043							
7	8,881634486	295,4737315	837,6919623	7,449753031							
8	7,640543881	296,9262464	1438,29419	6,889163576							
9	11,70316678	391,9634061	879,408867	8,766924701							
10	6,938591118	349,9387443	1782,618683	7,162940276							
11	8,855745038	345,1635449	1109,854627	7,819681297							
12	9,759330618	264,5271386	1036,177474	6,791258395							
13	9,255029057	240,9924005	1662,606169	7,627626285							
14	4,905933746	205,7007645	752,28249	5,467710229							
15	5,994445679	302,0662075	1723,128194	5,839591202							
16	13,39236437	337,5753516	1333,784327	9,962089752							
17	8,290749601	315,7575758	1028,516746	7,594577352							
18	9,903981043	316,7772512	842,985782	7,922654028							
19	12,78053239	307,734491	1347,389466	9,036558894							
20	6,486149584	190,1939058	901,0156971	5,3388735							
21	6,413850301	255,4945055	456,9044163	5,295044578							
22	4,878780367	160,7585523	341,5220625	3,715418939							
23	4,470726496	390,4059829	568,1837607	3,580128205							
60	7,855538611	265,31263	1013,1568	5,494939692							
61	4,72593801	267,3191952	1127,351822	5,215878195							
62	6,916878225	203,5050364	510,5888134	8,265498321							
63	4,04333089	153,7193111	864,9322096	3,961982411							
64	5,388862218	236,4378301	1147,191551	5,745919347							
65	8,275821458	356,1182428	1681,450948	7,629372277							
66											
							Y	X1	X2	X3	
							Среднее	8,511279	289,7707	1180,536	7,457692
							Стандартная ошибка	0,412123	9,935091	64,48195	0,324782
							Медиана	7,640544	293,2475	1071,671	7,071205
							Мода	#N/D	#N/D	#N/D	#N/D
							Стандартное отклонение	3,271127	78,85734	511,8096	2,577877
							Дисперсия выборки	10,70027	6218,481	261949,1	6,64545
							Экссесс	0,410778	-0,82352	-0,264	0,268729
							Асимметричность	1,029776	0,066946	0,562511	0,891476
							Интервал	12,94959	310,5475	2126,823	10,72132
							Минимум	4,043331	148,5877	341,5221	3,580128
							Максимум	16,99292	459,1351	2468,345	14,30145
							Сумма	536,2106	18255,55	74373,77	469,8346
							Счет	63	63	63	63

Рисунок 18 - Показатели шестидесяти организаций Краснодарского края

Таблица 5 - Описательная статистика

	Y	X1	X2	X3
Среднее	8,511279	289,7707	1180,536	7,457692
Стандартная ошибка	0,412123	9,935091	64,48195	0,324782
Медиана	7,640544	293,2475	1071,671	7,071205
Мода	#Н/Д	#Н/Д	#Н/Д	#Н/Д
Стандартное отклонение	3,271127	78,85734	511,8096	2,577877
Дисперсия выборки	10,70027	6218,481	261949,1	6,64545
Экссесс	0,410778	-0,82352	-0,264	0,268729
Асимметричность	1,029776	0,086946	0,562511	0,891476
Интервал	12,94959	310,5475	2126,823	10,72132
Минимум	4,043331	148,5877	341,5221	3,580128
Максимум	16,99292	459,1351	2468,345	14,30145
Сумма	536,2106	18255,55	74373,77	469,8346
Счет	63	63	63	63
Наибольший(1)	16,99292	459,1351	2468,345	14,30145
Наименьший(1)	4,043331	148,5877	341,5221	3,580128

Для расчета необходимых показателей Y, X₁, X₂, X₃ введём в ячейки диапазона A2:D2 соответствующие формулы. Выделим диапазон A2:D2 и с помощью **маркера заполнения** скопируем формулы до строки 64. В результате получим исходные данные для регрессионного анализа, изображенные на рис. 18.

Применим инструмент **Описательная статистика** (6). Исследуем корреляцию факторов, используя инструмент Пакета анализа **Корреляция** (4). Из полученной корреляционной матрицы следует, что факторы X₁, X₂ коррелированы между собой, поэтому МНК применять нельзя.

Таблица 6 - Корреляционная матрица

	Y	X1	X2	X3
Y	1			
X1	0,594210765	1		
X2	0,58585918	0,565050507	1	
X3	0,937325544	0,549271689	0,630009324	1

Однако, так как коэффициент корреляции менее 0,65, то этот факт можно игнорировать.

В противном случае один из факторов необходимо отбросить, либо использовать другие методы оценки параметров регрессии (модуль 4).

Заполним диалоговое окно инструмента Пакета анализа **Регрессия** (рис.19). В результате выбора **ОК** получим таблицу итогов регрессионного анализа.

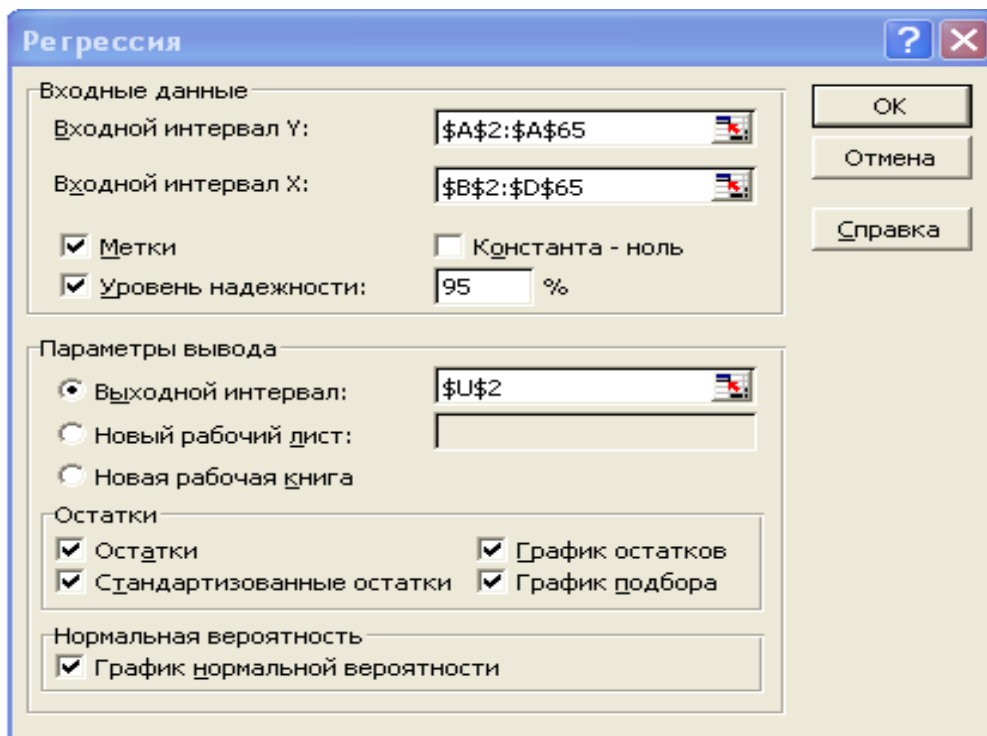


Рисунок 19 - Диалоговое окно Регрессия

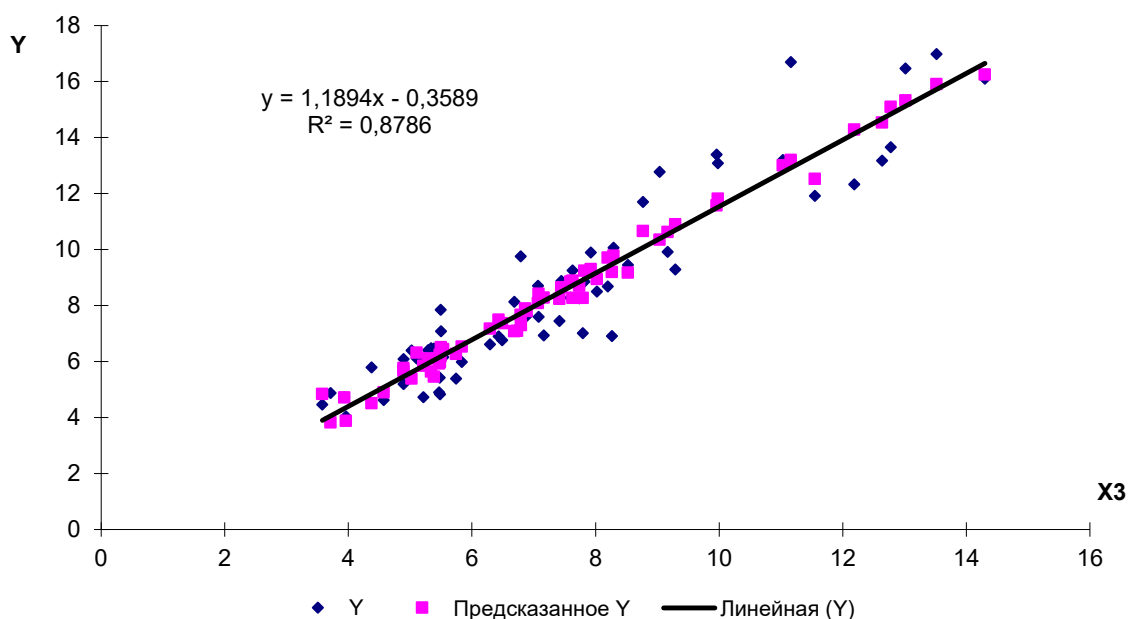


Рисунок 20 - График подбора

Таблица 7 – Регрессионная статистика

Регрессионная статистика					
Множественный R	0,942991				
R-квадрат	0,889233				
Нормированный R-квадрат	0,88360				
Стандартная ошибка	1,11601				
Наблюдения	63				
ДА	df	SS	MS	F	Значимость F
Регрессия	3	589,9325323	196,6441774	157,8845	3,7865E-28
Остаток	59	73,48415634	1,245494175		
Итого	62	663,4166886			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	-1,1649987	0,56045999	-2,07864738	0,04200737	-2,286477719	-0,04351966
X1	0,00543403	0,002284902	2,37823362	0,02065306	0,000861948	0,01000611
X2	-0,0003535	0,000378819	-0,93308841	0,35457729	-0,001111488	0,00040454
X3	1,14230254	0,07425756	15,3829794	2,59445E-22	0,993713347	1,29089173

Дисперсионный анализ показывает, что уравнение является значимым при уровне значимости $\alpha=3,7865E-28$. Множественный коэффициент корреляции R равен 0,942991, то есть полученное уравнение достаточно хорошо описывает изучаемую взаимосвязь между факторами. Коэффициент детерминации R^2 равен 0,889 - это означает, что 88,9% вариации результативного признака (Y) объясняется вариацией факторных переменных (X_1, X_2, X_3).

Согласно таблице 7, искомое уравнение регрессии имеет вид:

$$Y=0,0054X_1 - 0,00035X_2 + 1,1423X_3 - 1,165.$$

Причём доверительный интервал при уровне значимости 5%:

для коэффициента при X_1 : (0,00086; 0,010);

для коэффициента при X_2 : (-0,00111; 0,00040);

для коэффициента при X_3 : (0,99371; 1,29089);

для свободного члена: (-2,28647; -0,04352).

Excel позволяет также анализировать парные регрессионные зависимости (остатки и многое другое). Так в нашем примере зависимость Y от X_3 имеет вид, представленный инструментом **Регрессия** на рис. 20. Коэффициенты уравнения регрессии показывают на сколько изменится Y при изменении одной из факторных переменных на единицу (при условии, что остальные переменные не изменяются).

2.2 Группировка в Excel 2007 (Сводные таблицы)

Ценность данных определяется не их объемом, а возможностью их преобразования в значимую – релевантную информацию (релевантный <англ. relevant - существенный, уместный, относящийся к делу). Согласно современным представлениям наиболее удобным способом хранения, организации и поиска информации являются базы данных (БД) (рисунок 21).

БД-это фактически любой набор данных: телефонный справочник, список книг в библиотеке, данные о курсе доллара по дням в разных банках, урожайность различных культур в сельском хозяйстве по годам, список лиц работающих в коллективе (год рождения, состав семьи, адрес, стаж работы, телефон, e-mail).

№ п.п.	Среднегодовая численность работников, чел.	Численность тракторов, эт. ед.	Площадь сельскохозяйственных угодий, га	Энергетические мощности, л.с.	Основные фонды сельскохозяйственного назначения, тыс. руб.	Затраты на производство валовой продукции, тыс. руб.	Затраты на производство реализованной продукции, тыс. руб.	Валовая продукция, тыс. руб.	Реализованная продукция, тыс. руб.	
1										
2	1	591	102	12139	34503	74171	111276	80946	120456	90126
3	2	334	54	6773	14698	64382	30960	25670	31362	2607
4	3	335	45	8698	15506	69721	38056	29209	50375	41528
5	4	657	102	12926	32885	52744	63272	38176	78800	53704
6	5	541	75	11135	32901	93277	82953	68145	98897	84089
7	6	864	113	12135	36032	174537	83600	54719	92718	63837
8	7	370	68	7105	27849	62482	62289	56879	83151	77741
9	8	437	54	6530	22851	116405	46774	36995	45309	35530
10	9	410	76	7154	24693	79399	55942	49226	63354	56638
11	10	552	68	9083	24027	94116	61685	60013	88644	86972
12	11	246	48	4474	10782	74385	34126	29769	41407	37050
13	12	492	104	13735	28253	103326	75099	54292	67383	46576
14	13	217	53	4501	13596	47558	26284	19065	26981	19762
15	14	603	98	7465	25200	99567	74367	70913	99974	96520
16	15	400	58	6270	19798	64488	47618	25379	51983	29744
17	16	602	121	10550	33420	88935	83584	60564	104487	81467
18	17	768	88	8777	26974	117077	70007	61608	111628	87176
19	18	312	57	5217	12896	68042	28615	21611	25202	18198
20	19	117	22	2788	10238	26645	14237	11837	17034	14634
21	20	284	43	6638	15757	31237	43087	41277	44903	43093
22	21	304	32	5133	12985	22407	35222	23856	39517	28151
23	22	377	70	6084	20010	81056	39135	28504	41983	31354
24	23	341	67	7213	19137	73079	39635	29687	56662	46714
25	24	155	28	3678	9832	41464	19184	14300	17382	12498
26	25	525	74	12211	24850	62348	100930	58596	84462	42123
27	26	383	57	10916	16780	94416	43249	28381	44137	29269
28	27	372	73	8332	19700	95584	47875	33253	44900	30278
29	28	519	69	8491	30238	142772	64781	45353	70270	50842
30	29									
31	30									
32	31									
33	32									
34	33									
35	34									
36	35									
37	36									
38	37									
39	38									
40	39									
41	40									
42	41									
43	42									
44	43									
45	44									
46	45									
47	46									
48	47									
49	48									
50	49									
51	50									
52	51									
53	52									
54	53									
55	54									
56	55									
57	56									
58	57									
59	58									
60	59									
61	60									
62	61									
63	62									
64	63									
65										
66										
67										
68										

Рисунок 21 – База данных

Создание баз данных упрощает обработку данных и их анализ. Для этого в верхнюю строку необходимо ввести заголовки столбцов, а под ними без пропусков в каждую ячейку соответствующие данные.

При большом их количестве, для редактирования или отбора по некоторому критерию, удобно воспользоваться командой Данные-Фильтр.

Пример 5 Используя данные приложения А, создадим Лист 1 с исходными данными для группировки (таблица 8). Сохраним рабочую книгу под названием Группировка.

Важной задачей является задача разбиения на группы, удовлетворяющие определённым критериям. Она может быть решена в Excel различными способами: команда Данные - Фильтр; команда Данные – Группировать и т.д.

Важным средством Excel при решении задачи разбиения на группы являются *Сводные таблицы* (команда Вставка – Сводная таблица).

Таблица 8 – Расчетные показатели для группировки

	А	В	С	Д	Е	Ф	Г
1	Фондообеспеченность на 100 га с/х угодий, тыс. руб.	Энергообеспеченность на 100 га с/х угодий, л.с.	Энергетические мощности, л.с.	Основные фонды сельскохозяйственного назначения, тыс. руб.	Площадь сельскохозяйственных угодий, га	Валовая продукция, тыс. руб.	Среднегодовая численность работников, чел.
2							
3							
4							

Рассмотрим процесс построения группировки по шагам.

Шаг 1. Выберем команду Вставка – Сводная таблица – ОК.

Если курсор мыши находится в поле таблицы, то диапазон задается автоматически (рис.22).

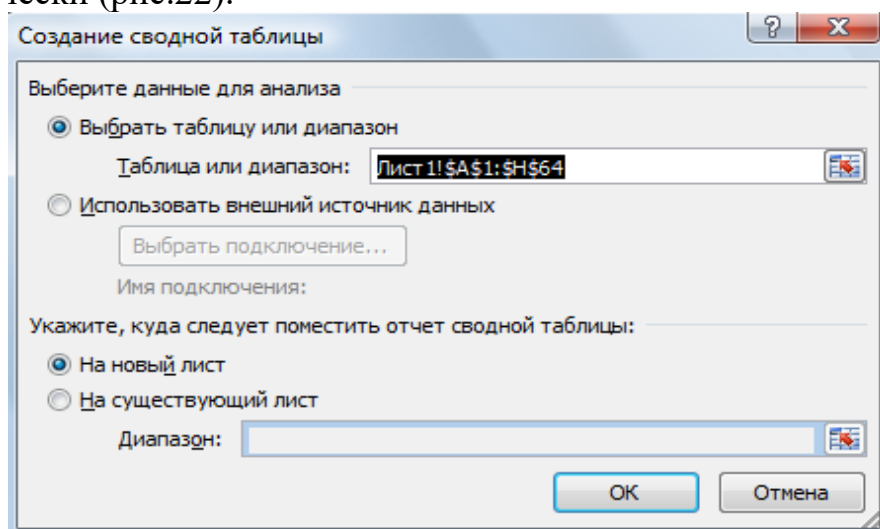


Рисунок 22 – Создание сводной таблицы

Шаг 2. Для группировки исходных данных по признаку энергообеспеченность на 100 га с/х угодий предварительно заполним макет сводной таблицы (рис. 23-24).

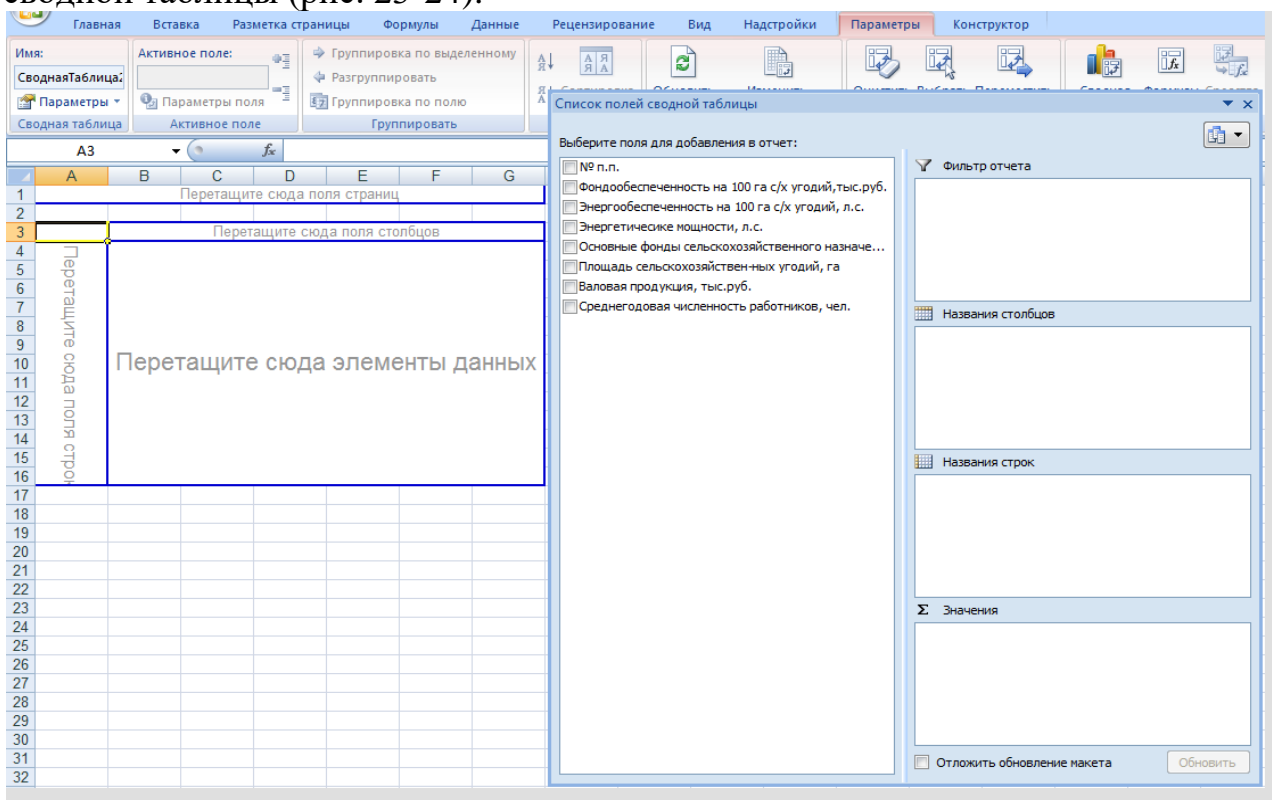


Рисунок 23 – Макет сводной таблицы

	A	B	C	D	E
1					
2	Энергообеспеченность	Данные	Итого		
3	148,5876814	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	1		
4		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	148,5876814		
5		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	947,1295933		
6		Сумма по полю Энергетические мощности, л.с.	22725		
7		Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	144854		
8		Сумма по полю Площадь сельскохозяйственных угодий, га	15294		
9		Сумма по полю Валовая продукция, тыс.руб.	98967		
10		Сумма по полю Среднегодовая численность работников, чел.	922		
11	153,7193111	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	1		
12		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	153,7193111		
13		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	864,9322096		
14		Сумма по полю Энергетические мощности, л.с.	16780		
15		Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	94416		
16		Сумма по полю Площадь сельскохозяйственных угодий, га	10916		
17		Сумма по полю Валовая продукция, тыс.руб.	44137		
18		Сумма по полю Среднегодовая численность работников, чел.	383		
19	156,5914397	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	1		
20		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	156,5914397		
21		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	782,5836576		
22		Сумма по полю Энергетические мощности, л.с.	10061		
23		Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	50281		
24		Сумма по полю Площадь сельскохозяйственных угодий, га	6425		
25		Сумма по полю Валовая продукция, тыс.руб.	46181		
26		Сумма по полю Среднегодовая численность работников, чел.	230		
27	160,7585523	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	1		
28		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	160,7585523		
29		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	341,5220625		
30		Сумма по полю Энергетические мощности, л.с.	6485		

Рисунок 24 – Пример заполнения макета сводной таблицы

Шаг 3. Для группировки щёлкнув правой кнопкой мыши по полю «Энергообеспеченность» и выберем из контекстного меню команду Группировать с шагом 103,52 (для разбиения на три группы) (рис. 25). В результате получим группировку по указанному признаку (рис.26).

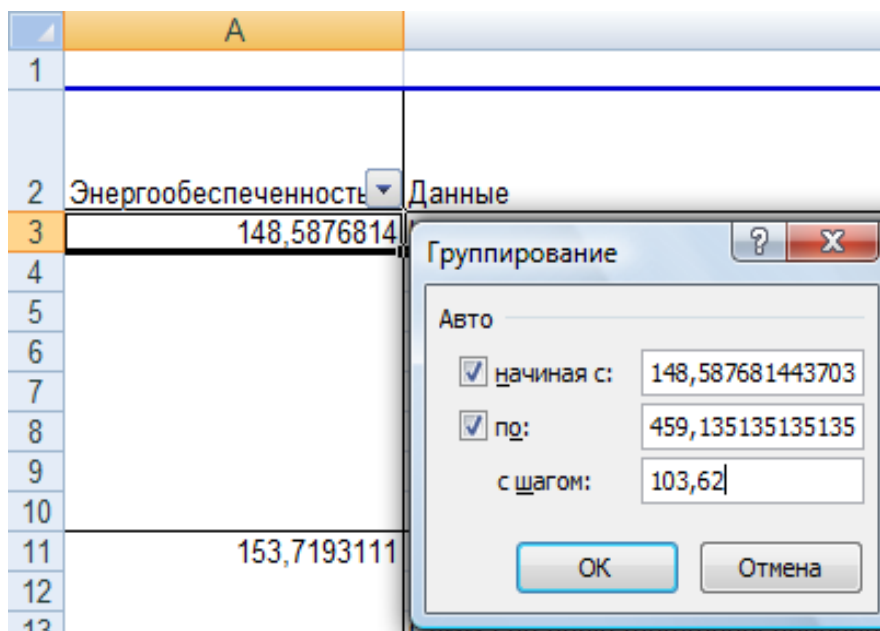


Рисунок 25 – Окно мастера «Группирование»

1			
2	Энергообеспеченность на 100 га с/х	Данные	Итого
3	148,587681443703-252,207681443703	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	21
4		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	4249,391613
5		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	21144,4866
6		Сумма по полю Энергетическая мощность, л.с.	377617
7		Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	1870700
8		Сумма по полю Площадь сельскохозяйственных угодий, га	188897
9		Сумма по полю Валовая продукция, тыс.руб.	1276704
10		Сумма по полю Среднегодовая численность работников, чел.	8942
11	252,207681443703-355,827681443703	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	28
12		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	8442,842411
13		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	30544,10321
14		Сумма по полю Энергетическая мощность, л.с.	698922
15		Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	2494482
16		Сумма по полю Площадь сельскохозяйственных угодий, га	231549
17		Сумма по полю Валовая продукция, тыс.руб.	1884415
18		Сумма по полю Среднегодовая численность работников, чел.	12922
19	355,827681443703-459,447681443703	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	14
20		Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	5563,320689
21		Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	22685,18232
22		Сумма по полю Энергетическая мощность, л.с.	506458
23		Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	2215299
24		Сумма по полю Площадь сельскохозяйственных угодий, га	124356
25		Сумма по полю Валовая продукция, тыс.руб.	1614693
26		Сумма по полю Среднегодовая численность работников, чел.	8429
27	Итого	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	63

Рисунок 26 – Сводная таблица после группировки по полю «Энергообеспеченность»

Шаг 4. Для вторичной группировки перетащим в поле название строк показатель «Фондообеспеченность», сделаем для него группировку с шагом 1063,45 (рис. 27, 28).

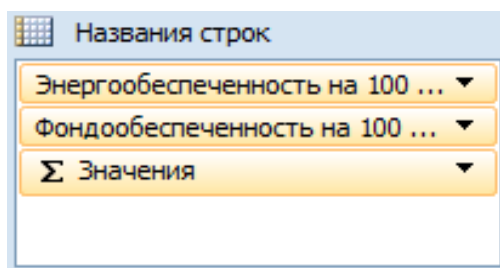


Рисунок 27 – Детализация списка полей сводной таблицы для вторичной группировки

1			
2	Энергообеспеченность на 100 га с/х угодий, л.с.	Фондообеспеченность на 100 га с/х угодий, тыс. руб.	Итого
3	148,587681443703-252,207681443703	341,522062469013-1404,97206246901	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.
4			Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.
5			Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс. руб.
6			Сумма по полю Энергетичесике мощности, л.с.
7			Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.
8			Сумма по полю Площадь сельскохозяйствен-ных угодий, га
9			Сумма по полю Валовая продукция, тыс.руб.
10			Сумма по полю Среднегодовая численность работников, чел.
11		1404,97206246901-2468,42206246901	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.
12			Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.
13			Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.
14			Сумма по полю Энергетичесике мощности, л.с.
15			Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.
16			Сумма по полю Площадь сельскохозяйствен-ных угодий, га
17			Сумма по полю Валовая продукция, тыс.руб.
18			Сумма по полю Среднегодовая численность работников, чел.
19	148,587681443703-252,207681443703	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.	21
20	148,587681443703-252,207681443703	Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.	4249,391613
21	148,587681443703-252,207681443703	Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.	21144,4866
22	148,587681443703-252,207681443703	Сумма по полю Энергетичесике мощности, л.с.	377617
23	148,587681443703-252,207681443703	Сумма по полю Основные фонды сельскохозяйственного назначения, тыс. руб.	1870700
24	148,587681443703-252,207681443703	Сумма по полю Площадь сельскохозяйствен-ных угодий, га	188897
25	148,587681443703-252,207681443703	Сумма по полю Валовая продукция, тыс.руб.	1276704
26	148,587681443703-252,207681443703	Сумма по полю Среднегодовая численность работников, чел.	8942
27	252,207681443703-355,827681443703	341,522062469013-1404,97206246901	Количество по полю Энергообеспеченность на 100 га с/х угодий, л.с.
28			Сумма по полю Энергообеспеченность на 100 га с/х угодий, л.с.
29			Сумма по полю Фондообеспеченность на 100 га с/х угодий, тыс.руб.
30			Сумма по полю Энергетичесике мощности, л.с.

Рисунок 28 – Сводная таблица после вторичной группировки

Полученную сводную таблицу скопируем на новый лист («Группировочная таблица»), используя команду *Правка-Специальная вставка - Значения*, чтобы иметь возможность редактирования полей и итогов. Достроим сводную таблицу, введя показатели, которые необходимо рассчитать на основании итогов сводной таблицы. Введем соответствующие формулы. С помощью контекстных меню строк и столбцов скроем строки и столбцы с лишними расчетами, отформатируем. В результате получим комбинационную группировку влияния энергообеспеченности и фондообеспеченности сельхозугодий на эффективность производства (рис. 29).

	A	B	D	E	F	G	H	I
1	Группы хозяйств по энергообеспеченности на 100 га с/х угодий, л.с.	Подгруппы по фондообеспеченности на 100 га с/х угодий, тыс.руб.	Итого	Численность хозяйств в группе	Энергообеспеченность в среднем по группе, л.с.	Фондообеспеченность в среднем по группе, тыс.руб.	Валовая продукция на 100 га с/х угодий, тыс.руб.	Валовая продукция на 1 работника, тыс.руб.
2	148,6-252,1	341,5-1404,9	18	18	2,1	9,5	652,1	141,3
10		1404,9-2468,4	3	3	3,2	24,9	865,4	152,3
18	В среднем по группе		21	21	2,2	11,2	675,9	142,8
26	252,1-355,6	341,5-1404,9	22	22	3,6	11,3	834,0	155,2
34		1404,9-2468,4	6	6	3,7	19,8	745,1	118,5
42	В среднем по группе		28	28	3,6	13,2	813,8	145,8
50	355,6-459,1	341,5-1404,9	5	5	6,9	18,0	952,9	178,0
58		1404,9-2468,4	9	9	3,8	18,3	1396,7	194,4
66	В среднем по группе		14	14	4,5	18,2	1298,4	191,6
74	Всего и в среднем		63	63	3,4	13,7	876,6	157,7

Рисунок 29 – Комбинационная группировка

Комбинационная группировка сельхозпредприятий по уровню энергообеспеченности и фондообеспеченности показала, что с ростом энергообеспеченности повышаются показатели экономической эффективности с/х производства.

В первой группе хозяйств с энергообеспеченностью от 148,6 до 252,1 л.с. производство валовой продукции на 100 га с/х угодий в среднем составляет 652,1 тыс.руб., а на среднегодового работника – 141,3 тыс.руб. Если обратиться ко второй группе хозяйств, уровень энергообеспеченности повышался от 252,1 до 355,6 л.с. производство валовой продукции увеличивается до 813,8 тыс.руб. на 100 га с/х угодий, а на 1 работника увеличивается на 2,1%. Увеличение производства валовой продукции проявляется и во второй группе, где повышение энергообеспеченности до 459,1 л.с. обеспечивает рост валовой продукции на 100 га с/х угодий до 1 298,4 тыс.руб.

Такая же закономерность повышения эффективности производства проявляется и в подгруппах с различным уровнем фондообеспеченности. Так в подгруппах с более высокой фондообеспеченностью увеличивается производство валовой продукции как на единицу площади, так и на одного работника.

Таким образом, повышение уровня энергообеспеченности и фондообеспеченности с/х предприятий являются важнейшими повышения эффективности сельскохозяйственного производства.

Вопросы для самоконтроля

1. Что характеризуют коэффициенты регрессии, корреляции, эластичности, детерминации.
2. Виды статистических группировок и их виды
3. Выбор группировочного признака, определение числа групп, расчет величины интервала, нижних и верхних границ интервалов.
4. Способы построения вторичных группировок

ЛАБОРАТОРНАЯ РАБОТА № 3

Пакет анализа

Цель работы: ознакомиться с возможностями пакета анализа данных в Excel 2007

Теоретические сведения

Однофакторный дисперсионный анализ позволяет статистически обосновать существенность влияния факторного признака A на результативный F .

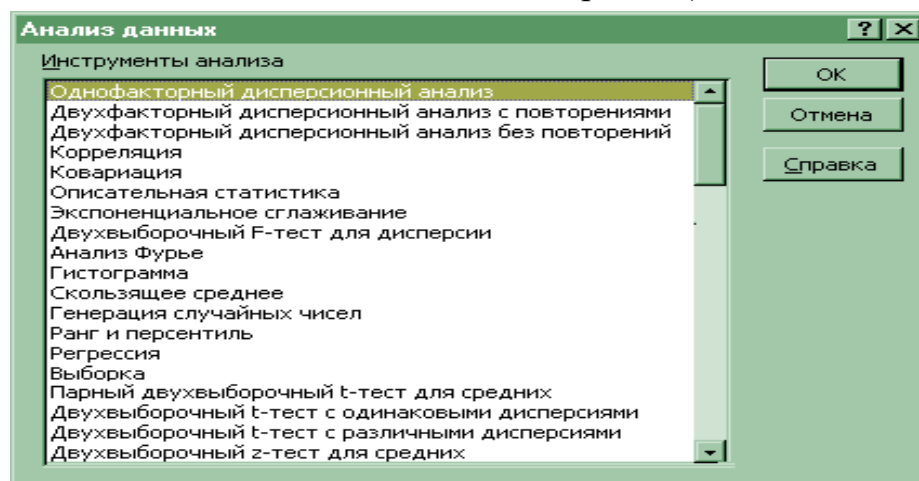
Замечание. Нет выделения моделей дисперсионного анализа (ДА) по виду факторов.

Однофакторный дисперсионный анализ используется для проверки гипотезы о сходстве средних значений двух или более выборок, принадлежащих одной и той же генеральной совокупности.

Этот метод распространяется также на тесты для двух средних (к которым относится, например, t -критерий). То есть если для разных уровней фактора A средние отличаются незначительно, то следует принять гипотезу о несущественном влиянии факторного признака A на результативный F .

Для активизации *Пакета анализа* необходимо выполнить команду *кнопка Office-параметры Excel-Надстройки* и выбрать *Пакет анализа*, после этого на ленте во вкладке *Данные* появится строка *Анализ данных*¹.

Пакет анализа позволяет решать в диалоговом режиме 19 задач, наиболее часто встречающихся в классической математической статистике (рис. 30) (хотя это не мешает подходить к ним творчески).



¹ Аналогично устанавливается Поиск решения – инструмент для решения задач линейного программирования. Примеры *C:\Program Files\Microsoft Office\Office12\SAMPLES\SOLVSAMP.xlsx*

Рисунок 30 - Инструменты анализа данных

Пример 6 Проверить статистическую существенность влияния катализатора А на химическую реакцию, результаты измерений при 5 уровнях фактора А приведены в таблице.

A1	A2	A3	A4	A5
3,2	2,6	2,9	3,7	3
3,1	3,1	2,6	3,4	3,4
3,1	2,7	3	3,2	3,2
2,8	2,9	3,1	3,3	3,5
3,3	2,7	3	3,5	2,9
3	2,8	2,8	3,3	3,1

Введём исходные данные в диапазон A1:A7 листа Excel. Заполним параметры диалогового окна (рис 31).

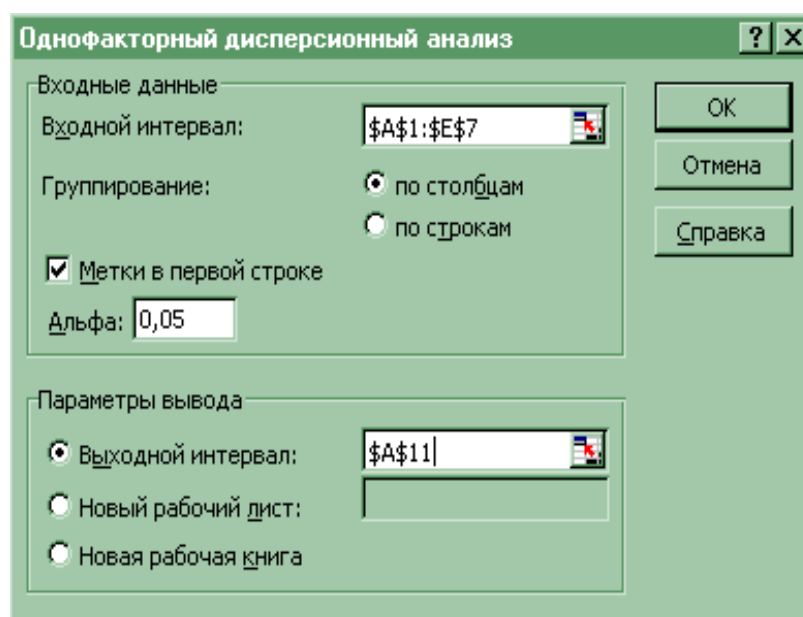


Рисунок 31 - Диалоговое окно однофакторного дисперсионного анализа

Группирование: по столбцам - именно такое расположение имеют уровни фактора А; отметим *Метки в первой строке* (там расположены уровни фактора А); в *Выходном интервале* достаточно отметить левую верхнюю ячейку выходного интервала \$A\$11. После нажатия кнопки ОК, получим таблицу однофакторного дисперсионного анализа (табл. 9).

Таблица 9 - Однофакторный дисперсионный анализ

Группы	Счет	Сумма	Среднее	Дисперсия
A1	6	18,5	3,083333333	0,029666667
A2	6	16,8	2,8	0,032
A3	6	17,4	2,9	0,032
A4	6	20,4	3,4	0,032
A5	6	19,1	3,183333333	0,053666667

Дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Между группами	1,342	4	0,335	9,35408	9,16424E-05	2,75871
Внутри групп	0,89667	25	0,035			
Итого	2,238667	29				

$F_{\text{расч.}}=9,3540 > F_{\text{кр.}}=2,7587$, поэтому гипотезу о несущественном влиянии фактора А на результат следует отвергнуть.

Двухфакторный дисперсионный анализ с повторениями. Он представляет собой более сложный вариант однофакторного анализа, включающего более чем одну выборку для каждой группы данных.

Двухфакторный дисперсионный анализ позволяет статистически обосновать существенность влияния факторных признаков А и В и взаимодействия факторов (А и В) на результативный фактор F.

Пример 7. У 60 рабочих фиксировалась среднечасовая выработка в натуральных единицах продукции. Данные обследования отражены в табл. 10.

Таблица 10 – Исходные данные

	А	В	С	Д
1	Стаж	Возраст		
2		от 25 до 35 лет	от 35 до 45 лет	от 45 до 55 лет
3	от 1 до 4 лет	19	19	18
4		20	20	19
5		20	20	20
6		20	23	21
7		22	25	23
8	от 4 до 7 лет	30	20	19
9		31	29	25
10		32	30	25
11		32	31	26
12		34	31	26
13	от 7 до 10 лет	35	36	24
14		35	40	24
15		39	41	24
16		40	42	25
17		41	45	25
18	свыше 10 лет	40	28	20
19		40	31	24
20		41	35	25

21		41	36	31
22		42	40	32

Оценить существенность влияния возраста и стажа на производительность труда.

Данные в табл. 10 приведены так, как они должны выглядеть на листе Excel. Заполним диалоговое окно согласно рис. 32.

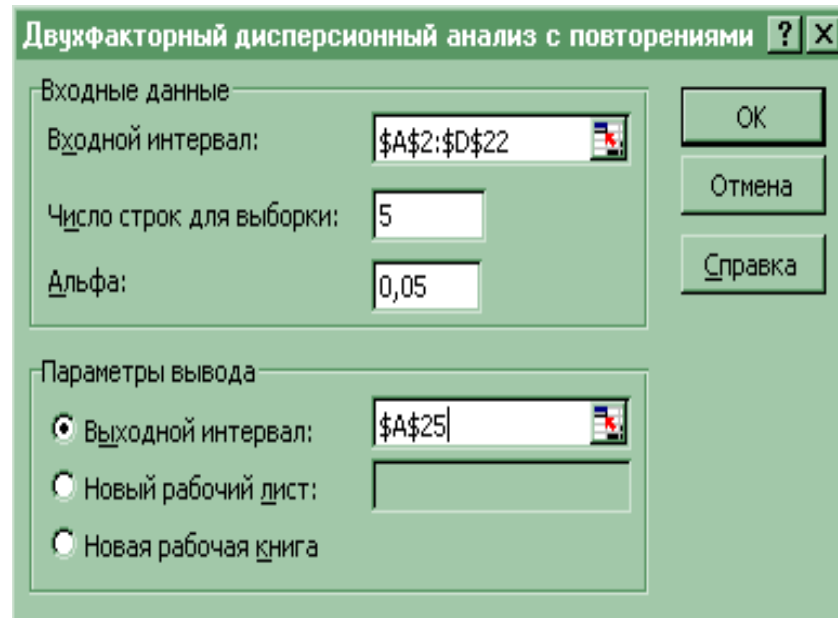


Рисунок 32 - Диалоговое окно двухфакторного дисперсионного анализа с повторениями

Следует отметить, что в исходных данных должно быть одинаковое число строк (повторений наблюдений), в противном случае рекомендуется ввести оценки пропущенных наблюдений, выбрав их таким образом, чтобы минимизировать остаточную дисперсию (иначе нужно ввести среднюю других наблюдений в ячейках), причём нельзя включать эти оценки при подсчёте соответствующих степеней свободы.

В случае пропуска данных в дисперсионном анализе без повторений необходимы более сложные методы. Как отмечали Кокрен и Кокс (Cochran and Cox): «единственное решение проблемы выпавших данных - не иметь их» [Л.П., 16].

После нажатия на кнопку **ОК** мы получим итоговую таблицу 11. Дисперсионный анализ с повторениями позволяет оценить существенность влияния фактора А (стажа), В (возраста), и их взаимодействия (факторов А и В) на среднечасовую выработку продукции в натуральных единицах.

$$\begin{aligned} \text{Так как: } F_{A \text{ расч.}} &= 66,8189 > F_{A \text{ кр.}} = 2,7980; \\ F_{B \text{ расч.}} &= 48,9791 > F_{B \text{ кр.}} = 3,1907; \\ F_{AB \text{ расч.}} &= 9,7456 > F_{AB \text{ кр.}} = 2,2945, \end{aligned}$$

Таблица 11 - Двухфакторный дисперсионный анализ с повторениями

ИТОГИ	от 25 до 35 лет	от 35 до 45 лет	от 45 до 55 лет	Итого
от 1 до 4 лет				
Счет	5	5	5	15
Сумма	101	107	101	309
Среднее	20,2	21,4	20,2	20,6
Дисперсия	1,2	6,3	3,7	3,54286
от 4 до 7 лет				
Счет	5	5	5	15
Сумма	159	141	121	421
Среднее	31,8	28,2	24,2	28,0667
Дисперсия	2,2	21,7	8,7	19,6381
от 4 до 7 лет				
Счет	5	5	5	15
Сумма	190	204	122	516
Среднее	38	40,8	24,4	34,4
Дисперсия	8	10,7	0,3	60,4
свыше 10 лет				
Счет	5	5	5	15
Сумма	204	170	132	506
Среднее	40,8	34	26,4	33,7333
Дисперсия	0,7	21,5	25,3	50,6381
Итого				
Счет	20	20	20	
Сумма	654	622	476	
Среднее	32,7	31,1	23,8	
Дисперсия	68,53684211	66,621053	13,3263158	

Дисперсионный анализ						
Источник вариации	SS	df	MS	F	P-Значение	F критическое
Выборка	1842,533333	3	614,177778	66,819	3,70226E-17	2,798
Столбцы	900,4	2	450,2	48,9791	2,56035E-12	3,190
Взаимодействие	537,4666667	6	89,5777778	9,74554	5,19758E-07	2,294
Внутри	441,2	48	9,19166667			
Итого	3721,6	59				

то следует признать статистически значимым влияние стажа (фактор А) возраста (фактор В) и их взаимодействия (факторы А и В) на производительность труда рабочих. Итоговая таблица позволяет более детально рассмотреть свойства отдельных групп (например, возраст от 35 до 45 лет и стаж от 4 до 7 лет). В Excel имеется хорошая возможность наглядного представления изучаемых процессов или явлений с помощью мастера диаграмм. Для этого: предварительно выделим данные табл.10; нажмём кнопку **Мастер диаграмм** панели **Стандартная**; выберем тип диаграммы *Стандартные – График*. В результате, после преобразований, получим рис. 33.

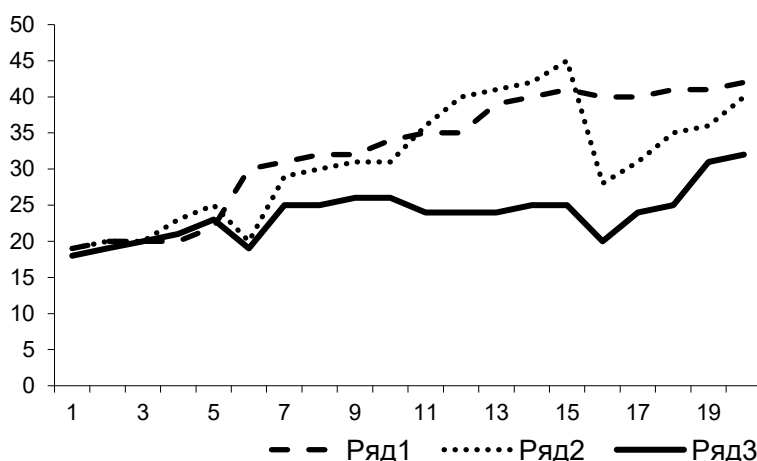


Рисунок 33 - Изменение среднечасовой выработки по категориям

Двухфакторный дисперсионный анализ без повторений позволяет оценить существенность воздействия факторов А и В на результирующий фактор без учёта воздействия взаимодействия факторов А и В.

Пример 8 Рассмотрим решение примера (данные из примера 7), взяв в каждой ячейке среднее значение наблюдений.

Стаж	Возраст		
	от 25 до 35 лет	от 25 до 35 лет	от 25 до 35 лет
от 1 до 4 лет	20,2	21,4	20,2
от 4 до 7 лет	31,8	28,2	24,2
от 7 до 10 лет	38	40,8	24,4
свыше 10 лет	40,8	34	26,4

В результате заполнения диалогового окна (аналогичного окну примера 8) получим итоговую таблицу двухфакторного дисперсионного анализ без повторений.

Таблица 12 - Двухфакторный дисперсионный анализ без повторений

ИТОГИ	Счет	Сумма	Среднее	Дисперсия
от 1 до 4 лет	3	61,8	20,6	0,48
от 4 до 7 лет	3	84,2	28,06667	14,4533333
от 7 до 10 лет	3	103,2	34,4	76,96
свыше 10 лет	3	101,2	33,73333	51,8933333
от 25 до 35 лет	4	130,8	32,7	83,5866667
от 25 до 35 лет	4	124,4	31,1	68,3333333
от 25 до 35 лет	4	95,2	23,8	6,74666667

Дисперсионный анализ

Источник вариации	SS	df	MS	F	P-Значение	F критическое
Строки	368,507	3	122,8356	6,85636319	0,02295241	4,757055194
Столбцы	180,08	2	90,04	5,02580005	0,05222744	5,143249382
Погрешность	107,493	6	17,91556			
Итого	656,08	11				

Данные дисперсионного анализа свидетельствуют о том, что фактор А (стаж) существенно влияет на производительность труда (так как $F_{\text{Арасч.}} = 6,8563 > F_{\text{Акр.}} = 4,7570$), а фактор В (возраст) статистически существенного влияния не оказывает (так как $F_{\text{Врасч.}} = 5,0258 < F_{\text{Вкр.}} = 5,1432$). Различие в выводах примеров 6 и 7 можно объяснить тем, что в примере 7 мы не учитывали конкретные наблюдения в ячейках, а рассматривали лишь их среднее значение. Поэтому результат дисперсионного анализа с повторениями является более значимым, что соответствует и самому смыслу задачи. Так как очевидно – на производительность труда влияют и стаж, и возраст, и их взаимодействие (стаж и возраст). Отсюда следует сделать вывод, что при организации наблюдений необходимо, для каждого уровня факторов, рассматривать, возможно, большее количество элементов в ячейках.

Корреляция используется для количественной оценки взаимосвязи двух наборов данных с помощью коэффициента корреляции. Коэффициент корреляции выборки представляет собой ковариацию двух наборов данных, делённую на произведение их стандартных отклонений (модуль 4).

Ковариация дает возможность установить, ассоциированы ли наборы данных по величине. Например, большие значения из одного набора данных связаны с большими значениями другого набора (положительная ковариация), или, наоборот, малые значения одного набора связаны с большими значениями другого (отрицательная ковариация), или данные двух диапазонов никак не связаны (ковариация близка к нулю).

Описательная статистика. Генерирует статистический отчёт для массива чисел: характеристики положения и вариации, наибольшее и

наименьшее значение, коэффициенты асимметрии и эксцесса, сумму, предельную ошибку доверительного интервала (уровень надёжности $(1-\alpha) \cdot 100\%$) для средней при заданном уровне надёжности (Модуль3).

Экспоненциальное сглаживание. Предназначается для предсказания значения на основе прогноза для предыдущего периода, скорректированного с учетом погрешностей в этом прогнозе. Использует константу сглаживания, по величине которой определяет, насколько сильно влияют на прогнозы погрешности в предыдущем прогнозе (Модуль 4). **Двухвыборочный тест для дисперсии.** Решает задачу сравнения дисперсий двух генеральных совокупностей при заданном уровне значимости α (Модуль3, приложение 5).

Анализ Фурье. Позволяет анализировать периодические данные посредством быстрого преобразования Фурье, которое, однако, предполагает, что количество наблюдений равно 2^{2n} ($n \in \mathbb{N}$). При большом количестве данных (порядка нескольких тысяч) рекомендуется добавлять нули для достижения 2^{2n} (причём наибольшее значение в Excel для $2^{2n} = 4096$).

Гистограмма. Вычисляет выборочные и накопленные частоты попадания ряда данных в интервалы, а также может выводить соответствующие графики (модуль 3). Границы интервалов (карманы) можно указать (обычно задают верхние границы интервалов). Если не задать границы, то Excel сам разобьёт анализируемый ряд на интервалы с шагом h :

$$h = \frac{x_{\max} - x_{\min}}{\sqrt{n}},$$

где x_{\max} -наибольший вариант,
 x_{\min} - наименьший вариант,
 n -ряда (количество членов).

Гистограмму можно построить автоматически для этого достаточно выделить необходимый диапазон и нажать F11.

Скользящее среднее. Прогнозирует значения временного ряда на основе предшествующих значений с помощью простой скользящей средней.

Генерация случайных чисел. Заполняет диапазон случайными числами, заданными по одному из законов: равномерному; нормальному; Бернулли (СВ X принимает значение 1 с вероятностью p и 0 с вероятностью $(1-p)$ (индикаторная СВ)); биномиальному; Пуассона; модельному (позволяющему генерировать последовательности случайных чисел от a до b с шагом c , и возможностью повторения каждого числа и последовательности); дискретному (решающему задачу, получения по имеющемуся распределению, новых значений того же распределения).

Замечание. Инструмент генерации случайных чисел позволяет решать целый ряд задач: численных методов (например, приближённого вычисления определённых интегралов методом статистических испытаний - методом Монте-Карло); имитационного моделирования изучаемых процессов и т.д.

Пример 8 Вычислить определённый интеграл $J = \int_0^1 x^2 dx$ методом Монте-Карло.

Вычисление определенного интеграла J равносильно нахождению площади D криволинейной трапеции функции $Y=f(X)=X^2$ (рис. 34).

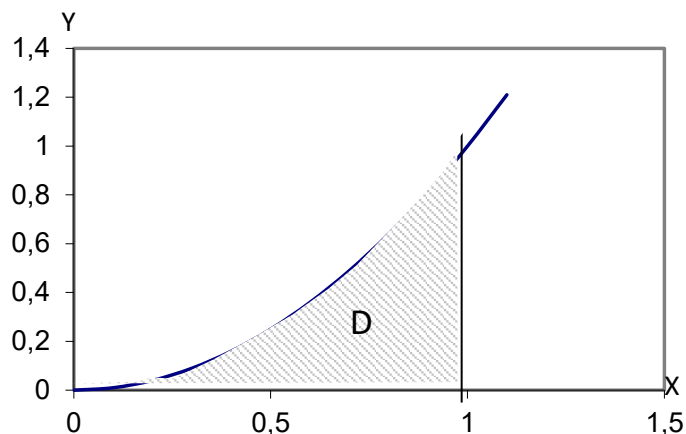


Рисунок 34 - Область D

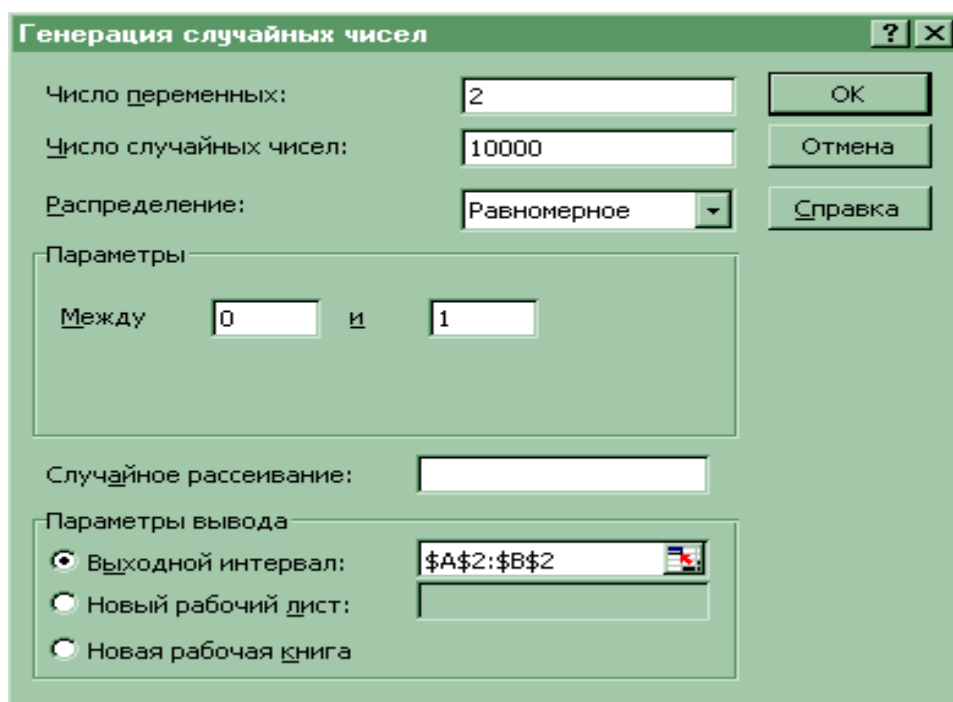


Рисунок 35 - Диалоговое окно генерации двумерных, равномерно распределённых случайных величин

Рассмотрим систему двумерных равномерно распределенных случайных величин (X, Y) на интервале от 0 до 1. При достаточно большом числе опытов N , площадь D будет приближенно равна относительной частоте попадания точек $M_i(x_i, y_i)$ в область D (в силу закона больших чисел):

$$J \approx \frac{n}{N}.$$

Для генерации системы двух равномерно распределенных на интервале от 0 до 1 случайных величин используем инструмент **Генерация случайных чисел**. Заполним диалоговое окно для генерации 10000 пар указанных случайных чисел (рис. 35). В результате в диапазоне A1:B10001 получим искомые пары случайных чисел. В ячейке C2 введем формулу: = A1^2; в ячейке D2: = Если (B2>C2;0;1). Последняя формула присваивает ячейке значение 0, если точка M_i не попадает в область D и значение 1 в противном случае. Выделим диапазон C2:D2 и скопируем вниз до строки 10001. Найдем сумму значений в диапазоне D2:D10001, в результате получим, что $n=\sum m_i=3337$ (рис. 36).

D2		= =ЕСЛИ(B2>C2;0;1)		
	A	B	C	D
1	X_i	Y_i	$f(X_i)$	m_i
2	0,382	0,100681	0,145924	1
3	0,596484	0,899106	0,355793	0
4	0,88461	0,958464	0,782534	0
5	0,014496	0,407422	0,00021	0
6	0,863247	0,138585	0,745195	1
7	0,245033	0,045473	0,060041	1
8	0,03238	0,164129	0,001048	0
9	0,219611	0,01709	0,048229	1
10	0,285043	0,343089	0,081249	0
9998	0,644215	0,951323	0,415013	0
9999	0,878994	0,310404	0,772631	1
10000	0,589068	0,885311	0,347001	0
10001	0,54677	0,571673	0,298957	0
10002	Итого	-	-	3337

Рисунок 36 - Результат применения метода Монте-Карло

Отсюда $J \approx \frac{n}{N} = \frac{3337}{10000} = 0,3337.$

Применяя неравенство Чебышева, имеем

$$P\left(\left|\frac{n}{N} - J\right| < \varepsilon\right) \geq 1 - \frac{J(1-J)}{\varepsilon^2 \cdot N} \geq 1 - \frac{1}{4\varepsilon^2 N}.$$

Если мы зададим уровень значимости α , то неравенство приведенное выше будет всегда верно с гарантийной вероятностью $p=1-\alpha$, при $\alpha = \frac{1}{4\varepsilon^2 N}$.

При заданных значениях ε и α можно определить необходимое число испытаний

$$N = \frac{1}{4\varepsilon^2\alpha}.$$

В силу того, что неравенство Чебышева дает нижнюю оценку вероятности значение N будет завышено, например, в нашем случае при $\varepsilon = 0,001$ и $\alpha = 0,01$: $N = 25000000$. Точное значение $J = 0,(3)$. В рассматриваемом примере точность $\varepsilon=0,001$ достигается уже при 10000 испытаний. (Существуют более точные методы оценки N^1 , основывающиеся на предельных теоремах теории вероятностей).

Ранг и перцентиль. Выводит таблицу с порядковым и процентным рангом для каждого значения в ряде данных.

Регрессия. Подбирает линейную функцию с помощью метода наименьших квадратов (МНК). Основные параметры диалогового окна: *входной интервал Y* - диапазон анализируемых зависимых данных (диапазон должен состоять из одного столбца); *входной интервал X* - диапазон независимых данных (находящихся в соседних столбцах), подлежащих анализу, Excel располагает независимые переменные этого диапазона слева направо в порядке возрастания.

Максимальное число входных диапазонов равно 16. При необходимости можно сформировать и нелинейную функцию. Например, для случая двух переменных

$$Y(X_1, X_2) = a_0 + a_1X_1 + a_2X_2 + a_3X_1^2 + a_4X_2^2 + a_5X_1X_2 -$$

формируется диапазон, содержащий X_1^2, X_2^2, X_1X_2 , рядом с диапазоном переменных X_1, X_2 .

Пример 9. Исследовать влияние фондовооруженности и энергообеспеченности на стоимость валовой продукции.

Результативным признаком (Y) являются стоимость валовой продукции на 100 га сельхозугодий, тыс. руб., которая характеризует затраты на производство продукции.

Факторные признаки:

x_1 – фондовооруженность 1-го работника, характеризующая стоимость основных фондов, в расчете на одного работника (тыс. руб.);

x_2 – энергообеспеченность на 100 га сельхозугодий, л.с., выражающая мощность энергетических ресурсов на 100 га сельхозугодий.

Требуется определить:

- параметры множественного уравнения регрессии в натуральной и стандартизованной форме;
- средние коэффициенты эластичности для каждого фактора;
- коэффициенты частной и множественной корреляции;

¹ Демидович Б.П., Марон И.А. Основы вычислительной математики. - М.:Физматгиз. 1963. - 660 с.

- общий и частные критерии F – Фишера.

Уравнение регрессии примет вид: $y=b_0+b_1x_1+b_2x_2+\varepsilon$

Рассмотрим применение пакета анализа в Excel MS Office 2007 (2010) для решения данной задачи. Исходные данные для анализа введем на листе MS Excel в виде, представленном на рисунке 37.

Для проведения анализа предварительно установим пакет анализа, выполнив последовательно действия: *кнопка Office (в MS Office 2010 – файл) – Параметры Excel – Настройки – Пакет анализа – Перейти* (выделим в окне доступных настроек Пакет анализа), после этого во вкладке Данные ленты появится инструмент Пакет анализа.

	A	B	C
1	Стоимость валовой продукции на 100 га сельхозугодий, тыс. руб.	Фондовооруженность 1-го работника, тыс.руб.	Энергообеспеченность на 100 га сельхозугодий, л.с.
2	Y	X1	X2
3	2715	315	255
4	4190	479	309
5	4219	467	315
6	3193	355	180
7	4699	450	255
8	3574	347	229
9	3806	503	307
10	4891	529	343
11	4120	581	199
12	2989	370	205
13	3722	477	243
14	4544	542	321
15	3005	404	212
16	5364	541	320
17	4216	425	246

Рисунок 37 – Вид исходных данных в MS Excel

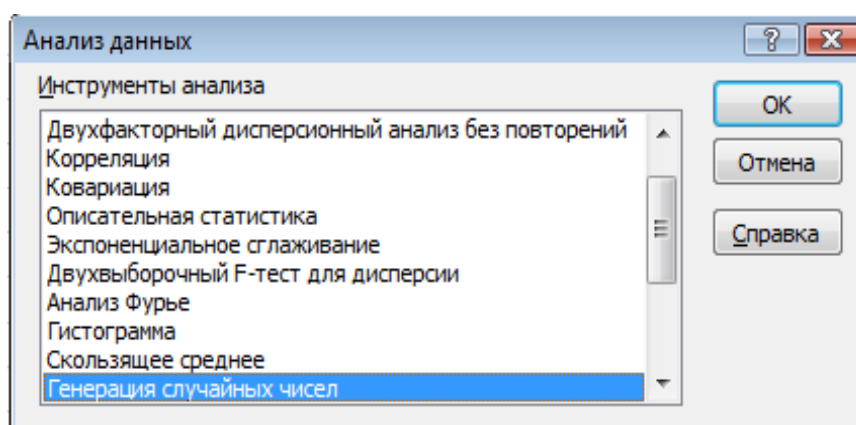


Рисунок 38 – Настройка Пакет анализа (Анализ данных)

Выберем в Пакете анализа инструмент *Описательная статистика* и заполним параметры диалогового окна (рисунки 38, 39).

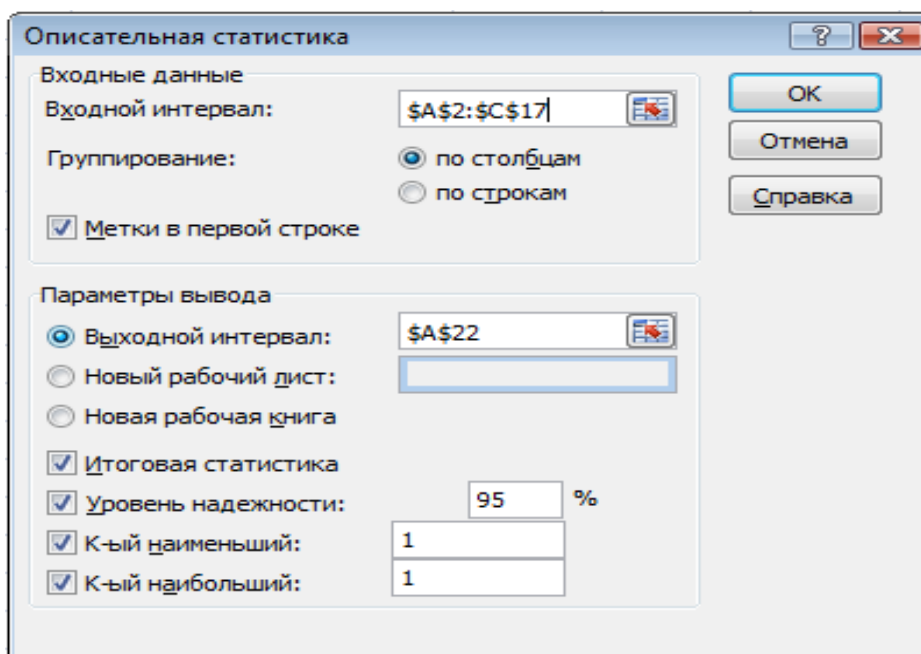


Рисунок 39 – Диалоговое окно инструмента Описательная статистика

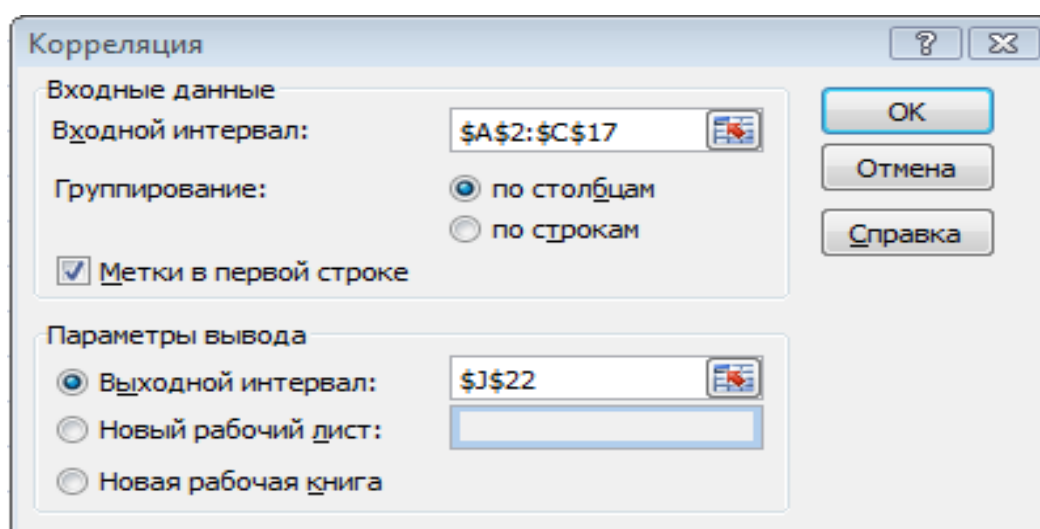


Рисунок 40 – Диалоговое окно инструмента Корреляция

Таблица 13 – Результаты применения инструмента Описательная статистика

	Y	X1	X2	Принятые обозначения
--	---	----	----	----------------------

Среднее	3949,8	452,333	262,6	$\bar{X} = \sum x_i n_i / n$
Стандартная ошибка	197,0888898	20,900	13,57336	$S_{\bar{X}} = S / \sqrt{n}$
Медиана	4120	467	255	Me
Мода	#Н/Д	#Н/Д	255	Mo
Стандартное отклонение	763,322	80,947	52,56941	S
Дисперсия выборки	582660,457	6552,381	2763,543	$S^2 = \sum (x_i - \bar{X})^2 n_i / (n-1)$
Экцесс	-0,674	-1,048	-1,43373	$Ex = \sum ((x_i - \bar{X}) / S)^4 n_i / n - 3$
Асимметричность	0,0436	-0,187	0,026622	$Sk = \sum ((x_i - \bar{X}) / S)^3 n_i / n$
Интервал	2649	266	163	$R = X_{\max} - X_{\min}$
Минимум	2715	315	180	X_{\min}
Максимум	5364	581	343	X_{\max}
Сумма	59247	6785	3939	$\sum x_i$
Счет	15	15	15	$n = \sum n_i$
Наибольший (1)	5364	581	343	-
Наименьший (1)	2715	315	180	-
Уровень надежности (95,0%)	422,714	44,827	29,112	$\Delta = t_{\alpha, n-1} S_{\bar{X}}$

Для нахождения парных коэффициентов корреляции применим инструмент пакета анализа Корреляция, для этого заполним параметры диалогового окна как на рисунке 40.

Таблица 14 – Парные коэффициенты корреляции между признаками

	Y	X1	X2
Y	1		
X1	0,773088239	1	
X2	0,679291686	0,543201832	1

$$r_{yx_1} = 0,7731; \quad r_{yx_2} = 0,6793; \quad r_{x_1x_2} = 0,5432.$$

Парные коэффициенты корреляции свидетельствуют о тесной связи между факторными и результативным признаками, что дает основание включить данные факторы в уравнение регрессии.

Линейное уравнение множественной регрессии в натуральной форме имеет вид $y = b_0 + b_1 x_1 + b_2 x_2 + \varepsilon$ найдем параметры этого уравнения, используя инструмент Пакета анализа – Регрессия. Заполним параметры диалогового окна (рис. 41).

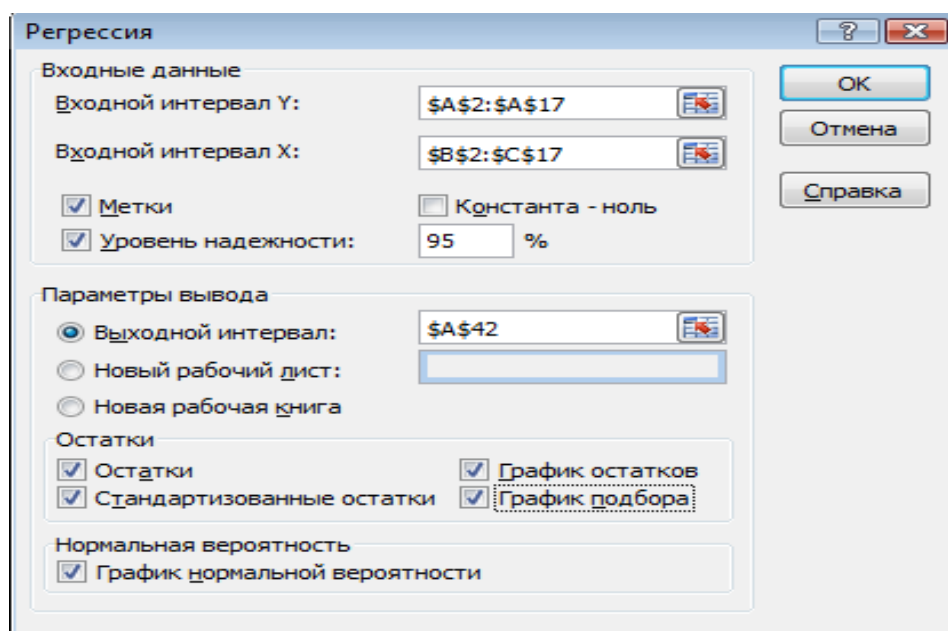


Рисунок 41 – Диалоговое окно инструмента Регрессия

Регрессионная статистика	
Множественный R	0,8325
R-квадрат	0,6931
Нормированный R-квадрат	0,6419
Стандартная ошибка	456,765
Наблюдения	15

Дисперсионный анализ

	df	SS	MS	F	Значимость F	
Регрессия	2	5653635,943	2826817,971	13,549	0,0008	
Остаток	12	2503610,457	208634,2048			
Итого	14	8157246,4				
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y-пересечение	101,8239	748,636	0,136	0,894	-1529,314	1732,962
X ₁	5,4056	1,796	3,009	0,011	1,492	9,319
X ₂	5,3421	2,7658	1,931	0,077	-0,6843	11,368

Рисунок 42 – Вывод итогов регрессионного анализа

Получим линейное уравнение множественной регрессии

$$y = 101,8239 + 5,4056x_1 + 5,3421x_2.$$

Коэффициенты множественной регрессии показывают, что при увеличении фондовооруженности на 1 тыс. руб. стоимость валовой продукции в среднем повысится на 5,406 тыс. руб., а при увеличении энергообеспеченности на 1 л.с. в расчете на 100 га с/х угодий стоимость производства валовой продукции увеличится на 5,342 тыс. руб.

В стандартизованной форме уравнение регрессии имеет вид:

$$t_y = \beta_1 \cdot t_{x_1} + \beta_2 t_{x_2},$$

где

$$t_y = \frac{y - \bar{y}}{\sigma_y}; t_{x_1} = \frac{x_1 - \bar{x}_1}{\sigma_{x_1}}; t_{x_2} = \frac{x_2 - \bar{x}_2}{\sigma_{x_2}}.$$

Найдем β – коэффициенты, используя их связь с коэффициентами b , уравнения регрессии в нормальной форме:

$$\beta_i = b_i \cdot \frac{\sigma_{x_i}}{\sigma_y}.$$

Имеем:

$$\beta_1 = b_1 \cdot \frac{\sigma_{x_1}}{\sigma_y} = 5,4056 \cdot \frac{80,9468}{763,3220} = 0,5732;$$

$$\beta_2 = b_2 \cdot \frac{\sigma_{x_2}}{\sigma_y} = 5,3421 \cdot \frac{52,5694}{763,3220} = 0,3679.$$

β – коэффициенты, можно также найти с помощью парных коэффициентов корреляции по формулам:

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,7731 - 0,6793 \cdot 0,5432}{1 - 0,5432^2} = 0,5732;$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r_{x_1x_2}^2} = \frac{0,6793 - 0,7731 \cdot 0,5432}{1 - 0,5432^2} = 0,3679.$$

Линейное уравнение множественной регрессии в стандартизованном масштабе имеет вид:

$$t_y = 0,5732t_{x_1} + 0,3679t_{x_2}.$$

По абсолютной величине β – коэффициентов можно сделать вывод об относительной силе влияния факторов на изменение результативного признака. На стоимость производства валовой продукции более сильное

влияние оказывает фондовооруженность одного работника, а влияние энергообеспеченности на результаты производства несколько ниже.

Средние коэффициенты эластичности находятся по формуле:

$$\mathcal{E}_{yx_i} = b_i \cdot \frac{\bar{x}_i}{\bar{y}};$$

$$\mathcal{E}_{yx_1} = b_1 \cdot \frac{\bar{x}_1}{\bar{y}} = 5,4056 \cdot \frac{452,333}{3949,8} = 0,619;$$

$$\mathcal{E}_{yx_2} = b_2 \cdot \frac{\bar{x}_2}{\bar{y}} = 5,3421 \cdot \frac{262,6}{3949,8} = 0,355$$

Значит, при увеличении фондовооруженности 1 работника на 1 % выход валовой продукции на 100 га сельхозугодий увеличивается в среднем на 0,619% при неизменности влияния второго фактора. Изменение энергообеспеченности предприятия на 1 % приводит к росту стоимости валовой продукции на 0,355%, при исключении влияния первого фактора

Коэффициенты частной корреляции определяются через парные коэффициенты корреляции по формулам:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_2}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,7731 - 0,6793 \cdot 0,5432}{\sqrt{(1 - 0,6793^2)(1 - 0,5432^2)}} = 0,6558$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1 x_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{x_1 x_2}^2)}} = \frac{0,6793 - 0,7731 \cdot 0,5432}{\sqrt{(1 - 0,7731^2)(1 - 0,5432^2)}} = 0,4871$$

$$r_{x_1 x_2 \cdot y} = \frac{r_{x_1 x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1 - r_{yx_1}^2) \cdot (1 - r_{yx_2}^2)}} = \frac{0,5432 - 0,7731 \cdot 0,6793}{\sqrt{(1 - 0,7731^2)(1 - 0,6793^2)}} = 0,0470.$$

Коэффициенты частной корреляции характеризуют тесноту связи между двумя переменными, исключив влияние третьей переменной. Значит связь между стоимостью валовой продукции и фондовооруженностью прямая и достаточно тесная, между стоимостью валовой продукции и энергообеспеченностью при исключении влияния фондовооруженности средняя. Связь между собой факторных признаков очень слабая.

Коэффициент множественной корреляции находится по формуле:

$$R_{yx_1x_2} = \sqrt{\beta_1 \cdot r_{yx_1} + \beta_2 r_{yx_2}} = \sqrt{0,6732 \cdot 0,7731 + 0,3679 \cdot 0,6793} = \\ = \sqrt{0,5205 + 0,2499} = \sqrt{0,7704} = 0,8777.$$

Более точный результат (без округлений) можно получить, согласно рисунка 42:

$$R = \sqrt{\frac{\sum(\bar{y} - \hat{y})^2}{\sum(y_i - \hat{y})^2}} = \sqrt{\frac{5653635,943}{8157246,4}} = \sqrt{0,6931} = 0,8325.$$

Величина коэффициента показывает, что связь между Y , X_1 и X_2 довольно тесная, причем 69,3 % вариации стоимости валовой продукции объясняется вариацией фондовооруженности и энергообеспеченности.

Оценим значимость уравнения регрессии и коэффициента R_2 с помощью критерия F – Фишера. Фактически рассматривается нулевая гипотеза $H_0: R_2 = 0 (b_1 = b_2 = 0)$ и альтернативная гипотеза $H_1: R_2 \neq 0, b_1 \neq 0, b_2 \neq 0$.

Наблюдаемое или фактическое значение критерия находится по формуле:

$$F_n = \frac{R_{yx_1x_2}^2}{1 - R_{yx_1x_2}^2} : \frac{m}{n - m - 1},$$

где m – число факторов в линейном уравнении регрессии;

n – число единиц наблюдения.

$$F_n = \frac{0,6931}{1 - 0,6931} : \frac{2}{15 - 2 - 1} = 13,12.$$

При уровне значимости $\alpha = 0,05$ и числе степеней свободы $k_1 = m = 2$, $k_2 = n - m - 1 = 15 - 2 - 1 = 12$ по таблице значений критерия F – Фишера критическое значения составляет 3,80, т.е. $F_{кр} = 3,80$. Сравниваем F_n с $F_{кр}$. Так как $F_n > F_{кр}$, то нулевую гипотезу о незначимости величины R^2 отклоним, т.е. уравнение множественной регрессии и R^2 статистически значимы.

В уравнении множественной регрессии не все факторы могут оказывать статистически существенное влияние на изменение результативного признака. Оценка значимости факторов в уравнении регрессии может быть дана с помощью частного F – критерия или критерия t – Стьюдента.

$$F_{nx_1} = \frac{R_{yx_1x_2}^2 - r_{yx_2}^2}{1 - R_{yx_1x_2}^2} \cdot \frac{n - m - 1}{1} = \frac{0,6931 - 0,6793^2}{1 - 0,6931} \cdot \frac{15 - 2 - 1}{1} = 8,772.$$

При $\alpha = 0,05$, $k_1 = 1$, $k_2 = 12$, $F_{кр} = 4,75$.

Так как $F_{\text{нк}1} > F_{\text{кр}}$, то в уравнение регрессии целесообразно включение фактора x_1 после x_2 . Фактор X_1 оказывает статистически значимое влияние на Y .

$$F_{\text{нк}2} = \frac{R^2_{y_{x_1x_2}} - r^2_{y_{x_1}}}{1 - R^2_{y_{x_1x_2}}} \cdot \frac{n - m - 1}{1} = \frac{0,6931 - 0,7731^2}{1 - 0,6931} \cdot \frac{15 - 2 - 1}{1} = 3,6131.$$

$F_{\text{нк}2} < F_{\text{кр}}$ - это свидетельствует о статистической незначимости влияния фактора X_2 и нецелесообразности включения его в уравнение множественной регрессии. В данной задаче на стоимость валовой продукции статистически значимое влияние оказывает фондовооруженность одного работника.

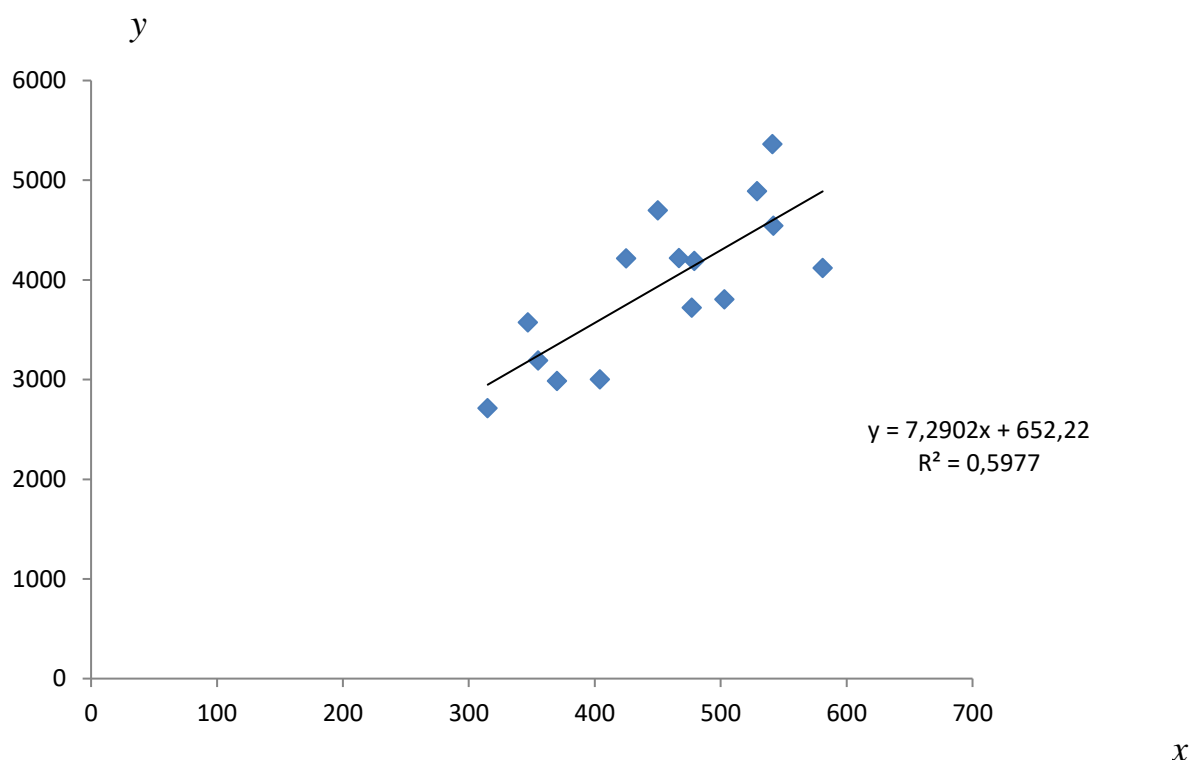


Рисунок 43 – Парное уравнение регрессии

Уравнение парной регрессии между X_1 и Y :

$$Y = 652,22 + 7,2902x_1.$$

Значит, при увеличении фондовооруженности одного работника на 1 тыс. руб. производство валовой продукции увеличивается на 7,29 тыс. руб.

Полученное уравнение объясняется 59,77 % различий в стоимости валовой продукции (что на 9,54 % меньше, чем уравнение с двумя факторами).

Выборка. Создаёт выборку случайным или периодическим (серийным) способом в предположении, что заданный диапазон является генеральной совокупностью.

Парный двухвыборочный тест для средних.

Двухвыборочный t - тест с одинаковыми дисперсиями.

Двухвыборочный t - тест с различными дисперсиями.

Двухвыборочный z- тест для средних.

Инструменты *Пакета анализа* 16-19 позволяют проверить существенность различий между двумя средними для совокупностей, заданных в двух диапазонах.

Задания

По данным приложения А определить:

- параметры множественного уравнения регрессии в натуральной и стандартизованной форме;
- средние коэффициенты эластичности для каждого фактора;
- коэффициенты частной и множественной корреляции;
- общий и частные критерии F – Фишера.

Вопросы для самоконтроля

1. Опишите возможности ввода информации.
2. Дисперсия, ее свойства и способы расчета.
3. Как проводится дисперсионный анализ.
4. Опишите возможности пакета анализа данных.

ЛАБОРАТОРНАЯ РАБОТА № 4

Анализ временных рядов в Excel 2007

Цель работы: ознакомиться с возможностями методик анализа временных рядов, получить навыки анализа данных с использованием Excel 2007.

Теоретические сведения

Одной из важнейших задач статистики является изучение изменения экономических явлений во времени путем построения и анализа рядов динамики. **Ряд динамики** представляет собой численные значения статистического показателя в последовательные моменты или периоды времени.

В ряду динамики выделяют два элемента: периоды или моменты

времени (t) и соответствующие им количественные значения показателя, называемые уровнями ряда (y_t). Если уровни ряда характеризуют значения показателя за определенный период времени, то ряд называется интервальным, а если на определенный момент времени – моментным. Уровни выражаются абсолютными, относительными и средними величинами.

Пример 10 Рассмотрим задачу прогнозирования урожайности подсолнечника, при наличии данных с 1975г по 2009г (за 35 лет), для этого рассмотрим ряд моделей и выберем ту из них, которая даёт наименьшую ошибку в 2010 году, хотя можно в качестве критерия отбора выбрать минимальную сумму модулей ошибок точек или минимальную дисперсию для последних k -точек и т. д. - в конечном счете, вид лучшего критерия мы узнаем только в 2010 году.

Предполагая, что тенденция изменения урожайности не изменится в ближайшие годы, для прогноза экстраполируем выбранную модель для первых 3-5 лет после 2009 года. Используя мастер диаграмм, получим график ряда динамики урожайности (рис. 44).

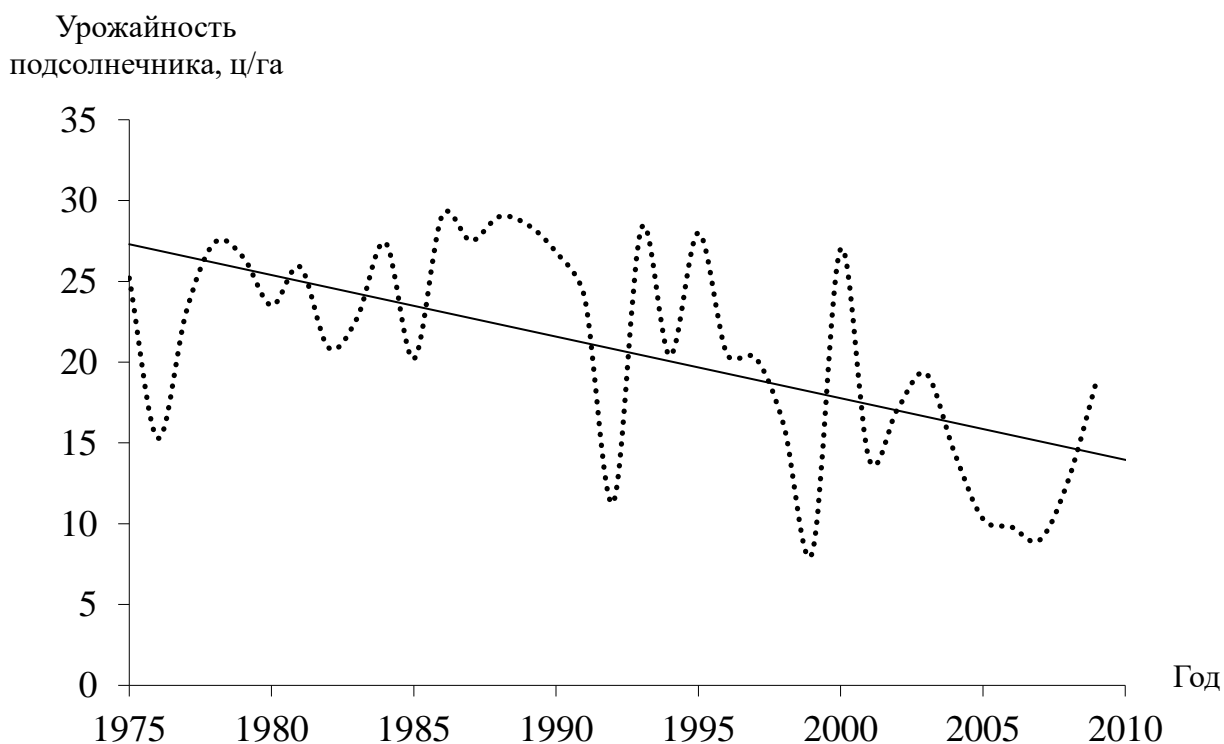


Рисунок 44 – Динамика урожайности подсолнечника

Для аналитического выравнивания и прогноза по уравнению прямой $Y_t = a_0 + bt$ можно использовать в категории Статистические следующие функции (причём для отсчёта времени перейдём к условным годам $t: 1, 2, \dots, 35$) (рис. 45):

а) ПРЕДСКАЗ (выделим диапазон D2:D40 и в ячейку D2 введём формулу массива: $\{=ПРЕДСКАЗ(B2:B40;C2:C36;B2:B36)\}$);

б) ТЕНДЕНЦИЯ (выделим диапазон E36:E40 и в ячейку E36 введём формулу массива: {=ТЕНДЕНЦИЯ(C2:C36;B2:V36;V36:V40;1)});

в) ЛИНЕЙН - позволяет получить коэффициенты уравнения регрессии помощью МНК, которые можно использовать в формуле для выравнивания и прогноза.

Для аналитического выравнивания по уравнению экспоненты $y=b*m^x$:

а) РОСТ (используется для предсказания или выравнивания по экспоненциальной кривой, выделим диапазон F2:F40 и введём формулу массива {=РОСТ(C2: C36;B2:V36;B2:V40;1)});

б) ЛГРФПРИБЛ может использоваться аналогично ЛИНЕЙН.

Самый простой способ прогнозирования на основе линейного или экспоненциального тренда заключается в использовании контекстного меню. Необходимо: 1) выделить диапазон данных; 2) при нажатой правой клавише мыши, протащить маркер заполнения на необходимый период прогнозирования; 3) в открывшемся контекстном меню выбрать вид приближения: Линейное, Экспоненциальное, Прогрессия (рис. 46).

D2		fx {=ПРЕДСКАЗ(B2:V40;C2:C36;B2:V36)}				
	A	B	C	D	E	F
	ГОД	Порядковый номер года, t	Урожайность подсолнечника, ц/га	ПРЕДСКАЗ	ТЕНДЕНЦИЯ	РОСТ
1						
2	1975	1	25,2	27,29571429		28,3618191
3	1976	2	15,3	26,91462185		27,75293051
4	1977	3	23,1	26,53352941		27,15711391
5	1978	4	27,4	26,15243897		26,57408865
6	1979	5	26,5	25,77134454		26,00358013
7	1980	6	23,5	25,3902521		25,44531963
8	1981	7	25,9	25,00915966		24,8990442
9	1982	8	20,9	24,62808723		24,36449655
10	1983	9	22,7	24,24697479		23,84142488
32	2005	31	10,3	15,86294118		14,79030229
33	2006	32	9,8	15,48184874		14,47277518
34	2007	33	9	15,1007563		14,16206494
35	2008	34	12,7	14,71966387		13,85802521
36	2009	35	18,9	14,33857143	14,33857143	13,56051279
37	2010	36		13,95747899	13,95747899	13,26938754
38	2011	37		13,57638655	13,57638655	12,98451235
39	2012	38		13,19529412	13,19529412	12,70575303
40	2013	39		12,81420168	12,81420168	12,43297828

Рисунок 45 - Результаты линейной и экспоненциальной экстраполяции

Замечание. Все приведённые выше формулы можно (даже нужно) вводить не в ручную, а используя, МАСТЕР ФУНКЦИЙ, категорию Статистические; так как рассматриваемые выше формулы обрабатывают массивы данных, то после их введения необходимо нажать Ctrl + Shift + Enter.

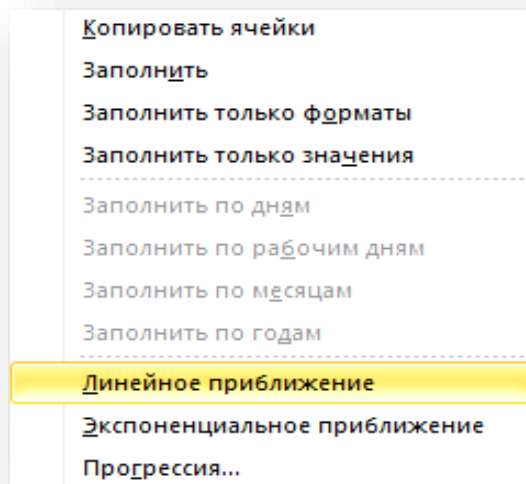


Рисунок 46 -

Контекстное

меню для экстраполяции

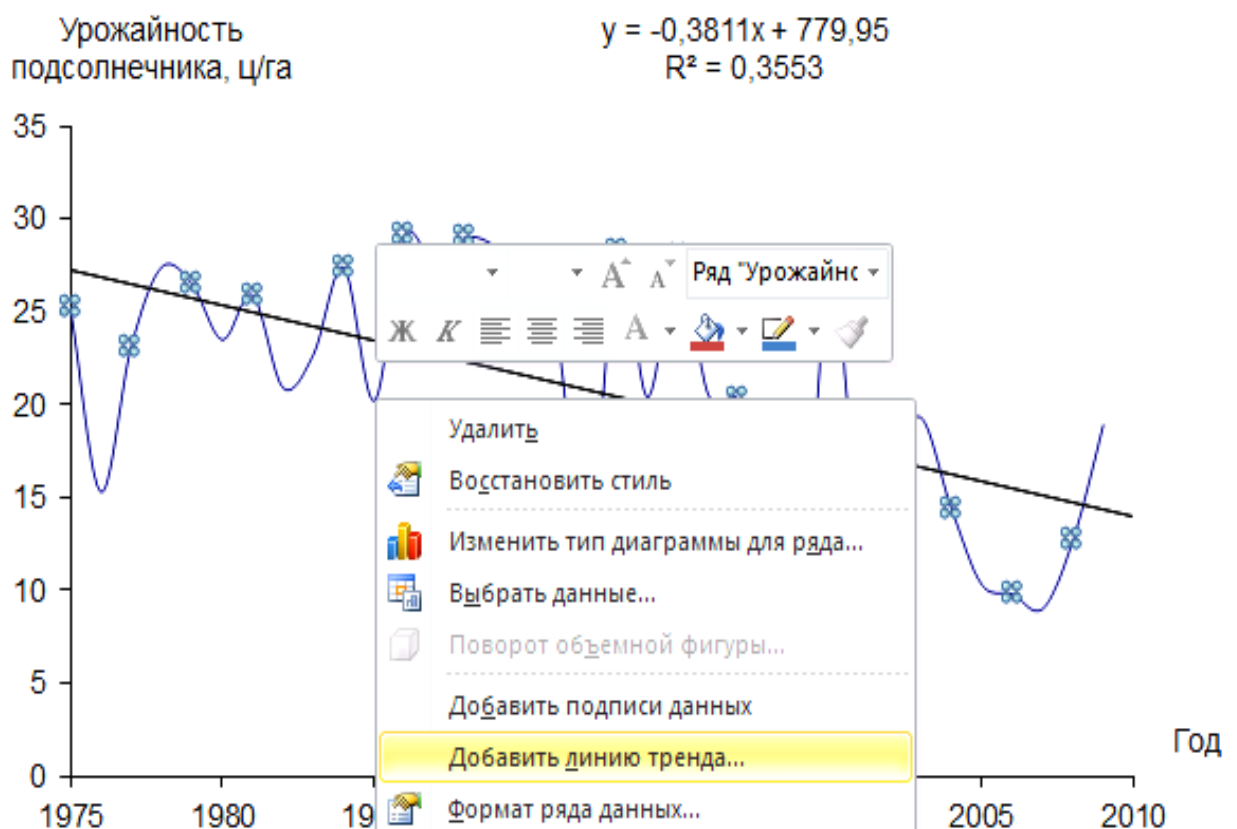


Рисунок 47 - Контекстное меню выделенных точек наблюдений

Важным методом анализа временных рядов в Excel являются диаграммы. Выделим на (рис.47) щелчком левой клавиши мыши маркеры наблюдений урожайности подсолнечника по годам; с помощью правой клавиши откроем контекстное меню (рис.48) и выберем одну из перечисленных линий трендов

(рис.48):

Линейная,
Логарифмическая,
Полиномиальная,
Степенная,
Экспоненциальная,
Линейная фильтрация (Скользящая средняя).

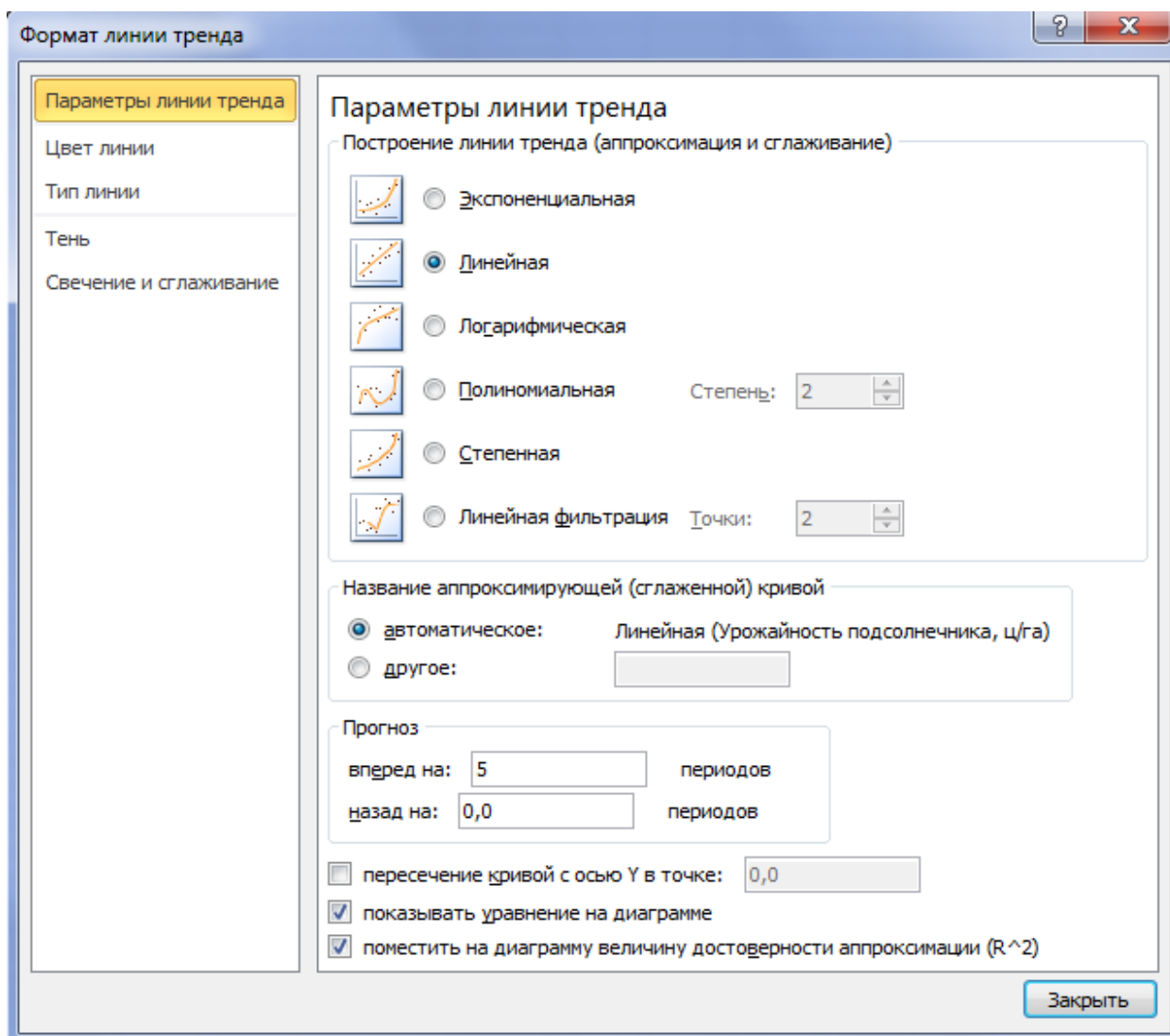


Рисунок 48 - Диалоговое окно выбора линии тренда

После выбора одного из трендов, например, линейного - отметим «показывать уравнение на диаграмме» и «поместить на диаграмму коэффициент достоверности аппроксимации (R^2)» (рис. 49).

Можно выбрать название (назвать тренд самостоятельно) или оставить автоматически предлагаемое Excel; для прогноза согласно выбранной линии тренда на 5 лет вперёд выберем соответствующее значение в диалоговом окне; Для отображения на диаграмме уравнения тренда и коэффициента

детерминации отметим соответствующие элементы вкладки Параметры (рис. 48). Далее выберем ОК.

При выборе других типов линии тренда получим рис.49.

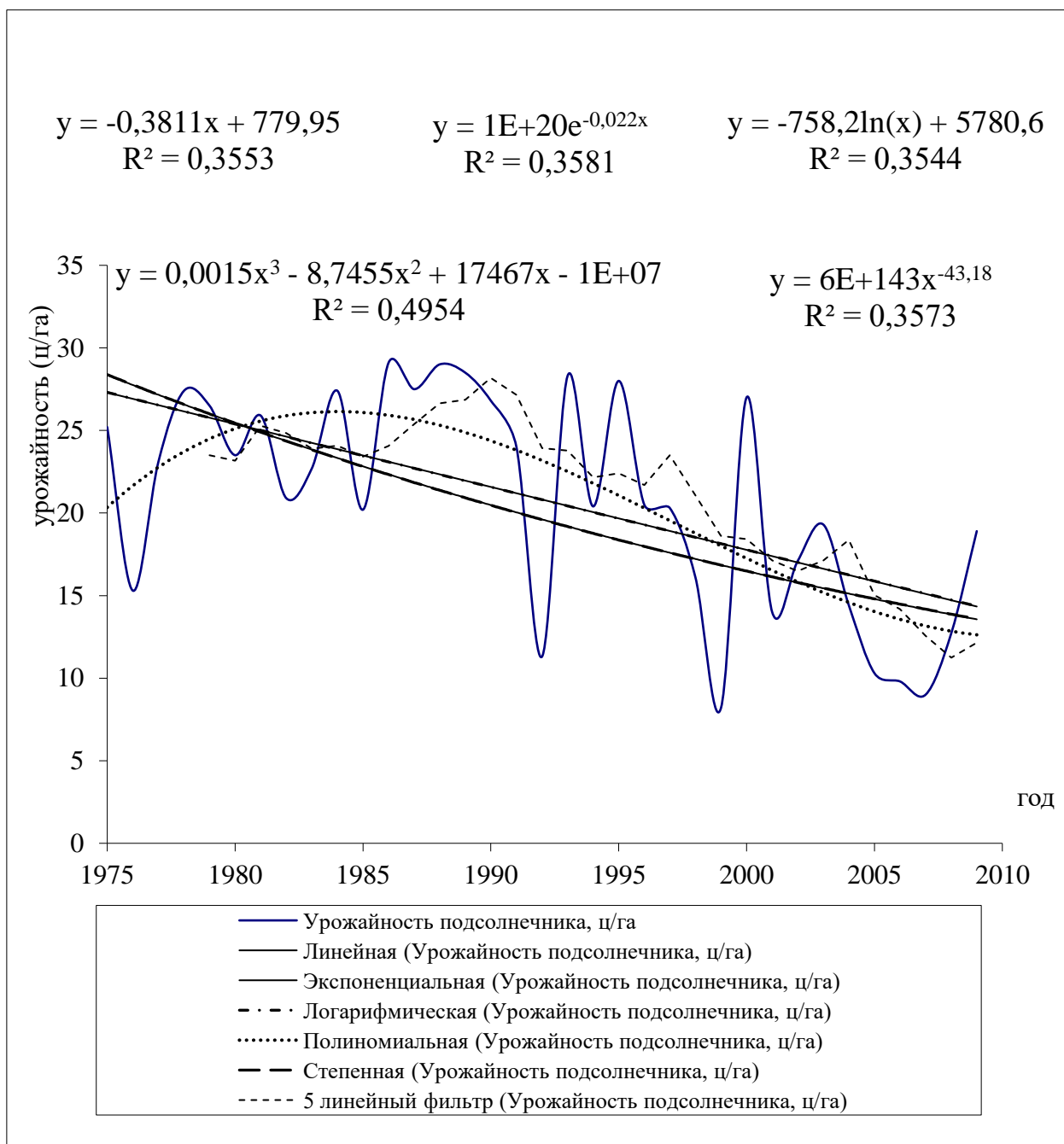


Рисунок 49 - Линии тренда

Рассмотрим другие типы моделей (рис.50).

Прогноз по уравнению третьей степени можно получить, используя рисунок расположенный ранее: $Y = 0,0015 \cdot T^3 - 8,7017 \cdot T^2 + 17292 \cdot T - 1E+07$, где T год. Но лучше (так как погрешность расчётов меньше) $Y = 0,0015 \cdot t^3 - 0,102 \cdot t^2 + 1,6043 \cdot t + 18,838$, где t - порядковый номер года (для прогноза на 2010 - 2013 годы, $t = 36, 37, 38, 39$).

Выравнивание и прогноз по СС - скользящим средним проведем с использованием пакета анализа (инструмент Скользящее среднее).

Взвешенную скользящую среднюю (ВСС) для 5 точек определим для полинома 3 степени (модуль 4) по формуле:

$$z_0 = \frac{1}{35}(-3z_{-2} + 12z_{-1} + 17z_0 + 12z_1 - 3z_2)$$

Для прогнозирования и определения последних 2-х точек, которые нельзя получить с использованием ВСС, с помощью контекстного меню можно определить уравнение полинома третьей степени по последним 5 точкам (предварительно изобразив их на отдельной диаграмме): $y = 0,2333x^3 - 0,8214x^2 - 0,1548x + 11,14$, где k - порядковый номер года (для прогноза на 2000-2003 годы k=6,7,8,9).

Год	Подсол нечник	Прогноз по линейному уравнению	Прогноз по уравнению полинома 3 степени	СС 3точки	СС 4точки	СС 5точек	СС 11точек	ВСС 5точек
1975	25.2	23.5	20.3	-	-	-	-	-
1976	15.3	23.1	21.7	-	-	-	-	-
1977	23.1	22.7	22.8	21.2	-	-	-	21.4
1978	27.4	22.3	23.7	21.9	22.8	-	-	27.0
1979	26.5	21.9	24.5	25.7	23.1	23.5	-	26.1
1980	23.5	21.6	25.1	25.8	25.1	23.2	-	25.2
1981	25.9	21.2	25.6	25.3	25.8	25.3	-	23.6
1982	20.9	20.8	25.9	23.4	24.2	24.8	-	22.5
1983	22.7	20.4	26.1	23.2	23.3	23.9	-	23.6
1984	27.4	20.0	26.1	23.7	24.2	24.1	-	23.7
1985	20.2	19.7	26.1	23.4	22.8	23.4	23.5	24.9
1986	29.1	19.3	25.9	25.6	24.9	24.1	23.8	25.7
1987	27.5	18.9	25.7	25.6	26.1	25.4	24.9	29.1
1988	29.0	18.5	25.3	28.5	26.5	26.6	25.5	28.5
1989	28.5	18.1	24.9	28.3	28.5	26.9	25.6	28.6
1990	26.8	17.8	24.4	28.1	28.0	28.2	25.6	27.6
1991	24.0	17.4	23.8	26.4	27.1	27.2	25.6	19.9
1992	11.3	17.0	23.2	20.7	22.7	23.9	24.3	19.4
1993	28.3	16.6	22.5	21.2	22.6	23.8	25.0	20.2
1994	20.4	16.2	21.8	20.0	21.0	22.2	24.8	26.5
1995	28.0	15.8	21.1	25.6	22.0	22.4	24.8	23.5
1996	20.5	15.5	20.3	23.0	24.3	21.7	24.9	23.4
1997	20.3	15.1	19.5	22.9	22.3	23.5	24.1	19.3
1998	16.1	14.7	18.8	19.0	21.2	21.1	23.0	13.5
1999	8.2	14.3	18.0	14.9	16.3	18.6	21.1	15.8
2000	27.0	13.9	17.3	17.1	17.9	18.4	21.0	17.9
2001	14.0	13.6	16.5	16.4	16.3	17.1	19.8	19.6
2002	17.1	13.2	15.8	19.4	16.6	16.5	19.2	16.2
2005	10.3	12.0	14.0	14.7	15.3	15.0	17.7	10.9
2006	9.8	11.7	13.6	11.5	13.5	14.2	16.1	9.1
2007	9.0	11.3	13.2	9.7	10.9	12.6	15.0	9.6
2008	12.7	10.9	12.8	10.5	10.5	11.2	14.4	12.3
2009	18.9	10.5	12.6	13.5	12.6	12.1	14.6	19.0
2010	13.3	10.1	12.5	15.0	13.5	12.7	15.1	31.0
2011	13.0	9.7	12.5	15.1	14.5	13.4	13.8	49.8
2012	12.7	9.4	12.6	13.0	14.5	14.1	13.7	76.8
2013	12.4	9.0	12.9	12.7	12.8	14.1	13.3	113.3

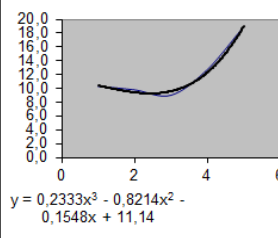


Рисунок 50 - Прогнозирование по линиям тренда

В качестве модели прогноза целесообразно выбрать модель, имеющую наименьшую ошибку в последней точке (у нас это модель линейной зависимости и 11-летней скользящей средней), исходя из предположения о сохранении тенденции в дальнейшем.

Проверим автокорреляцию временного ряда (корреляцию между

соседними значениями ряда, то есть, влияют ли предыдущие значения на последующие).

Для этого расположим рядом пять столбцов урожайности, сдвинутых каждый относительно предыдущего на единицу (в случае предположения о стационарности ряда смещённые данные каждого столбца можно перенести в его начало).

В результате применения инструмента Корреляция (Пакет анализа) получим корреляционную матрицу (табл. 15).

Таблица 15 - Корреляционная матрица

	Столбец 1	Столбец 2	Столбец 3	Столбец 4	Столбец 5
Столбец 1	1				
Столбец 2	0,39447398	1			
Столбец 3	0,4662682	0,39447398	1		
Столбец 4	0,3200895	0,4662682	0,394474	1	
Столбец 5	0,19072011	0,3200895	0,4662682	0,394474	1

Очевидно, что уравнение авторегрессии будет иметь вид:

$$Y_t = a_0 + a_1 Y_{t-1} + a_2 Y_{t-2}$$

Так как соответствующие коэффициенты автокорреляции значительно отличны от нуля по сравнению с другими. Его легко построить, используя аппарат регрессионного анализа (однако, результат будет не совсем достоверным потому, что не выполняется пятое условие применимости МНК - отсутствие автокорреляции (модуль 4), адекватное решение можно получить, решая систему уравнений Юла - Уокера [6]).

Рассмотрим получение основных параметров распределения на примере урожайности подсолнечника (ц/га).

Выполним команду Анализ данных - Описательная статистика, заполним параметры диалогового окна (при уровне значимости $\alpha=0,05$). В результате получим первые два столбца табл. 16, в третьем столбце нами указаны принятые обозначения для статистик.

Таблица 16 - Описательная статистика

Урожайность подсолнечника, ц/га		Принятые обозначения
Среднее	20,81714286	$\bar{X} = \sum x_i n_i / n$
Стандартная ошибка	1,10738913	$S_{\bar{x}} = S / \sqrt{n}$
Медиана	20,9	Me
Мода	27,4	Mo
Стандартное отклонение	6,551402441	S
Дисперсия выборки	42,92087395	$S^2 = \sum (x_i - \bar{X})^2 n_i / (n-1)$
Экссесс	-1,015633106	Ex
Асимметричность	-0,479752816	Sk
Интервал	20,9	W=xmax - xmin
Минимум	8,2	xmin
Максимум	29,1	xmax
Сумма	728,6	$\sum x_i$
Счет	35	$n = \sum n_i$
Наибольший(1)	29,1	-
Наименьший(1)	8,2	-
Уровень надежности(95,0%)	2,250483999	$\Delta = t_{\alpha;n-1} S_{\bar{x}}$

Проведем выравнивание по ряду Фурье для кукурузы (z_t).

Предварительно введя данные и перейдя к условным годам, введём формулы (рис. 51):

C2:=(ATAN(1)*8/35)*(A2-1);
D2:=COS(C2)*B2; E2:=COS(2*C2)*B2;
F2:=COS(3*C2)*B2; G2:=COS(4*C2)*B2;
H2:=SIN(C2)*B2; I2:=SIN(2*C2)*B2;
J2:=SIN(3*C2)*B2; K2:=SIN(4*C2)*B2.

Формулу в ячейке C2 скопируем для диапазона C3:C42 (последние пять ячеек соответствуют пяти годам (2000-2004) на которые мы будем давать прогноз); формулы диапазона D2:K2 можно выделить и протаскив мышью маркер заполнения скопировать для диапазона C3:K36.

Просуммируем элементы диапазонов B2:B36 и D2:K36 по столбцам (для этого достаточно выделить диапазон и нажать Σ - автосумма).

Введём в ячейке B38 " $=B37/35$ ", а в D38 " $=(2/35)*D37$ " и скопируем последнюю формулу для диапазона E38:K38, в результате получим коэффициенты разложения Фурье:

$$z_t = a_0 + \sum_{j=1}^k (a_j \cos jt + b_j \sin jt)$$

(в настоящем примере рассматривается $k = 4$ – четыре гармоники), где

$$a_0 = \frac{1}{n} \sum_{t=1}^n z_t, \quad a_j = \frac{2}{n} \sum_{t=1}^n z_t \cos jt, \quad b_j = \frac{2}{n} \sum_{t=1}^n z_t \sin jt.$$

▲	A	B	C	D	E	F	G	H	I	J	K	L
1	n	Кукуруза	t	cost	cos2t	cos3t	cos4t	sint	sin2t	sin3t	sin4t	y
2	1	70,3	0,000	70,300	70,300	70,300	70,300	0,000	0,000	0,000	0,000	47,428
3	2	44,2	0,180	43,490	41,382	37,943	33,286	7,892	15,531	22,670	29,081	53,175
4	3	48,4	0,359	45,314	36,449	22,935	6,497	17,007	31,844	42,621	47,962	52,411
5	4	50,9	0,539	43,695	24,120	-2,284	-28,041	26,107	44,822	50,849	42,480	45,434
6	5	40,3	0,718	30,349	5,410	-22,201	-38,848	26,515	39,935	33,633	10,721	35,153
7	6	22,5	0,898	14,029	-5,007	-20,272	-20,272	17,591	21,936	9,762	-9,762	25,855
8	7	20	1,077	9,477	-11,018	-19,919	-7,861	17,612	16,691	-1,793	-18,391	21,356
9	8	16,3	1,257	5,037	-13,187	-13,187	5,037	15,502	9,581	-9,581	-15,502	23,374
10	9	34,6	1,436	4,644	-33,353	-13,599	29,702	34,287	9,205	-31,816	-17,746	30,869
11	10	56,9	1,616	-2,553	-56,671	7,638	55,986	56,843	-5,100	-56,385	10,160	40,631
12	11	20,5	1,795	-4,562	-18,470	12,782	12,782	19,986	-8,895	-16,028	16,028	48,822
13	12	65,9	1,975	-25,900	-45,541	61,698	-2,957	60,597	-47,632	-23,156	65,834	52,732
30	29	52,7	5,027	16,285	-42,635	-42,635	16,285	-50,121	-30,976	30,976	50,121	39,832
31	30	11,9	5,206	5,639	-6,556	-11,852	-4,677	-10,479	-9,931	1,067	10,942	32,228
32	31	30,1	5,386	18,767	-6,698	-27,119	-27,119	-23,533	-29,345	-13,060	13,060	24,702
33	32	32,9	5,565	24,776	4,416	-18,125	-31,714	-21,646	-32,602	-27,457	-8,753	20,201
34	33	35,7	5,745	30,647	16,917	-1,602	-19,667	-18,311	-31,437	-35,664	-29,794	21,013
35	34	6,2	5,924	5,805	4,669	2,938	0,832	-2,179	-4,079	-5,460	-6,144	27,482
36	35	24,6	6,104	24,205	23,031	21,118	18,526	-4,392	-8,644	-12,617	-16,185	37,547
37	36	1558	6,283	-232,344	93,608	122,577	67,146	-55,895	106,688	28,242	144,008	47,428
38	37	44,51428571	6,463	-13,277	5,349	7,004	3,837	-3,194	6,096	1,614	8,229	53,175
39	38	a0	6,642	a1	a2	a3	a4	b1	b2	b3	b4	52,411
40	39		6,822									45,434
41	40		7,001									35,153
42	41		7,181									25,855

Рисунок 51 - Экстраполяция урожайности кукурузы (по ряду Фурье)

В ячейку L2 введем формулу (*), которая у нас будет иметь вид:

"=B\$38+\$D\$38*COS(C2)+\$E\$38*COS(C2*2)+\$F\$38*COS(C2*3)+\$G\$38*COS(C2*4)+\$H\$38*SIN(C2)+\$I\$38*SIN(2*C2)+\$J\$38*SIN(3*C2)+\$K\$38*SIN(4*C2)".

Скопируем формулу из L2 для диапазона L3:L42, в результате в диапазоне L2:L36 получим аппроксимацию с помощью ряда Фурье (1972-2006 гг.), а в диапазоне L37:L42 - прогноз (экстраполяцию) (2008-2012 гг.) (рис. 51).

Построим графики наблюдений (урожайности кукурузы) и её аппроксимации с помощью ряда Фурье. Для этого используем кнопку Мастер диаграмм, расположенную на панели инструментов Стандартная. Выберем График с маркерами, введём диапазоны соответствующих значений (чтобы подписи оси X были годами необходимо ввести ссылку ("Подписи оси X") на диапазон, содержащий соответствующие года).

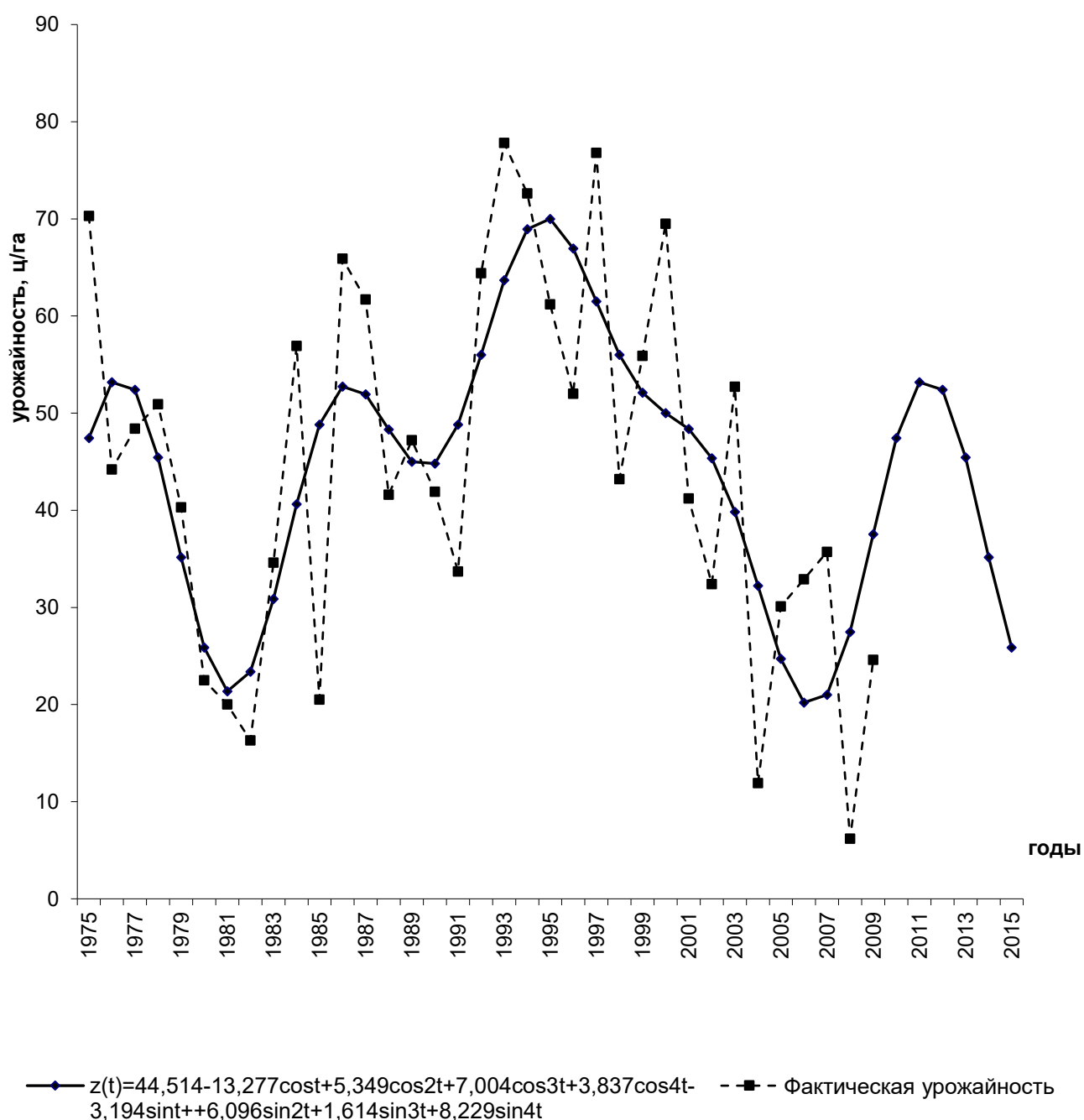


Рисунок 52 – Динамика урожайности кукурузы на зерно в ОАО «Заря»

Задание

По данным приложения провести анализ и прогнозирование временных рядов с использованием различных моделей. Сделать вывод.

Вопросы для самоконтроля

1. С какой целью проводится анализ временных рядов.
2. Как может быть выявлена основная тенденция в изменениях уровней временных рядов.
3. Как выполнить прогноз на будущее с помощью Excel 2007.
4. Что понимается под интерполяцией и экстраполяцией.

ЛАБОРАТОРНАЯ РАБОТА № 5

Финансовые вычисления в Excel 2007

Цель работы: Дать целостную концепцию количественного финансового анализа условий и результатов финансово-кредитных и коммерческих сделок, связанных с предоставлением денег в долг.

Теоретические сведения

Финансово-экономические расчеты - это область знаний, в которой излагается методология количественного финансового анализа условий и результатов финансово-кредитных и коммерческих сделок. Они представляют собой совокупность методов и приемов определения изменения стоимости денег, происходящего вследствие их возвратного движения. Процент рассматривается как плата за пользование заемными средствами, так и показатель доходности любого вложения капитала.

Существует два метода начисления процентов: декурсивный и антисипативный.

Пример 11 Расчет наибольшей суммы кредита, который, может Вам выдать банк, проводится по формуле

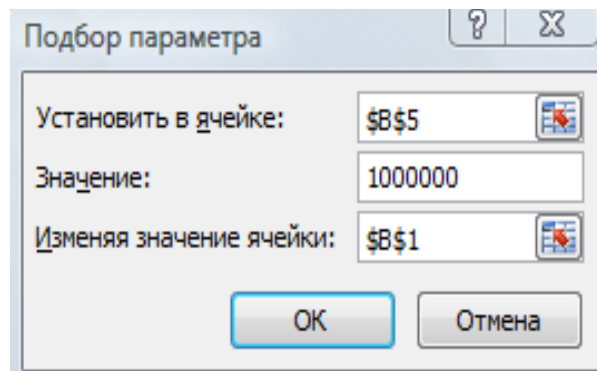
$$\text{Максимальный кредит} = \frac{\text{заработная плата} \times \text{срок кредита (мес)} \times 0,6}{1 + \frac{(\text{срок кредита (мес)} + 1) \times \text{процент. ст.}}{2 \times 12 \times 100}}$$

- 1) У Вас заработная плата 25000 руб. Найдите наибольшую сумму кредита. Заполним лист Excel

	A	B
1	заработная плата, руб.	25000
2	срок кредита, мес.	60
3	процентная ставка, %	19
4		
5	Максимальная сумма кредита	<code>=(B1*B2*0,6)/(1+(B2+1)*B3/(2*12*100))</code>

В результате получим, что наибольшая сумма кредита на пять лет 606912 руб.

- 2) Найдём необходимую ежемесячную заработную плату для получения кредита 1000000 руб. на три года. Для этого изменим срок кредита на 36 мес., на вкладке Данные в группе Работа с данными выберем команду Анализ «что-если», а затем – в списке пункт Подбор параметра и заполним параметры диалогового окна.



В результате получим необходимую заработную плату в размере 41192,12 руб.

Пример 12 01.03.2010г. Вы хотите взять в банке «АЛЬФА» кредит 100 000 руб. на неотложные нужды под 17 % годовых на 5 лет. Составьте примерный график платежей по кредиту и дайте на его основе оценку эффективной кредитной ставки.

Заполните на лист Excel согласно рисунку 53. Обратите внимание:

D2		01.03.2010						
A	B	C	D	E	F	G		
1	Кредит, руб.	Срок кредита, мес.	Процентная ставка кредита, %	Дата выдачи кредита				
2	100000	60	0,17	40238				
3								
4	Примерный график платежей							
5	Платеж за расчётный период, ед. валюты					Остаток задолженности по ссуде, ед. валюты	Денежный поток (расходы) получателя ссуды, ед. валюты	
6	в том числе							
7	Дата платежа	Сумма платежа	Проценты	Погашение основной суммы ссуды	Комиссии и другие платежи			
8	40238	=-A2+E8			1000	=A2	=B8	
9	40269	=C9+D9+E9	=(A9-A8)/365*F8*SCS2	=SAS2/SBS2	0	=F8-D9	=B9	
10	40299	=C10+D10+E10	=(A10-A9)/365*F9*SCS2	=SAS2/SBS2	0	=F9-D10	=B10	
67	42036	=C67+D67+E67	=(A67-A66)/365*F66*SCS2	=SAS2/SBS2	0	=F66-D67	=B67	
68	42064	=C68+D68+E68	=(A68-A67)/365*F67*SCS2	=SAS2/SBS2	0	=F67-D68	=B68	
69	ИТОГО:		=СУММ(C9:C68)	=СУММ(D9:D6)	=СУММ(E		=СУММ(G9:G68)	
70	Эффективная процентная ставка:						=ЧИСТВНДОХ(G8:G68;A8:A68)	
71					Переплата	=G69/A2-1		

Рисунок 53 – Примерный график платежей по кредиту в режиме отображения формул

- одинаковые по внешнему виду формулы вводятся один раз, а затем копируются с помощью маркера заполнения;

A	B	C	D	E	F	G	
1	Кредит, руб.	Срок кредита, мес.	Процентная ставка кредита, %	Дата выдачи кредита			
2	100000	60	17%	01.03.2010			
3							
4	Примерный график платежей						
5	Платеж за расчётный период, ед. валюты					Остаток задолженности по ссуде, ед. валюты	Денежный поток (расходы) получателя ссуды, ед. валюты
6	в том числе						
7	Дата платежа	Сумма платежа	Проценты	Погашение основной суммы ссуды	Комиссии и другие платежи		
8	01.03.2010	- 99 000,00			1 000,00	100 000,00	- 99 000,00
9	01.04.2010	3 110,50	1 443,84	1 666,67	-	98 333,33	3 110,50
10	01.05.2010	3 040,64	1 373,97	1 666,67	-	96 666,67	3 040,64
67	01.02.2015	1 714,79	48,13	1 666,67	-	1 666,67	1 714,79
68	01.03.2015	1 688,40	21,74	1 666,67	-	0,00	1 688,40
69	ИТОГО:		43 260,35	100 000,00	1 000,00		143 260,35
70	Эффективная процентная ставка:						19,0%
71					Переплата	43%	

Рисунок 54 – Примерный график платежей по кредиту

- поле дата платежа содержит даты в формате ###.###.#####, которые представлены в виде чисел, после включения режима отображения формул (Ctrl+` (левая кавычка, расположенная на одной клавише со знаком ~ «ТИЛЬДА»));

- в формуле процентов нужно учесть, что в високосном году 366 дней;
- $\$C\2 – абсолютная ссылка на ячейку C2, которая при копировании не изменяется и получается нажатием клавиши F4;
- с помощью справки Excel опишите функцию **ЧИСТВНДОХ()**.

Пример 13 У Вас есть возможность получить большой кредит в других банках на тот же срок, но с другими процентами:

Банк	Кредит, руб.	Срок кредита, мес.	Процентная ставка кредита, %
«XYZ»	130 000	60	23
«GEO»	150 000	60	27
«И.ГРЕКОВ»	200 000	60	29

Сравните эти предложения и выберите для себя более выгодное.

Рисунок 55 – Диспетчер сценариев

Для решения этой задачи воспользуемся сценарным подходом.

Сценарий в Excel представляет собой некоторое множество исходных значений, предназначенных для подстановки в выбранные зависимости для получения вариантных отчетов.

На вкладке *Данные* в группе *Работа с данными* выберите команду *Анализ «что-если»*, а затем выберите в списке пункт *Диспетчер сценариев*.

Выбрав *Добавить* введем название сценария «XYZ», изменяемые ячейки и затем значения для изменяемых ячеек (рисунок 55).

Затем аналогично добавим сценарии «GEO» и «И.ГРЕКОВ» и выберем *Отчет – Тип отчета – Структура* (рисунок 56) – ячейки результата G69:G71.

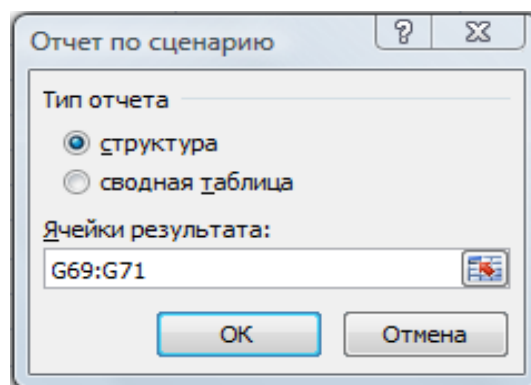
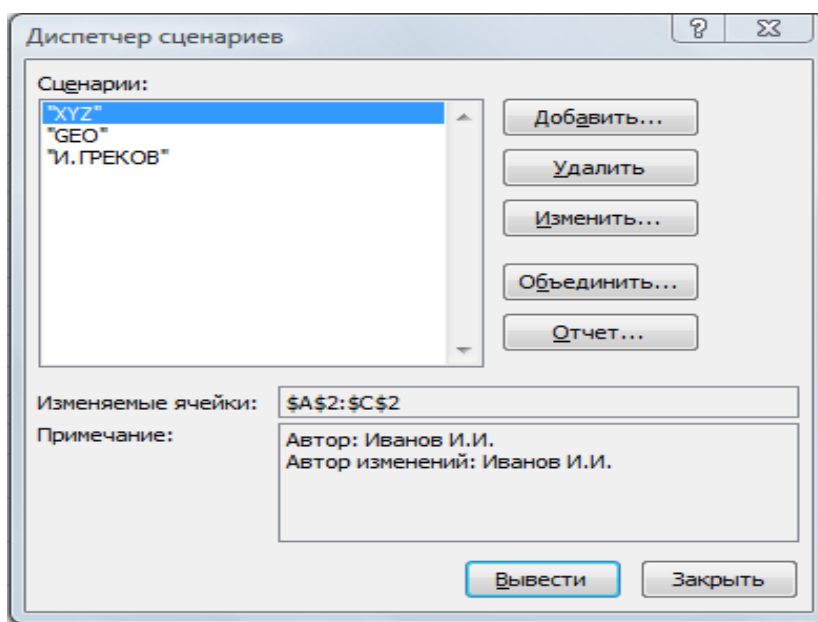


Рисунок 56 – Вывод отчета

После выбора ОК, получим четыре возможных сценария получения кредита (рисунок 57).

Структура сценария					
		Текущие значения:	"XYZ"	"GEO"	"И.ГРЕКОВ"
Изменяемые:					
	\$A\$2	100000	130000	150000	200000
	\$B\$2	60	60	60	60
	\$C\$2	17%	23%	27%	29%
Результат:					
	\$G\$69	143 260,35	206 087,33	253 061,43	347 594,15
	\$G\$70	19,0%	26,1%	31,1%	33,6%
	\$G\$71	43%	59%	69%	74%

Рисунок 57 – Сценарии кредитования в разных банках

Переплата при увеличении суммы кредита и процентной ставки естественно растет. Возникает вопрос – какой срок кредитования позволит в других банках переплатить не более 43%?

Задание

1. Вкладка *Данные* позволяет удалять дубликаты (повторяющиеся строки, например, введите представленную ниже таблицу и выберите *Удалить дубликаты*. Рассмотрим возможности использования инструмента Консолидация.

1	1
2	2
3	3
4	4
5	5
3	3

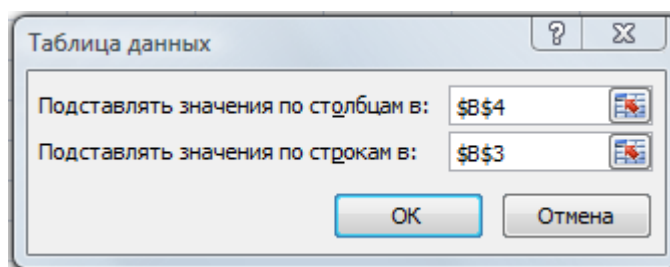
2. По аналогии с примером 12 построить примерный график платежей ипотечного кредита на 20 лет в сумме 1000 000 руб. под 12 % годовых до регистрации ипотеки, и 11 % годовых после регистрации ипотеки (через шесть месяцев). Дайте оценку эффективной кредитной ставки и процент переплаты.

3. (Финансовые функции). Рассмотрим функцию ПЛТ, которая возвращает сумму периодического платежа для аннуитета на основе постоянства сумм платежей и постоянства процентной ставки.

В таблице данных с двумя переменными может быть показано влияние на размер ежемесячных выплат по закладной различных процентных ставок и сроков займа. В следующем примере ячейка С2 содержит формулу вычисления платежа, =ПЛТ(В3/12;В4;-В5), которая ссылается на ячейки ввода В3 и В4.

	A	B	C	D	E
1	Ссуды на недвижимость				
2	Первый взнос	Нет	=ПЛТ(В3/12;В4;-В5)	180	360
3	Процентная ставка	0,095	0,09	=ТАБЛИЦА(В4;В3)	=ТАБЛИЦА(В4;В3)
4	Срок (месяцы)	360	0,0925	=ТАБЛИЦА(В4;В3)	=ТАБЛИЦА(В4;В3)
5	Сумма ссуды	100000	0,095	=ТАБЛИЦА(В4;В3)	=ТАБЛИЦА(В4;В3)

Далее выделим диапазон С2:Е5 и выполним команду Данные – Анализ – «что-если» - Таблица данных и заполним параметры диалогового окна:



В результате получим таблицу сумм периодического платежа для аннуитета на основе постоянства сумм платежей и постоянства процентной ставки.

Вопросы для самоконтроля

1. Что представляет собой потребительский кредит.
2. Сущность ломбардного кредита.
3. Практики, используемые при расчете количества дней ссуды.
4. Виды финансовых рент

ЛАБОРАТОРНАЯ РАБОТА № 6

Управление проектами в MS Project

Цель работы: Ознакомиться с возможностями управления проектами и их анализа в MS Project

Теоретические сведения

Планирование мероприятий (проектов), для реализации которых необходимо выполнять некоторое множество работ – обязательная составляющая практически любой деятельности человека. Проект можно определить как подлежащую решению задачу, состоящую из обособленных работ, которые выполняются в заданной последовательности и в установленные сроки. Обычно желательно реализовать проект в наиболее короткий срок. Для

этого лучше управлять деятельностью всех работ, а также временем реализации всего проекта. Поэтому приобретает особое значение временной график реализации проекта. Специфика управления проектами отражается в проектном треугольнике:



Изменение любой из составляющих проекта, так или иначе, влияет на качество.

MS Project – представляет собой средство составления временных расписаний и управления проектами на базе сетевых моделей.

Каждый процесс (работа) проекта обозначается дугой, ориентированной по направлению выполнения проекта. Узлы сети (события) устанавливают отношения предшествования среди процессов проекта. Существуют правила построения сети проекта:

- каждый процесс в проекте представляется одной дугой;
- каждый процесс идентифицируется двумя концевыми узлами;
- необходимо определить процессы, которые выполняются до и после текущего процесса, который конкурирует (выполняется параллельно) с текущим.

Для выполнения правил (например, при рассмотрении параллельных (конкурирующих) процессов) можно рассматривать фиктивные работы, не поглощающие временные или другие ресурсы.

В программе реализованы наиболее известные методы управления проектами:

- метод Критического пути позволяет определить самые ранние и самые поздние моменты начала и окончания работ (*CPM-critical path method*);
- метод *PERT (program evolutionary and Review Technigue)* – позволяет, используя вероятностный подход, дать для каждой работы оптимистические, наиболее вероятные и пессимистические оценки. В методе *CPM* предполагается, что длительность каждого этапа детерминирования, а в *PERT*- стохастическая.

CPM. Процесс является критическим, если он не имеет «зазора» для времени своего начала и завершения. Таким образом, чтобы весь проект завершился без задержек, необходимо, чтобы все критические процессы начинались и заканчивались в строго определенное время. Для некритических процессов возможен «дрейф» во времени, но в определенных границах, когда время его начала не влияет на длительность выполнения проекта.

Критическое время – наименьшее время реализации проекта (ввиду того, что многие работы проводятся параллельно, критическое время намного меньше суммы времени выполнения всех работ проекта)

Последовательность работ, длительность выполнения которых, определяет критическое время – критический путь, а соответствующие работы – критические работы. Критическим работам необходимо уделять особое внимание.

Конечным результатом применения метода критического пути будет построение временного графика выполнения проекта:

PERT. Используется три оценки времени работ проекта:

a – оптимистическая длительность выполнения работ в благоприятных условиях;

b – наиболее вероятная длительность выполнения работы в нормальных условиях;

c – пессимистическая длительность выполнения работы в неблагоприятных условиях;

Предполагается, что вероятности длительности выполнения работ имеют β – распределение, так что математическое ожидание *m* длительности выполнения работы и ее дисперсия σ^2 выражаются формулами:

$$m = \frac{a+4b+c}{6};$$
$$\sigma^2 = \left(\frac{c-a}{6}\right)^2.$$

Применение метода *PERT* позволяет ответить на вопросы:

- чему равны математическое ожидание и дисперсия реализации проекта в целом;

- какова вероятность реализации проекта в заданное время;

- как определить момент завершения реализации проекта с заданной вероятностью.

Интерфейс *MS Project*, изображенный на рисунке 39 состоит из стандартных частей *MS Office* (сверху вниз) – строки меню, панели инструментов. Ниже которых находятся три модуля (прямоугольные области слева направо):

– *панель представлений* (представляет информацию графически, в виде таблицы или формы, аналогично печатной – содержит восемь представлений из более чем 20 доступных, выбирая которые можно увидеть разные варианты отображения проекта),

– *консультант* (интерактивный помощник *Project* для: создания нового проекта, ввода задач, назначения ресурсов, отслеживания хода выполнения проекта, составления отчета о состоянии проекта),

– *рабочая область* (по умолчанию отображается представление под названием *Диаграмма Ганта* – основное место работы).

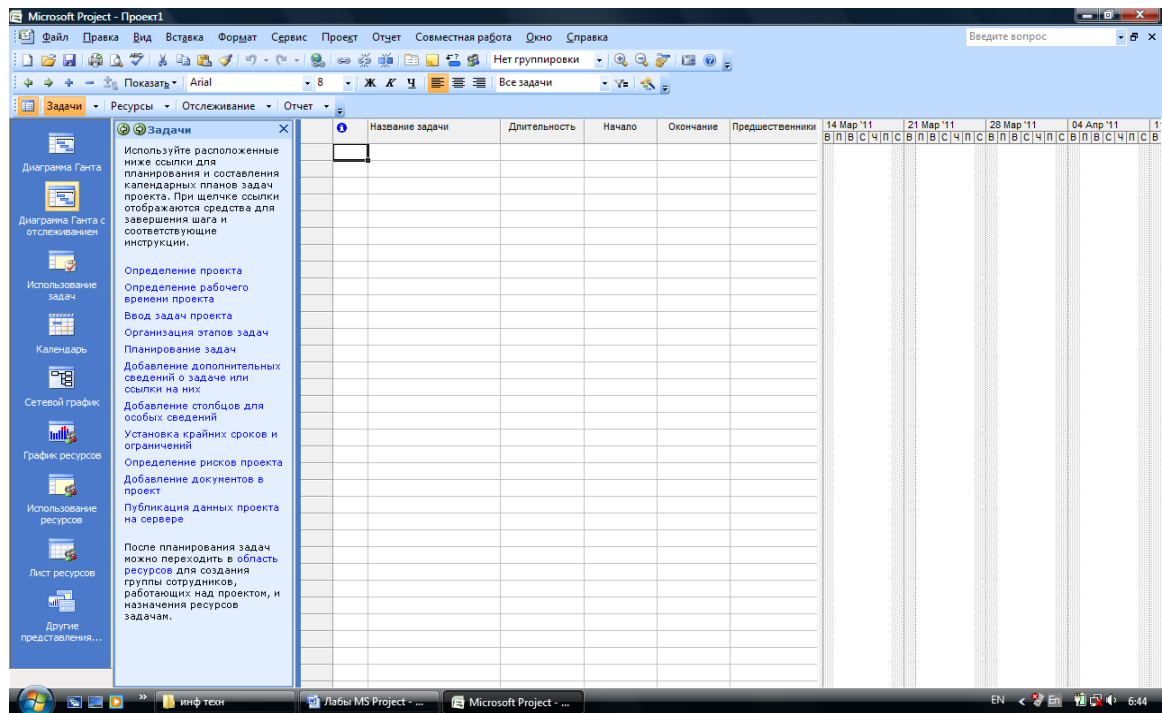


Рисунок 58 - Интерфейс MS Project

Панели инструментов можно включать, настраивать и выключать самостоятельно используя команду *Вид-Панели инструментов*.

Необходимые настройки в MS Office осуществляются посредством команды *Сервис*. Например, выполните команды:

- *Сервис - Вид - Представление по умолчанию* (выберите самостоятельно)
- *Ok*,
- *Сервис-Настройка - Панели инструментов - Всегда показывать полные меню*.

Основное представление проекта, задаваемое по умолчанию - Диаграмма Ганта состоит из двух частей: табличной и графической. Она предназначена для ввода данных о задачах проекта и их параметрах (дате начала, длительности). Результатом работы с диаграммой Ганта является календарный план.

Данные представления проектных данных в MS Project это таблицы с информацией о задачах и ресурсах проекта. Работа с проектом может начинаться с использованием Консультанта или самостоятельно.

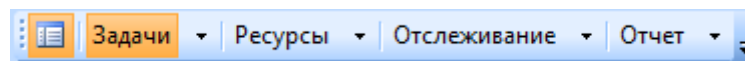


Рисунок 59 – Панель Консультант

Рассмотрим пример проекта– «Яблочный пирог на скорую руку».

Проект. Приготовление яблочного пирога. Откроем новый проект и сохраним его под названием «Внучка идет в гости к бабушке».

Таблица 17 – План реализации проекта «В гости к бабушке»

№ задачи	Описание задачи	Срок
1	Покупка продуктов (ингредиентов яблочного пирога)	1 час
2	Приготовление теста	20 мин
3	Подготовить форму и залить тесто	10 мин
4	Нарезать дольками яблоки	5 мин
5	Прогрев духовки	20 мин
6	Размещение ингредиентов на тесте	10 мин
7	Приготовление пирога	45 мин
8	Пирог готов	Века (о мин)
9	Упаковка пирога	5 мин
10	Собраться	1ч 30 мин
11	Доставить бабушке	40 мин

Рассмотрим ручной вариант настройки проекта, начнем со свойств проекта.
Для этого выполним команду *Проект-Сведения о проекте*

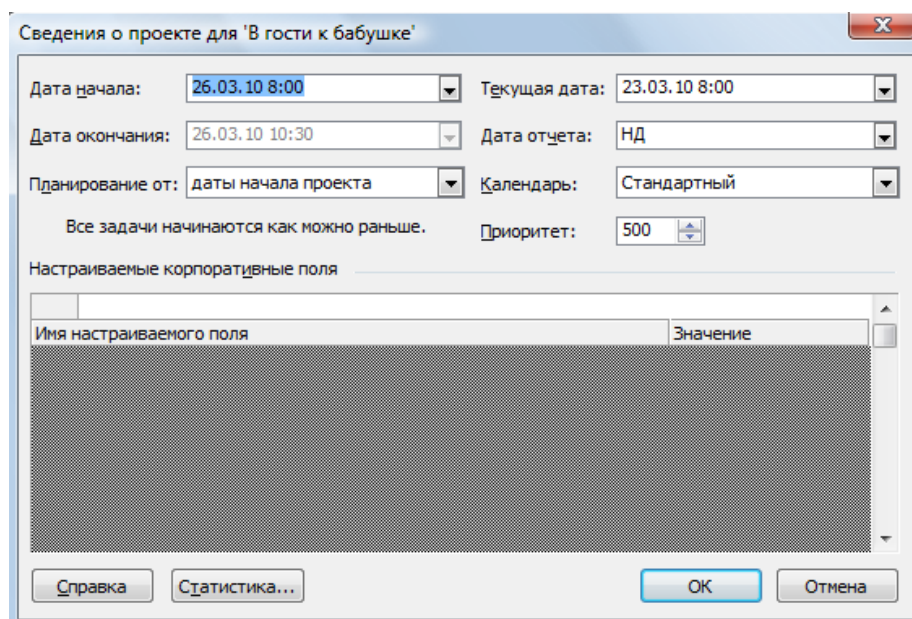


Рисунок 57 – Сведения о проекте

Возможно два варианта планирования (элемент Планирование от: даты начала проекта, даты окончания проекта). Выбор одного из вариантов блокирует другой. Второй вариант удобен, когда вы точно знаете конечный срок к которому должно быть закончен проект, например, корпоративная вечеринка к 8 марта или новому году, начало соревнований любого уровня (например, Олимпийские игры Сочи-2014). Сохраним планирование от даты начала проекта 26.03.2010 г.

После этого можно настроить календарь, задав время начала и окончания рабочего дня, начало финансового года, рабочей недели (команда *Сервис-Параметры-Календарь*). По умолчанию длительность выполнения измеряется в днях. С помощью команды (*Сервис-Параметры*) изменим эту – настройку

Планирование-Длительность вводится в часах,. На вкладке *Расчет* выберем *Автоматически*.

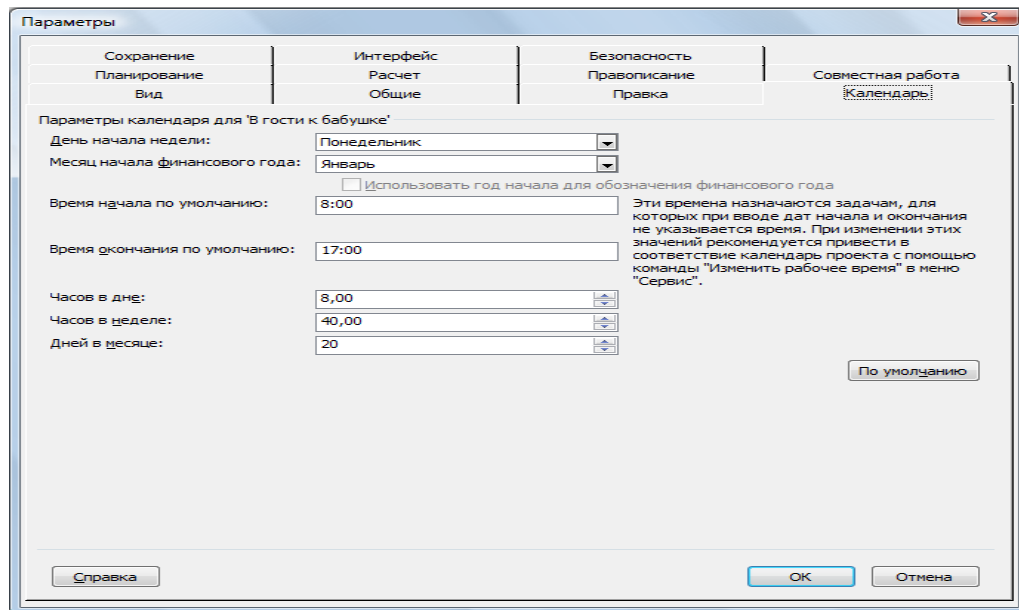


Рисунок 58 – Настройка календаря


Длительность задачи можно поменять с помощью диаграммы Ганта, потянув за правый край отрезка, графически представляющего задачу. Ведем задачи табл.10 либо непосредственным вводом, либо открыв двойным щелчком левой кнопки мыши Сведения о задаче. Для удобства графического представления воспользуемся функцией масштабирования, которая вызывается с помощью кнопок в виде лупы на панели инструментов Стандартная . Второй вариант – использование команды *Вид-масштаб*, с помощью которого можно задать точное количество отображенных на графике дней. После ввода данных проекта, и изменения масштаба получим результат, отображенный на рисунке 59.



Рисунок 59 – Все задачи проекта «Приготовление пиццы»

Задача № 8 «Пирог готов», изображенная в виде ромба является вехой, так как она означает ключевое событие проекта – момент окончания

приготовления пирога (нулевой срок – достаточное условие для автоматического определения вехи в MS Project).

Выполним команду Сервис – Параметры – Вид – Показывать суммарную задачу – ОК. Проект приобретет вид, представленный на рисунке 60.

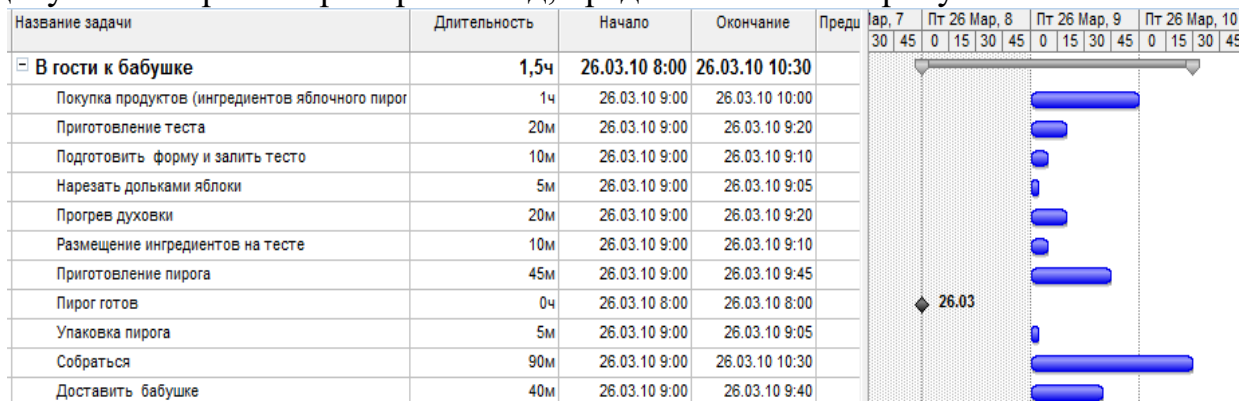


Рисунок 60 – Создание общей суммарной задачи

Все задачи начинаются в одно время и общий срок проекта 1,5ч – т.е. он равен длительности самой продолжительной задачи №10 «Собраться». Для просмотра информации по проекту выполните команду Проект – Информация о проекте – Статистика (Рисунок 61).

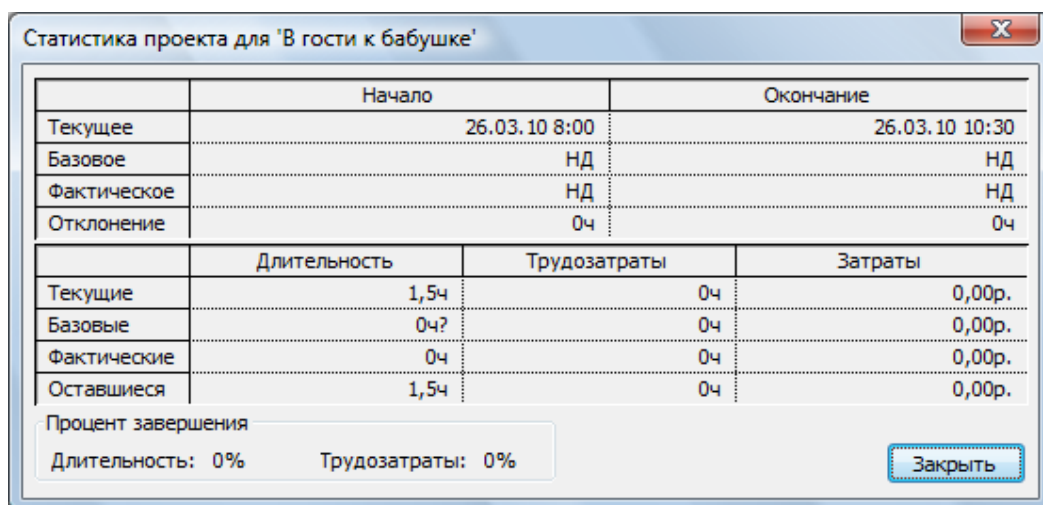


Рисунок 61 – Статистика проекта

Необходимо задать порядок взаимодействия задач между собой. Одна из задач – последователь, вторая – предшественник. У каждого последователя может быть несколько предшественников и наоборот. Обычно рассматривается 4 типа связей: окончание-начало (нельзя заливать в форму неготовое тесто), начало-начало (можно поставить пирог готовится и начать собираться), окончание-окончание (обе задачи оканчиваются одновременно), начало-окончание (цель связи – акцент на то, что после окончания одной задачи и началом второй не должно пройти много времени). Следует отметить, что можно влиять на дату начала задачи последователя задавая время задержки или опережения.

Настройка связей. Выделим левой кнопкой мыши задачу №2 и выполним команду *Проект-Сведения о задаче*, выбрав предшественник *Покупка продуктов*

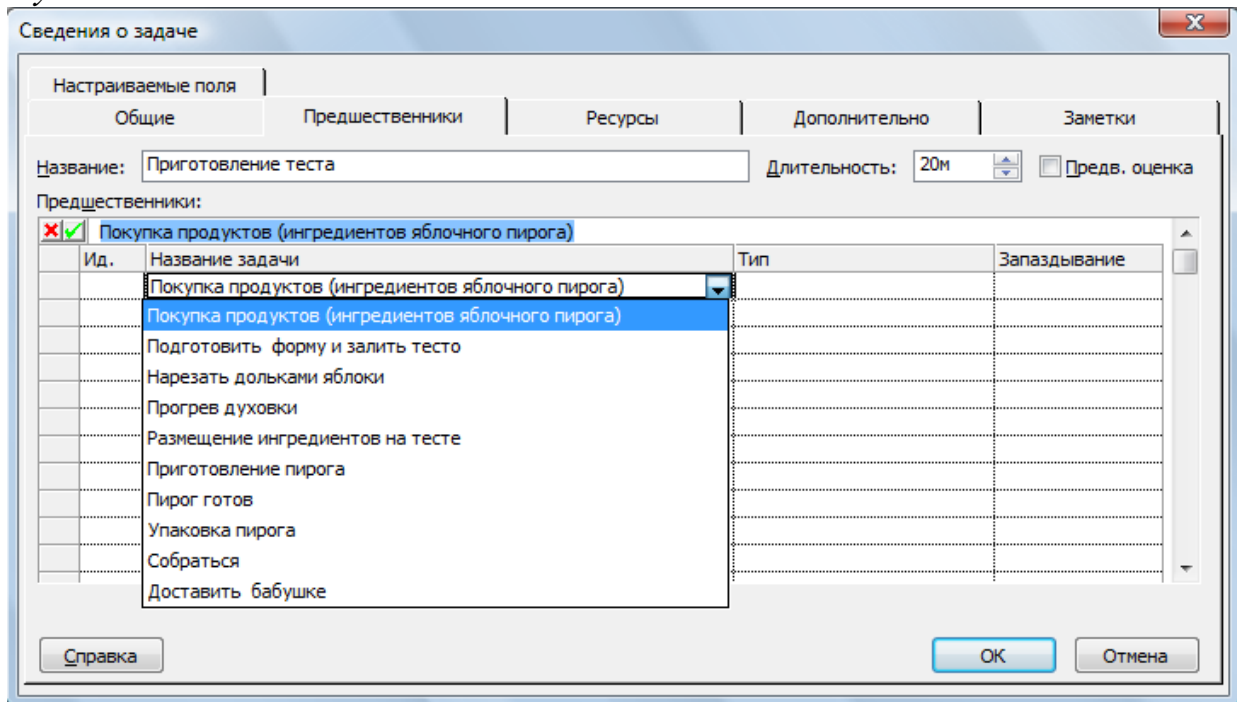


Рисунок 62 – Сведения о задаче

Последовательно связывая задачи по типу «окончание -начало» получим результат графического изображения диаграммы Ганта, изображенный на рисунке 47, в промежутке времени между 12:00 и 13:00 задача не выполняется, т.к. в выбранном нами календаре в это время перерыв.

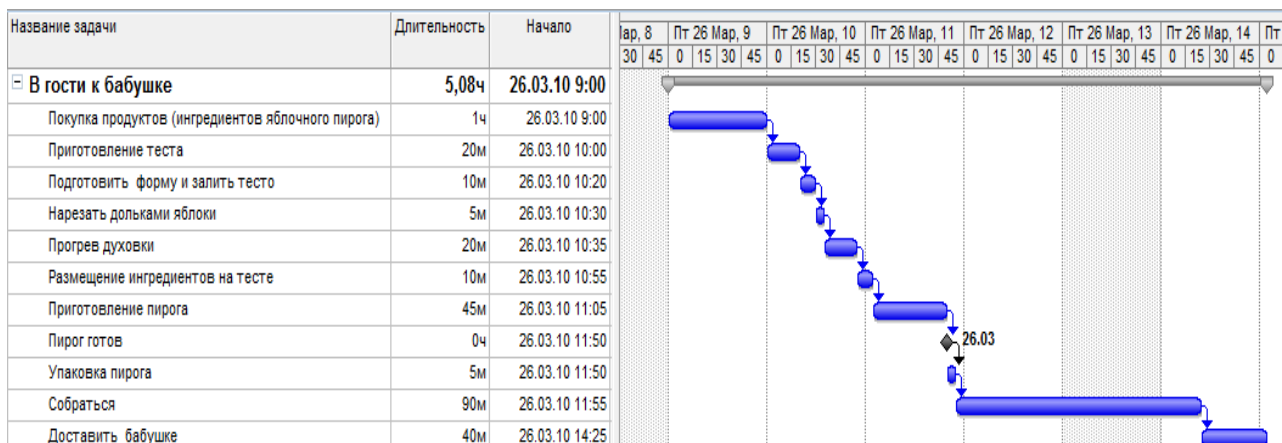


Рисунок 63 – Наименование задачи

Заметим, что прогревать духовку можно начать после приготовления теста, размещение ингредиентов возможно после нарезки яблок, приготовление пирога возможно только после размещения ингредиентов и прогрева духовки, собираться можно начать поставив пирог в духовку, а упаковать после того как

собрались, выполнив соответствующие преобразования (с помощью команды *Проект- Сведения о задаче*) мы в результате экономим 1,16ч и попадаем к бабушке до обеда (рисунок 63).

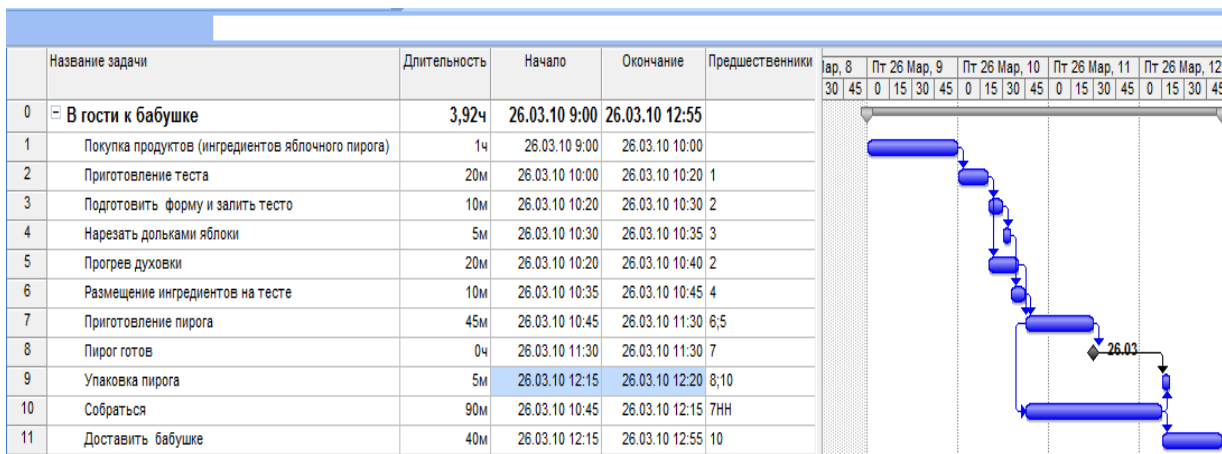


Рисунок 64 – Наименование задачи

Выполнив команду *Файл – Предварительный просмотр* мы получим календарный план проекта готовый к печати

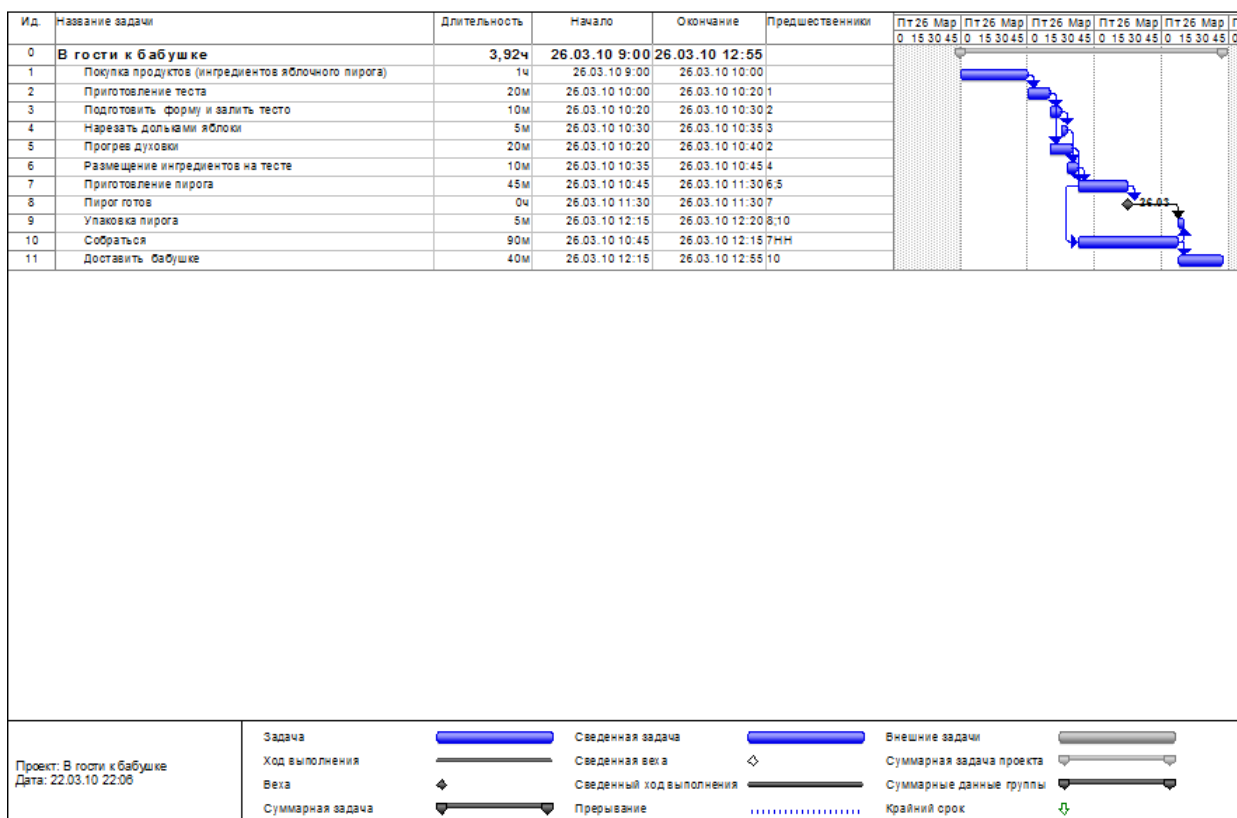


Рисунок 65 – План к печати

Приготовление яблочного пирога предполагает затрату ресурсов. В MS Project используются следующие типы ресурсов: материальные, которые оплачиваются за количество; трудовые, оплата за отработанные часы; затраты –

финансовые ресурсы, необходимые для оплаты трудовых и материальных ресурсов.

Таблица 18 – Трудовые ресурсы

№	Название	Цена	Комментарии
1	Водитель такси	250 руб/час	Поездка за продуктами к бабушке
2	Повар (в вашем лице)	150 руб/час	Плата символическая по сравнению с доброжелательностью бабушки...

Таблица 19 – Материальные ресурсы

№	Название	Объем/цена	Комментарии
1	1 стакан муки	200 гр (30 руб/ кг)	Ингредиент для теста
2	3 яйца	3 шт 30 руб /десяток	Ингредиент для теста Покупается десяток
3	1 стакан сахара	200г 25 руб /кг	Ингредиент для теста (покупается 1 кг)
4	5-6 яблок	1кг 50 руб/кг	Ингредиент для пирога
5	Кисель	1 упаковка 20 руб/шт	Ингредиент для пирога
6	Ягоды	100 гр 150 руб/кг	Ингредиент для пирога
7	Термокоробка	100 руб	Чтобы пирог не остыл

Выполним команду Ресурсы – Выбор людей и оборудования для проекта и заполним ресурсы

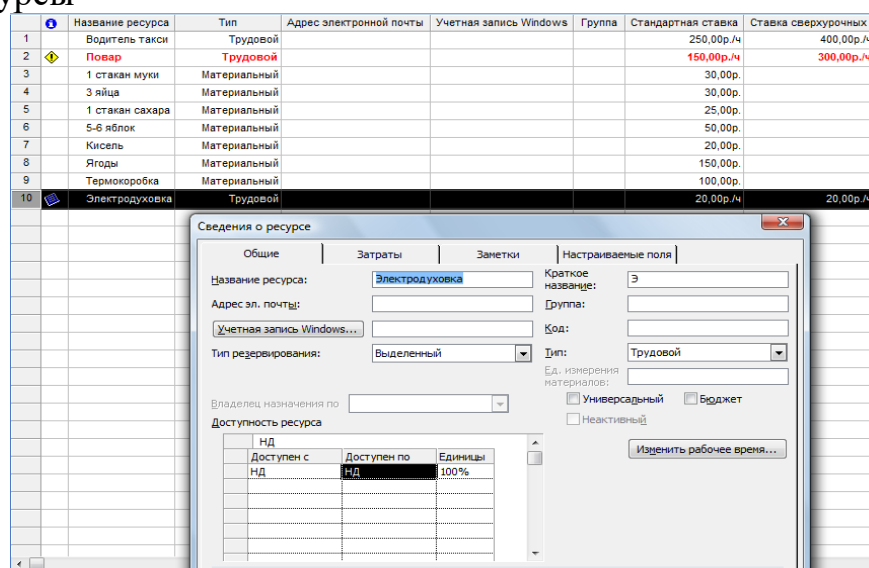


Рисунок 66 – Выбор параметра для проекта

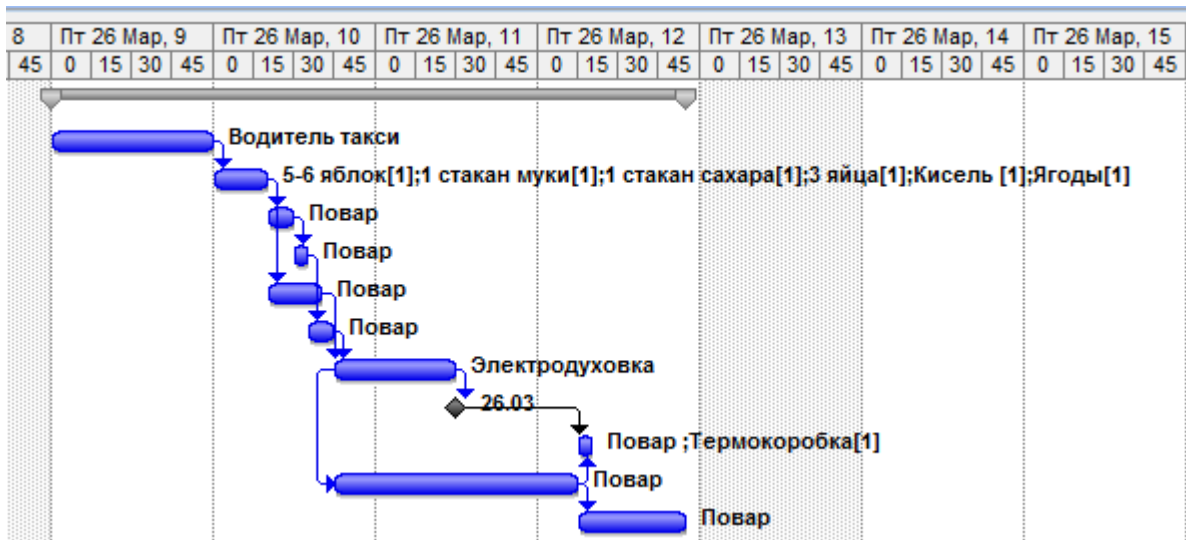


Рисунок 67 – Диаграмма Ганта после назначения ресурсов

Статистика проекта для 'В гости к бабушке'

	Начало	Окончание
Текущее	26.03.10 9:00	26.03.10 12:55
Базовое	НД	НД
Фактическое	НД	НД
Отклонение	0ч	0ч

	Длительность	Трудозатраты	Затраты
Текущие	3,92ч	4,75ч	1 120,00р.
Базовые	0ч?	0ч	0,00р.
Фактические	0ч	0ч	0,00р.
Оставшиеся	3,92ч	4,75ч	1 120,00р.

Процент завершения
 Длительность: 0% Трудозатраты: 0%

Закреть

Рисунок 68 – Статистика проекта

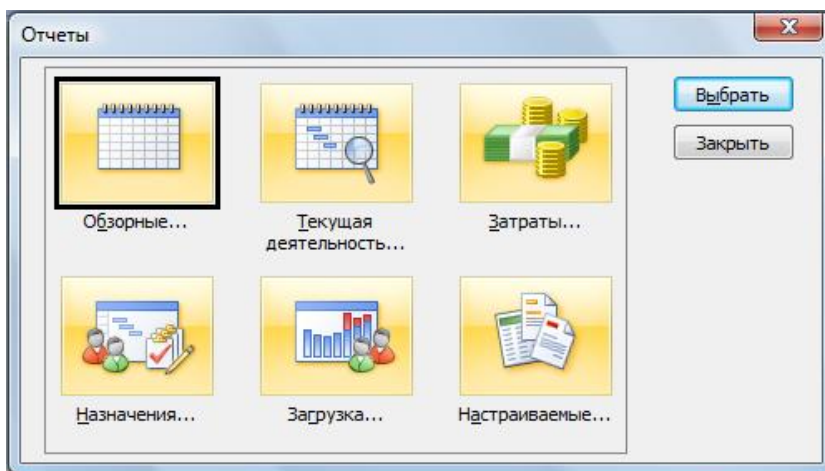


Рисунок 69 – Отчеты

Назначим ресурсы задачам *Ресурсы – Назначение людей и оборудования*

задачам. Изучим Статистику проекта и отчеты.

Критическое время – наименьшее время реализации проекта (ввиду того, что многие работы проводятся параллельно, критическое время намного меньше суммы времени выполнения всех работ проекта).

Последовательность работ длительность выполнения которых определяет критическое время – критический путь, а соответствующие работы – критические работы. Критическим работам необходимо уделять особое внимание.

Метод Критического пути позволяет определить самые ранние и самые поздние моменты начала и окончания работ (CPM – critical path method)

Метод PERT (program evaluational and Review Technique) – метод используя вероятностный подход дает для каждой работы Оптимистические, наиболее вероятные и пессимистические оценки.

Выполнив команду Формат – Мастер диаграмм Ганта можно настроить критический путь (рисунок 70).

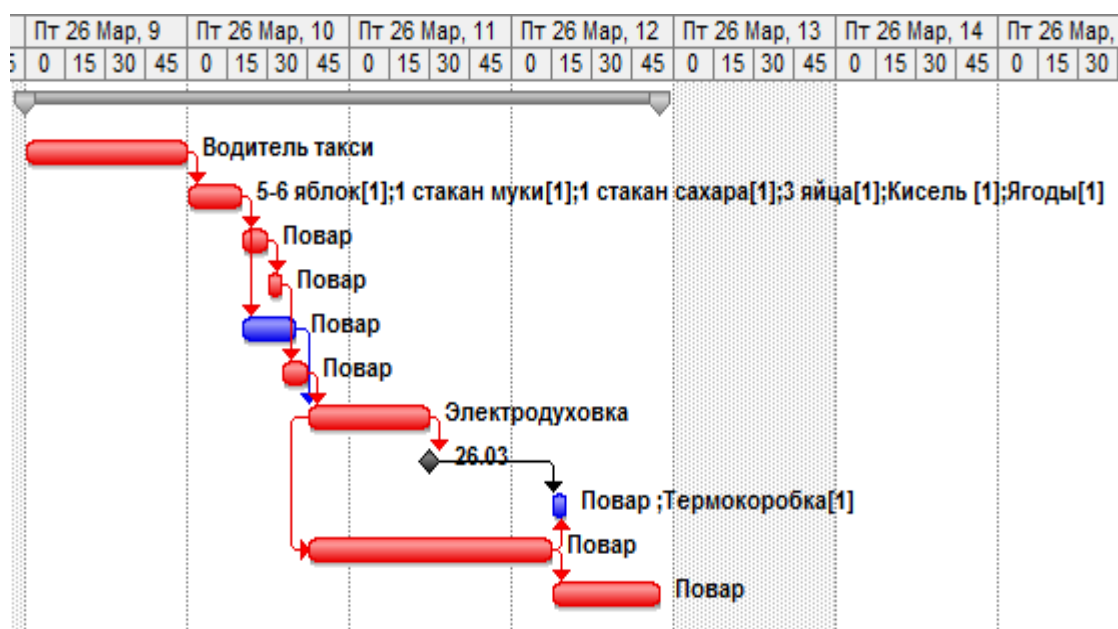


Рисунок 70 – Критический путь

Применение фильтра позволяет указать только определенные задачи в данном представлении

При возникновении опасений, что проект может завершиться позже, чем запланировано, можно скорректировать календарный план различными способами, а именно:

- оптимизировать задачи, образующие критический путь. Иногда при оценке календарного плана критических задач обнаруживаются ненужные зависимости или ограничения, отодвигающие дату окончания;
- изменить сведения о ресурсах или их назначения. Можно изменить календари ресурсов, предусмотрев сверхурочную работу для выполнения критических задач, или назначить дополнительные ресурсы задачам,

требующим скорейшего выполнения;

- уменьшить область охвата проекта. Для соблюдения сроков окончания можно пойти на сокращение объемов выполняемых работ. В этом случае можно удалить некоторые задачи или соответствующим образом сократить их длительности.

Вычисление критического пути. Критическими являются задачи, не имеющие временного резерва – его можно увеличить... Считается, что количество задач, образующих критический путь не должно превышать 50-70% от общего количества задач проекта.

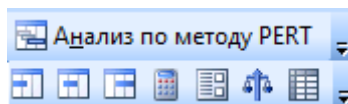


Рисунок 71 – Панель инструментов Анализ по методу PERT

Задание

Проведите самостоятельно анализ по методу *PERT*.

Вопросы для самоконтроля

1. Сущность анализа по методу *PERT*
2. Что представляет собой метод Критического пути
3. Способы корректировки плана

ПРИЛОЖЕНИЕ

Основные показатели производства в сельскохозяйственных предприятиях

Краснодарского края

№ п.п.	Среднегодовая численность работников,	Численность тракторов, эт. ед.	Площадь сельскохозяйственных	Энергетические мощности, л.с.	Основные фонды сельскохозяйственного	Загрты на производство валовой продукции, тыс.	Загрты на производство реализованной продукции,	Валовая продукция, тыс. руб.	Реализованная продукция, тыс. руб.
1	591	102	12139	34503	74171	111276	80946	120456	90126
2	334	54	6773	14698	64382	30960	25670	31362	26072
3	335	45	8698	15506	69721	38056	29209	50375	41528
4	657	102	12926	32885	52744	63272	38176	78800	53704
5	541	75	11135	32901	93277	82953	68145	98897	84089
6	864	113	12135	36032	174537	83600	54719	92718	63837
7	370	68	7105	27849	62482	62289	56879	83151	77741
8	437	54	6530	22851	116405	46774	36995	45309	35530
9	410	76	7154	24693	79399	55942	49226	63354	56638
10	552	68	9083	24027	94116	61685	60013	88644	86972
11	246	48	4474	10782	74385	34126	29769	41407	37050
12	492	104	13735	28253	103326	75099	54292	67383	46576
13	217	53	4501	13596	77558	26284	19065	26981	19762
14	603	98	7465	25200	99567	74367	70913	99974	96520
15	400	58	6270	19798	64488	47618	25379	51983	29744
16	602	121	10550	33420	88935	83584	60564	104487	81467
17	389	89	8753	26936	117937	79097	54599	111868	87370
18	435	45	10830	20598	97580	57820	47952	70245	60377
19	422	82	9646	24645	44073	51076	41828	61868	52620
20	100	17	4034	6485	13777	14988	12304	19681	16997

Продолжение приложения

№ п.п.	Среднегодовая численность	Численность тракторов, эт. ед.	Площадь сельскохозяйственных	Энергетические мощности, л.с.	Основные фонды сельскохозяйственного	Затраты на производство валовой	Затраты на производство	Валовая продукция, тыс. руб.	Реализованная продукция, тыс. руб.
22	395	79	9074	17540	58851	45593	36571	58057	49035
23	618	124	12730	39763	106264	70006	63853	90189	84036
24	308	77	8059	25510	75552	31741	26238	38159	32656
25	421	51	9912	18970	120849	66304	49863	80708	64267
26	615	75	10131	24848	108571	81290	55057	86152	59919
27	590	101	11576	24604	202088	90212	56178	81232	47198
28	230	43	6425	10061	50281	43276	36242	46181	39147
29	961	100	10533	41544	181137	128388	94954	129971	96537
30	414	42	6990	20498	61996	38719	26757	43039	31077
31	247	52	8160	15656	78086	55440	41135	61574	47269
32	605	100	11345	28422	137723	84156	67322	84577	67743
33	434	68	7671	26512	85796	63608	53064	77256	66712
34	741	82	10154	18016	103640	86587	61015	95941	70369
35	319	82	7740	26535	77899	54788	39825	58821	43858
36	441	58	5566	21576	103400	51722	42155	51716	42149
37	855	129	10276	43163	191108	114683	117335	171615	174267
38	1385	246	16816	72473	336464	227368	158245	285753	216630
39	320	78	7203	25810	87836	59059	33775	62608	37324
40	654	156	13313	56635	277550	146888	126933	175841	155886
41	1137	179	14800	67952	365315	192684	146848	243738	197902
42	650	102	11175	34110	159956	70307	57192	73950	60835
43	354	55	5931	22306	130280	74955	47242	78175	50462
44	649	113	11093	45756	207540	110701	90913	145212	125424

Продолжение приложения

№ п.п.	Среднегодовая численность	Численность тракторов,	Площадь сельскохозяйстве	Энергетические мощности, л.с.	Основные фонды сельскохозяйстве	Загрты на производство валовой	Загрты на производство реализованной	Валовая продукция, тыс.	Реализованная продукция, тыс.
45	358	68	4968	10470	90959	57390	34531	59258	36399
46	922	108	15294	22725	144854	82418	55884	98967	72433
47	433	59	5761	22647	76179	73604	58443	78686	63525
48	486	48	4767	13138	55665	68175	58911	76802	67538
49	218	52	4789	15454	78164	33864	31961	41700	39797
50	189	22	4776	12747	40077	26161	16736	25950	16525
51	911	126	13325	45990	317847	103085	78571	110162	85648
52	513	83	9132	25214	40053	44710	37169	47502	39961
53	312	57	5217	12896	68042	28615	21611	25202	18198
54	117	22	2788	10238	26645	14237	11837	17034	14634
55	284	43	6638	15757	31237	43087	41277	44903	43093
56	304	32	5133	12985	22407	35222	23856	39517	28151
57	377	70	6084	20010	81056	39135	28504	41985	31354
58	341	67	7213	19137	73079	39635	29687	56662	46714
59	155	28	3678	9832	41464	19184	14300	17382	12498
60	525	74	12211	24850	62348	100930	58596	84462	42128
61	383	57	10916	16780	94416	43249	28381	44137	29269
62	372	73	8332	19700	95584	47875	33253	44900	30278
63	519	69	8491	30238	142772	64781	45353	70270	50842

Рекомендуемая литература

1. Горелова Г.В. Теория вероятностей и математическая статистика в примерах и задачах с применением EXCEL / Г.В. Горелова, И.А. Кацко. Учеб. пособие для вузов. Изд. 4-е испр. и доп. – Ростов-на-Дону: «Феникс», 2006.
2. Статистика. Теория и практика в Excel В. С. Лялин, И. Г. Зверева, Н. Г. Никифорова. Финансы и статистика, Инфра-М 2010
3. Практикум по статистике в Excel. Учебное пособие Соболев Б.В. Феникс 2009
4. Microsoft Office Project 2007. Библия пользователя Элейн Мармел Диалектика, Вильямс, 2008 г.
5. Управление проектами в Microsoft Project 2007 Г. И. Сингаевская Диалектика, Вильямс, 2008 г.

**КОМПЬЮТЕРНЫЙ АНАЛИЗ
СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ДАННЫХ**

Методические указания
для студентов очной и заочной формы обучения

Бондаренко Петр Сергеевич
Ляховецкий Алексей Михайлович
Крепышев Дмитрий Александрович
Кацко Светлана Александровна
Сенникова Алина Евгеньевна
Жминько Альбина Евгеньевна

Подписано к печати _____
Бумага офсетная. Формат бумаги 60x84 1/16
Гарнитура «Times New Roman»
Тираж _____ Заказ _____

Типография Кубанского государственного аграрного университета.
350044, Краснодар, ул. им. Калинина, 13