

Лабораторная работа № 20 СТАТИСТИЧЕСКИЙ АНАЛИЗ ДАННЫХ В MICROSOFT EXCEL

Краткий комментарий

Статистический анализ данных в Microsoft Excel. Для решения задач статистического анализа данных в Excel применяются инструменты, входящие в состав пакета **Анализ данных** (команды **Данные** > **Анализ** > **Анализ данных**, а также набор статистических функций. При изложении материала данного раздела предполагается, что пользователь знаком с основными понятиями математической статистики.

Практическая работа

Построение таблиц частот и гистограмм

Для построения таблиц частот и гистограмм на основе выборки в Excel применяется инструмент **Гистограмма** из пакета **Анализ данных**.

Задание 1. Имеются значения показателя прочности 25 образцов некоторого материала:

32	48	62	54	59	31	29	47	42	37	51	52	43
47	61	54	53	41	42	36	48	41	49	57	46	

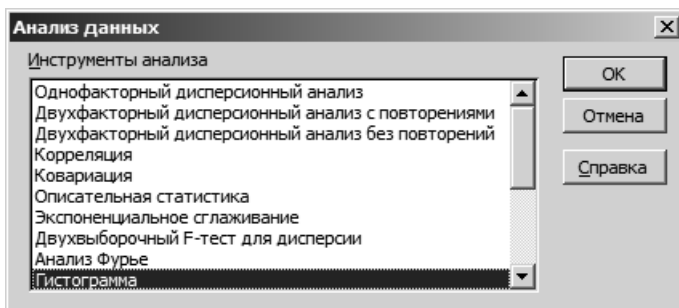
Построить таблицу частот и гистограмму.

1. В ячейку A1 ввести заголовок «Материал А». В ячейки A2:A26 ввести данные (значения прочности образцов материала).

2. В ячейку B1 ввести заголовок «Прочность»; в ячейки B2:B7 ввести границы интервалов гистограммы (20, 30, 40, 50, 60, 70).

3. Проверить, установлен ли пакет **Анализ данных** и при необходимости установить его. Для этого выбрать вкладку **Данные**, убедиться, что в группе **Анализ** присутствует команда **Анализ данных**.

4. Получить таблицу частот и гистограмму. Для этого выполнить команды **Данные** > **Анализ** > **Анализ данных**. Из списка **Инструменты анализа** выбрать инструмент **Гистограмма**.



В диалоговом окне «Гистограмма» установить следующие параметры:

Входной интервал: A1:A26, **Интервал карманов:** B1:B7. Установить флажки **Метки** (так как в ячейках A1 и B1 указаны не данные, а заголовки), **Интегральный процент** (для получения накопленных относительных частот) и **Вывод графика**

(для построения гистограммы). Чтобы результаты были выведены на отдельный лист, установить переключатель **Новый рабочий лист**. Для получения результатов нажать кнопку <OK>.

Гистограмма

Входные данные

Входной интервал: \$A\$1:\$A\$26

Интервал карманов: \$B\$1:\$B\$7

Метки

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Парето (отсортированная гистограмма)

Интегральный процент

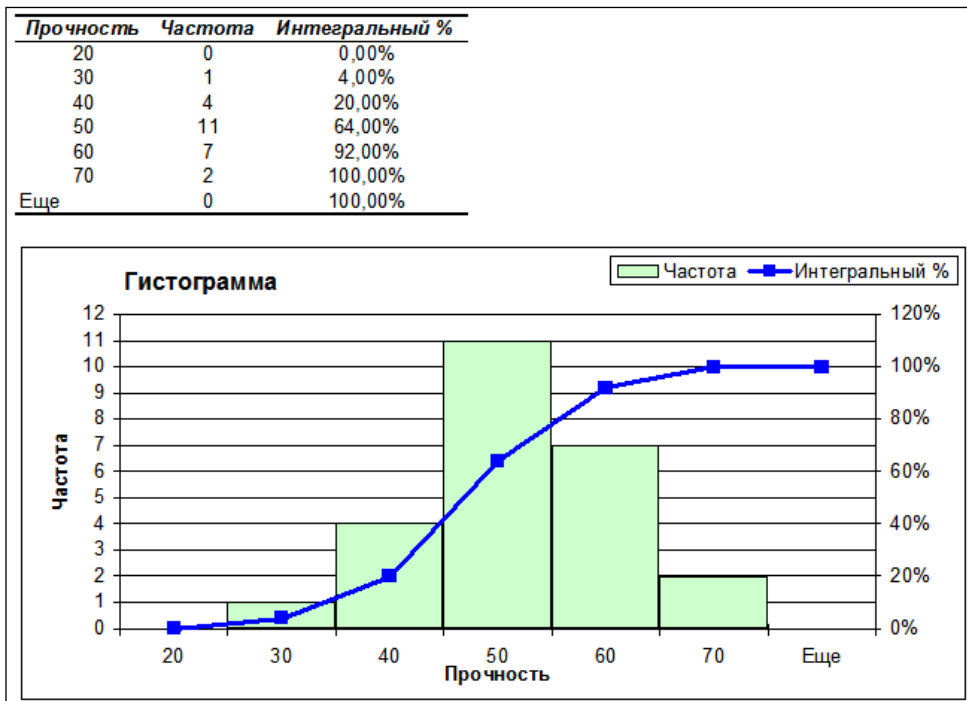
Вывод графика

OK

Отмена

Справка

Результат выполнения задания после форматирования таблицы и диаграммы представлен на рисунке.



Вычисление статистических показателей

Для вычисления отдельных статистических показателей по выборке применяются имеющиеся в Excel статистические функции, а для вычисления всех статистических показателей – инструмент **Описательная статистика** из пакета **Анализ данных**.

Задание 2. По данным примера 1 найти основные статистические показатели прочности испытываемого материала: среднее, стандартное отклонение, дисперсию, а также статистическую ошибку и 95-процентный доверительный интервал для среднего.

Статистические показатели	
Среднее	46,48
Стандартное отклонение	9,138381
Дисперсия	83,51
Объем выборки	25
Статистическая ошибка	1,827676
Уровень надежности	95%
Уровень значимости	0,05
Коэффициент доверия	2,063899
Предельная ошибка	3,772138
Доверительный интервал:	
нижняя граница	42,70786
верхняя граница	50,25214

1. В ячейки A1, A2, A3 ввести обозначения «Среднее», «Стандартное отклонение», «Дисперсия». В ячейках B1, B2, B3 найти среднее, стандартное отклонение и дисперсию, используя функции **СРЗНАЧ()**, **СТАНДОТКЛОН()** и **ДИСП()** соответственно.

2. В ячейку A6 ввести обозначение «Объем выборки». В ячейке B6 найти объем выборочной совокупности, используя функцию **СЧЕТ()**.

3. В ячейку A7 ввести обозначение «Статистическая ошибка». В ячейке B7 найти статистическую ошибку для среднего по формуле: $S_{\bar{X}} = \sigma / \sqrt{n}$, где σ – стандартное отклонение, n – объем выборки.

4. Найти нижнюю и верхнюю границы 95-процентного доверительного интервала по формуле: $\bar{X} \pm t_{\alpha;n-1} \cdot S_{\bar{X}}$, где \bar{X} – среднее; $t_{\alpha;n-1}$ – коэффициент доверия (квантиль распределения Стьюдента с уровнем значимости α и $n-1$ степенью свободы); $S_{\bar{X}}$ – статистическая ошибка; $t_{\alpha;n-1} \cdot S_{\bar{X}}$ – предельная ошибка. Для этого выполнить следующее:

– в ячейки A8, A9 ввести обозначения «Уровень надежности» (γ), «Уровень значимости» (α). В B8 – 95%, а в ячейке B9 найти уровень значимости по формуле $1-\gamma$.

– в ячейку A10 ввести обозначение «Коэффициент доверия», в ячейке B10 найти квантиль распределения Стьюдента $t_{0,5; 24}$, используя функцию **СТЮДРАСПОБР()** с параметрами **Вероятность**: α , **Степени свободы**: $n-1$;

– в ячейку A11 ввести обозначение «Предельная ошибка» (Уровень надежности (95%), в ячейке A11 найти предельную ошибку по формуле $t_{\alpha;n-1} \cdot S_{\bar{X}}$;

– в ячейки A12, A13, A14 ввести обозначения «Доверительный интервал:», «нижняя граница», «верхняя граница», в ячейке B13 найти нижнюю границу доверительного интервала, B14 – верхнюю границу.

Задание 3. Найти статистические показатели по данным из примера 1, используя инструмент **Описательная статистика**.

1. Выполнить команды **Данные**►**Анализ**►**Анализ данных**. Из списка инструментов выбрать инструмент **Описательная статистика**. Установить следующие параметры:

Входной интервал – A1:A26, **Группирование** – По столбцам. Установить флажки **Метки** в первой строке (так как в ячейке A1 указан заголовок), **Итоговая статистика** (для получения всех статистических показателей) и **Уровень надежности** (для получения доверительного интервала). В поле **Уровень надежности** указать доверительную вероятность в процентах: **95**. Установить переключатель **Новый рабочий лист**. Для получения результатов нажать кнопку <ОК>. На новый рабочий лист выводятся значения статистических показателей, вычисленные по выборке значений прочности материала.

2. Определить нижнюю и верхнюю границы доверительного интервала (см. задание 2).

Материал А	
Среднее	46,48
Стандартная ошибка	1,82767612
Медиана	47
Мода	48
Стандартное отклонение	9,138380601
Дисперсия выборки	83,51
Экссесс	-0,601945339
Асимметричность	-0,196012938
Интервал	33
Минимум	29
Максимум	62
Сумма	1162
Счет	25
Наибольший(1)	62
Наименьший(1)	29
Уровень надежности(95,0%)	3,772138089
Доверительный интервал:	
нижняя граница	42,70786191
верхняя граница	50,25213809

Оценка значимости различий между выборками

Для решения задач, связанных с оценкой статистической значимости различий между выборками, применяются инструменты, входящие в состав пакета **Анализ данных**: **Двухвыборочный F-тест для дисперсии**, **Двухвыборочный t-тест с одинаковыми дисперсиями**, **Двухвыборочный t-тест с различными дисперсиями**.

Задание 4. Имеются значения показателя прочности 25-ти образцов материала А (см. задание 1) и 19-ти образцов материала В:

48	29	37	32	47	52	38	34	41	32
47	53	42	35	37	29	51	42	53	

Выполнить сравнение показателей прочности двух материалов.

Для определения, статистической значимости различий между материалами А и В по среднему значению показателя прочности применяют инструменты **Двухвыборочный t-тест с одинаковыми дисперсиями** или **Двухвыборочный t-тест с различными дисперсиями**. **Двухвыборочный t-тест с одинаковыми дисперсиями** используют, если различие между дисперсиями выборок статистически незначимо, **Двухвыборочный t-тест с различными дисперсиями** – если различие между дисперсиями статистически значимо.

Для определения статистической значимости различия между значениями дисперсии исследуемой величины (в данном примере – прочности) в двух анализируемых выборках используют инструмент **Двухвыборочный F-тест для дисперсии**.

1. Перейти на новый рабочий лист Excel. Скопировать на него столбец с данными о прочности образцов материала А (см. пример 1).

2. В ячейку В1 ввести заголовок «Материал В». В столбец В, ячейки В2:В20, ввести значения показателя прочности образцов материала В.

3. В диалоговом окне **Анализ данных** выбрать инструмент **Двухвыборочный F-тест для дисперсии**:

– установить параметры: *Интервал переменной 1* – A1:A26, *Интервал переменной 2* – B1:B20, *Альфа* 0,05. Установить флажок *Метки*. Выбрать переключатель *Новый рабочий лист*.

Для получения результатов нажать кнопку <OK>.

<i>Двухвыборочный F-тест для дисперсии</i>		
	<i>Материал А</i>	<i>Материал В</i>
Среднее	46,48	41
Дисперсия	83,51	67,11
Наблюдения	25	19
df	24	18
F	1,244354305	
P(F<=f) одностороннее	0,320703044	
F критическое одностороннее	2,149664535	

Значимость различия по изменчивости (т.е. значимость различия между величинами дисперсии) оценивается по следующим показателям:

F – F-критерий Фишера для оценки значимости различия между дисперсиями;

P (F <= f) одностороннее – расчетный уровень значимости: вероятность того, что различие между величинами дисперсии двух анализируемых выборок статистически незначимо;

F критическое одностороннее – квантиль распределения Фишера $F_{\alpha; n1-1; n2-1}$.

Если величина F больше, чем F критическое одностороннее и P (F <= f) одностороннее меньше заданного уровня значимости α , то можно считать, что **выборки значимо различаются по дисперсии** прочности. Обычно применяется значение $\alpha = 0,05$.

4. В ячейку A12 ввести результат анализа в виде: «Различие между дисперсиями статистически незначимо (значимо)».

5. В **Анализе данных** выбрать инструмент **Двухвыборочный t-тест с одинаковыми дисперсиями** или **Двухвыборочный t-тест с различными дисперсиями**, в зависимости от предыдущего результата. Установить следующие параметры: *Интервал переменной 1* – A1:A26, *Интервал переменной 2* – B1:B20, *Гипотетиче*

ская разность – 0, Альфа – 0,05. Установить флажок **Метки** (так как в первых ячейках каждого столбца указаны не данные, а заголовки). Установить переключатель **Новый рабочий лист**. Нажать <ОК>.

Для оценки значимости различия между средними используются показатели:

t-статистика – t-критерий Стьюдента;

t критическое двухстороннее – квантиль распределения Стьюдента t_{α, n_1+n_2-2});

P (T <= t) двухстороннее – расчетный уровень значимости P, представляющий собой вероятность того, что различие между средними значениями выборок статистически незначимо.

Если величина t-статистика больше, чем **t критическое двухстороннее** и P (T <= t) двухстороннее меньше заданного уровня значимости α , обычно $\alpha = 0,05$, то можно считать, что выборки значимо отличаются по среднему значению прочностных образцов материала.

6. В ячейку A16 ввести результат анализа в виде: «Различие между средними статистически незначимо (значимо)».

Корреляционный анализ в Excel

Корреляционный анализ позволяет выяснить, имеется ли значимая связь между двумя анализируемыми величинами. Для решения таких задач вычисляется коэффициент парной корреляции между анализируемыми величинами. Для этого применяется функция **КОРРЕЛ()** или инструмент **Корреляция** из пакета **Анализ данных**.

Задание 5. Анализируется зависимость прочности некоторого материала от содержания примесей и температуры обработки. Получены значения исследуемых величин для девяти образцов материала:

Содержание примесей, %	13,5	17,0	11,5	16,0	23,0	18,0	9,0	21,0	15,0
Температура обработки, °C	200	320	290	200	170	230	280	300	260
Удлинение образца при разрыве, мм	6,2	8,7	11,2	4,9	3,1	4,5	9,7	4,2	5,1

Найти коэффициенты парной корреляции между удлинением при разрыве и содержанием примесей, между удлинением при разрыве и температурой обработки.

1. Перейти на свободный рабочий лист. В ячейки A1, B1, C1 ввести заголовки «Содержание примесей», «Температура обработки», «Удлинение при разрыве». В столбцы A, B и C ввести исходные данные.

2. В ячейку E2 ввести обозначение «R=». В ячейке F2 найти коэффициент парной корреляции (r) между удлинением при разрыве и содержанием примесей, используя функцию **КОРРЕЛ()** со следующими аргументами: **Массив1 – A2:A10, Массив2 – C2:C10**.

3. Выполнить проверку статистической значимости коэффициента парной корреляции между удлинением при разрыве и содержанием примесей. По ее результатам сделать выводы о связи между исследуемыми величинами. Для этого выполнить следующее:

– в ячейку E3 ввести обозначение «n=», в F3 – найти объем выборки (см. задание 2, пункт 2);

– в ячейку E4 ввести обозначение « t ». В ячейке F4 рассчитать значение t -критерия Стьюдента по формуле $t = r \frac{\sqrt{n-2}}{\sqrt{1-r^2}}$ где r – коэффициент парной корреляции, найденный на шаге 1; n – объем анализируемой выборки;

– в ячейки E5, E6 ввести обозначения « α », « $t_{\alpha, n-2}$ ». В ячейку F5 ввести – 0,05, F6 – найти квантиль распределения Стьюдента $t_{\alpha, n-2}$ - Для этого используется функция **СТЮДРАСПОБР**();

– в ячейку E7 ввести обозначение « P ». В ячейке F7 найти расчетный уровень значимости P – вероятность незначимости корреляции между исследуемыми величинами. Для этого используется функция **СТЮДРАСП**() со следующими аргументами: **X** – **ABS(F4)**, **Степени свободы** – **7**, **Хвосты** – **2**. Использование функции **ABS** (модуль) необходимо из-за того, что аргумент **X** функции **СТЮДРАСП** должен быть положительным.

Содержание примесей	Температура обработки	Удлинение при разрыве	
13,5	200	6,2	$R = -0,796023$
17,0	320	8,7	$n = 9$
11,5	290	11,2	$t = -3,934507$
16,0	200	4,9	$\alpha = 0,05$
23,0	170	3,1	$t_{\alpha, n-2} = 2,364624$
18,0	230	4,5	$P = 0,005643$
9,0	280	9,7	
21,0	300	4,2	
15,0	260	5,1	

– по найденным значениям t -критерия Стьюдента и квантиля распределения Стьюдента, а также по расчетному уровню значимости сделать выводы о связи между исследуемыми величинами (удлинением при разрыве и содержанием примесей).

Если $t > t_{\alpha, n-2}$ и $P < \alpha$ (в данном примере используется $\alpha = 0,05$), то коэффициент парной корреляции между исследуемыми величинами статистически значим. Это означает, что имеется достаточно существенная линейная связь между исследуемыми величинами. При этом если $r > 0$, то связь между исследуемыми величинами положительная: с ростом одной величины увеличивается и другая. Значение $r < 0$ указывает на отрицательную связь между величинами: с ростом одной величины другая уменьшается.

Если коэффициент парной корреляции статистически незначим, это означает, что линейная связь между исследуемыми величинами выражена слабо. Это может означать, что связь между величинами является более сложной (нелинейной) или вообще отсутствует.

4. Найти коэффициент парной корреляции между удлинением при разрыве и температурой обработки материала (аналогично тому, как показано выше). Проверить его статистическую значимость. Сделать выводы о связи между исследуемыми величинами.

Регрессионный анализ в Excel

Методы регрессионного анализа позволяют по выборкам значений двух величин (X и Y) или нескольких величин (X_1, X_2, \dots, X_m, Y) построить модель (уравнение), описывающую связь между этими величинами. Основным математический метод, применяемый для построения таких моделей – метод наименьших квадратов.

В Excel основным средством решения задач регрессионного анализа является инструмент **Регрессия** из пакета **Анализ данных**. Кроме того, модели связи между двумя величинами могут строиться на основе диаграмм.

Задание 6. Исследуется зависимость между некоторыми двумя величинами (X и Y). Имеются значения этих величин, полученные в десяти наблюдениях:

X	2,2	2,8	3,5	2,6	2	2,9	3,1	3,4	1,9	3,5
Y	21,4	28,6	35,7	21,7	18,5	27,1	34,1	31,4	18,1	21,2

Построить линейную модель связи между исследуемыми величинами: $Y = A_0 + A_1 X$, где A_0, A_1 – коэффициенты модели, которые требуется определить, используя метод наименьших квадратов. Выполнить проверку адекватности модели и статистической значимости ее коэффициентов.

Решение задачи с помощью инструмента Регрессия

1. Перейти на свободный рабочий лист. В ячейку A1 ввести заголовок «X», в ячейку B1 – заголовок «Y». В ячейки A2:A11 и B2:B11 ввести исходные данные.
2. Окне **Анализ данных** выбрать инструмент **Регрессия**.

Регрессия

Входные данные

Входной интервал Y:

Входной интервал X:

Метки Константа - ноль

Уровень надежности: %

Параметры вывода

Выходной интервал:

Новый рабочий лист:

Новая рабочая книга

Остатки

Остатки График остатков

Стандартизованные остатки График подбора

Нормальная вероятность

График нормальной вероятности

OK Отмена Справка

3. Установить параметры: **Входной интервал Y** – B1:B11, **Входной интервал X** – A1:A11. Установить флажок **Метки**. В области **Параметры вывода** выбрать переключатель **Выходной интервал** и указать ячейку, с которой будет начинаться вывод результатов на текущий рабочий лист (например, ячейку A14). Для получе-

ния результатов нажать кнопку <ОК>. Результаты будут иметь примерно такой вид, как показано на рисунке.

Выводы итогов							
<i>Регрессионная статистика</i>							
Множественный R		0,735357212					
R-квадрат		0,540750229					
Нормированный R-квадрат		0,483344008					
Стандартная ошибка		4,656404212					
Наблюдения		10					
<i>Дисперсионный анализ</i>							
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>		
Регрессия	1	204,2391985	204,2391985	9,419714733	0,015367589		
Остаток	8	173,4568015	21,68210018				
Итого	9	377,696					
	<i>Коэффициенты</i>	<i>Стандартная</i>	<i>t-статистика</i>	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>
Y-пересечение	3,794223168	7,313234083	0,518816043	0,617931139	-13,07012486	20,65857119	-13,07012486
X	7,88020675	2,567548977	3,069155378	0,015367589	1,959428197	13,8009853	1,959428197

Результаты применения инструмента **Регрессия** занимают достаточно много места. Поэтому в поле **Выходной интервал** необходимо указывать такую ячейку, чтобы снизу и справа от нее имелось достаточное количество свободных ячеек.

Искомые коэффициенты модели, вычисленные по методу наименьших квадратов, находятся в столбце **Коэффициенты**. Сначала указывается коэффициент A_0 затем – A_1 . В данном примере $A_0 = 3,79$, $A_1 = 7,88$ (значения коэффициентов указаны с округлениями). Таким образом, связь между величинами X и Y может быть описана следующим уравнением: $Y = 3,79 + 7,88 \times X$.

Проверка адекватности модели выполняется по расчетному уровню значимости, указанному в столбце **Значимость F**. Эта величина представляет собой вероятность того, что построенная модель является неадекватной (недостаточно точной). Если расчетный уровень значимости меньше заданного уровня значимости α (обычно используется $\alpha = 0,05$), то модель адекватна исходным данным (т. е. достаточно точная). В данном примере величина **Значимость F** равна (округленно) 0,015, что меньше, чем 0,05. Таким образом, построенная модель связи между величинами X и Y является адекватной. Упрощенно говоря, это означает следующее: если в построенную модель ($Y = 3,79 + 7,88 \times X$) подставить известные значения X (т. е. значения 2,2; 2,8; ...; 3,5), то будут получены модельные значения Y , достаточно близкие к фактическим (21,4; 28,6; ...; 21,2).

Если построенная линейная модель оказывается неадекватной (величина **Значимость F** превышает заданный уровень значимости), это означает, что связь между исследуемыми величинами X и Y не может быть достаточно точно описана линейным уравнением $Y = A_0 + A_1 X$, т. е. является более сложной (нелинейной) или вообще отсутствует.

Проверка статистической значимости коэффициентов модели выполняется по расчетным уровням значимости, указанным в столбце **P-Значение**. Если расчетный уровень значимости меньше заданного уровня значимости α , то соответствующий коэффициент модели статистически значим. В данном примере **P-значение** для коэффициента A_0 составляет $0,618 > 0,05$, для коэффициента $A_1 - 0,015 < 0,05$. Таким образом, коэффициент A_0 статистически незначим, коэффициент $A_1 -$ статистически значим. Значимость коэффициента модели при входной переменной (в данном при-

мере – значимость коэффициента A_1 при переменной X означает, что имеется статистически значимая (т. е. достаточно четко выраженная) линейная связь величин Y и X .

Объяснение смысла статистической значимости (или незначимости) коэффициента A_0 , а также других величин, вычисляемых при использовании инструмента **Регрессия** требует глубокого изучения математической статистики (в частности, регрессионного анализа). В данной работе эти величины не рассматриваются.

Решение задачи с использованием диаграммы

Требуется построить точечную диаграмму, отражающую значения исследуемых величин, и нанести на нее график линейного уравнения связи между этими величинами $Y = A_0 + A_1 X$. Для этого выполнить следующее:

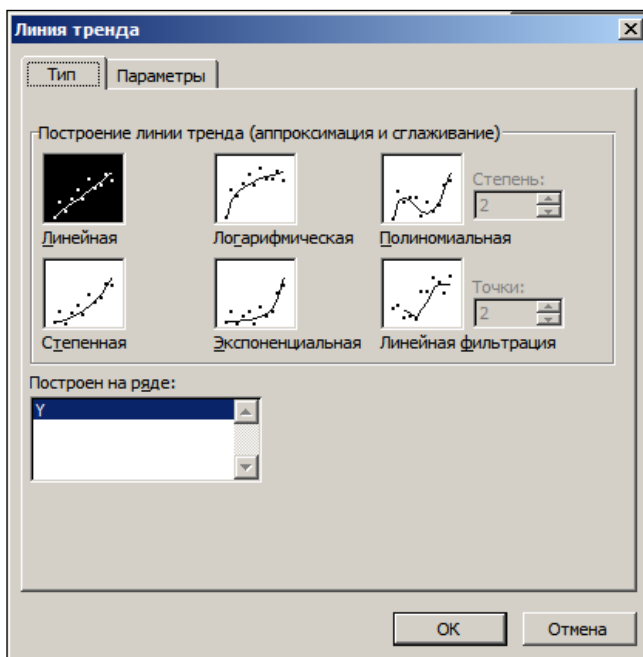
1. Построить точечную диаграмму, отражающую связь между исследуемыми величинами. Диаграмма строится в следующем порядке:

- выделить ячейки A1:B11;
- выполнить команды **Вставка**►**Диаграммы**►**Точечная**►**Точечная с маркерами**;

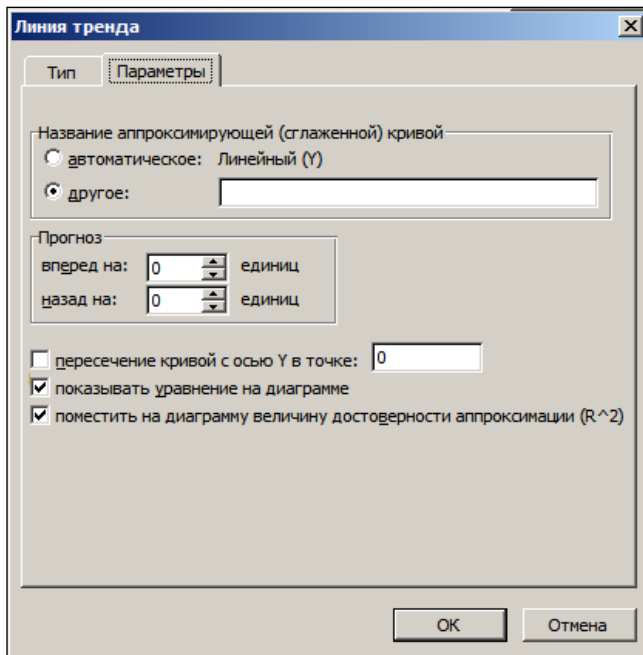
– в окне *Размещение диаграммы* выбрать кнопку *Поместить диаграмму на листе – имеющемся*. Нажать кнопку **<Готово>**. На текущем листе будет построена диаграмма.

2. Нанести на диаграмму линейную модель связи между исследуемыми величинами (линейный тренд). Тренд строится в следующем порядке:

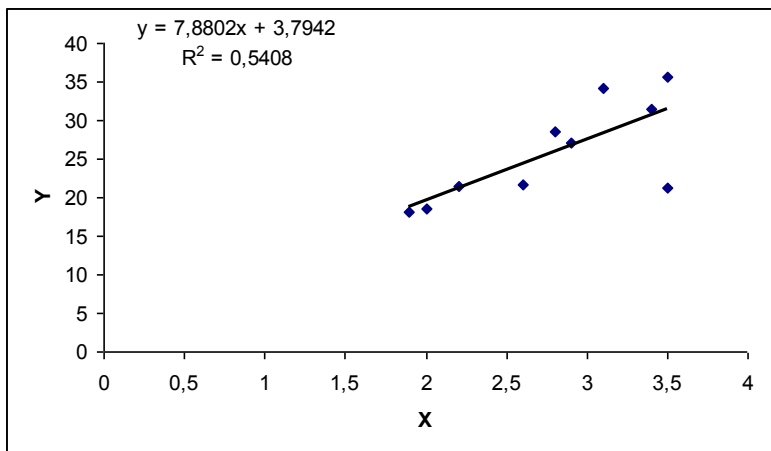
- выделить диаграмму щелчком мыши;
- из меню **Диаграмма** выбрать команду **Добавить линию тренда**; в окне *Линия тренда*, на вкладке **Тип**, выбрать вид линии тренда – **Линейная**.



На вкладке **Параметры** выбрать переключатель **Название аппроксимирующей сглаживающей кривой – другое**; установить флажки **Показывать уравнение на диаграмме** и **Поместить на диаграмму величину достоверности аппроксимации (R^2)**. Нажать кнопку <ОК>.



На диаграмме будет построен линейный тренд, указано уравнение модели, а также коэффициент детерминации, отражающий точность построенной модели (чем ближе эта величина к единице, тем точнее построенная модель).



Задание 7. По данным задания 5 построить линейную модель связи удлинения образца при разрыве с содержанием примесей и температурой обработки материала: $Y = A_0 + A_1 \cdot X_1 + A_2 \cdot X_2$, где Y – удлинение при разрыве (выходная переменная); X_1 – содержание примесей, X_2 – температура обработки материала (X_1, X_2 – входные переменные); A_0, A_1, A_2 – коэффициенты модели, которые требуется определить, используя метод наименьших квадратов. Выполнить проверку адекватности модели и статистической значимости ее коэффициентов.

Для решения задачи используется инструмент **Регрессия** точно так же, как в примере 6. При использовании инструмента **Регрессия** в параметре **Входной интервал X** указывается диапазон ячеек, включающий все значения входных величин (т. е. значения содержания примесей и температуры обработки материала).

Результаты решения данного примера приведены на рисунке.

Вывод итогов							
Регрессионная статистика							
Множественный R		0,887070707					
R-квадрат		0,786894438					
Нормированный R-квадрат		0,715859251					
Стандартная ошибка		1,490392668					
Наблюдения		9					
Дисперсионный анализ							
	df	SS	MS	F	Значимость F		
Регрессия	2	49,21237818	24,60618909	11,0775303	0,009677972		
Остаток	6	13,32762182	2,221270304				
Итого	8	62,54					
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%
Y-пересечение	7,723900223	3,804490171	2,030206381	0,088638449	-1,585351846	17,03315229	-1,585351846
Примесь	-0,426337863	0,125349164	-3,401202279	0,014474486	-0,733056217	-0,119619508	-0,733056217
Температура	0,021990022	0,010588646	2,077147216	0,083061963	-0,003914567	0,047894612	-0,003914567

Из полученных результатов видно, что модель связи удлинения образца при разрыве (Y) с содержанием примесей (X_1) и температурой обработки материала (X_2) имеет следующий вид: $Y = 7,72 - 0,426X_1 + 0,02X_2$. Модель является адекватной: величина **Значимость F** равна 0,01, т. е. меньше, чем 0,05. Из величин, указанных в столбце **P-значение**, видно, что коэффициент A_0 статистически незначим, коэффициенты A_1 и A_2 – статистически значимы. Из того, что коэффициент A_1 статистически значим, следует, что имеется значимая линейная связь между содержанием примеси и прочностью образца при разрыве. Следует также отметить, что эта связь – обратная (чем больше содержание примеси, тем меньше удлинение образца при разрыве), так как коэффициент A_1 – отрицательный. Статистическая значимость коэффициента A_2 указывает на линейную связь между температурой обработки материала и удлинением при разрыве. Следует также отметить, что эта связь – прямопропорциональная (чем больше температура, тем больше удлинение образца при разрыве), так как коэффициент A_2 – положительный.

Задание 8. Исследуется зависимость между некоторыми двумя величинами (X и Y). Имеются значения этих величин, полученные в семи наблюдениях:

X	5	10	20	30	40	50	60
Y	48	32	17	14	9	5	2

Построить модели связи между величинами Y и X : логарифмическую, степенную, экспоненциальную, линейную. Определить наиболее точную модель.

1. Для ввода исходных данных перейти на свободный рабочий лист. В ячейку A1 ввести заголовок «X», в ячейку B1 – «Y». В ячейки A2:A8 и B2:B8 ввести исходные данные.

2. Построить точечную диаграмму, отражающую связь между исследуемыми величинами (см. задание 6).

3. Сделать три копии построенной диаграммы. В результате на рабочем листе должны находиться четыре диаграммы (так как будут строиться четыре модели). Построить на диаграммах линейный, логарифмический, степенной и экспоненциальный тренды (по одному тренду на каждой диаграмме), как показано в задании 6. По максимальному значению коэффициента детерминации (R^2) определяется наиболее точная модель.

Самостоятельная работа

Небольшое частное кондитерское предприятие занимается производством фирменного печенья. Постоянные издержки производства (ПИ) составляют 20 000 руб. в месяц, а средние переменные издержки (СПИ) – 12 руб. на один килограмм произведенной продукции.

Если предприятие будет производить ОП кг печенья в месяц, то его общие издержки (ОИ) составят $20\,000 + 12 \times \text{ОП}$ руб., а общая выручка (ОВ) – $\text{ОП} \times \text{Ц}$ руб., где Ц – цена одного килограмма печенья (все произведенное печенье продается). Тогда прибыль предприятия (П) составит $\text{ОВ} - \text{ОИ} = \text{ОП} \times \text{Ц} - 20\,000 - 12 \times \text{ОП}$ руб.

С целью получения максимальной прибыли предприятие каждый месяц изменяет цену Ц и анализирует, как на изменение цены реагирует спрос населения. Сведения о цене и объеме продаж за первые шесть месяцев деятельности предприятия представлены в таблице.

Постоянные издержки 20 000 руб.
Средние переменные издержки 12 руб.

	январь	февраль	март	апрель	май	июнь
Цена, руб./кг	20,00	20,50	23,50	23,00	22,50	21,80
Продано, кг	6300	6400	4850	4950	5500	5700
Общая выручка, руб.						
Общие издержки, руб.						
Прибыль, руб.						

Рассчитайте показатели издержек, выручки и прибыли.

По имеющимся данным определите, по какой цене и в каком объеме следует производить продукцию в июле, чтобы получить наибольшую прибыль.

Идея решения задачи состоит в нахождении аналитической зависимости между спросом населения и ценой на покупаемую продукцию, т. е. в нахождении вида функции $\text{Спрос} = F(\text{Цена})$.

Для решения задачи используйте материалы задания 8. Следует заметить, что при выборе вида приближающего уравнения, прежде всего, должна приниматься во внимание экономическая (физическая, социальная и т.п.) сущность исследуемого явления или процесса, иначе в большинстве случаев будет получаться, что наилучшим приближающим уравнением является полином 6-й степени. Для рассматри-

ваемой задачи уравнения, которые отвечают ее экономической сущности, имеют линейный или степенной вид.

Получив наилучшую модель, подставьте уравнение в уравнение прибыли, и рассчитайте ее теоретические значения.

Для определения максимальной прибыли, цены и объема производства которые ее обеспечат, примените мастер диаграмм. После подстановки уравнения спроса в уравнение прибыли последнее приобретает квадратичный вид, т.е. является параболой, или, иными словами, полиномом 2-й степени. Постройте точечную диаграмму, взяв в качестве независимой переменной (X) – цену, зависимой переменной (Y) – рассчитанную прибыль, значения которой аппроксимируйте полиномом 2-й степени по рассмотренной выше технологии. По результату аппроксимации определите максимальную прибыль, цену, которая ее обеспечивает. Подставьте значение цены в уравнение спроса, получите оптимальный объем производства.

Контрольные вопросы

1. Многомерный статистический анализ в Microsoft Excel.
2. Методы и средства анализа данных в среде Excel 2007.
3. Деловая графика Excel и ее применение в статистическом анализе.
4. Построение гистограммы и полигона частот в Excel.
5. Типы статистических таблиц и их создание в Excel.
6. Надстройка Анализ данных и ее инструменты.
7. Статистические функции и их применение в статистическом анализе данных.
8. Факторный анализ в Excel.
9. Дискриминантный в Excel.
10. Выполнение кластерного анализа в Excel.