

ГОУ ВПО
Оренбургская Государственная медицинская академия
федерального агентства по социальному развитию и
здравоохранению.

Кафедра медицинской и биологической физики.

Колосова Н.И.
Бахарева Г.В.
Денисов Е.Н.

Пособие для самостоятельной работы студентов
(лечебного, педиатрического, медико-профилактического,
стоматологического, фармацевтического, ВСО факультетов)

Тема: Использование электронных таблиц Excel для
решения задач по статистике.

Оренбург 2009 г

Цель: 1. Научиться применять электронные таблицы Excel для статистической обработки и графического отображения полученных данных.

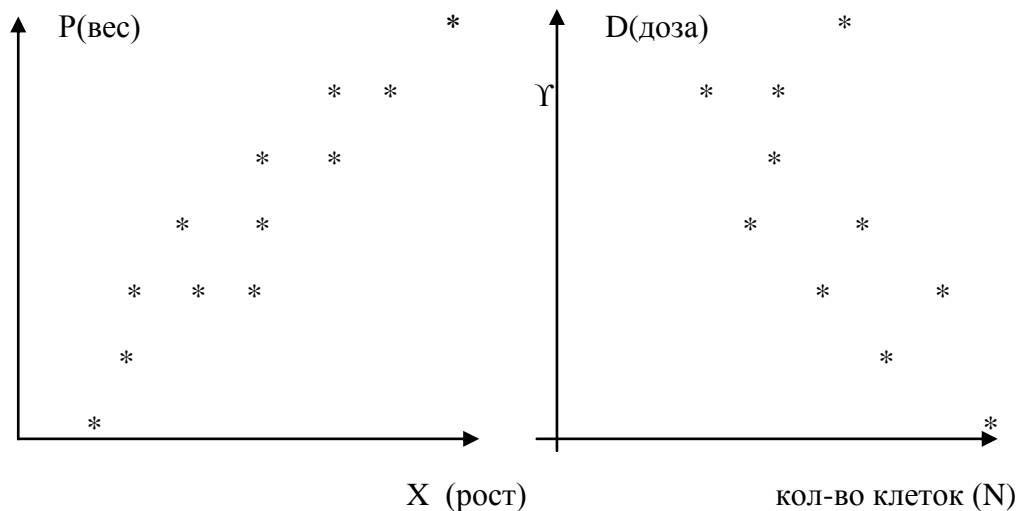
Т Е М А: КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗ.

1. Понятие корреляции.

Существуют две категории связей между признаками:

1. **Функциональные** - каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной (высота столба ртути соответствует определённой температуре);
2. **Корреляционные** - (статистические) - численному значению одной переменной соответствует много значений другой переменной (одному росту соответствует множество значений веса).

Если есть результаты наблюдения, то первый шаг в анализе процесса состоит в построении различного рода графиков, с помощью которых можно было бы исследовать его основные характеристики. Наиболее простую иллюстрацию парных наблюдений даёт график (диаграмма) рассеяния.



Для количественной оценки корреляционной зависимости вычисляют коэффициент корреляции.

Коэффициент корреляции-это число показывающее степень зависимости одной переменной величины от другой.

Вычисляется по формуле:

$$R = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (1)$$

Свойства коэффициента корреляции:

1. r - число; лежащее в интервале от -1 до $+1$ ($-1 \leq r \leq 1$).
2. если $r = \pm 1$, то точки лежат на одной прямой, следовательно, зависимость между x и y – функциональная.
3. если $r \geq 0.7$, существует сильная линейная зависимость между переменными.
4. если $r = 0.5-0.6$ то зависимость между переменными средняя.
5. если $r \leq 0.5$, то корреляция между переменными слабая.
6. О тесной связи можно говорить, когда $r \geq 0,7$

РЕГРЕССИОННЫЙ АНАЛИЗ.

Коэффициент корреляции указывает лишь на степень связи в вариации двух переменных величин, т.е. дает меру тесноты этой связи, но не дает возможность судить о том, как количественно меняется одна величина по мере изменения другой. На этот вопрос позволяет ответить другой метод определения связи между вариационными признаками - метод регрессии.

При простой корреляции изучается зависимость между изменчивостью двух признаков X и Y . С помощью регрессии ставится дополнительная задача: установить, как количественно меняется одна величина при изменении другой на единицу.

Линейная регрессия.

Зависимость между биологическими признаками может быть самой разнообразной. В большем числе случаев эмпирические регрессии выражаются простыми уравнениями линейной регрессии: $y = ax + b$

Математическими преобразованиями можно получить формулы для вычисления коэффициентов a и b .

$$a = \frac{n \sum x_i y_i - \sum x_i \cdot \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (2)$$

$$b = \frac{\sum y_i \sum x_i^2 - \sum x_i \cdot \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (3)$$

Определив коэффициенты a и b , находим уравнение регрессии:
 $y = ax + b$.

Задача:

В анализах крови определяли: X -содержание гемоглобина(%), Y -оседание эритроцитов крови за 2 часа(мм). Построить график рассеяния. Найти уравнение регрессии. Найти коэффициент корреляции.

X	77	80	82	79	84	75	82	79	87	87	87	90	97	96	92
Y	32	33	33	34	34	34	34	35	36	37	37	38	40	40	40

В программе Excel создаём таблицу для решения задачи.

Ход решения задачи:

Задача построения таблиц разбивается на несколько этапов. В первую очередь необходимо ввести данные в таблицу Excel.

Ввод данных в таблицу Excel:

1. Установить табличный курсор на первую ячейку создаваемой таблицы. (например, на **B1**).
2. Ввести заголовок-« X_i » и нажать клавишу **Enter**.
3. Установить табличный курсор на ячейку **C1** таблицы, ввести заголовок - « Y_i » и нажать клавишу **Enter**.
4. Аналогично вводятся названия остальных ячеек.
5. В ячейки «**B2-B16**» вводятся значения первого ряда данных.
6. В ячейки «**C2-C16**» вводятся значения второго ряда данных.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K
1		X_i	Y_i	$X_i - X_{cp}$	$Y_i - Y_{cp}$	$(X_i - X_{cp})^2$	$(Y_i - Y_{cp})^2$	$(X_i - X_{cp}) * (Y_i - Y_{cp})$	X_i^2	$X_i * Y_i$	
2		77	32								
3		80	33								
4		82	33								
5		79	34								
6		84	34								
7		75	34								
8		82	34								
9		79	35								
10		87	36								
11		87	37								
12		87	37								
13		90	38								
14		97	40								
15		96	40								
16		92	40								
17											
18											
19											
20											
21											

Вычисление сумм:

1. Установить табличный курсор на ячейку «**B17**» данной таблицы.
2. Указателем мыши нажать кнопку «**Атосумма**» на панели инструментов.
3. Если указываемый пунктиром блок соответствует требуемому диапазону суммирования, нажать клавишу **Enter**.
4. Аналогично найти сумму второго ряда данных.

Ввод формул:

Ввод формулы должен начинаться со знака «=» (равно). Все записи должны осуществляться латинскими буквами. Для нахождения среднего значения сделаем следующее:

1. Установить табличный курсор на ячейку «B18» и ввести с клавиатуры следующую формулу «=B17/15», нажать клавишу **Enter** (15-количество данных задачи).
2. Аналогично найти среднее значение второго ряда данных. В результате в ячейках «B18» и «C18» появились средние значения.

	A	B	C	D	E	F	G	H	I	J	K
1		X_i	Y_i	$X_i - X_{cp}$	$Y_i - Y_{cp}$	$(X_i - X_{cp})^2$	$(Y_i - Y_{cp})^2$	$(X_i - X_{cp}) * (Y_i - Y_{cp})$	X_i^2	$X_i * Y_i$	
2		77	32								
3		80	33								
4		82	33								
5		79	34								
6		84	34								
7		75	34								
8		82	34								
9		79	35								
10		87	36								
11		87	37								
12		87	37								
13		90	38								
14		97	40								
15		96	40								
16		92	40								
17	Сумма	1274	537								
18	Ср.знач.	85	35,8								
19											
20											
21											

Для выполнения следующего этапа решения задачи необходимо найти разницу между каждым значением X_i и средним значением выборки. Для этого сделаем следующее:

1. Установить табличный курсор на ячейку «D2» и ввести с клавиатуры следующую формулу «=B2-\$B\$16», нажать клавишу **Enter**. В ячейке «D2» появится число «-8» равное разности между первым значением X_i и средним значением выборки.
2. Снова установить табличный курсор на правый нижний угол ячейки «D2», нажать левую кнопку мыши (курсор должен приобрести вид креста +) и выделить блок ячеек, в которые будет осуществляться копирование. Отпускаем кнопку мыши и в ячейках появляются численные значения разницы между каждым значением X_i и средним значением выборки. Получим следующую таблицу.

The screenshot shows a Microsoft Excel spreadsheet with the following data:

	A	B	C	D	E	F	G	H	I	J	K
1		Xi	Yi	Xi-Xcp	Yi-Ycp	(Xi-Xcp) ²	(Yi-Ycp) ²	(Xi-Xcp)*(Yi-Ycp)	Xi ²	Xi*Yi	
2		77	32	-8							
3		80	33	-5							
4		82	33	-3							
5		79	34	-6							
6		84	34	-1							
7		75	34	-10							
8		82	34	-3							
9		79	35	-6							
10		87	36	2							
11		87	37	2							
12		87	37	2							
13		90	38	5							
14		97	40	12							
15		96	40	11							
16		92	40	7							
17	Сумма	1274	537								
18	Ср.знач.	85	35,8								

3. Аналогичные действия выполним для второго ряда данных « Y_i ».
4. Далее найдём квадраты разностей $(X_i - X_{cp})^2$. Для этого необходимо установить табличный курсор на ячейку «F2» и ввести с клавиатуры следующую формулу «=D2*D2», нажать клавишу **Enter**. В ячейке «F2» появится число «63».
5. Далее скопируем содержимое ячейки «F2» во все остальные ячейки «F3-F16». Аналогичные действия выполним для второго ряда данных. Возведём в квадрат $(Y_i - Y_{cp})$.
6. Для выполнения действия « $(X_i - X_{cp}) * (Y_i - Y_{cp})$ » установить табличный курсор на ячейку «H2» и ввести с клавиатуры следующую формулу «=D2*E2», нажать клавишу **Enter**. Скопировать содержимое ячейки «H2» во все остальные ячейки «H3-H16».
7. Для выполнения действия « $(X_i)^2$ » установить табличный курсор на ячейку «I2» и ввести с клавиатуры следующую формулу «=B2*B2», нажать клавишу **Enter**. Скопировать содержимое ячейки «I2» во все остальные ячейки «I3-I16».
8. Для выполнения действия « $(X_i * Y_i)$ » установить табличный курсор на ячейку «J2» и ввести с клавиатуры следующую формулу «=B2*C2», нажать клавишу **Enter**. Скопировать содержимое ячейки «J2» во все остальные ячейки «J3-J16».
9. Просуммируем содержимое ячеек «F, G, H, I, J» и получим таблицу.

	A	B	C	D	E	F	G	H	I	J	K
1		Xi	Yi	Xi-Xcp	Yi-Ycp	(Xi-Xcp) ²	(Yi-Ycp) ²	(Xi-Xcp)*(Yi-Ycp)	Xi ²	Xi*Yi	
2		77	32	-8	-4	63	14	30	5929	2464	
3		80	33	-5	-3	24	8	14	6400	2640	
4		82	33	-3	-3	9	8	8	6724	2706	
5		79	34	-6	-2	35	3	11	6241	2686	
6		84	34	-1	-2	1	3	2	7056	2856	
7		75	34	-10	-2	99	3	18	5625	2550	
8		82	34	-3	-2	9	3	5	6724	2788	
9		79	35	-6	-1	35	1	5	6241	2765	
10		87	36	2	0	4	0	0	7569	3132	
11		87	37	2	1	4	1	2	7569	3219	
12		87	37	2	1	4	1	2	7569	3219	
13		90	38	5	2	26	5	11	8100	3420	
14		97	40	12	4	146	18	51	9409	3880	
15		96	40	11	4	122	18	46	9216	3840	
16		92	40	7	4	50	18	30	8464	3680	
17	Сумма	1274	537			631	104	236	108836	45845	
18	Ср.знач.	85	35,8								
19											
20											
21											
22											
23											

Расчёт коэффициента корреляции и коэффициентов a и b для нахождения уравнения регрессии.

1. Установить табличный курсор на ячейку в любую свободную ячейку, например, в ячейку «B20» и ввести с клавиатуры следующую формулу «=H17/корень(F17*G17)», нажать клавишу *Enter*. (для нахождения корня квадратного необходимо воспользоваться *Мастером функций*). В ячейке «B20» появится число «0,92». Это значение коэффициента корреляции.
2. Делают вывод: $R=0,92$ – зависимость сильная, прямопропорциональная.
3. Для нахождения коэффициента a установить табличный курсор на ячейку «B21» и ввести с клавиатуры следующую формулу «=(15*J17-B17*C17)/(15*I17-B17*B17)», нажать клавишу *Enter*. В ячейке «B21» появится значение коэффициента a ($a=0,37$).
4. Для нахождения коэффициента b установить табличный курсор на ячейку «B22» и ввести с клавиатуры следующую формулу «=(C17*I17-B17*J17)/(15*I17-B17*B17)», нажать клавишу *Enter*. В ячейке «B22» появится значение коэффициента b ($b=4,1$).

В результате получим таблицу.

Microsoft Excel

Файл Правка Вид Вставка Формат Сервис Данные Окно Справка

Введите вопрос

Calibri 11 Ж К Ч

ЛВ fx

Книга2-10

	A	B	C	D	E	F	G	H	I	J	K	L
1		X_i	Y_i	$X_i - X_{cp}$	$Y_i - Y_{cp}$	$(X_i - X_{cp})^2$	$(Y_i - Y_{cp})^2$	$(X_i - X_{cp}) * (Y_i - Y_{cp})$	X_i^2	$X_i * Y_i$		
2		77	32	-8	-4	63	14	30	5929	2464		
3		80	33	-5	-3	24	8	14	6400	2640		
4		82	33	-3	-3	9	8	8	6724	2706		
5		79	34	-6	-2	35	3	11	6241	2686		
6		84	34	-1	-2	1	3	2	7056	2856		
7		75	34	-10	-2	99	3	18	5625	2550		
8		82	34	-3	-2	9	3	5	6724	2788		
9		79	35	-6	-1	35	1	5	6241	2765		
10		87	36	2	0	4	0	0	7569	3132		
11		87	37	2	1	4	1	2	7569	3219		
12		87	37	2	1	4	1	2	7569	3219		
13		90	38	5	2	26	5	11	8100	3420		
14		97	40	12	4	146	18	51	9409	3880		
15		96	40	11	4	122	18	46	9216	3840		
16		92	40	7	4	50	18	30	8464	3680		
17	Сумма	1274	537			631	104	236	108836	45845		
18	Ср.знач.	85	35,8									
19												
20	R= 0,92											
21	a= 0,37											
22	b= 4,1											

Готово

Пуск биометр Microsoft Excel Рабочий стол Мои документы 12:20

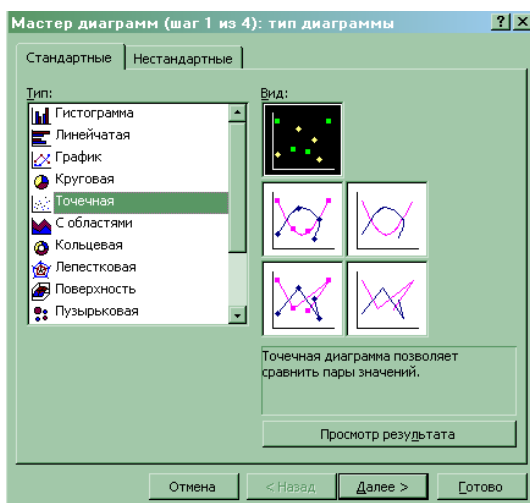
Построение графика рассеяния и уравнения регрессии.

Этап 1. Ввод данных.

Прежде чем строить диаграмму, необходимо ввести данные в таблицу Excel.

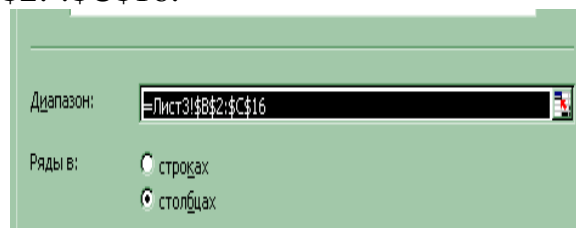
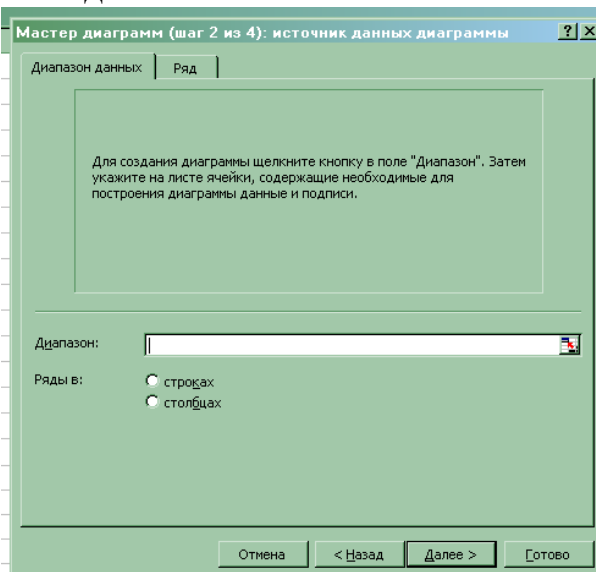
Этап 2. Выбор типа диаграмм.

На панели инструментов необходимо нажать кнопку *Мастер диаграмм*. В появившемся диалоговом окне *Мастер диаграмм шаг 1 из 4* указать тип диаграммы. Выбираем точечную диаграмму и нажимаем кнопку *Далее*.



Этап 3. Указание диапазонов.

В появившемся диалоговом окне *Мастер диаграмм шаг 2 из 4* необходимо выбрать вкладку *Диапазон* и в поле *Диапазон* указать интервал данных, т.е. ввести ссылку на ячейки, содержащие данные, которые необходимо представить на диаграмме. Для этого с помощью клавиши *Delete* необходимо очистить рабочее поле *Диапазон* и, убедившись, что в нём остался только мигающий курсор, следует навести указатель мыши на левую верхнюю ячейку данных (B2), нажать левую кнопку мыши и, не отпуская её, протянуть указатель мыши к правой нижней ячейке, содержащей выносимые на диаграмму данные (C16), затем отпустить левую кнопку мыши. В рабочем поле должна появиться запись: `=Лист1!B2: !C16`.

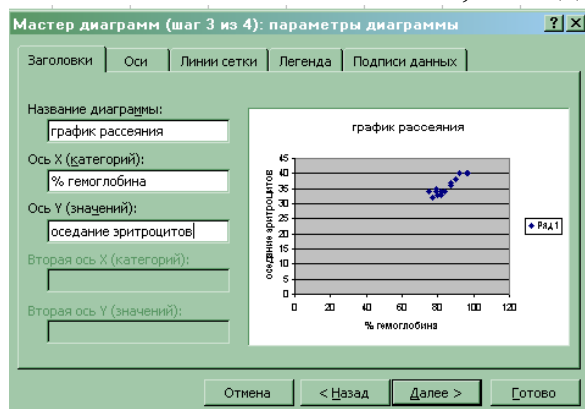


После появления требуемой записи диапазона необходимо нажать кнопку *Далее*.

Этап 4. Ввод подписей по осям и введение заголовка.

В появившемся диалоговом окне *Мастер диаграмм шаг 3 из 4* необходимо выбрать вкладку *Заголовк*, щёлкнув по ней указателем мыши.

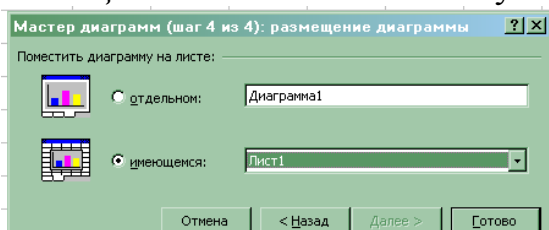
В поле *Название диаграммы* ввести название: «График рассеяния». Затем аналогичным образом ввести в рабочие поля *Ось X* и *Ось Y*: соответствующие названия: «% гемоглобина», «оседание эритроцитов».



Нажать кнопку *Далее*.

Этап 5. Выбор места размещения.

В четвёртом окне необходимо указать место размещения диаграммы. В появившемся диалоговом окне *Мастер диаграмм шаг 4 из 4* необходимо переключатель *Поместить диаграмму на листе*: установить в нужное положение. В нашем примере устанавливаем переключатель в положение *имеющемся* и нажимаем кнопку *Готово*.

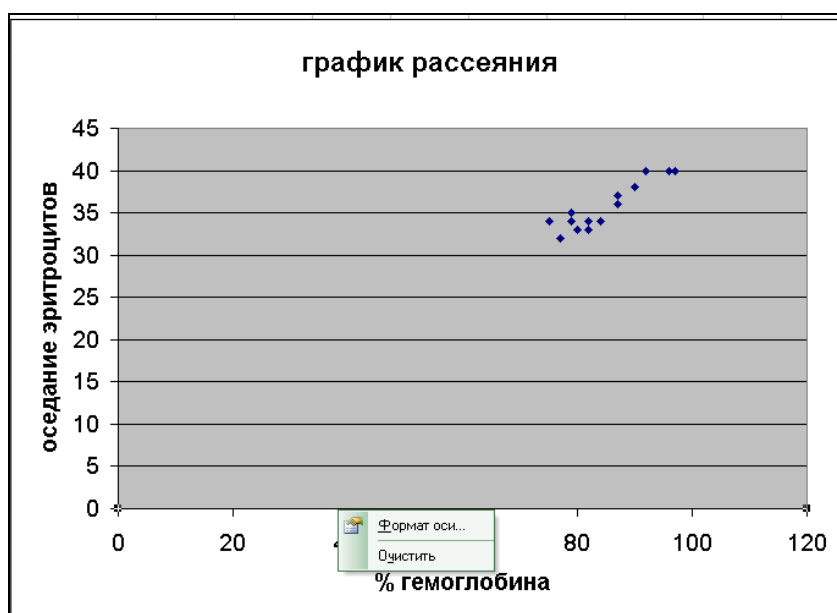


На текущем листе появляется следующая диаграмма.

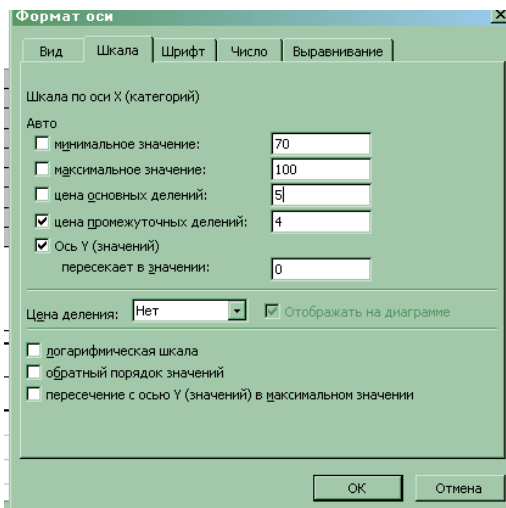
Этап 6. Редактирование диаграммы.

1. Формат оси.

Её необходимо отредактировать таким образом, чтобы точки были расположены не в углу, а на всём поле. Для этого необходимо изменить масштаб осей. Для этой цели выделяем *ось X* и после щелчка правой кнопкой появляется диалоговое окно, в котором выбираем функцию *Формат оси*.



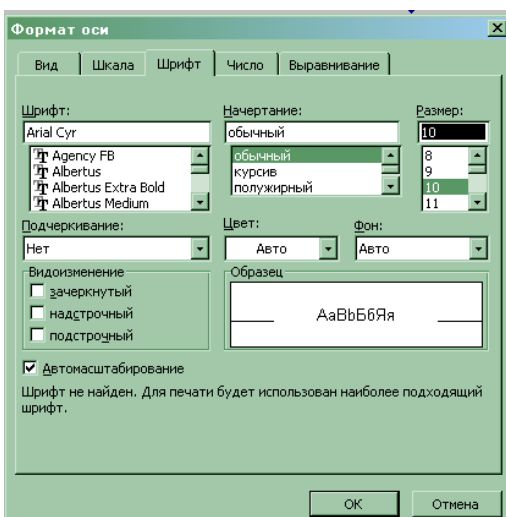
Появляется диалоговое окно *Формат оси*. Выбираем вкладку *Шкала* (щелкнув по ней указателем мыши) и в поле *Шкала по оси X* ввести *минимальное значение* «70» (меньше меньшего) из первого ряда данных, *максимальное значение* «100» (больше большего) и *цену одного деления* «5».



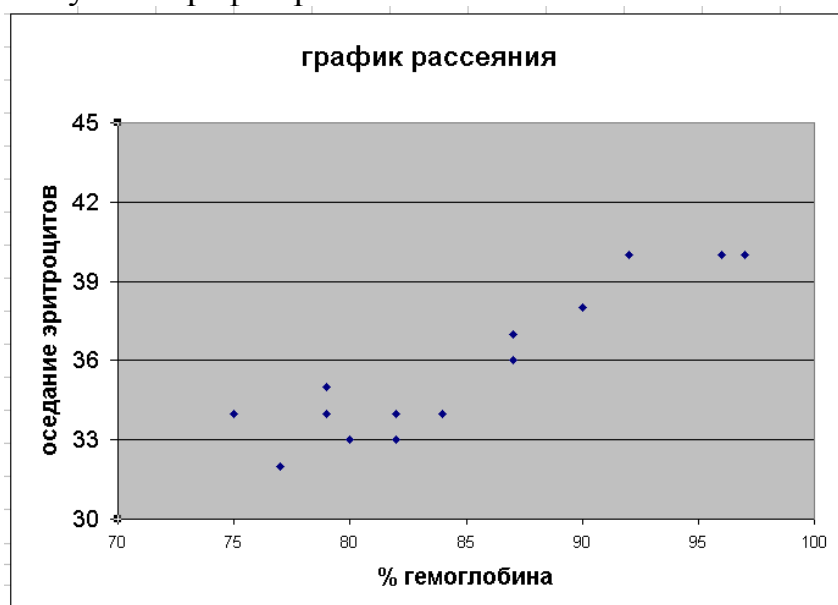
2. Формат размера шрифта.

Выбираем вкладку *Шрифт* (щелкнув по ней указателем мыши) и в поле *Размер* выбираем размер равный «10» и нажимаем кнопку *Готово*.


Аналогичным образом форматируем *Ось Y*.

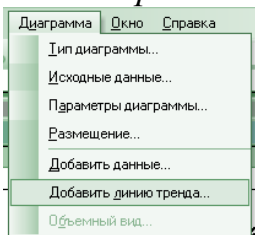


Получаем график рассеяния.

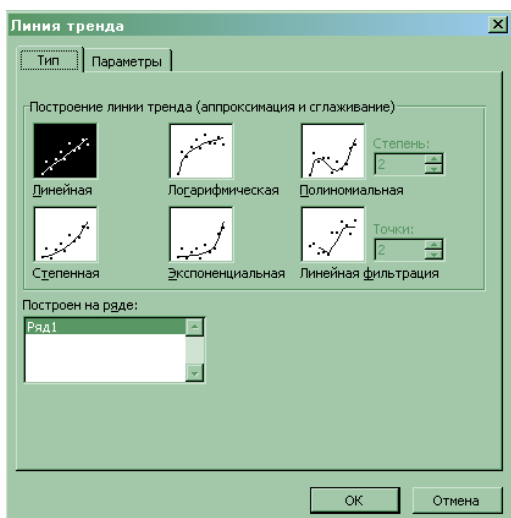


Этап 7. Построение линии регрессии.

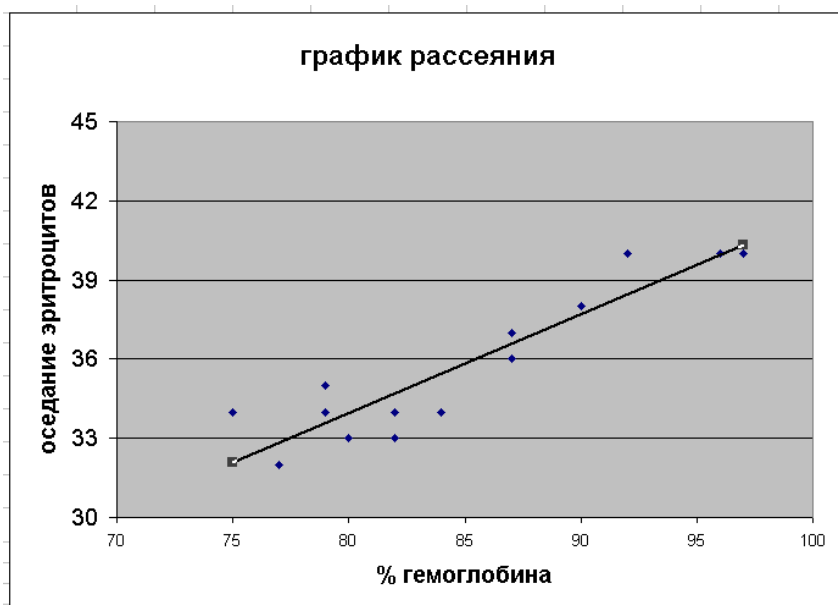
В строке *меню* выбираем кнопку *Диаграмма*  (щелкнув по ней указателем мыши), появляется диалоговое меню, в котором выбираем функцию *Добавить линию тренда*.



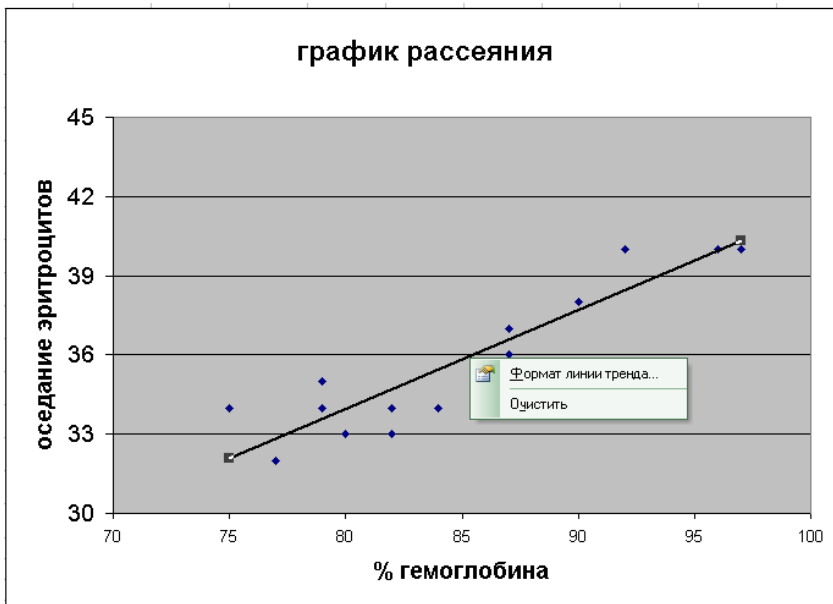
Появляется диалоговое окно *Добавить линию тренда*, в котором выбираем тип линии *Линейная*.



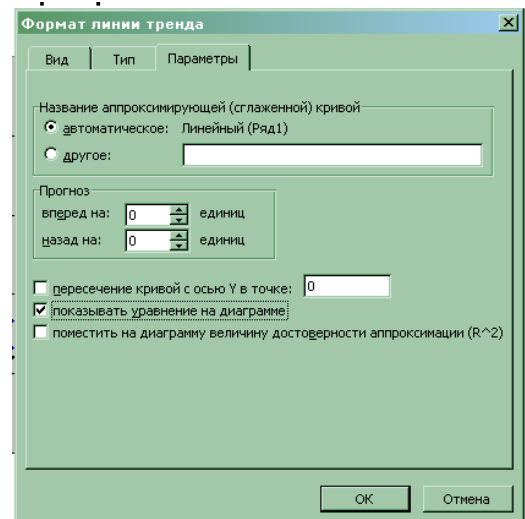
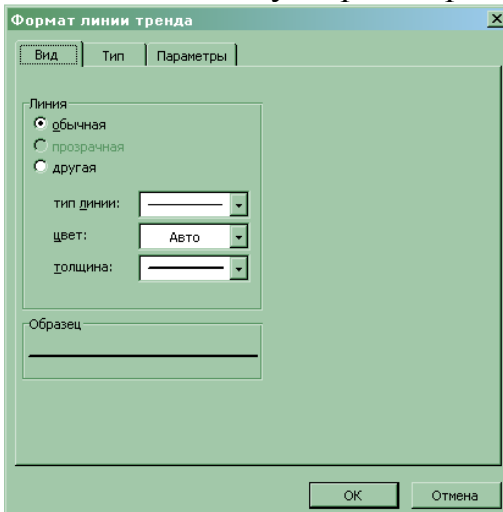
Далее нажимаем кнопку *OK* и появляется *График рассеяния* с линией регрессии.



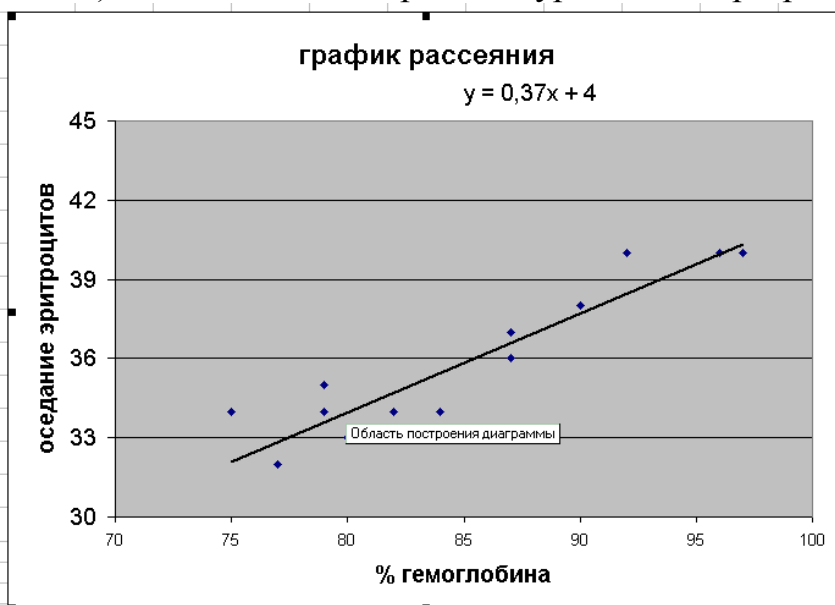
Осталось получить формулу этого уравнения. Для этой цели необходимо выделить линию (щелкнув по ней указателем мыши) и выбрать пункт *Формат линии тренда*.



В результате появляется диалоговое окно *Формат линии тренда*, в котором нажимаем кнопку *Параметры*.



В поле *Показывать линию тренда* ставим галочку (щелкнув по ней указателем мыши). Появляется диаграмма с уравнением регрессии: $Y=0,37X+4$.



Т Е М А: Построение гистограммы

Исходным понятием статистики является понятие **генеральная совокупность**, объединяющее обычно какое-либо множество испытуемых (учащихся) по одному или нескольким интересующим признакам. Главное требование к выделению изучаемой совокупности — это ее качественная однородность, например, по уровню знаний, росту, весу и другим признакам.

Применение большинства статистических методов основано на идее использования небольшой случайной совокупности испытуемых из общего числа тех, на которых можно было бы распространить (генерализовать) выводы, полученные в результате изучения совокупности. Эта небольшая совокупность в статистике называется **выборочной совокупностью** (или короче — **выборкой**).

Любая статистическая работа предполагает выбор генеральной совокупности. Генеральная совокупность – совокупность однородных статистических данных, отличающихся друг от друга, но при этом имеющих сходные свойства. Но, с генеральной совокупностью на практике работать невозможно из-за огромного числа объектов, входящих в нее. Поэтому на практике формируют выборку. Выборка – часть генеральной совокупности. На выборке измеряют переменные.

Переменная – это то, что можно измерять, контролировать или чем можно манипулировать в исследованиях. Иными словами, переменная – это то, что варьируется, изменяется, а не остается постоянным. Например, измеряя давление или содержание лейкоцитов в крови можно получить различные значения у различных пациентов или значения для одного пациента в разное время суток.

Можно привести примеры разных переменных. И, понятно, что все они будут отличаться своими характеристиками.

Так как значения переменных не постоянны, нужно научиться описывать их изменчивость.

Для этого придуманы описательные статистики: минимум, максимум, среднее, дисперсия, стандартное отклонение, медиана, квартили, мода и т.д. Они дают общие представления о значениях, которые принимает переменная.

Минимум и максимум – это минимальное и максимальное значения переменной.

Среднее – это сумма всех значений переменной, деленное на количество переменных.

$$\bar{x} = \frac{\sum x_i}{n}$$

Дисперсия и стандартное отклонение - наиболее часто используемые меры изменчивости переменной.

Дисперсия рассчитывается по формуле:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

Стандартное отклонение равно квадратному корню из дисперсии. Формально имеем:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Чем выше стандартное отклонение и дисперсия, тем сильнее разбросаны значения переменной относительно среднего

Медиана- это величина, относительно которой ряд распределения делится на две равные части: в обе стороны от медианы располагается одинаковое число вариант. При наличии небольшого числа вариант медиана определяется довольно просто. Для этого все данные нумеруют, и при нечетном числе вариант центральная варианта и будет его медианой. При четном числе определяется как среднее значение двух величин, находящихся в середине ряда. Для величин, по которым построена гистограмма, медиану можно определить следующим способом. Необходимо найти класс, в котором содержится медиана. Для этого необходимо складывать частоты встречаемости по классам до тех пор, пока сумма частот не превзойдет половину всех членов ряда. Данный класс называется медианным. Тогда медиану можно найти по формуле:

$$Me = x_n + \lambda \left(\frac{\frac{n}{2} - \sum f_i}{f_{Me}} \right)$$

где x_n - нижняя граница интервала, содержащего медиану, $\sum f_i$ - сумма накопленных частот, стоящая перед медианным классом, λ - величина классового интервала, f_{Me} – частота медианного класса, n - общее число наблюдений.

Мода- это величина, наиболее часто встречающаяся в данной совокупности. Класс с наибольшей частотой называется модальным. Моду можно найти по формуле:

$$Mo = x_n + \lambda \left(\frac{f_2 - f_1}{2f_2 - f_1 + f_3} \right)$$

где x_n - нижняя граница модального класса, f_2, f_1 - частота класса, предшествующего модальному, f_3 - частота класса, следующего за модальным, λ - ширина классового интервала.

Коэффициент асимметрии характеризует симметричность графика распределения относительно среднего значения. Он принимает значения от -1 до 1.

Коэффициент асимметрии определяется по формуле:

$$As = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^3}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^{\frac{3}{2}}}$$

Экссесс, в свою очередь, характеризует узковершинность или плосковершинность распределения. Экссесс определяется по формуле:

$$\Theta = \frac{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^4}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2} - 3$$

Далее необходимо характеристики, полученные на выборке, перенести на всю генеральную совокупность. Для этого вводят понятие доверительного интервала.

Доверительный интервал- это интервал, в который попадает среднее значение генеральной совокупности с той или иной вероятностью.

$$\bar{x} - t \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{x} + t \frac{\sigma}{\sqrt{n}}$$

В качестве наглядного представления выборочной совокупности можно рассмотреть гистограмму. Гистограмма- это фигура, состоящая из прямоугольников, ширина которых- ширина класса, а высота- функция плотности вероятности.

Рассмотрим алгоритм построения гистограммы на конкретной задаче.

Дан рост группы детей. Согласно полученным данным построить гистограмму.


70 69 72 73 71 66 73 67 68 73
 71 69 67 74 71 70 70 67 71 69
 70 70 70 71 74 74 71 69 72 71

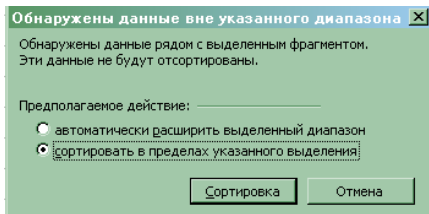
Ввод данных

1. В ячейку A1 ввести заголовок *РОСТ*
2. В ячейки A2-A31 ввести данные задачи.

Построение вариационного ряда

1. В ячейку B1 ввести заголовок *вариаци. ряд*
2. Скопировать данные ячеек A2-A31 в ячейки B2-B31. Для этого выделить ячейки A2-A31, щелкнуть правой кнопкой мыши по выделенным ячейкам, выбрать пункт меню *копировать*, встать на ячейку B2, щелкнуть правой кнопкой мыши, выбрать пункт меню *вставить*.
3. Выделить ячейки B2-B31. На панели инструментов выбрать кнопку

сортировать по возрастанию,  в появившемся окне выбрать предполагаемое действие *сортировать в пределах указанного диапазона*, нажать *сортировка*.



Расчет необходимых величин

1. В ячейке C1 вводим заголовок *Размах*.
2. В ячейке C2 вводим формулу «=B31-B2». Для этого необходимо в ячейке C2 с клавиатуры ввести знак =, щелкнуть мышкой по ячейке B31, с клавиатуры ввести знак -, щелкнуть мышкой по ячейке B2. После нажатия Enter, в ячейке C2 появится число 8.
3. В ячейке D1 вводим заголовок *число Кл.*
4. В ячейке D2 вводим число классов 4.
5. В ячейку E1 вводим заголовок *Ширина*.
6. В ячейку E2 вводим формулу «=C2/D2». Для этого необходимо в ячейке E2 ввести с клавиатуры знак =, щелкнуть мышкой по ячейке C2, с клавиатуры ввести знак /, щелкнуть мышкой по ячейке D2. После нажатия Enter, в ячейке E2 появится число 2.
7. В ячейку F1 вводим заголовок *Объем*.
8. В ячейку F2 вводим объем выборки 30.

Определение границ интервалов

1. В ячейках E5, E6, E7, E8 вводим заголовки 1кл, 2кл, 3кл, 4кл.
2. В ячейках F4, G4 ввести заголовки нижняя, верхняя.
3. В ячейке F5 ввести минимальное число, т.е. содержимое ячейки B2. Для этого в ячейку F5 ввести формулу «=B2». После нажатия Enter, в ячейке F5 появится число 66.
4. В ячейку G5 вводим формулу «=F5+\$E\$2». Для этого необходимо в ячейке G5 ввести с клавиатуры знак =, щелкнуть мышкой по ячейке G5, с клавиатуры ввести знак +, щелкнуть мышкой по ячейке E2. Для того, чтобы зафиксировать ячейку E2 для процесса копирования, необходимо ее адрес окружить знаками \$. После нажатия Enter, в ячейке G5 появится число 68.
5. В ячейку F6 необходимо перенести данные ячейки G5. Для этого в ячейку F6 необходимо ввести формулу «=G5». После нажатия Enter, в ячейке F6 появится число 68.
6. В ячейки F7, F8 копируем содержимое ячейки F6. Для этого необходимо выделить ячейку F6, подвести указатель мыши к правому нижнему углу ячейки F6, нажать правую кнопку мыши и, не отпуская ее, протянуть до ячейки F8.
7. В ячейки G6-G8 аналогичным способом скопировать ячейку G5. Выполнив эти действия получим:

	E	F	G	H
		нижняя	верхняя	
1кл		66	68	
2кл		68	70	
3кл		70	72	
4кл		72	74	

Расчет логических функций.

Для подсчета частоты попадания в каждый класс необходимо ввести логическую функцию, которая проверяла бы, попадает данное число в данный класс или нет. В качестве такой функции можно использовать логическую функцию ЕСЛИ.

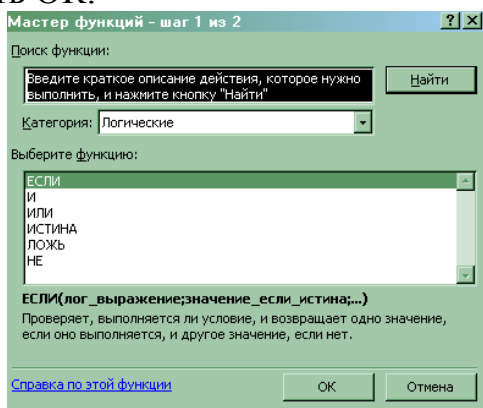
Структура данной функции

ЕСЛИ(лог_выражение; Значение_если_истина;...)

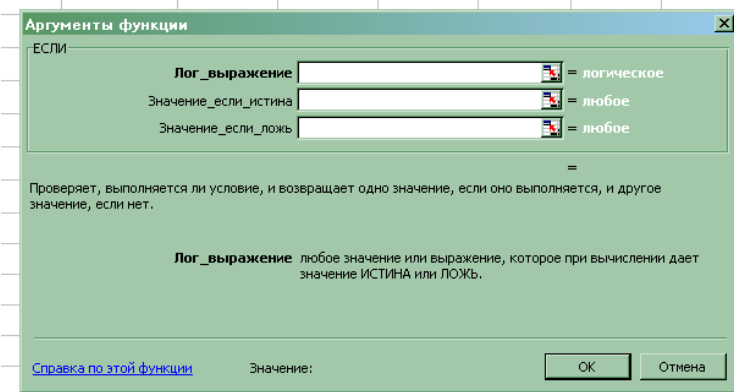
Данная функция проверяет, выполняется ли условие, и возвращает одно значение, если оно выполняется, и другое значение, если нет.

1. В ячейку H1 вводим заголовок 1 класс, I1- 2 класс, J1- 3 класс, K1- 4 класс.

2. В ячейку H2 вводим логическую функцию ЕСЛИ. Для этого в ячейке H2 вводим с клавиатуры знак =, в пункте меню *Вставка* выбрать *Функция*, в категории выбрать *Логические* из предлагаемых функций выбрать ЕСЛИ. нажать ОК.



3. В окне *Аргументы функции* в пункт *Лог_выражение* ввести «B2<\$G\$5», в пункт *Значение_если_истина* 1, в пункт *Значение_если_ложь* 0.



После чего нажать ОК. При этом происходит проверка: если число вариационного ряда меньше, чем последнее число первого класса, то в ячейке Н2 появится число 1, иначе число 0.

4. Скопировать ячейку Н2 в ячейки Н3-Н31. Для этого выделить ячейку Н2, подвести указатель мыши к правому нижнему углу ячейки Н2, протянуть указатель мыши до ячейки Н31.

5. В столбце I рассчитываем частоту попадания во второй класс. Для этого в ячейку I2 вводим формулу: «=ЕСЛИ(Н2=1;0;ЕСЛИ(В2<\$G\$6;1;0))». Первое условие для чисел, которые уже попали в первый класс, выставляет значение 0. Логическая функция ЕСЛИ(В2<\$G\$6;1;0) проверяет условие: если число вариационного ряда меньше, чем последнее число второго класса, то в ячейке I2 появится число 1, иначе число 0.

6. Скопировать ячейку I2 в ячейки I3-I31. Для этого выделить ячейку I2, подвести указатель мыши к правому нижнему углу ячейки I2, протянуть указатель мыши до ячейки I31.

7. В столбце J рассчитываем частоту попадания в третий класс. Для этого в ячейку J2 вводим формулу: «=ЕСЛИ(Н2+I2=1;0;ЕСЛИ(В2<\$G\$7;1;0))». Первое условие для чисел, которые уже попали в первый и второй класс, выставляет значение 0. Логическая функция ЕСЛИ(В2<\$G\$7;1;0) проверяет условие: если число вариационного ряда меньше, чем последнее число третьего класса, то в ячейке J2 появится число 1, иначе число 0.

8. Скопировать ячейку J2 в ячейки J3-J31. Для этого выделить ячейку J2, подвести указатель мыши к правому нижнему углу ячейки J2, протянуть указатель мыши до ячейки J31.

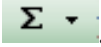
9. В столбце K рассчитываем частоту попадания в четвертый класс. Для этого в ячейку K2 вводим формулу: «=ЕСЛИ(Н2+I2+J2=1;0;ЕСЛИ(В2<=\$G\$8;1;0))». Первое условие для чисел, которые уже попали в первый, второй или третий класс, выставляет значение 0. Логическая функция ЕСЛИ(В2<=\$G\$8;1;0) проверяет условие: если число вариационного ряда меньше или равно, чем последнее число четвертого класса, то в ячейке K2 появится число 1, иначе число 0.

10. Скопировать ячейку K2 в ячейки K3-K31. Для этого выделить ячейку K2, подвести указатель мыши к правому нижнему углу ячейки K2, протянуть указатель мыши до ячейки K31.

Подсчет частоты попадания в каждый класс

Для нахождения частоты попадания в каждый класс, необходимо просуммировать значения столбцов Н, I, J, K.

1. Выделить ячейки Н2-Н31.

2. На панели инструментов выбрать кнопку *Автосумма* . После чего в ячейке Н32 появится число 4.

3. Выделить ячейки I2-I31.

4. На панели инструментов выбрать кнопку *Автосумма*. После чего в ячейке I32 появится число 5.

5. Аналогичные действия для ячеек J и K приведут к тому, что в ячейках I32 и J32 появятся числа 13 и 8.

Расчет функции плотности вероятности


1. В ячейку A34 вводим заголовок *Частота*.
2. Для удобства расчетов перенесем данные из ячеек H32-K32 в ячейки A35-A38. Для этого в ячейку A35 вводим формулу «=H32», в ячейку A36 – «=I32», в ячейку A37 – «=J32», в ячейку A38 – «=K32».
3. В ячейку B34 вводим заголовок: *Функция*.
4. В ячейке B35 вводим формулу: «=A35/(\$F\$2*\$E\$2)». Для этого необходимо в ячейке B35 с клавиатуры ввести знак =, щелкнуть мышкой по ячейке A35, с клавиатуры ввести знак /, щелкнуть мышкой по ячейке F2, с клавиатуры ввести знак *, щелкнуть мышкой по ячейке E2, с клавиатуры ввести знак), добавить знаки \$. После нажатия Enter, в ячейке B35 появится число 0.666667.
5. Скопировать ячейку B35 в ячейки B36-B38. Для этого необходимо выделить ячейку B35, подвести указатель мыши к правому нижнему углу ячейки B35, протянуть указатель мыши до ячейки B38. Выделяем ячейки B35-B38. С помощью пункта меню *Формат/ячейки* выбрать числовой формат *Числовой*, число десятичных знаков 2. При этом появится таблица

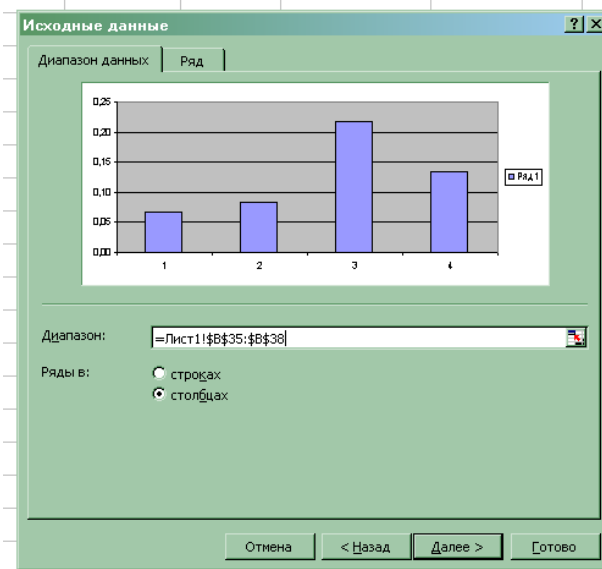
33			
34	частота	функция	
35	4	0,07	
36	5	0,08	
37	13	0,22	
38	8	0,13	
39			

Построение гистограммы

1. Для более наглядного построения гистограммы, перенесем границы интервалов в такой форме:

39				
40				
41	частота	функция	границы	
42	4	0,07	66	68
43	5	0,08		70
44	13	0,22		72
45	8	0,13		74
46				
47				

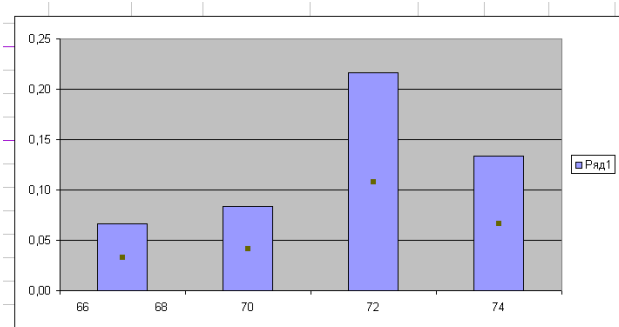
1. На панели инструментов выбрать кнопку *Мастер диаграмм* .
- В появившемся окне выбрать тип *Гистограмма*, затем нажать *Далее*.
2. Появится окно *Мастер диаграмм*.



В поле *Диапазон* удалить появившуюся запись, щелкнуть по ячейке B42, протянуть мышку до ячейки B45.

3. Зайти во вкладку *Ряд*. Щелкнуть мышкой в поле *Подписи оси X*, затем щелкнуть мышкой по ячейке C42, протянуть указатель мыши до ячейки C45. Затем нажать *Далее* и *Готово*.

4. Щелкнуть мышкой по одному из прямоугольников гистограммы так, чтобы в центре каждого из прямоугольников появилась точка. Щелкнуть правой кнопкой мыши и в появившемся окне выбрать пункт *Формат рядов данных*.



5. В появившемся окне выбрать вкладку *Параметры*. Затем уменьшить ширину зазора до 0, нажать *Ок*. При этом получим гистограмму

Тема: Ряды динамики.

Динамический ряд – это совокупность однородных статистических величин, показывающих изменения какого-либо явления на протяжении определенного промежутка времени.

Для динамических рядов можно рассчитать следующие показатели:

1. **Абсолютный прирост**- разница между значением данного года и предыдущим.

$$\Delta y = y_i - y_{i-1}$$

2. **Коэффициент роста** – отношение данного уровня к базисному. В качестве базисного уровня принимается уровень первого года.

$$k = \frac{y_i}{y_0}$$

3. Темп роста- коэффициент роста, выраженный в процентах.

$$k' = \frac{y_i}{y_0} \cdot 100\%$$

4. Темп прироста- величина, показывающая на сколько процентов данный уровень больше или меньше базисного.

$$k'' = \frac{y_i}{y_0} \cdot 100\% - 100\%$$

Изучение изменения явлений во времени является одной из важнейших задач статистики. Решается эта задача при помощи составления и анализа так называемых рядов динамики или временных рядов.

Тренд – основная тенденция изменения уровней.

Анализ рядов динамики начинается с выявления формы тренда. Чтобы сделать правильные выводы о закономерностях развития того или иного показателя, надо суметь отделить главную тенденцию изменения от колебаний, вызванных влиянием случайных кратковременных причин.

Статистика дает возможность количественно охарактеризовать влияние, оказываемое этими двумя группами факторов, на изменение изучаемых явлений, то есть определить, в какой мере эти изменения вызваны длительно действующими и в какой мере временно действующими факторами.

В целях выравнивания используются следующие методы:

1. Графический метод
2. Метод удлинения периодов
3. Метод скользящей средней
4. Метод наименьших квадратов

Рассмотрим данные методы на задаче

Задача: даны значения детской рождаемости по годам.

Год	1986	1987	1988	1989	1990	1991	1992	1993	1994
Y	41	35	32	32	31	29	27	26	26

Расчет основных показателей

Вводим в таблицу данные задачи:

1. Установить табличный курсор на первую ячейку создаваемой таблицы (например на **B1**).
2. Ввести заголовок – «Год» и нажать клавишу **Enter**.
3. Установить табличный курсор на ячейку **C1** таблицы, ввести заголовок – «Yi» и нажать клавишу **Enter**.
4. Аналогично вводится название остальных ячеек.
5. В ячейке «B2 – B10» вводятся года.
6. В ячейке «C2 – C10» вводятся значения детской рождаемости.

The screenshot shows a Microsoft Excel window titled "Microsoft Excel - Временные ряды". The menu bar includes "Файл", "Правка", "Вид", and "Вставка". The toolbar contains various icons for file operations and editing. The active cell is A16. The table below is displayed in the spreadsheet:

	А	В	С
1		Год	У
2		1986	41
3		1987	35
4		1988	32
5		1989	32
6		1990	31
7		1991	29
8		1992	27
9		1993	26
10		1994	26
11			

Расчет абсолютного прироста

1. В ячейку **D1** вводим заголовок ΔY .
2. Для 1986 года абсолютный прирост не рассчитываем, так как в ряду отсутствует значение для 1985 года.
3. В ячейку **D3** вводим формулу: « $=C3-C2$ ». При этом необходимо следить за тем, чтобы язык ввода был английский.
Формулу можно вводить другим способом: с клавиатуры ввести знак $=$, затем щелкнуть мышкой по ячейке **C3**, с клавиатуры ввести знак $-$, щелкнуть мышкой по ячейке **C2**. После чего нажать **Enter**. При этом в ячейке **D3** появится число -6
4. Подвести указатель мыши к правому нижнему углу ячейки **D3**, нажать левую кнопку мыши и не отпуская ее протянуть до ячейки **D10**. При этом абсолютный прирост будет рассчитан для всего ряда.

Расчет коэффициента роста

1. В ячейку **E1** ввести заголовок **K**.
2. В ячейку **E2** ввести формулу « $=C2/SC2$ ». Для этого с клавиатуры ввести знак $=$, затем щелкнуть мышкой по ячейке **C2**, с клавиатуры ввести знак $/$, щелкнуть мышкой по ячейке **C2**. Добавить знак $\$$. после нажатия **Enter**, в ячейке **E2** появится число **1**.
3. Подвести указатель мыши к правому нижнему углу ячейки **E2**, нажать левую кнопку мыши и не отпуская ее протянуть до ячейки **E10**. При этом коэффициент роста будет рассчитан для всего ряда.

Расчет темпа роста

1. В ячейку **F1** ввести заголовок **K'**.
2. В ячейку **F2** ввести формулу « $=E2*100$ ». Для этого с клавиатуры ввести знак $=$, затем щелкнуть мышкой по ячейке **E2**, с клавиатуры ввести $*100$. После нажатия **Enter**, в ячейке **E2** появится число **100**.
3. Подвести указатель мыши к правому нижнему углу ячейки **F2**, нажать левую кнопку мыши и не отпуская ее протянуть до ячейки **F10**. При этом темп роста будет рассчитан для всего ряда.

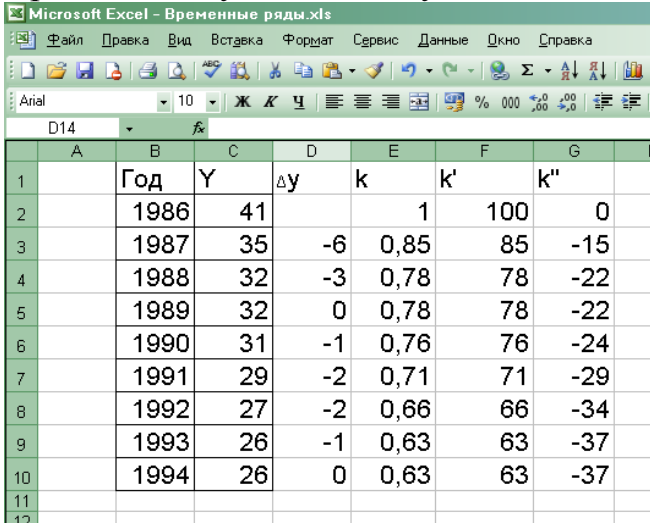
Расчет темпа прироста

1. В ячейку **G1** ввести заголовок **К''**.
2. В ячейку **G2** ввести формулу «**=F2-100**». Для этого с клавиатуры ввести знак **=**, затем щелкнуть мышкой по ячейке **F2**, с клавиатуры ввести **-100**. После нажатия **Enter**, в ячейке **G2** появится число **0**.
3. Подвести указатель мыши к правому нижнему углу ячейки **G2**, нажать левую кнопку мыши и не отпуская ее протянуть до ячейки **G10**. При этом темп прироста будет рассчитан для всего ряда.

Форматирование таблицы.

1. Выделить ячейки с E2 до E10.
2. В пункте меню выбрать *Формат, Ячейки*. Где выбрать *числовой*, число десятичных знаков 2.
3. Выделить ячейки с F2 до G10.
4. В пункте меню выбрать *Формат, Ячейки*. Где выбрать *числовой*, число десятичных знаков 0.

При этом получим таблицу:



	A	B	C	D	E	F	G	I
1		Год	Y	Δy	k	k'	k''	
2		1986	41		1	100	0	
3		1987	35	-6	0,85	85	-15	
4		1988	32	-3	0,78	78	-22	
5		1989	32	0	0,78	78	-22	
6		1990	31	-1	0,76	76	-24	
7		1991	29	-2	0,71	71	-29	
8		1992	27	-2	0,66	66	-34	
9		1993	26	-1	0,63	63	-37	
10		1994	26	0	0,63	63	-37	
11								
12								

Графический метод

Суть данного метода заключается в построении графика по данным и построении приближенной линии тренда.

Выявить тенденцию развития данного ряда, используя графический метод.

Сделать вывод о тенденции развития динамического ряда.

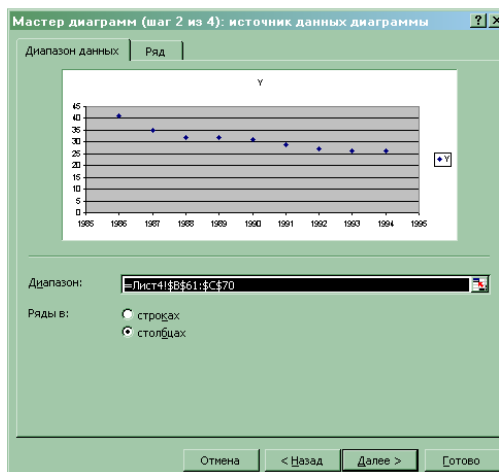
Вводим в таблицу данные задачи:

1. Установить табличный курсор на первую ячейку создаваемой таблицы (например, на **B1**).
2. Ввести заголовок – «Год» и нажать клавишу **Enter**.
3. Установить табличный курсор на ячейку **C1** таблицы, ввести заголовок – «Y_i» и нажать клавишу **Enter**.
4. Аналогично вводится название остальных ячеек.
5. В ячейке «**B2 – B10**» вводятся года.
6. В ячейке «**C2 – C10**» вводятся значения детской рождаемости.

	А	В	С
1		Год	У
2		1986	41
3		1987	35
4		1988	32
5		1989	32
6		1990	31
7		1991	29
8		1992	27
9		1993	26
10		1994	26
11			

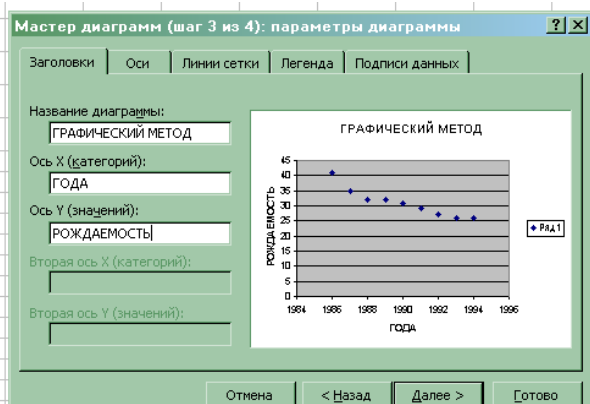
Строим график

1. На панели инструментов выбираем **Мастер диаграмм**. В появившемся окне выбираем тип **Точечная**. Нажимаем **Далее**.



2. В пункте **Диапазон** удалить появившуюся запись, щелкнуть мышкой по ячейке **B2** и не отпуская ее протянуть до ячейки **C10**. Нажать **Далее**.

3. В появившемся окне, ввести название диаграммы *Графический метод*, Ось X *Года*, Ось Y *Рождаемость*.



Затем нажать **Далее** и **Готово**.

Редактируем диаграмму.

	A	B	C	D	E
1		Год	Y	Y'	
2		1986	41		
3		1987	35	36	
4		1988	32		
5		1989	32		
6		1990	31	30,7	
7		1991	29		
8		1992	27		
9		1993	26	26,3	
10		1994	26		
11					
12					

Необходимо изменить масштаб осей. Необходимо щелкнуть ось Y правой кнопкой появляется диалоговое окно, в котором выбираем функцию Формат оси. Выбираем вкладку шкала и в поле шкала по оси Y ввести минимальное значение «20», максимальное значение «45» и цену одного деления «5».

Метод укрупнения периодов.

Суть данного метода заключается в нахождении средних значений по нескольким (например, по трем годам) и построении графика по полученным данным, по которым также можно построить линию тренда.

Данный метод рассмотрим на том же примере.

Рассчитаем среднее значение по трем годам

1. В ячейку D1 вводим заголовок Y'.

2. В ячейку D3 вводим формулу « $=(C2+C3+C4)/3$ ». Для этого с клавиатуры ввести знак =, затем щелкнуть мышкой по ячейке C2, с клавиатуры ввести +, щелкнуть мышкой по ячейке C3, с клавиатуры ввести (+, щелкнуть мышкой по ячейке C4, с клавиатуры ввести $/3$. После нажатия Enter, в ячейке D3 появится число 36.

3. Скопировать содержимое ячейки D3 в ячейки D6 и D9. Для этого выделить ячейку D3, в пункте меню *Правка* выбрать пункт *Копировать*, встать на ячейку D6, в пункте меню *Правка* выбрать пункт *Вставить*. Аналогичные действия произвести в ячейке D9.

При этом получим Таблицу:

Строим график

1. График по данным метода укрупнения периодов можно добавить к первому графику. Для этого в пункте меню *Диаграмма* выбрать пункт *Добавить данные*. В появившемся окне, в качестве диапазона поместить данные. Для этого щелкнуть мышкой по ячейке D2, и не отпуская ее протянуть до ячейки D10. после чего нажать *Ок*. дабавленный график будет выделен другим цветом.

Метод скользящей средней

Суть данного метода заключается в нахождении средних значений по нескольким (например, по трем годам) и построении графика по полученным данным, по которым также можно построить линию тренда.

Данный метод рассмотрим на том же примере.

Рассчитаем среднее значение по трем годам

1. В ячейку E1 вводим заголовок Y''.

2. В ячейку E3 вводим формулу « $=(C2+C3+C4)/3$ ». Для этого с клавиатуры ввести знак =, затем щелкнуть мышкой по ячейке C2, с клавиатуры ввести +, щелкнуть мышкой по ячейке C3, с клавиатуры ввести (+, щелкнуть мышкой по ячейке C4, с клавиатуры ввести $) / 3$. После нажатия Enter, в ячейке E3 появится число 36.

3. Скопировать содержимое ячейки E3 в ячейки от E4 до E9. Для этого необходимо подвести указатель мыши к правому нижнему углу ячейки E3, нажать левую кнопку мыши и, не отпуская ее, протянуть до ячейки E9.

При этом получим Таблицу:

	B	C	D	E	F
Год	Y	Y'	Y''	Y'''	
1986	41				
1987	35	36,00	36,00		
1988	32		33,00		
1989	32		31,67		
1990	31	30,67	30,67		
1991	29		29,00		
1992	27		27,33		
1993	26	26,33	26,33		
1994	26				

Строим график

График по данным метода скользящей средней можно добавить к уже построенным графикам. Для этого в пункте меню *Диаграмма* выбрать пункт *Добавить данные*. В появившемся окне, в качестве диапазона поместить данные. Для этого щелкнуть мышкой по ячейке E2, и не отпуская ее протянуть до ячейки E10. после чего нажать *Ок*. Добавленный график будет выделен другим цветом.

Метод наименьших квадратов.

Более совершенным способом обработки динамических рядов с целью установления тенденции развития является выравнивание по аналитическим формулам. При этом способе на основе фактических данных ряда подбирается наиболее подходящая математическая формула (аппроксимирующая функция), по которой рассчитывают выровненные значения.

Этот метод преследует ту же цель, что и описанные выше методы – устранения влияния временно действующих причин и выявить тенденцию развития, вызванные только действием длительно действующих факторов. Тенденцию

развития лучше всего можно выразить линией, наиболее близкой к фактическим данным, это достигается методом наименьших квадратов. Суть метода в том, что: сумма квадратов отклонений точек от линии должна быть минимальной, т.е. $(y-y_i)^2 = \min$.

Вводим в таблицу данные задачи:

Откроем новый лист.

1. Установить табличный курсор на первую ячейку создаваемой таблицы (например, на **B1**).
2. Ввести заголовок – «Год» и нажать клавишу **Enter**.
3. Установить табличный курсор на ячейку **C1** таблицы, ввести заголовок – «Y_i» и нажать клавишу **Enter**.
4. Аналогично вводится название остальных ячеек.
5. В ячейке «B2 – B10» вводятся годы.
6. В ячейке «C2 – C10» вводятся значения детской рождаемости.

	А	В	С	Д
1		Год	Y	
2		1986	41	
3		1987	35	
4		1988	32	
5		1989	32	
6		1990	31	
7		1991	29	
8		1992	27	
9		1993	26	
10		1994	26	
11	n=	9	279	

Вычисление сумм:

1. Установит табличный курсор на ячейку «**B11**» таблицы.
2. Указателем мыши нажать кнопку «автосумма» на панели инструментов.
3. Аналогично найти сумму детской рождаемости (Y_i).

Ввод формул:

Ввод формулы должен начинаться со знака «=». Все записи должны осуществляться латинскими буквами.

Вводим отсчет времени:

1. Установить табличный курсор на ячейку «**D6**» и установить значение «0».
2. В ячейках «**D5 – D2**» вводим значения от -1 до -4.
3. В ячейках «**D7 – D10**» вводим значения от 1 до 4.
4. Установить табличный курсор на ячейку «**D11**» и нажать «автосумма» на панели инструментов. $\sum t_i = 0$.

	A	B	C	D	E
1		Год	Y	Ti	
2		1986	41	-4	
3		1987	35	-3	
4		1988	32	-2	
5		1989	32	-1	
6		1990	31	0	
7		1991	29	1	
8		1992	27	2	
9		1993	26	3	
10		1994	26	4	
11	n=	9	279	0	
12					

Находим произведение показателей детской рождаемости на соответствующие времени ($Y \cdot t$):

1. Установить табличный курсор на ячейку «E2» и ввести с клавиатуры следующую формулу «=C2*D2», нажать клавишу **Enter**. В ячейке «E2» появится число «-164».
2. Далее скопируем содержимое ячейки «E2» во все остальные ячейки «E3 –E10».
3. Установить табличный курсор на ячейку «E11» и нажать кнопку «автосумма».

	A	B	C	D	E	F
1		Год	Y	Ti	Yi * Ti	
2		1986	41	-4	-164	
3		1987	35	-3	-105	
4		1988	32	-2	-64	
5		1989	32	-1	-32	
6		1990	31	0	0	
7		1991	29	1	29	
8		1992	27	2	54	
9		1993	26	3	78	
10		1994	26	4	104	
11	n=	9	279	0	-100	
12						

Возводим каждое значение t в квадрат:

1. Установить табличный курсор в ячейку «F2» и ввести с клавиатуры следующую формулу «=D2*D2», нажать клавишу **Enter**.
2. Далее скопируем содержимое ячейки «F2» во все остальные ячейки «F3 – F10».
3. Установить табличный курсор на ячейку «F11» и нажать кнопку «автосумма».

	A	B	C	D	E	F	G
1		Год	Y	Ti	Yi*Ti	Ti^2	
2		1986	41	-4	-164	16	
3		1987	35	-3	-105	9	
4		1988	32	-2	-64	4	
5		1989	32	-1	-32	1	
6		1990	31	0	0	0	
7		1991	29	1	29	1	
8		1992	27	2	54	4	
9		1993	26	3	78	9	
10		1994	26	4	104	16	
11	n=	9	279	0	-100	60	
12							

По формулам находят коэффициенты **a** и **b**:

$$b = \frac{\sum Y_i}{n}$$

Установить табличный курсор в ячейку «B13» и ввести следующую формулу «=C11/B11» и нажать клавишу **Enter**. Появится число «31».

$$a = \frac{\sum Y_i * t}{\sum t^2}$$

Установить табличный курсор в ячейку «B14» и ввести следующую формулу «=E11/F11» и нажать клавишу **Enter**. Появится число «-1,7».

	A	B	C	D	E	F	G
1		Год	Y	Ti	Yi*Ti	Ti^2	
2		1986	41	-4	-164	16	
3		1987	35	-3	-105	9	
4		1988	32	-2	-64	4	
5		1989	32	-1	-32	1	
6		1990	31	0	0	0	
7		1991	29	1	29	1	
8		1992	27	2	54	4	
9		1993	26	3	78	9	
10		1994	26	4	104	16	
11	n=	9	279	0	-100	60	
12							
13	b=	31					
14	a=	-1,67					
15							
16							

Получаем уравнение регрессии: $y=1.7t+31$.

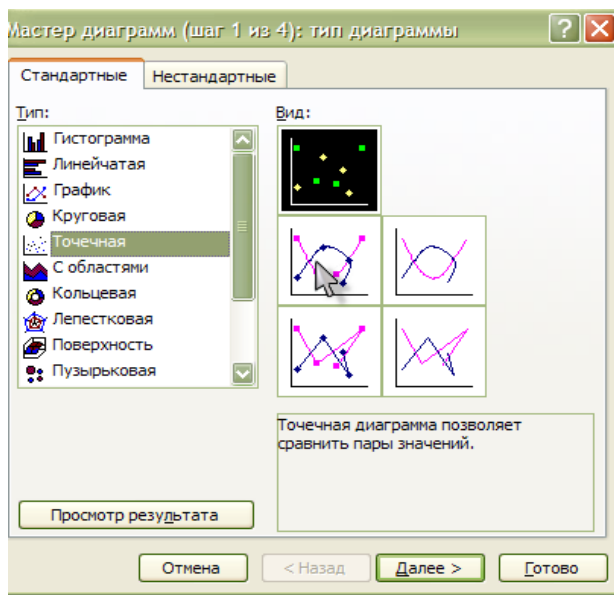
Строим график:

1. Установить табличный курсор на ячейку «G2» и ввести следующую формулу « = \$B\$14*\$D2+\$B\$13». Появится число «37,7».
2. Скопируем содержимое ячейки «G2» в ячейки «G3 – G10».

	A	B	C	D	E	F	G	H
1		Год	Y	Ti	Yi*Ti	Ti^2	Y'	
2		1986	41	-4	-164	16	38	
3		1987	35	-3	-105	9	36	
4		1988	32	-2	-64	4	34	
5		1989	32	-1	-32	1	33	
6		1990	31	0	0	0	31	
7		1991	29	1	29	1	29	
8		1992	27	2	54	4	28	
9		1993	26	3	78	9	26	
10		1994	26	4	104	16	24	
11	n=	9	279	0	-100	60		
12								
13	b=	31						
14	a=	-1,67						
15								

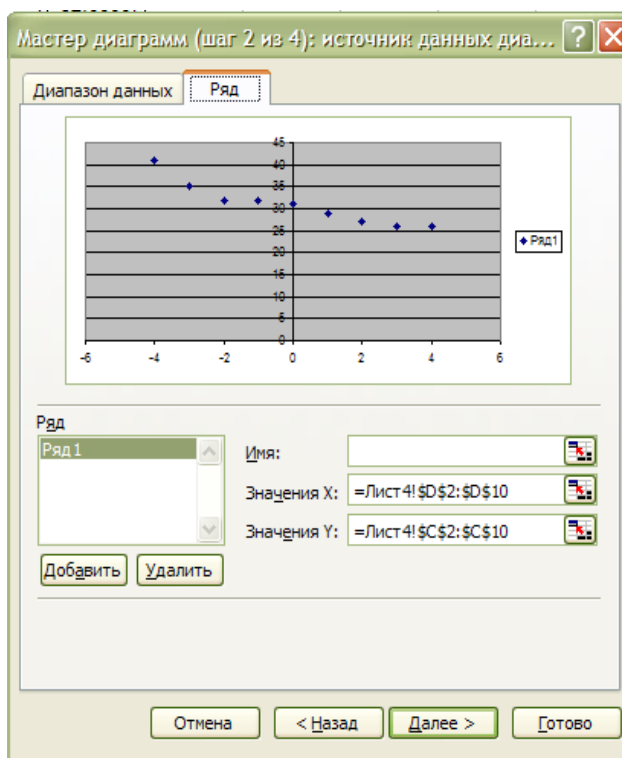
Выбор типа диаграмм.

На панели инструментов нажать кнопку *Мастер диаграмм*. В появившемся окне *Мастер диаграмм шаг 1 из 4* указать точечную диаграмму и нажимаем кнопку *далее*.



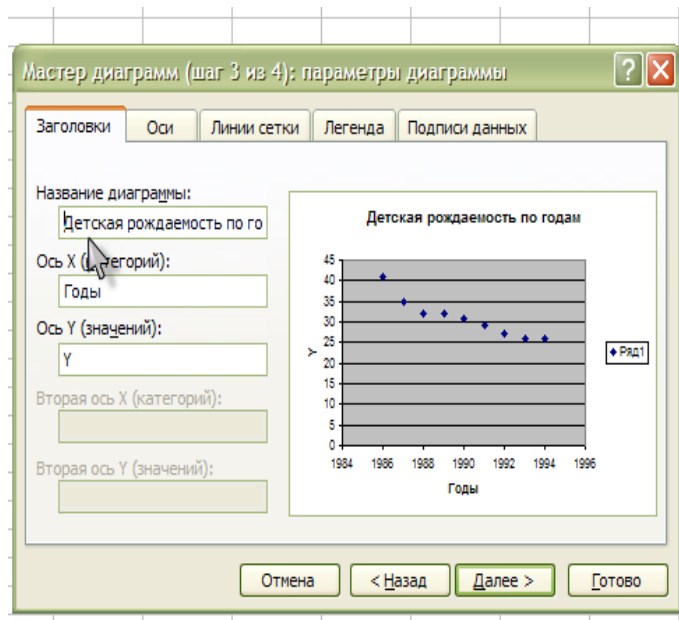
Указание диапазонов.

В окне *Мастер диаграмм шаг 2 из 4* выбрать вкладку *диапазон* и указать интервал данных. В рабочем поле появится запись «=Лист 4!\$D\$2:\$D\$10».



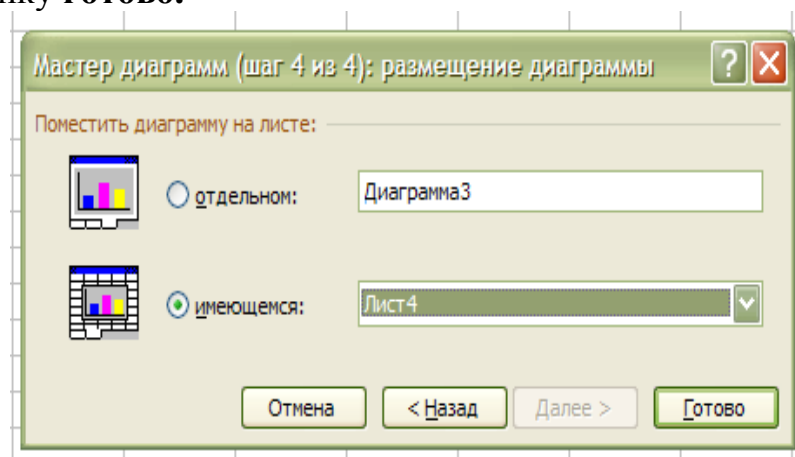
Ввод подписей по осям и введение заголовка.

В появившемся окне *Мастер диаграмм шаг 3 из 4* выбрать вкладку *заголовок*. В поле *название диаграммы* ввести название: «Детская рождаемость по годам». Затем ввести в рабочие поля *ось X* и *ось Y* соответствующие названия: «Годы», «Y».



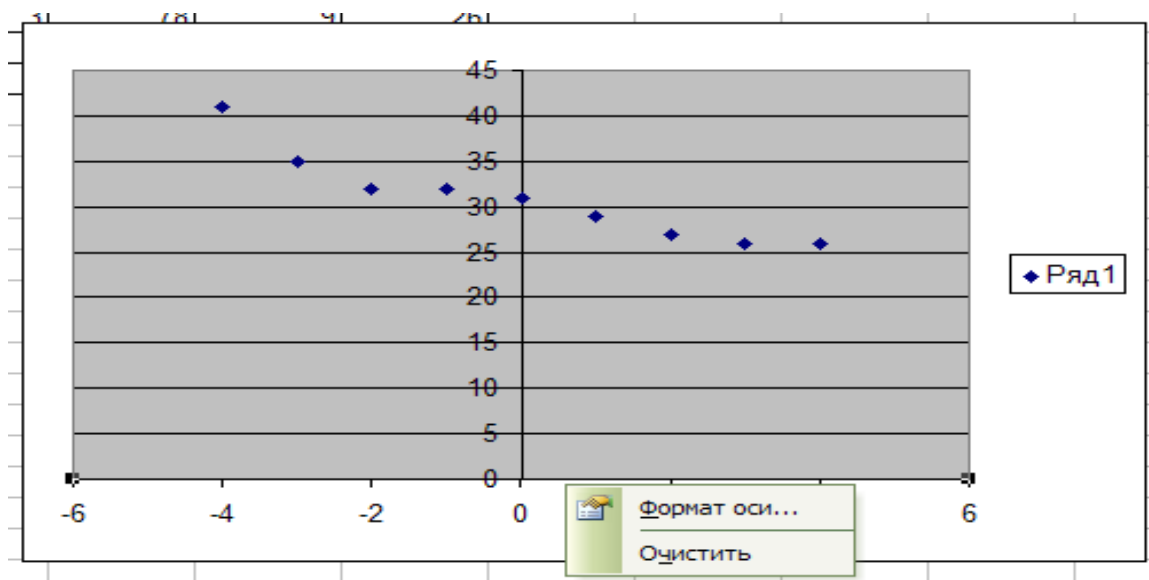
Выбор места размещения.

В появившемся окне *Мастер диаграмм шаг 4 из 4* необходимо переключатель *Поместить диаграмму на листе*: установить в положение *имеющемся* и нажимаем кнопку **ГОТОВО**.

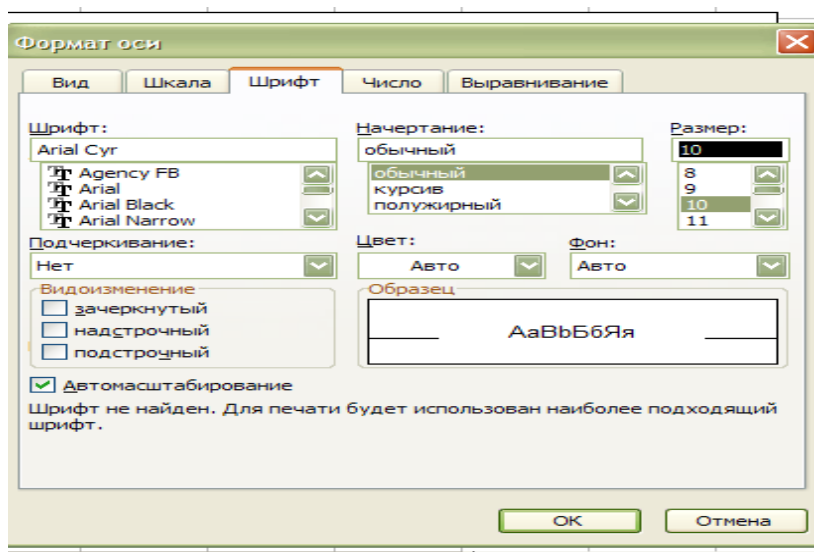


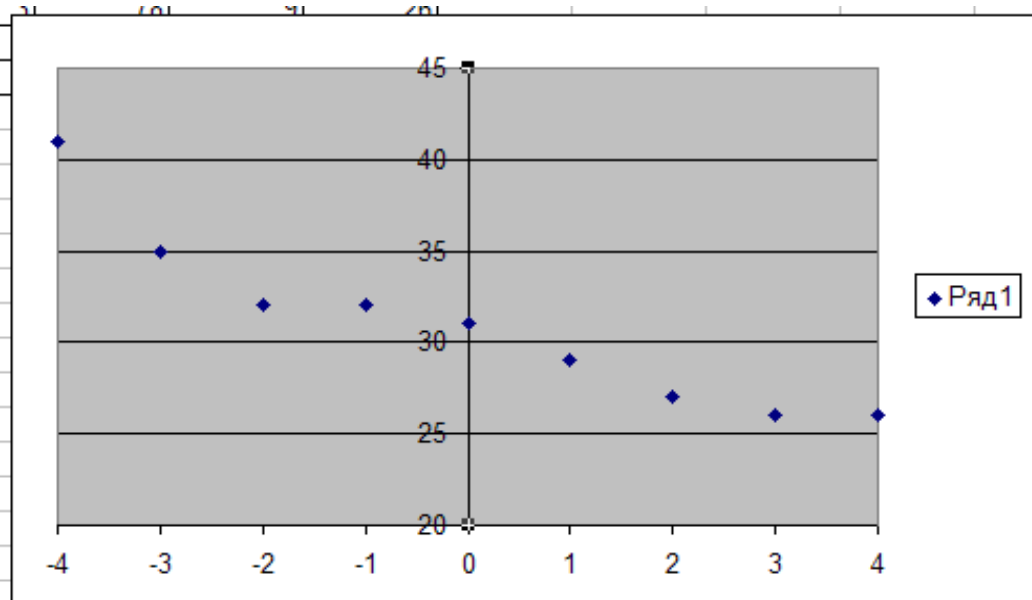
Редактирование диаграммы.

Необходимо изменить масштаб осей. Выделяем *ось X* и после щелчка правой кнопкой появляется диалоговое окно, в котором выбираем функцию *Формат оси*. Выбираем вкладку *шкала* и в поле *шкала по оси X* ввести минимальное значение «20», максимальное значение «45» и цену одного деления «5».

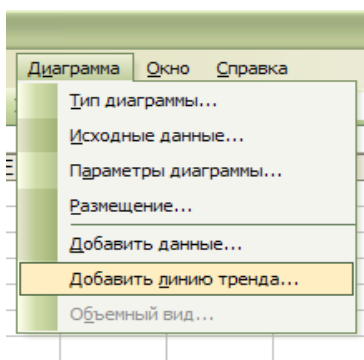


Выбираем вкладку шрифт и в поле размер выбираем равный «10» и нажимаем кнопку готово.

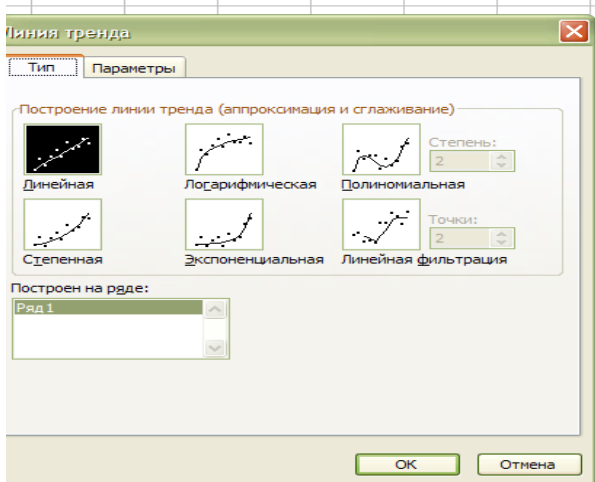




В строке меню выбираем диаграммы, появляется диалоговое меню, в котором выбираем функцию добавить линию тренда

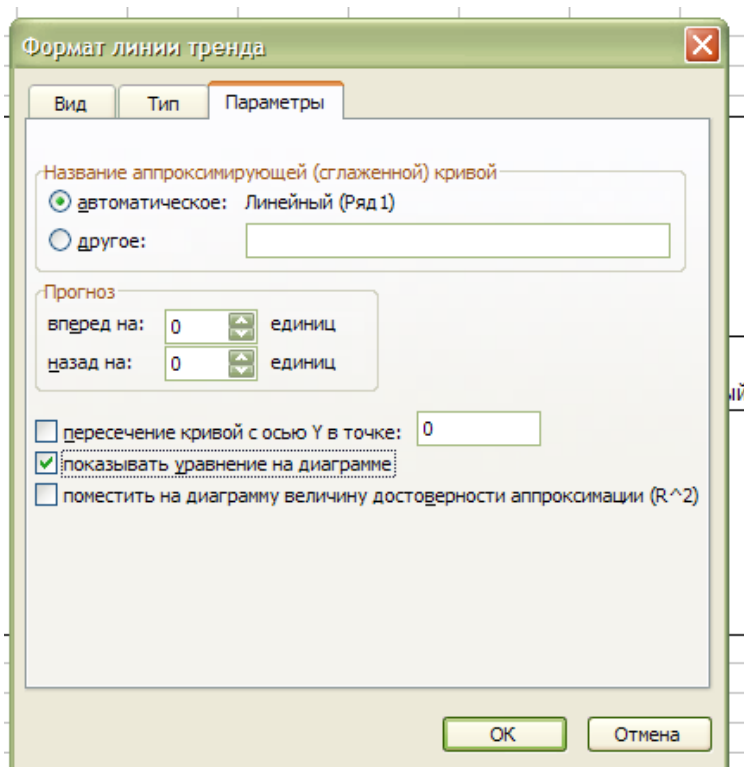
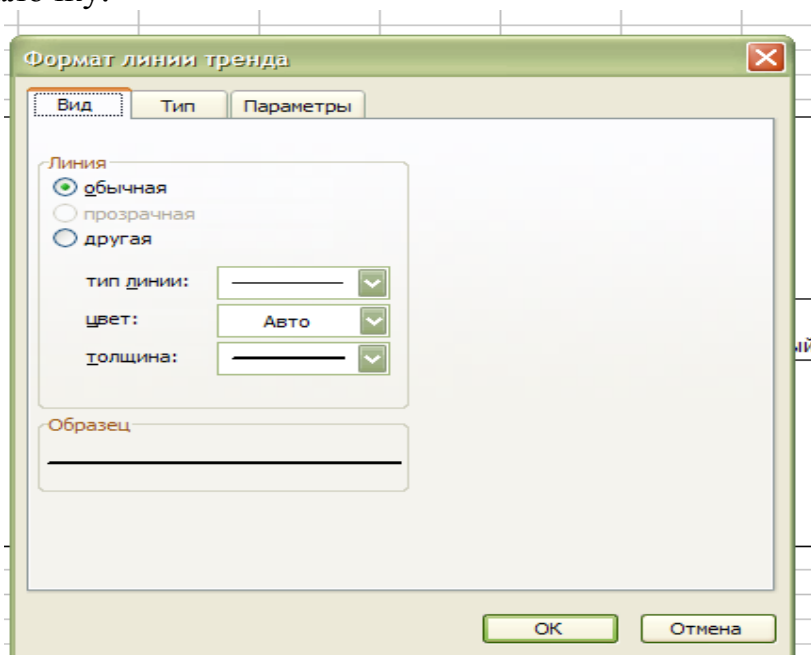


Выбираем тип линии линейная, нажимаем кнопку Ok.

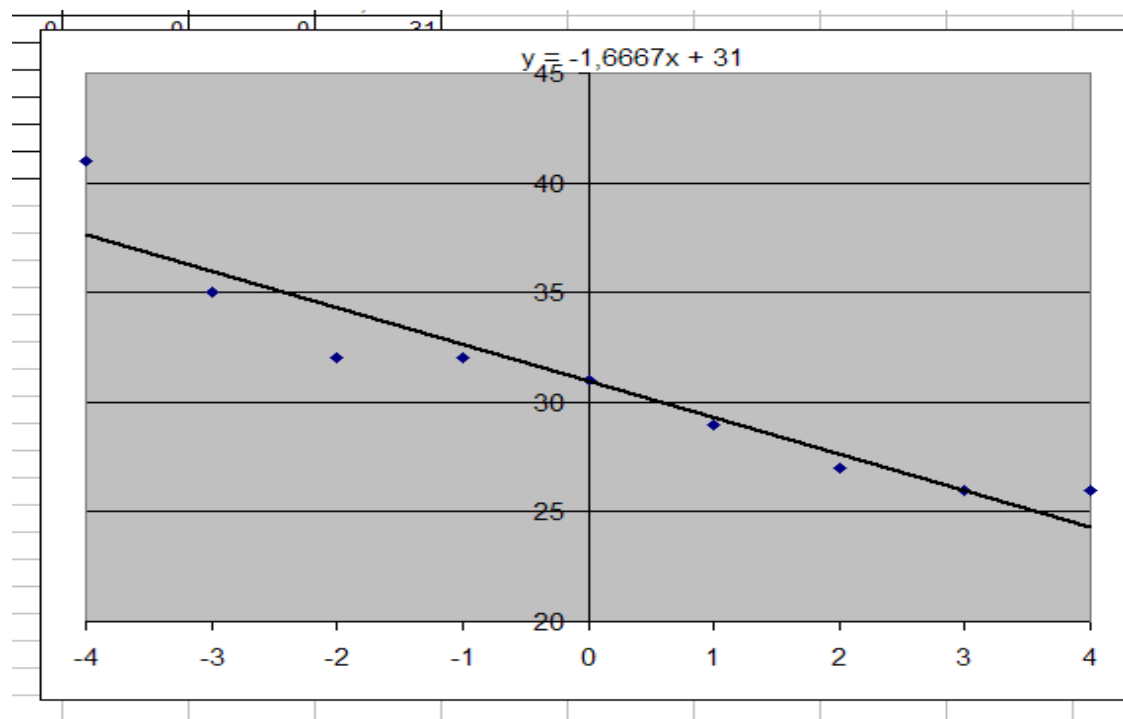


Получаем график.

Выделяем линию и выбираем опцию *Формат линии тренда*, в котором нажимаем кнопку *Параметры*. В поле *Показывать линию тренда* ставим галочку.



Появляется диаграмма с уравнением: $Y=1,7X+31$.



Вывод: данный метод является наиболее точным, так как уравнение линии найдено методом наименьших квадратов, то есть теоретически рассчитано.

Глоссарий.

1. **Функциональные** - каждому значению одной переменной величины соответствует одно вполне определенное значение другой переменной.
2. **Корреляционные** - (статистические) - численному значению одной переменной соответствует много значений другой переменной.
3. **Коэффициент корреляции**-это число показывающее степень зависимости одной переменной величины от другой.
4. **Линейная регрессия**-показывает как количественно меняется одна величина при изменении другой на единицу.

Генеральная совокупность – совокупность однородных статистических данных, отличающихся друг от друга, но при этом имеющих сходные свойства.

Выборка – часть генеральной совокупности. На выборке измеряют переменные.

Переменная – это то, что можно измерять, контролировать или чем можно манипулировать в исследованиях. Иными словами, переменная – это то, что варьируется, изменяется, а не остается постоянным. Можно привести примеры разных переменных. И, понятно, что все они будут отличаться своими характеристиками.

Минимум и максимум – это минимальное и максимальное значения переменной.

Среднее – это сумма всех значений переменной, деленное на количество переменных.

$$\bar{x} = \frac{\sum x_i}{n}$$

Дисперсия рассчитывается по формуле:

$$\sigma^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}$$

Стандартное отклонение равно квадратному корню из дисперсии. Формально имеем:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}}$$

Медиана- это величина, относительно которой ряд распределения делится на две равные части: в обе стороны от медианы располагается одинаковое число вариант.

Мода- это величина, наиболее часто встречающаяся в данной совокупности. Класс с наибольшей частотой называется модальным.

Коэффициент асимметрии характеризует симметричность графика распределения относительно среднего значения

Эксцесс, в свою очередь, характеризует узковершинность или плосковершинность распределения.

Доверительный интервал- это интервал, в который попадает среднее значение генеральной совокупности с той или иной вероятностью.

Динамический ряд – это совокупность однородных статистических величин, показывающих изменения какого-либо явления на протяжении определенного промежутка времени.

Абсолютный прирост- разница между значением данного года и предыдущим.

$$\Delta y = y_i - y_{i-1}$$

Коэффициент роста – отношение данного уровня к базисному. В качестве базисного уровня принимается уровень первого года.

$$k = \frac{y_i}{y_0}$$

Темп роста- коэффициент роста, выраженный в процентах.

$$k' = \frac{y_i}{y_0} \cdot 100\%$$

Темп прироста- величина, показывающая на сколько процентов данный уровень больше или меньше базисного.

$$k'' = \frac{y_i}{y_0} \cdot 100\% - 100\%$$

Тренд –основная тенденция изменения уровней.

Литература:

1. В.Я. Гольдман Медицинская информатика, Санкт-Петербург, 2001г..
2. Г.Ф. Лакин «Биометрия», Москва, 1990г.