Начальник отдела кадров крупной компании, выпускающей сложные промыш­ленные инструменты, пытается применить регрессионные модели для правильно­го выбора менеджеров по продажам, возглавляющих торговые представительства в 45 регионах. Многие из них имеют научные степени в области электротехники, поэтому с учетом технического профиля компании некоторые руководители счи­тают, что на должность менеджера по продажам следует принимать только спе­циалистов с учеными степенями. Для этого во время интервью кандидатам пред­лагают пройти два теста. Поскольку проведение этих экзаменов требует дополни­тельных затрат, появилось предложение отменить один из них или оба. Для нача­ла отдел кадров собрал информацию о каждом из 45 менеджеров по продажам, возглавляющих торговые представительства в данный момент. В эту информацию входит стаж работы, образование, а также оценки, полученные на обоих экзаме­нах. В качестве зависимой переменной выбран "индекс продаж", представляющий собой отношение фактического объема продаж в регионе к плановому объему про­даж, который каждый год устанавливается руководством компании на основании предыдущих результатов и емкости рынка в каждом регионе. В качестве независимых пере­менных выбраны следующие величины.

Продажи — отношение фактического объема продаж к плановому. Считается, что плановый объем продаж устанавливается на основе "реалистичных предпо­ложений".

Тест — оценки, полученные менеджерами при сдаче второго теста (SCIIT). Чем выше оценка, тем выше заинтере­сованность менеджера в увеличении объема продаж.

1. Разработайте наиболее подходящую модель для прогнозирования объема продаж.

|  |  |
| --- | --- |
| **Объем продаж** | **Тест SC** |
| 96 | 42 |
| 90 | 46 |
| 113 | 55 |
| 98 | 47 |
| 76 | 45 |
| 117 | 56 |
| 118 | 63 |
| 101 | 50 |
| 95 | 54 |
| 94 | 41 |
| 119 | 62 |
| 120 | 79 |
| 115 | 52 |
| 131 | 62 |
| 99 | 45 |
| 102 | 59 |
| 66 | 40 |
| 129 | 64 |
| 100 | 39 |
| 111 | 52 |
| 128 | 74 |
| 104 | 45 |
| 133 | 61 |
| 125 | 66 |
| 99 | 46 |
| 90 | 46 |
| 122 | 63 |
| 62 | 54 |
| 98 | 46 |
| 100 | 49 |
| 123 | 62 |
| 120 | 57 |
| 83 | 41 |
| 71 | 34 |
| 102 | 54 |
| 89 | 48 |
| 75 | 53 |
| 106 | 54 |
| 80 | 36 |
| 99 | 49 |
| 104 | 56 |
| 105 | 55 |
| 87 | 43 |
| 105 | 50 |
| 90 | 41 |

Указание к работе:

**Указания (аналитический способ)**

**1. Построить линейное уравнение парной регрессии** у **от** х (исходные данные приведены в табл.)

Даны исходные данные: у - зависимая величина, х - независимая.

* Чтобы выполнить п.1 создаем таблицу ниже.

В этой таблице заголовки таблиц следующие: *х*, *х*2, *у*, *ху*, у2,

(*у* - ȳ)2 - разность между значением *у* и средним значением ȳ, например, в первой строке (124 - 141,5)2 = - 17,5832 = 309,17. Протянуть до конца столбца.[[1]](#footnote-1) Не забываем про абсолютные адреса! То же самое выполняем для переменной *х*.

- вычисляемое значение уравнения регрессии. Коэффициенты *a* и *b* уравнения у = а + bх рассчитать по формулам и подставить в уравнение. Например, в первой строке

 = 0,947\*81 + 60,279 = 136,99 ≈137.

Ā (%) - средняя ошибка аппроксимации (заголовок последнего столбца) рассчитывается по формуле (ниже). Модуль функции в формуле можно заменить абсолютным значением abs.



= -13,01/124 = 0,1049 (первая строка таблицы).

Формулы для расчета дисперсии и среднеквадратического отклонения даны.

**- среднее квадрата х,

 - квадрат среднего значения х.

* Вычислить значение -критерия Фишера.



Сумма равна 1334,597.

Вводим дополнительный столбец

По таблице Фишера находим табличное значение на пересечении столбца *k*1 = 1 и строки *k*2 = 10, *α =*0,05: *F*= 4,96. *F* р> *F* табл .

Уравнение значимо в целом.

* Оценить значимость каждого параметра уравнения регрессии

По таблице Стъюдента tтабл = 2,23.

* Определить ошибки параметров *ma, mb*
* Найти t-распределение для коэффициента корреляции.
* Коэффициент корреляции.
* Коэффициент детерминации.
* Считаем доверительный интервал.

2. **Построить точечную диаграмму**

Точечная диаграмма. Вставка - Диаграмма - точечная.

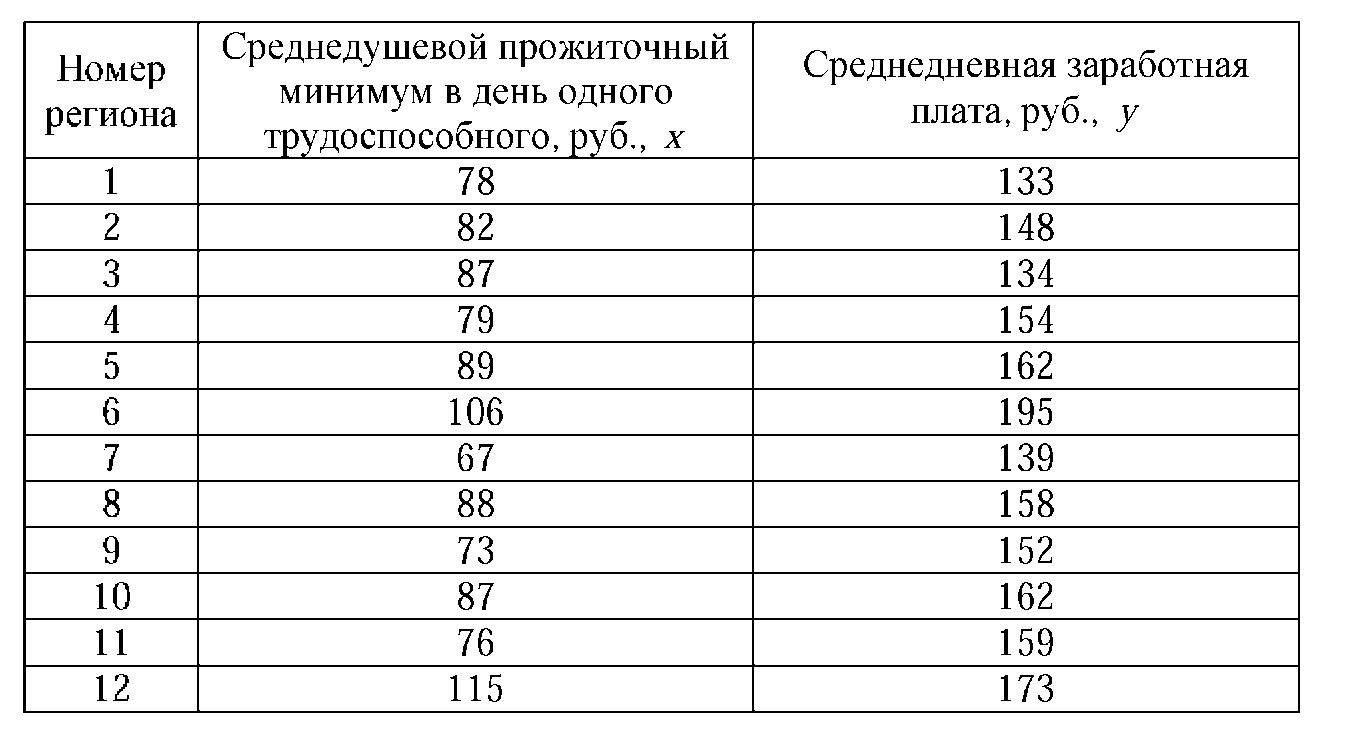
Выделить диаграмму: Работа с диаграммами - Макет - Анализ - Линия тренда - Дополнительные параметры линии тренда - флажки (показывать уравнение на диаграмме, поместить на диаграмму величину достоверности аппроксимации R2).

* Вывод.

# Пример: Решение типовой задачи

**Таблица 1**

**Данные за 199Х г. по территориям региона**



Требуется:

1. Построить линейное уравнение парной регрессии у по х.

1. Рассчитать линейный коэффициент парной корреляции, коэффициент детерминации и среднюю ошибку аппроксимации.
2. Оценить статистическую значимость уравнения регрессии в целом и отдельных параметров регрессии и корреляции с помощью F-критерия Фишера и t-критерия Стьюдента.

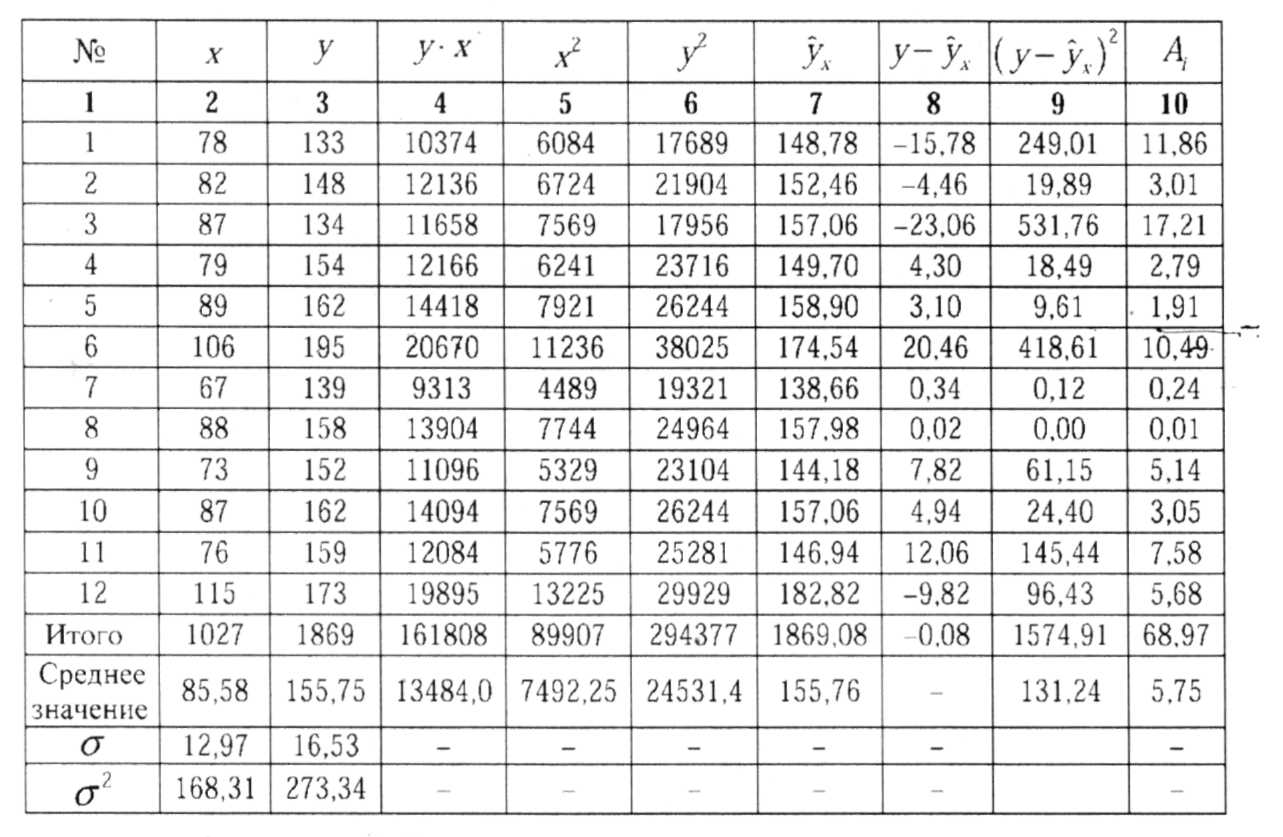
4. Выполнить прогноз заработной платы у при значении среднедушевого прожиточного минимума х, составляющем 107% от среднего уровня.

1. Оценить точность прогноза, рассчитав ошибку прогноза и его доверительный интервал.
2. На одном графике отложить исходные данные и теоретическую прямую.

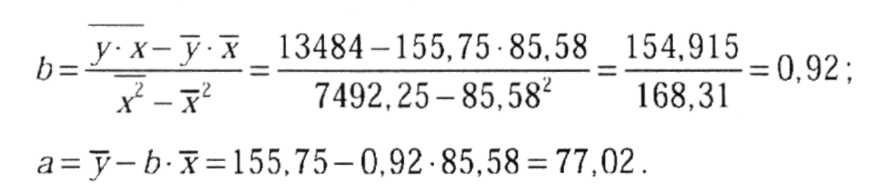
Решение

1. Для расчета параметров уравнения линейной регрессии строим расчетную таблицу 2.

Таблица 2



По формулам находим параметры регрессии



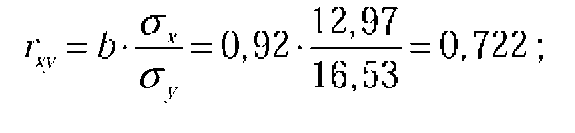
Получено уравнение регрессии:

у = 77,02 + 0,92\*х.

Параметр регрессии позволяет сделать вывод, что с увеличением среднедушевого прожиточного минимума на 1 руб. среднедневная заработная плата возрастает в среднем на 0,92 руб. (или 92 коп.).

После нахождения уравнения регрессии заполняем столбцы 7-10 таблицы 3.

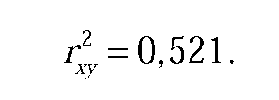
2. Тесноту линейной связи оценит коэффициент корреляции:



Т.к. значение коэффициента корреляции больше 0,7, то это говорит о

наличии весьма тесной линейной связи между признаками.

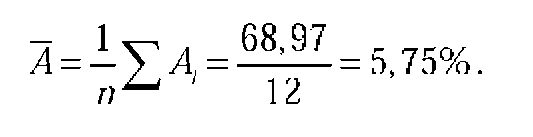
Коэффициент детерминации:



Это означает, что 52% вариации заработной платы (у) объясняется

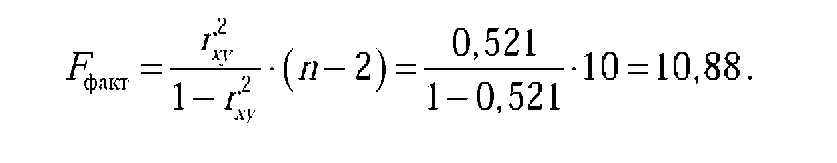
вариацией фактора х - среднедушевого прожиточного минимума.

Качество модели определяет средняя ошибка аппроксимации:



Качество построенной модели оценивается как хорошее, так как Ặ не превышает 10%.

3. Оценку статистической значимости уравнения регрессии в целом проведем с помощью *F*-критерия Фишера. Фактическое значение F-критерия составит

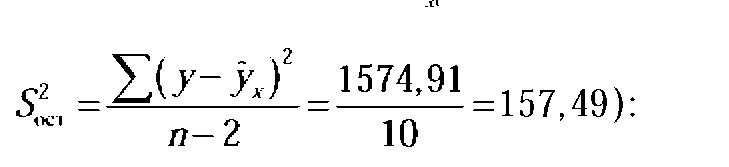


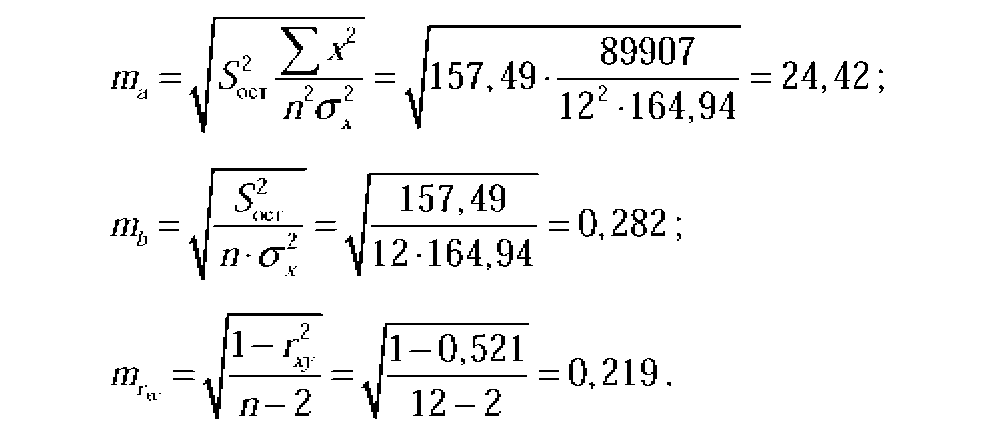
Табличное значение критерия при пятипроцентном уровне значимости и степенях свободы *k*1 = 1 и *k*12 = 12-2 = 10 составляет

*F*табл = 4,96. Так как *F*факт = 10,41 > Fтабл = 4,96, то уравнение регрессии признается статистически значимым. Оценку статистической значимости параметров регрессии и корреляции проведем с помощью t -статистики Стьюдента и путем расчета доверительного интервала каждого из параметров.

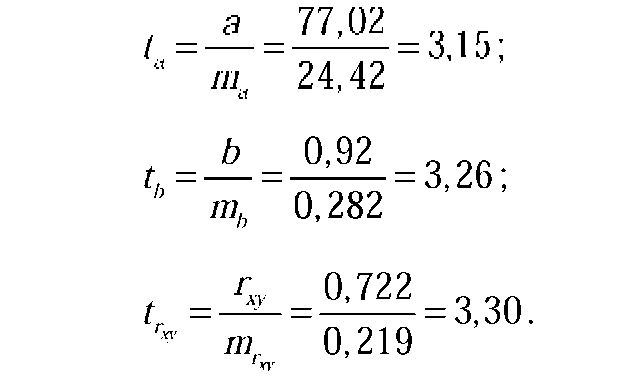
Табличное значение t-критерия для числа степеней свободы

*df* = *n* -2 = 12 - 2 = 10 и уровня значимости а = 0,05 составит *t*табл = 2,23.

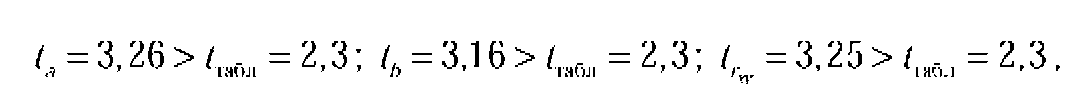
Определим стандартные ошибки тa, тb, mrxy  (остаточная дисперсия на одну степень свободы



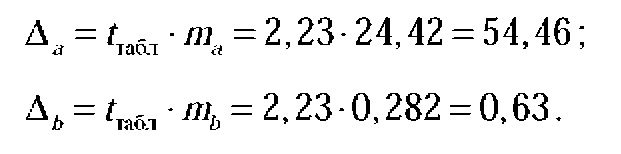
Тогда

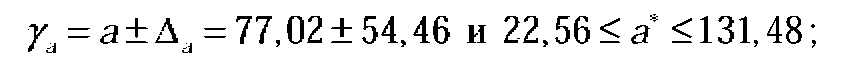


Фактические значения t-статистики превосходят табличное значение:



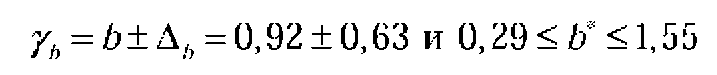
поэтому параметры *a*,b и rxy не случайно отличаются от нуля, а статистически значимы.





Рассчитаем доверительные интервалы для параметров регрессии a и *b*. Для этого определим предельную ошибку для каждого показателя:

Доверительные интервалы

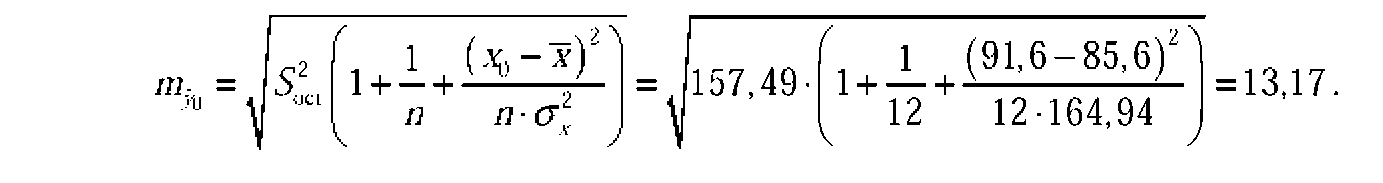


Анализ верхней и нижней границ доверительных интервалов приводит к выводу о том, что с вероятностью р = 1 – а = 0,95 параметры *а*

и b, находясь в указанных границах, не принимают нулевых значений, т.е. являются статистически значимыми и существенно отличны от нуля.

4. Полученные оценки уравнения регрессии позволяют использовать  
его для прогноза. Если прогнозное значение прожиточного минимума  
составит: руб., тогда индивидуальное прогнозное значение заработной платы составит:

*ŷ* = 77,02 + 0,92-91,6 = 161,29 руб.



5. Ошибка прогноза составит:

Предельная ошибка прогноза, которая в 95% случаев не будет превышена



Доверительный интервал прогноза составит:



Выполненный прогноз среднемесячной заработной платы является надежным (*p* = 1 – *a* = 1 – 0,05 = 0,95) и находится в пределах от 131,92 руб. до 190,66 руб.

6. В заключение решения задачи построим на одном графике исходные данные и теоретическую прямую (рис. 1):

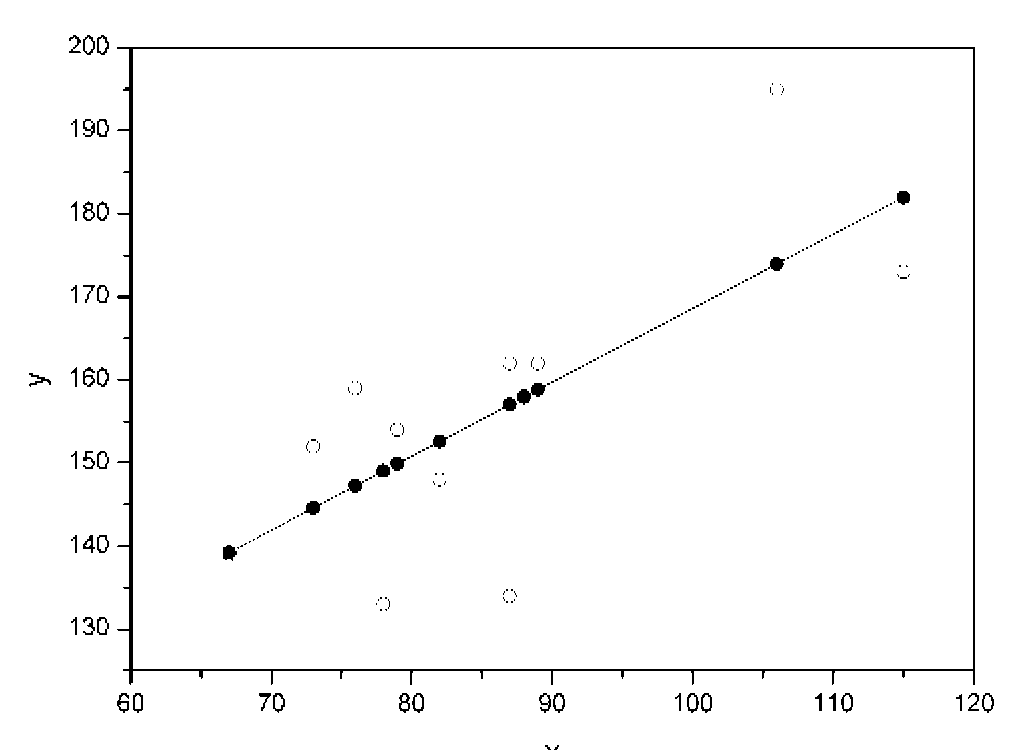


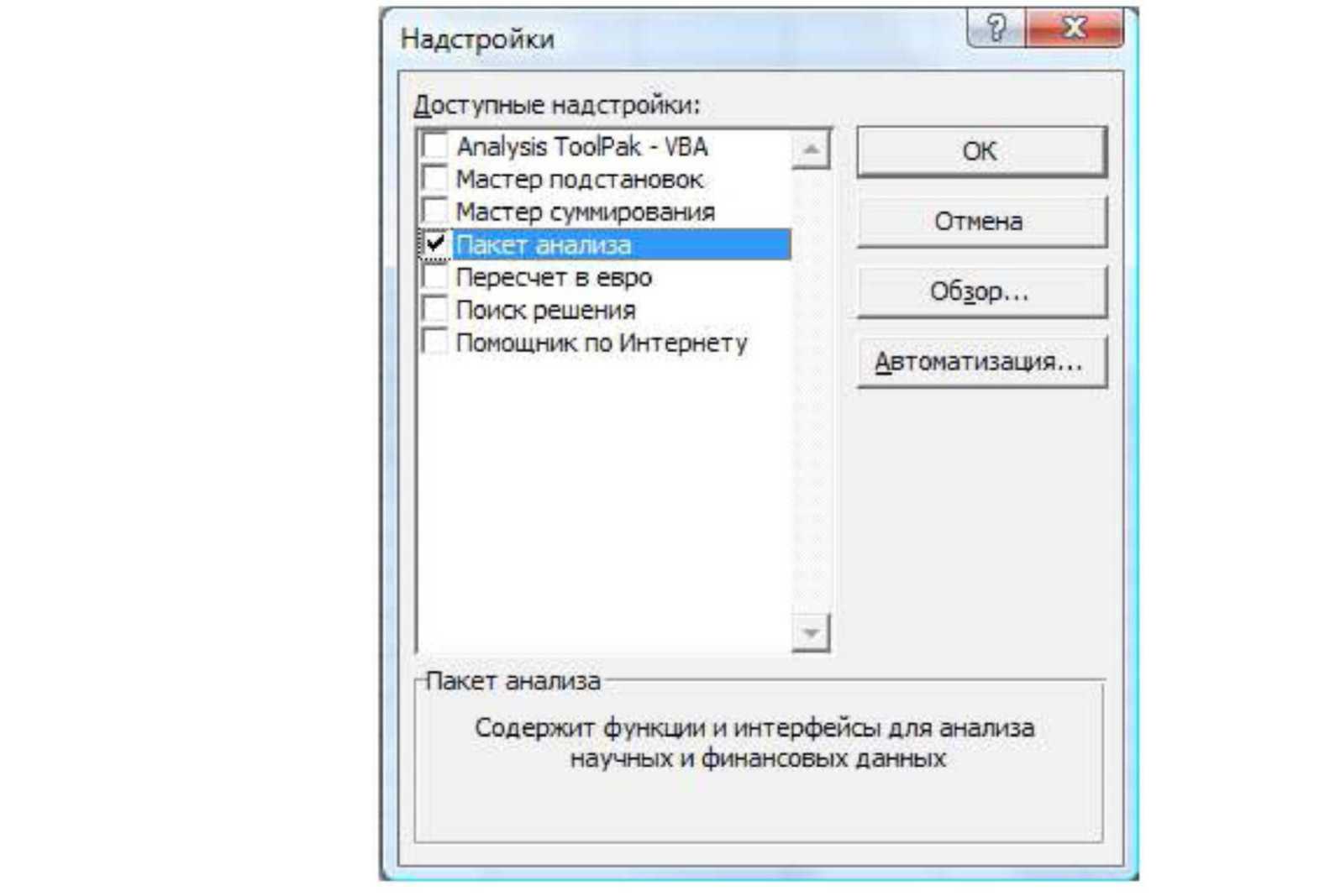
Рис. 1.

Рис. 1. Исходные данные и теоретическая прямая

2. Решение типовой задачи в MS Excel

С помощью инструмента анализа данных Регрессия можно получить результаты регрессионной статистики, дисперсионного анализа, доверительных интервалов, остатки и графики подбора линии регрессии.

Если в меню сервис еще нет команды Анализ данных, то необходимо сделать следующее. В главном меню последовательно выбираем Сервис → Надстройки и устанавливаем «флажок» в строке Пакет анализа (рис. 2).



:

Рис. 2. Надстройки - пакет анализа

Далее следуем по следующему плану.

1. Если исходные данные уже внесены, то выбираем Сервис → Анализ данных → Регрессия.

2. Заполняем диалоговое окно ввода данных и параметров вывода (рис. 3)

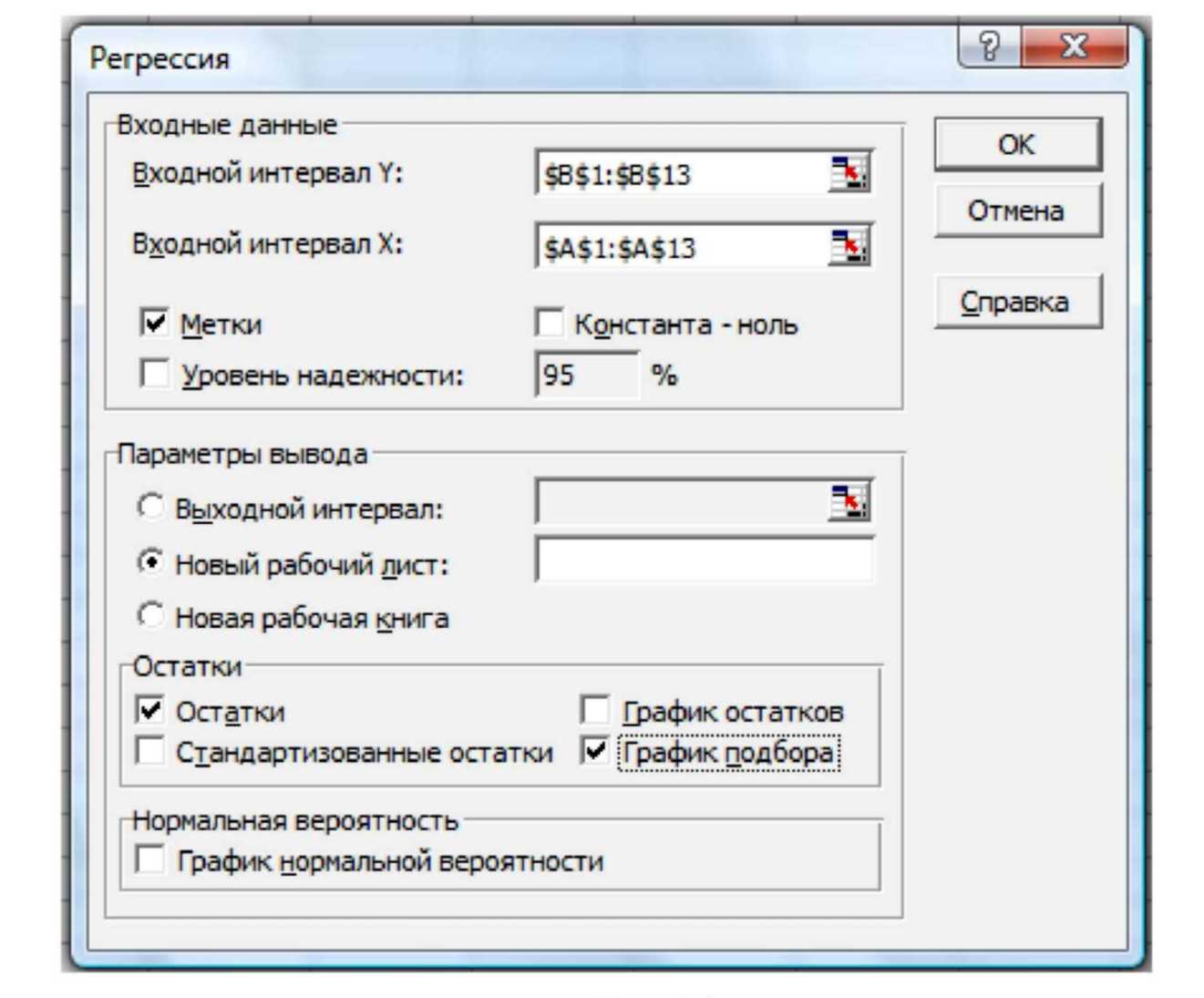


Рис.3. Анализ данных - Регрессия

Здесь:

Входной интервал Y - диапазон, содержащий данные результативного признака;

Входной интервал X - диапазон, содержащий данные признака-фактора;

Метки - «флажок», который указывает, содержит ли первая строка названия столбцов;

Константа - ноль - «флажок», указывающий на наличие или отсутствие свободного члена в уравнении;

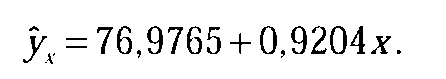
Выходной интервал - достаточно указать левую верхнюю ячейку будущего диапазона;

Новый рабочий лист - можно указать произвольное имя нового листа (или не указывать, тогда результаты выводятся на вновь созданный лист).

Получаем следующие результаты для рассмотренного выше примера:

Откуда выписываем, округляя до 4 знаков после запятой и переходя к нашим обозначениям:

Уравнение регрессии:



Коэффициент корреляции:

Коэффициент детерминации:

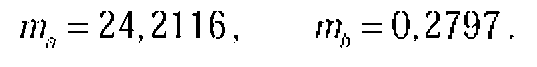


Фактическое значение F -критерия Фишера:

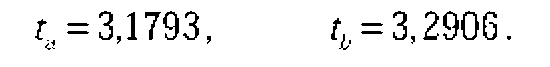


Остаточная дисперсия на одну степень свободы:

Корень квадратный из остаточной дисперсии (стандартная ошибка):

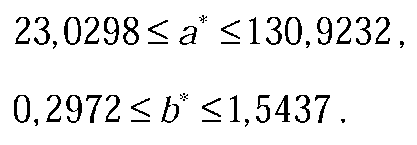


Стандартные ошибки для параметров регрессии:



Фактические значения t -критерия Стьюдента

Доверительные интервалы:



Как видно, найдены все рассмотренные выше параметры и характеристики уравнения регрессии, за исключением средней ошибки аппроксимации (значение t -критерия Стьюдента для коэффициента корреляции совпадает с tb). Результаты «ручного счета» от машинного отличаются незначительно (отличия связаны с ошибками округления).

1. Средние значения округлены до одного знака после запятой. Если рассчитать до 3 знаков - результаты расчетов совпадают до сотой. [↑](#footnote-ref-1)