

## ***Лабораторная работа №2***

### **Регрессионный анализ**

**Цель работы:** получить навыки проведения регрессионного анализа и прогнозирования

#### ***Уравнение линейной парной регрессии***

Уравнение линейной парной регрессии выглядит следующим образом:  $Y=a_0+a_1X$

При помощи этого уравнения переменная  $Y$  выражается через константу  $a_0$  и угол наклона прямой (или угловой коэффициент)  $a_1$ , умноженный на значение переменной  $X$ . Константу  $a_0$  также называют свободным членом, а угловой коэффициент - коэффициентом регрессии. Параметры уравнения могут быть определены с помощью метода наименьших квадратов (МНК)

#### ***Метод наименьших квадратов***

(в справочных системах англоязычных программ - Least Squares Method, LS) является одним из основных методов определения параметров регрессионных уравнений, дающий наилучшие линейные несмещенные оценки. Именно он используется в MS Excel. Линейные – относятся к характеру взаимосвязи переменных. Несмещенные значит, что ожидаемые значения коэффициентов регрессии должны быть истинными коэффициентами. То есть точки, построенные по исходным данным  $(x_i, y_i)$ , должны лежать как можно ближе к точкам линии регрессии. Сущность данного метода заключается в нахождении параметров модели, при которых сумма квадратов отклонений эмпирических (фактических) значений результирующего признака от теоретических, полученных по выбранному уравнению регрессии, то есть:

$$S = \sum_{i=1}^n (y_i^{\partial} - y_i)^2 = \sum_{i=1}^n (y_i^p - a_0 - a_1 x)^2 \rightarrow \min ,$$

где  $y_i^p$  – значение, вычисленное по уравнению регрессии;  $(y_i^p - y_i)$  – **отклонение  $\varepsilon$**  (ошибка, остаток) (рис. 1);  $n$  – количество пар исходных данных.

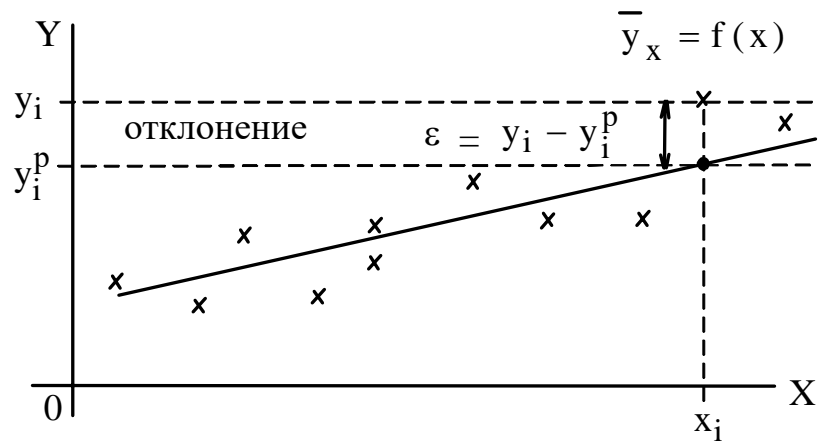


Рис. 1 Понятие отклонения  $\varepsilon$  для случая линейной регрессии

В регрессионном анализе предполагается, что математическое ожидание случайной величины  $\varepsilon$  равно нулю и ее дисперсия одинакова для всех наблюдаемых значений  $Y$ . Отсюда следует, что рассеяние данных возле линии регрессии должно быть одинаково при всех значениях параметра  $X$ . В случае, показанном на рис. 2 данные распределяются вдоль линии регрессии неравномерно, поэтому метод наименьших квадратов в этом случае неприменим.

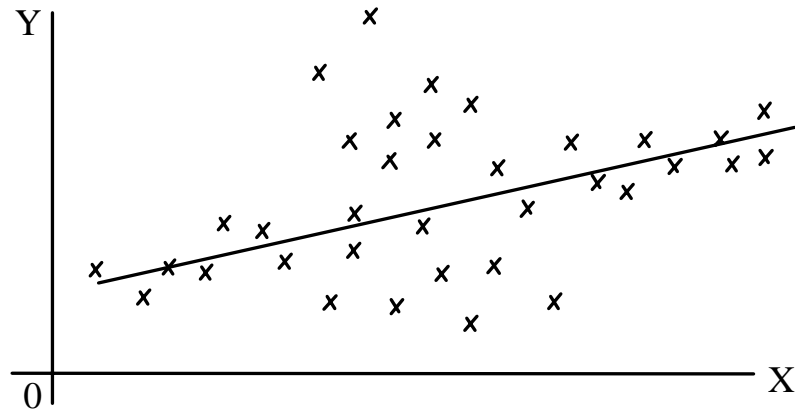


Рис.2. Неравномерное распределение исходных точек вдоль линии регрессии

Проведя необходимые преобразования, получим систему двух уравнений с двумя неизвестными  $a_0$  и  $a_1$ , которые найдем решив систему.

$$a_1 = \frac{n(\sum y_i x_i) - \sum y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2}; \quad (1)$$

$$a_0 = \frac{1}{n}(\sum y_i - a_1 \sum x_i) \quad (2)$$

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициента регрессии (коэффициента  $a_1$ ).

Если знак при коэффициенте регрессии - положительный, связь зависимой переменной с независимой будет положительной. В нашем случае знак коэффициента регрессии положительный, следовательно, связь также является положительной.

Если знак при коэффициенте регрессии - отрицательный, связь зависимой переменной с независимой является отрицательной (обратной).

Для анализа общего качества уравнения регрессии используют обычно *множественный коэффициент детерминации*  $R^2$ , называемый также квадратом коэффициента

множественной корреляции  $R$ .  $R^2$  (мера определенности) всегда находится в пределах интервала  $[0;1]$ .

Если значение  $R^2$  близко к единице, это означает, что построенная модель объясняет почти всю изменчивость соответствующих переменных. И наоборот, значение  $R$ -квадрата, близкое к нулю, означает плохое качество построенной модели.

Коэффициент детерминации  $R^2$  показывает, на сколько процентов ( $R^2 \cdot 100\%$ ) найденная функция регрессии описывает связь между исходными значениями факторов  $X$  и  $Y$

$$R^2 = \frac{\sum_{i=1}^n (y_i^p - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

где  $(y_i^p - \bar{y})^2$  – объясненная вариация;  $(y_i - \bar{y})^2$  – общая вариация (рис.3).



Рис. 3 Графическая интерпретация коэффициента детерминации для случая линейной регрессии. Соответственно, величина  $(1 - R^2) \cdot 100\%$  показывает, сколько процентов вариации параметра  $Y$  обусловлены факторами, не включенными в регрессионную модель. При высоком ( $R^2 \geq 75\%$ ) значении коэффициента детерминации можно делать прогноз  $y^* = f(x^*)$  для конкретного значения  $x^*$ .

### Нелинейная регрессия

Рассмотрим наиболее простые случаи *нелинейной* регрессии: гиперболу, экспоненту и параболу. При нахождении коэффициентов гиперболы и экспоненты используют прием приведения нелинейной регрессионной зависимости к линейному виду. Это позволяет использовать для вычисления коэффициентов функций регрессии выше приведенные формулы.

**Гипербола.** Для приведения уравнения вида  $y = a_0 + \frac{a_1}{x}$  к линейному виду вводят новую переменную  $z = \frac{1}{x}$ , тогда уравнение гиперболы принимает линейный вид  $y = a_0 + a_1 z$ . После этого используют формулы (1) и (2) для нахождения линейной функции, но вместо значений  $x_i$  используются значения  $z_i = \frac{1}{x_i}$ :

$$a_1 = \frac{n(\sum y_i z_i) - \sum y_i \sum z_i}{n(\sum z_i^2) - (\sum z_i)^2}; \quad a_0 = \frac{1}{n}(\sum y_i - a_1 \sum z_i). \quad (3)$$

**Экспонента.** Для приведения к линейному виду уравнения экспоненты  $y = a_0 e^{a_1 x}$  проведем логарифмирование:

$$\begin{aligned} \ln y &= \ln(a_0 e^{a_1 x}); \\ \ln y &= \ln a_0 + \ln(e^{a_1 x}); \\ \ln y &= \ln a_0 + a_1 x. \end{aligned}$$

Введем переменные  $b_0 = \ln a_0$  и  $b_1 = a_1$ , тогда  $\ln y = b_0 + b_1 x$ , откуда следует, что можно применять формулы (1) и (2), в которых вместо значений  $y_i$  надо использовать  $\ln y_i$ :

$$b_1 = \frac{n(\sum [\ln y_i] x_i) - \sum \ln y_i \sum x_i}{n(\sum x_i^2) - (\sum x_i)^2}; \quad b_0 = \frac{1}{n}(\sum \ln y_i - b_1 \sum x_i) \quad (4)$$

При этом мы получим численные значения коэффициентов  $b_0$  и  $b_1$ , от которых надо перейти к  $a_0$  и  $a_1$ , используемых в модели экспоненты. Исходя из введенных обозначений и определения логарифма, получаем

$$a_0 = e^{b_0}, \quad a_1 = b_1.$$

**Парабола.** Для нахождения коэффициентов уравнения параболы  $y = a_0 + a_1 x + a_2 x^2$  необходимо решить линейную систему из трех уравнений:

$$\begin{cases} n \cdot a_0 + (\sum x_i) a_1 + (\sum x_i^2) a_2 = \sum y_i, \\ (\sum x_i) a_0 + (\sum x_i^2) a_1 + (\sum x_i^3) a_2 = \sum (y_i x_i), \\ (\sum x_i^2) a_0 + (\sum x_i^3) a_1 + (\sum x_i^4) a_2 = \sum (y_i x_i^2). \end{cases}$$

Сила регрессионной связи для гиперболы и параболы определяется непосредственно по той же формуле что и для линейной модели. При вычислении коэффициента детерминации для экспоненты все значения параметра  $Y$  (исходные, регрессионные, среднее) необходимо заменить на их логарифмы, например,  $y_i^p$  – на  $\ln(y_i^p)$  и т.д.

Если функция регрессии определена, интерпретирована и обоснована, и оценка точности регрессионного анализа соответствует требованиям, можно считать, что построенная модель и прогнозные значения обладают достаточной надежностью.

Прогнозные значения, полученные таким способом, являются средними значениями, которые можно ожидать.

## **Методические рекомендации**

Для проведения регрессионного анализа и прогнозирования необходимо:

- 1) **построить график** исходных данных и попытаться зрительно, приближенно определить характер зависимости;
- 2) **выбрать вид функции** регрессии, которая может описывать связь исходных данных;
- 3) **определить численные коэффициенты** функции регрессии методом наименьших квадратов;
- 4) **оценить силу** найденной регрессионной зависимости на основе коэффициента детерминации  $R^2$ ;
- 5) **сделать прогноз** (при  $R^2 \geq 75\%$ ) или сделать вывод о невозможности прогнозирования с помощью найденной регрессионной зависимости. При этом не рекомендуется использовать модель регрессии для тех значений независимого параметра  $X$ , которые не принадлежат интервалу, заданному в исходных данных.

### **Справочная информация по технологии работы с режимом "Регрессия" настройки Пакет анализа MS Excel**

Режим работы "Регрессия" служит для расчета параметров уравнения линейной регрессии и проверки его адекватности исследуемому процессу.

Для решения задачи регрессионного анализа в MS Excel выбираем в меню **Сервис** команду **Анализ данных** и инструмент анализа **"Регрессия"**.

В появившемся диалоговом окне задаем следующие параметры:

1. **Входной интервал  $Y$**  - это диапазон данных по результативному признаку. Он должен состоять из одного столбца.
2. **Входной интервал  $X$**  - это диапазон ячеек, содержащих значения факторов (независимых переменных). Число входных диапазонов (столбцов) должно быть не больше 16.
3. **Флажок *Метки***, устанавливается в том случае, если в первой строке диапазона стоит заголовок.
4. **Флажок *Уровень надежности*** активизируется, если в поле, находящееся рядом с ним необходимо ввести уровень надежности, отличный от установленного по умолчанию. Используется для проверки значимости коэффициента детерминации  $R^2$  и коэффициентов регрессии.
5. **Константа ноль**. Данный флажок необходимо установить, если линия регрессии должна пройти через начало координат ( $a_0=0$ ).
6. **Выходной интервал/ Новый рабочий лист/ Новая рабочая книга** – указать адрес верхней левой ячейки выходного диапазона.

7. Флажки в группе *Остатки* устанавливаются, если необходимо включить в выходной диапазон соответствующие столбцы или графики.

8. Флажок График нормальной вероятности необходимо сделать активным, если требуется вывести на лист точечный график зависимости наблюдаемых значений  $Y$  от автоматически формируемых интервалов перцентилей.

После нажатия кнопки ОК в выходном диапазоне получаем отчет.

### ***Пример выполнения лабораторной работы***

**Задача:** Некоторая фирма занимается поставками различных грузов на короткие расстояния внутри города. Оценить стоимость таких услуг, зависящую от затрачиваемого на поставку времени. В качестве наиболее важного фактора, влияющего на время поставки, выбрано пройденное расстояние. Были собраны исходные данные о десяти поставках (таблица 1)

Таблица 1

Расстояние, миль	3,5	2,4	4,9	4,2	3,0	1,3	1,0	3,0	1,5	4,1
Время, мин	16	13	19	18	12	11	8	14	9	16

Определите характер зависимости между расстоянием и затраченным временем, используя мастер диаграмм MS Excel, проанализируйте применимость метода наименьших квадратов, постройте уравнение регрессии, используя МНК, проанализируйте силу регрессионной связи. Проведите регрессионный анализ, используя режим работы "Регрессия" в MS Excel и сравните с результатами, полученными ранее. Сделайте прогноз времени поездки на 2 мили. Посчитать и построить графически меру ошибки регрессионной модели используя табличный процессор Excel.

### ***Решение***

На графике строим исходные данные по десяти поездкам.

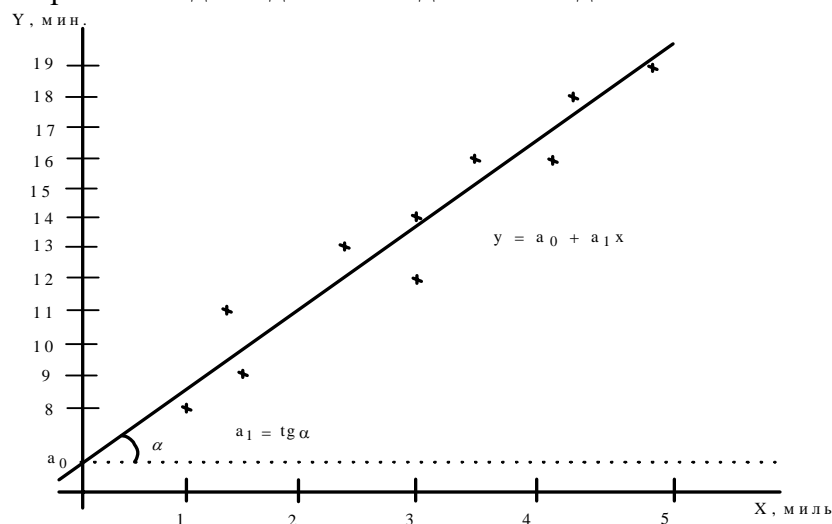


Рис.4. График исходных данных и предполагаемая линия регрессии

Помимо расстояния на время поставки влияют пробки на дорогах, время суток, дорожные работы, погода, квалификация водителя, вид транспорта. Построенные точки не находятся точно на линии, что обусловлено описанными выше факторами. Но эти точки собраны вокруг прямой линии, поэтому можно предположить линейную связь между параметрами. Все исходные точки равномерно распределены вдоль предполагаемой прямой линии, что позволяет применить метод наименьших квадратов.

Вычислим суммы, необходимые для расчета коэффициентов уравнения линейной регрессии и коэффициента детерминации  $R^2$  с помощью вспомогательной таблицы (таблица 2).

Таблица 2

$x_i$	$y_i$	$x_i^2$	$x_i y_i$	$y_i^p$	$(y_i^p - \bar{y})^2$	$(y_i - \bar{y})^2$
3,5	16	12,25	56,00	15,223	2,634129	5,76
2,4	13	5,76	31,2	12,297	1,697809	0,36
4,9	19	24,01	93,1	18,947	28,59041	29,16
4,2	18	17,64	75,60	17,085	12,14523	19,36
3,0	12	9,00	36,00	13,893	0,085849	2,56
1,3	11	1,69	14,30	9,371	17,88444	6,76
1,0	8	1,00	8,00	8,573	25,27073	31,36
3,0	14	9,00	42,00	13,893	0,085849	0,16
1,5	9	2,25	13,50	9,903	13,66781	21,16
4,1	16	16,81	65,60	16,819	10,36196	5,76
$\Sigma = 28,9$	$\Sigma = 136$	$\Sigma = 99,41$	$\Sigma = 435,30$	–	112,4242	122,4

$$\bar{y} = \frac{\sum y_i}{n} = \frac{16 + 13 + 19 + 18 + 12 + 11 + 8 + 14 + 9 + 16}{10} = 13,6.$$

Вычислим коэффициенты линейной регрессии по формулам (1) и (2):

$$a_1 = \frac{10 \cdot 435,30 - 136 \cdot 28,9}{10 \cdot 99,41 - 835,21} = 2,660 ;$$

$$a_0 = 0,1 \cdot (136 - 2,660 \cdot 28,9) = 5,913.$$

Таким образом, искомая регрессионная зависимость имеет вид:

$$y^p = 5,913 + 2,660x.$$

Наклон линии регрессии  $a_1 = 2,66$  минут на милю – это количество минут, приходящееся на одну милю расстояния. Координата точки пересечения прямой с осью  $Y$   $a_0 = 5,913$  минут – это время, которое не зависит от пройденного расстояния, а обуславливается всеми остальными возможными факторами, явно не учтенными при анализе.

Вычислим коэффициент детерминации:

$$R^2 = \frac{112,424}{122,400} = 0,918 \text{ или } 91,8\%.$$

Проведем регрессионный анализ с использованием режима Регрессия MS Excel. Значения параметров, установленных в одноименном диалоговом окне, представлены на рис.6.

Рис. 6.

Сгенерируются результаты по регрессионной статистике, представленные в таблице 3.

Таблица 3.

#### ВЫВОД ИТОГОВ

<i>Регрессионная статистика</i>	
Множественный R	0,958275757
R-квадрат	0,918292427
Нормированный R-квадрат	0,90807898
Стандартная ошибка	1,11809028
Наблюдения	10

Рассмотрим представленную в таблице 3 регрессионную статистику. Величина R-квадрат, называемая также мерой определенности, характеризует качество полученной регрессионной прямой. Это качество выражается степенью соответствия между исходными данными и регрессионной моделью (расчетными данными). Мера определенности всегда находится в пределах интервала  $[0;1]$ . В нашем примере мера определенности равна 0,91829, что говорит об очень хорошей подгонке регрессионной прямой к исходным данным и совпадает с коэффициентом детерминации  $R^2$ , вычисленным по формуле.

Таким образом, линейная модель объясняет 91,8% вариации времени доставки, что означает правильность выбора фактора (расстояния). Не объясняется  $100\% - 91,8\% = 8,2\%$  вариации времени поездки, которые обусловлены остальными факторами, влияющими на время поставки, но не включенными в линейную модель регрессии.

Рассчитанный уровень значимости  $\alpha_p = 1,26E-05 < 0,05$  (показатель значимости F в таблице Дисперсионный анализ) подтверждает значимость  $R^2$ .

Множественный R - коэффициент множественной корреляции R - выражает степень зависимости независимых переменных (X) и зависимой переменной (Y) и равен квадратному корню из коэффициента детерминации, эта величина принимает значения в интервале от нуля до единицы. В простом линейном регрессионном анализе множественный R равен



коэффициенту корреляции Пирсона. Действительно, множественный R в нашем случае равен коэффициенту корреляции Пирсона (0,95827), который вычисляется по формуле:

$$r_{xy} = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2] \cdot [n \sum y^2 - (\sum y)^2]}}$$

Теперь рассмотрим среднюю часть расчетов, представленную в таблице 4 (приведена в сокращенном варианте). Здесь даны коэффициент регрессии  $a_1$  (2,65970168) и смещение по оси ординат, т.е. константа  $a_0$  (5,913462144).

Таблица 4

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-Значение</i>
Y-пересечение	5,913462144	0,884389599	6,686489927	0,00015485
Переменная X 1	2,65970168	0,280497238	9,482095791	1,26072E-05

Исходя из расчетов, можем записать уравнение регрессии таким образом:

$$y^{\hat{}} = 5,91346 + 2,65970x. \quad (*)$$

Видим, что это уравнение, совпадает с уравнением, полученным нами при расчете по МНК вручную с точностью до ошибки округления.

Направление связи между переменными определяется на основании знаков (отрицательный или положительный) коэффициента регрессии (коэффициента  $a_1$ ). В нашем случае знак коэффициента регрессии положительный, следовательно, связь также является положительной.

Далее проверим значимость коэффициентов регрессии:  $a_0$  и  $a_1$ . Сравнивая попарно значения столбцов *Коэффициенты* и *Стандартная ошибка* в таблице 4, видим, что абсолютные значения коэффициентов больше чем их стандартные ошибки. К тому же эти коэффициенты являются значимыми, о чем можно судить по значениям показателя *P-значение* в таблице 4, которые меньше заданного уровня значимости  $\alpha=0,05$ .

Таблица 5

ВЫВОД ОСТАТКА			
<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
1	15,22241803	0,777581975	0,737641894
2	12,29674618	0,703253823	0,667131568
3	18,94600038	0,053999622	0,051225961
4	17,0842092	0,915790799	0,868751695
5	13,89256718	-1,892567185	-1,795356486
6	9,371074328	1,628925672	1,545256778
7	8,573163824	-0,573163824	-0,543723571
8	13,89256718	0,107432815	0,101914586
9	9,903014664	-0,903014664	-0,8566318
10	16,81823903	-0,818239033	-0,776210624

В таблице 5. представлены результаты вывода остатков. При помощи этой части отчета мы можем видеть отклонения каждой точки от построенной линии регрессии. Наибольшее абсолютное значение остатка в нашем случае - 1,89256, наименьшее - 0,05399. Для лучшей интерпретации этих данных воспользуемся графиком исходных данных и построенной линией

регрессии, представленными на рис. 7. Как видим, линия регрессии хорошо "подогнана" под значения исходных данных.

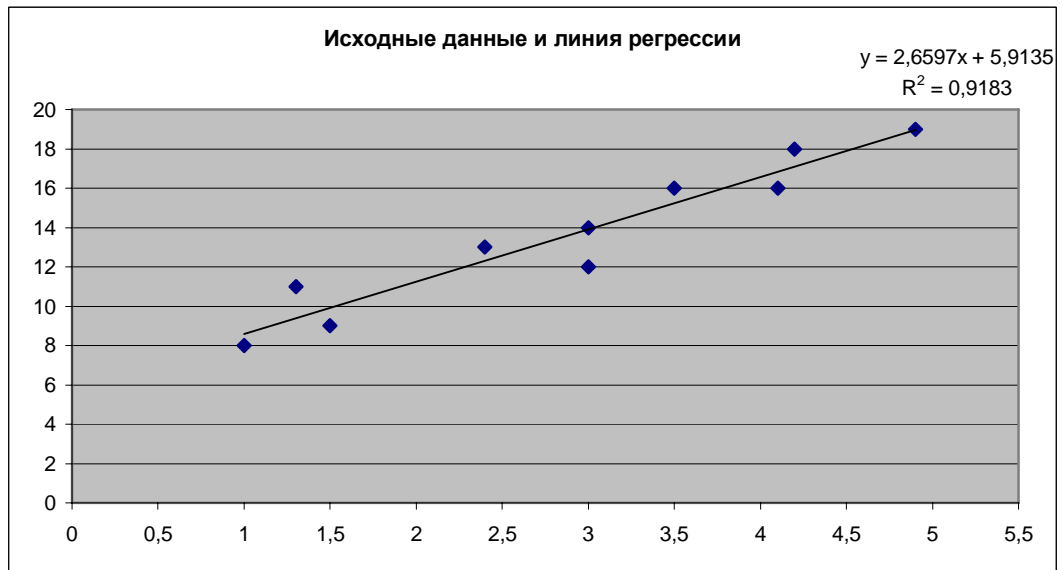


Рис.7.

Приблизительным, но самым простым и наглядным способом проверки удовлетворительности регрессионной модели является графическое представление

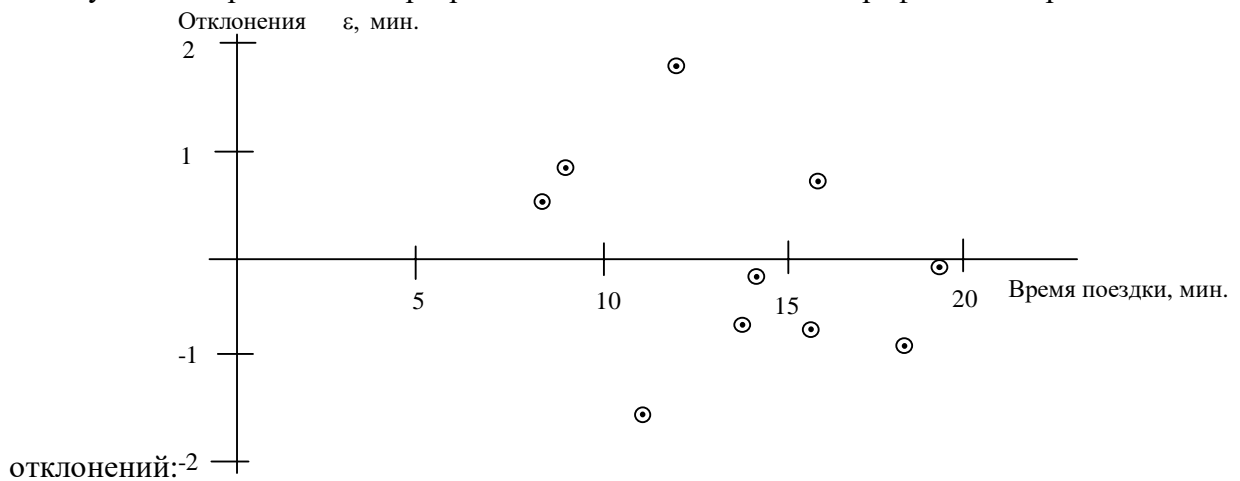
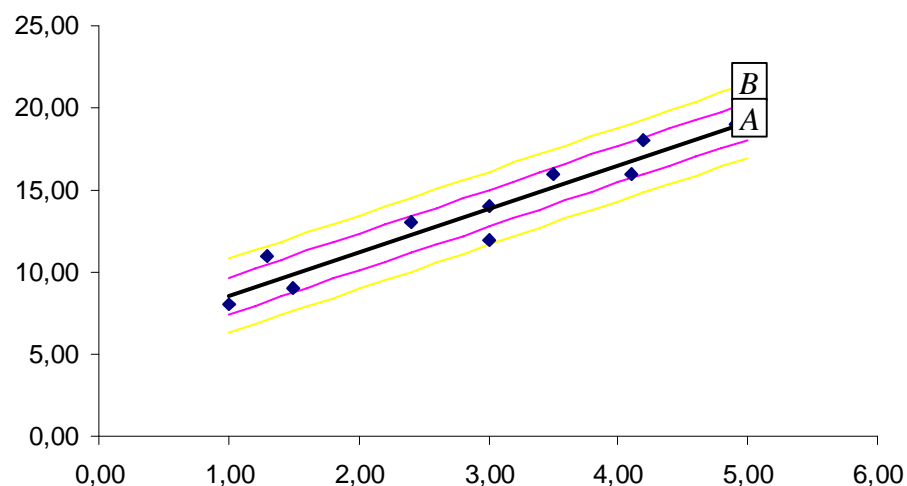


Рис. 8. График отклонений

Отложим отклонения  $(y_i^p - y_i)$  по оси Y, для каждого значения  $y_i$ . Если регрессионная модель близка к реальной зависимости, то отклонения будут носить случайный характер и их



сумма будет близка к нулю. В рассмотренном примере  $\sum_{i=1}^n (y_i^p - y_i) = 0,004$ .

Обычно мерой ошибки регрессионной модели служит среднее квадратическое отклонение

$$\sigma_{\varepsilon} = \left[ \left( \sum_{i=1}^n \varepsilon_i^2 \right) / (n-2) \right]^{1/2} = \left\{ \left[ \sum_{i=1}^n (y_i^p - y_i)^2 \right] / (n-2) \right\}^{1/2}$$

Для нормально распределенных процессов приблизительно 67% точек находится в пределах одного отклонения  $\sigma_{\varepsilon}$  от линии регрессии и 95% - в пределах  $2\sigma_{\varepsilon}$  (на рисунке 8 трубки А и В соответственно).

Решим задачу прогнозирования. Поскольку коэффициент детерминации  $R^2$  имеет достаточно высокое значение и расстояние 2 мили, для которого надо сделать прогноз, находится в пределах диапазона исходных данных (таблица 1), то мы можем использовать полученное уравнение линейной регрессии для прогнозирования

$$y^*(2 \text{ мили}) = 5,913 + 2,660 \cdot 2 = 11,2 \text{ минут.}$$

При прогнозах на расстояния, не входящие в диапазон исходных данных, нельзя гарантировать справедливость полученной модели. Это объясняется тем, что связь между временем и расстоянием может изменяться по мере увеличения расстояния. На время дальних перевозок могут влиять новые факторы такие, как использование скоростных шоссе, остановки на отдых, обед и т.п.

Таким образом, в результате использования регрессионного анализа в пакете Microsoft Excel мы:

- построили уравнение регрессии;
- установили форму зависимости и направление связи между переменными - положительная линейная регрессия, которая выражается в равномерном росте функции;
- установили направление связи между переменными;
- оценили качество полученной регрессионной прямой;
- смогли увидеть отклонения расчетных данных от данных исходного набора;
- предсказали будущее значение зависимой переменной.

## Варианты задач для самостоятельного решения

### Задача №1

Постройте регрессионную модель (линейную) для исходных данных приведенных в таблице. Для облегчения расчетов исходные данные содержат только четыре пары значений  $(x_i, y_i)$ .

### Исходные данные задачи №1

№ варианта	Координаты	Точки				$x^*$
1	X	1	2	3	4	1.6
	Y	30	7	8	1	?
2	X	1	2	3	4	2.3
	Y	25	7	7	2	
3	X	9	5	2	3	2.9

	Y	25	7	7	2	?
4	X	1	2	3	4	2.6
	Y	15	10	7	0.5	?
5	X	10	3	6	4	8
	Y	25	7	7	2	?
6	X	9	5	2	3	2.5
	Y	15	8.5	7.5	5	?
7	X	2	3	7	8	7.5
	Y	11	8.5	6.5	5	?
8	X	10	3	6	4	9
	Y	15	7	8	6	?
9	X	2	3	4	5	4.5
	Y	13	9	8	7	?
10	X	1	2	3	4	1.5
	Y	7.5	7	5	3.5	?
11	X	1	2	3	4	3.6
	Y	13	9	8	7	?
12	X	3	4	6	10	8
	Y	7.5	7	6.5	3.5	?
13	X	3	4	5	6	7.8
	Y	9	7	5	3	?
14	X	7	5.6	13	14.7	15
	Y	7.5	7	5	3.5	?
15	X	9	5	2	3	5.7
	Y	13	9	8	7	?
16	X	3	4	6	8	5
	Y	7.5	7	6.5	5	?
17	X	2	3	7	8	7.5
	Y	9	9	8	7	?
18	X	9	10	11	12	10.5
	Y	13	9	8	7	?
19	X	1	2	3	4	3.5
	Y	5	4.5	3	3	?
20	X	11	12	13	16	13.6
	Y	7.6	8	6.5	4.2	?
21	X	5	6	7	8	6.5
	Y	5	4.5	3	3	?
22	X	9	10	12	14	12.5
	Y	8	7	6.5	4.2	?
23	X	7	8	9	10	9.6
	Y	8	7	6	4.2	?
24	X	1.5	2.5	3.5	4.5	3.9
	Y	5	4.5	3	3	?
25	X	1	2	5	6	3.9
	Y	5	4	3	3	?
26	X	1.5	2.4	3.8	6.9	4.1
	Y	5.5	5.5	4.8	1.1	?
27	X	1	2	3	4	3.6
	Y	12	3	9	5	?
28	X	1	2	3	7	2.8
	Y	5	5.5	4.8	1.1	?
29	X	11	12	13	16	14.1
	Y	0.25	0.19	5.2	8	?
30	X	1	2	3	4	3.4

	Y	13	4	10	6	?
--	---	----	---	----	---	---

Исследуйте модель с помощью режима Регрессия в MS Excel и сделайте прогноз для  $x^*$ .

### ***Задача № 2***

Для исходных данных, представленных в таблице, были построены следующие регрессионные модели:

- $y = 6,067 - 0,085x$  ;
- $y = -2,017 + 3,957x - 0,367x^2$  ;
- $y = 5,918e^{-0,043x}$  .

### ***Исходные данные задачи №2***

X	3	8	5	10	7	6	4	9	1	2
Y	6	5	9	1	8	9	8	4	2	4

С помощью графика отклонений выберите удовлетворительную модель и проверьте свой выбор с помощью соответствующего расчета.

### ***Задача №3***

В таблице представлены данные о ценах на комплектующие для ПЭВМ. Комплектующие производятся различными компаниями-производителями и разбиты на группы по своим функциональным возможностям.

### ***Исходные данные задачи №3***

Группа	1	1	2	2	2	3	3	3	4	4
Цена, \$	50	60	70	80	95	100	115	120	105	120
Группа	4	5	5	5	6	6	6	7	7	7
Цена, \$	130	110	150	190	120	130	220	145	265	270

Постройте график исходных данных и с его помощью проанализируйте применимость метода наименьших квадратов. Подтвердите свои выводы с помощью расчета (для линейной модели). Прокомментируйте экономические причины полученного результата.