

Лабораторная работа №1

Метод наименьших квадратов

Цель работы: получить навыки расчета значений коэффициентов линейной зависимости по заданным эмпирическим данным.

№1

Постановка задачи:

Задача 1. Для изучения зависимости октанового числа бензина от чистоты катализатора (%) провели 11 измерений, приведенных ниже в таблице:

| | | | | | | | | | | | |
|----------------------|------|------|------|------|------|------|------|------|------|------|------|
| Чистота катализатора | 98.8 | 98.9 | 99.0 | 99.1 | 99.2 | 99.3 | 99.4 | 99.5 | 99.6 | 99.7 | 99.8 |
| Октановое число | 87.1 | 86.1 | 86.4 | 87.3 | 86.1 | 86.8 | 87.2 | 88.4 | 87.2 | 86.4 | 88.6 |

1). Найдите коэффициенты a , b линейной зависимости $y(x)$

= $ax + b$ октанового числа от чистоты катализатора

2). Постройте график зависимости.

3). Вычислите значение октанового числа для чистоты катализатора 87%.

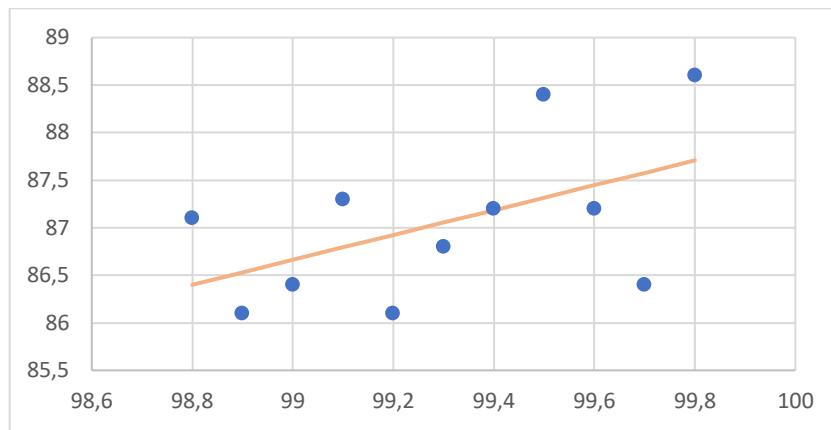
1. Нахождение коэффициентов a , b линейной зависимости $y(x) = ax + b$ октанового числа от чистоты катализатора:

$$a = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{8644,647 - 99,3 * 87,0545}{9860,59 - 99,3^2} = 1,309091$$

$$b = \bar{y} - a\bar{x} = 87,05455 - 1,309091 * 99,3 = -42,93818$$

$$y' = 1,309091x - 42,93818$$

2. График зависимости:



3. Вычисление значения октанового числа для чистоты катализатора 87%:

$$y' = 1,309091 * 87 - 42,93818 = 70,9527$$

4. Критерии для оценки полученной регрессионной модели:

1) средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \tilde{y}_i}{y_i} * 100\%,$$

где n — количество измерений;

y_i — заданные значения;

\tilde{y}_i — рассчитанные значения;

$$\bar{A} = 0,6667\%$$

2) коэффициент эластичности:

$$\Xi = a \frac{\bar{x}}{\bar{y}} = 1,49$$

3) коэффициент корреляции:

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2},$$

где σ_x — среднее квадратическое отклонение по x ;

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2},$$

где σ_y — среднее квадратическое отклонение по y ;

$$r = \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} = \frac{8644,647 - 99,3 * 87,0545}{0,316228 * 0,800413} = 0,517197$$

4) коэффициент детерминации:

$$r^2 = r^2 * 100\% = 26,75\%$$

5) t – критерий Стьюдента:

Т.к. исходные данные выборочные, то необходимо оценить значимость величины коэффициента корреляции.

Выдвигаем гипотезу H_0 :

коэффициент корреляции генеральной совокупности равен 0, и чистота катализатора не оказывает существенного влияния на октановое число бензина.

Тогда

$$H_0: r = 0$$

$$H_1: r \neq 0$$

t – критерий Стьюдента применяется для проверки гипотезы H_0 .

Расчетное значение t – критерия Стьюдента:

$$t_p = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}}$$

где $(n - 2)$ – количество степеней свободы;

$$t_p = 1,812887$$

Критическое значение t находится из таблицы “ t – распределение Стьюдента ” (уровень значимости $\alpha = 0,05$):

$$t_{кр} = 2,26$$

Т.к. $t_p = 1,812887 < t_{кр} = 2,26$, то H_0 подтверждается.

5. Выводы:

Коэффициент регрессии показывает, что в среднем при увеличении чистоты катализатора на 1% октановое число бензина в среднем увеличивается на 1,309091.

Т.к. $\bar{A} = 0,6667\% < 10\%$, то уравнение регрессии считается качественным.

Коэффициент эластичности показывает, что при увеличении чистоты катализатора на 1% октановое число бензина в среднем возрастает на 149%.

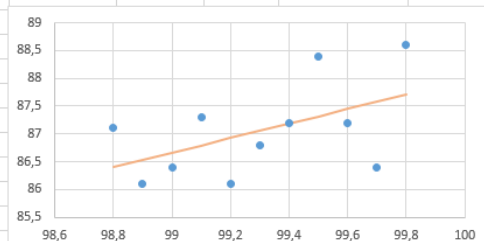
Т.к. $0,5 < r = 0,517197 < 0,7$, то взаимосвязь между признаками заметная.

Из значения $r^2 = 26,75\%$ можно сделать вывод, что октановое число бензина на 26,75% напрямую зависит от чистоты катализатора и при этом октановое число на 73,25% зависит от других факторов.

Т.к. подтвердилась гипотеза H_0 , значит чистота катализатора не оказывает существенного влияния на октановое число бензина.

6. Результаты:

| № | x | y | xy | x ² | y ² | y' | A' | ср(xy) | 8644,647 | A | 0,6667% |
|---------|--------|---------|---------|----------------|----------------|---------|---------|----------------------|-----------|----------------|----------|
| 1 | 98,8 | 87,1 | 8605,48 | 9761,44 | 7586,41 | 86,4 | 0,00804 | ср(x)*ср(y) | 8644,516 | Э | 149% |
| 2 | 98,9 | 86,1 | 8515,29 | 9781,21 | 7413,21 | 86,5309 | 0,005 | ср(x ²) | 9860,59 | ср кв откл X | 0,316228 |
| 3 | 99 | 86,4 | 8553,6 | 9801 | 7464,96 | 86,6618 | 0,00303 | (ср(x)) ² | 9860,49 | ср кв откл Y | 0,800413 |
| 4 | 99,1 | 87,3 | 8651,43 | 9820,81 | 7621,29 | 86,7927 | 0,00581 | b1 | 1,309091 | r | 0,517197 |
| 5 | 99,2 | 86,1 | 8541,12 | 9840,64 | 7413,21 | 86,9236 | 0,00957 | ср(x) | 99,3 | r ² | 26,75% |
| 6 | 99,3 | 86,8 | 8619,24 | 9860,49 | 7534,24 | 87,0545 | 0,00293 | ср(y) | 87,05455 | t p | 1,812887 |
| 7 | 99,4 | 87,2 | 8667,68 | 9880,36 | 7603,84 | 87,1855 | 0,00017 | b0 | -42,93818 | t кр | 2,26 |
| 8 | 99,5 | 88,4 | 8795,8 | 9900,25 | 7814,56 | 87,3164 | 0,01226 | | | | |
| 9 | 99,6 | 87,2 | 8685,12 | 9920,16 | 7603,84 | 87,4473 | 0,00284 | | | | |
| 10 | 99,7 | 86,4 | 8614,08 | 9940,09 | 7464,96 | 87,5782 | 0,01364 | | | | |
| 11 | 99,8 | 88,6 | 8842,28 | 9960,04 | 7849,96 | 87,7091 | 0,01006 | | | | |
| Итого | 1092,3 | 957,6 | 95091,1 | 108466 | 83370,5 | 957,6 | 0,07333 | | | | |
| Среднее | 99,3 | 87,0545 | 8644,65 | 9860,59 | 7579,13 | 87,0545 | 0,00667 | | | | |
| | 87 | 70,9527 | | | | | | | | | |



№2

Постановка задачи:

Задача 2. Имеются следующие выборочные данные о стоимости квартир и общей их площади в некотором городе.

| | | | | | | | | | | | | | | | |
|---|------|------|----|------|----|----|----|------|----|------|----|----|----|------|------|
| y | 13.8 | 13.8 | 14 | 22.5 | 24 | 28 | 32 | 20.9 | 22 | 21.5 | 32 | 35 | 24 | 37.9 | 27.5 |
| x | 33 | 40 | 36 | 60 | 55 | 80 | 95 | 70 | 48 | 53 | 95 | 75 | 63 | 112 | 70 |

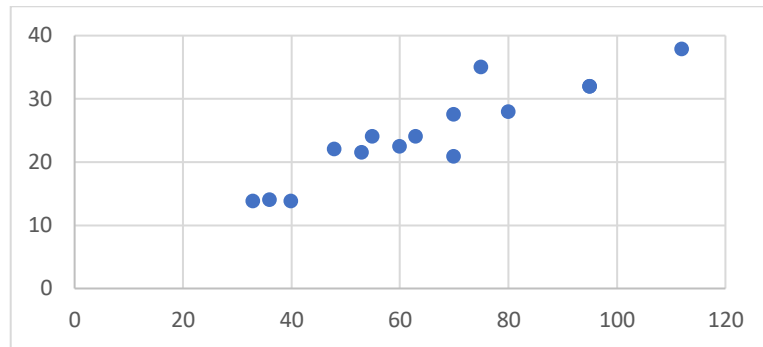
x - общая площадь квартиры в кв. м;

y — рыночная стоимость квартиры в тыс. у.е.

Требуется:

1. Построить график зависимости между переменными, по которому необходимо подобрать модель регрессии.
2. Рассчитать параметры (коэффициенты) уравнения методом наименьших квадратов.
3. С учетом вычисленных параметров построить модель, описывающую зависимость стоимости квартиры от ее площади.
4. Найдите рыночную стоимость квартиры, площадью 150 кв м.

1. Построить график зависимости между переменными, по которому необходимо подобрать модель регрессии:



По графику можно предположить наличие линейной зависимости.

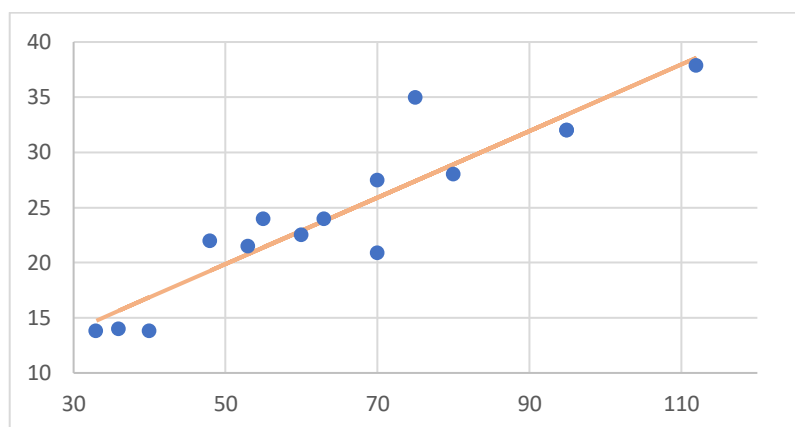
2. Рассчитать параметры (коэффициентов) уравнения методом наименьших квадратов:

$$b_1 = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{1764,4467 - 65,6667 * 24,5933}{4807,4 - 65,6667^2} = 0,30181$$

$$b_0 = \bar{y} - a\bar{x} = 24,5933 - 0,30181 * 65,6667 = 4,7743$$

$$y' = 0,30181x + 4,7743$$

3. С учетом вычисленных параметров построить модель, описывающую зависимость стоимости квартиры от ее площади:



4. Найти рыночную стоимость квартиры, площадью 150 кв. м:

$$y' = 0,30181 * 150 + 4,7743 = 50,0462$$

5. Критерии для оценки полученной регрессионной модели:

1) средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \tilde{y}_i}{y_i} * 100\%,$$

где n — количество измерений;

y_i — заданные значения;

\tilde{y}_i — рассчитанные значения;

$$\bar{A} = 9,0310\%$$

2) коэффициент эластичности:

$$\Xi = a \frac{\bar{x}}{\bar{y}} = 0,81$$

3) коэффициент корреляции:

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2},$$

где σ_x — среднеквадратическое отклонение по x ;

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2},$$

где σ_y — среднеквадратическое отклонение по y ;

$$r = \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} = \frac{1764,4467 - 65,6667 * 24,5933}{22,2550868 * 7,28180075} = 0,92241832$$

4) коэффициент детерминации:

$$r^2 = r^2 * 100\% = 85,09\%$$

5) t — критерий Стьюдента:

Т.к. исходные данные выборочные, то необходимо оценить значимость величины коэффициента корреляции.

Выдвигаем гипотезу H_0 :

коэффициент корреляции генеральной совокупности равен 0, и общая площадь квартиры не оказывает существенного влияния на рыночную стоимость квартиры.

Тогда

$$H_0: r = 0$$

$$H_1: r \neq 0$$

t — критерий Стьюдента применяется для проверки гипотезы H_0 .

Расчетное значение t — критерия Стьюдента:

$$t_p = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}},$$

где $(n - 2)$ – количество степеней свободы;

$$t_p = 8,611841789$$

Критическое значение t находится из таблицы “ t – распределение Стьюдента ” (уровень значимости $\alpha = 0,05$):

$$t_{кр} = 2,16$$

Т.к. $t_p = 8,611841789 > t_{кр} = 2,16$, то H_0 опровергается, а следовательно верна гипотеза H_1 .

6. Выводы:

Коэффициент регрессии показывает, что при увеличении общей площади квартиры на 1% ее рыночная стоимость в среднем увеличивается на 0,30181.

Т.к. $8\% < \bar{A} = 9,0310\% < 10\%$, то уравнение регрессии считается качественным.

Коэффициент эластичности показывает, что в среднем при увеличении общей площади квартиры на 1% ее рыночная стоимость в среднем возрастает на 0,81%.

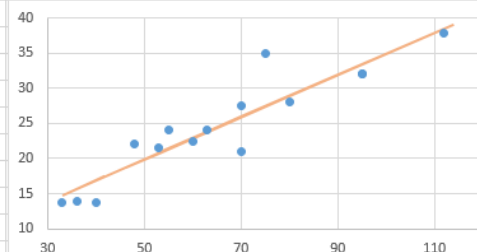
Т.к. $0,9 < r = 0,92241832 < 1$, то взаимосвязь между признаками высокая.

Из значения $r^2 = 85,09\%$ можно сделать вывод, что рыночная стоимость квартиры на 85,09% напрямую зависит от ее общей площади и при этом рыночная стоимость квартиры на 14,91% зависит от других факторов.

Т.к. подтвердилась гипотеза H_1 , значит общая площадь квартиры оказывает существенное влияние на ее рыночную стоимость.

7. Результаты:

| № | x | y | xy | x ² | y ² | y' | A' | | ср(xy) | 1764,45 | | A | 9,0310% |
|---------|---------|---------|---------|----------------|----------------|---------|---------|--|----------------------|---------|--|----------------|------------|
| 1 | 33 | 13,8 | 455,4 | 1089 | 190,44 | 14,7341 | 0,06769 | | ср(x)*ср(y) | 1614,96 | | Э | 81% |
| 2 | 40 | 13,8 | 552 | 1600 | 190,44 | 16,8468 | 0,22078 | | ср(x ²) | 4807,4 | | | |
| 3 | 36 | 14 | 504 | 1296 | 196 | 15,6396 | 0,11711 | | (ср(x)) ² | 4312,11 | | ср кв откл X | 22,2550868 |
| 4 | 60 | 22,5 | 1350 | 3600 | 506,25 | 22,8831 | 0,01702 | | b1 | 0,30181 | | ср кв откл Y | 7,28180075 |
| 5 | 55 | 24 | 1320 | 3025 | 576 | 21,374 | 0,10942 | | | | | r | 0,92241832 |
| 6 | 80 | 28 | 2240 | 6400 | 784 | 28,9193 | 0,03283 | | ср(x) | 65,6667 | | r ² | 85,09% |
| 7 | 95 | 32 | 3040 | 9025 | 1024 | 33,4465 | 0,0452 | | ср(y) | 24,5933 | | t p | 8,61184179 |
| 8 | 70 | 20,9 | 1463 | 4900 | 436,81 | 25,9012 | 0,23929 | | b0 | 4,7743 | | t кр | 2,16 |
| 9 | 48 | 22 | 1056 | 2304 | 484 | 19,2613 | 0,12449 | | | | | | |
| 10 | 53 | 21,5 | 1139,5 | 2809 | 462,25 | 20,7704 | 0,03394 | | | | | | |
| 11 | 95 | 32 | 3040 | 9025 | 1024 | 33,4465 | 0,0452 | | | | | | |
| 12 | 75 | 35 | 2625 | 5625 | 1225 | 27,4103 | 0,21685 | | | | | | |
| 13 | 63 | 24 | 1512 | 3969 | 576 | 23,7885 | 0,00881 | | | | | | |
| 14 | 112 | 37,9 | 4244,8 | 12544 | 1436,41 | 38,5773 | 0,01787 | | | | | | |
| 15 | 70 | 27,5 | 1925 | 4900 | 756,25 | 25,9012 | 0,05814 | | | | | | |
| Итого | 985 | 368,9 | 26466,7 | 72111 | 9867,85 | 368,9 | 1,35465 | | | | | | |
| Среднее | 65,6667 | 24,5933 | 1764,45 | 4807,4 | 657,857 | 24,5933 | 0,01935 | | | | | | |
| | 150 | 50,0462 | | | | | | | | | | | |



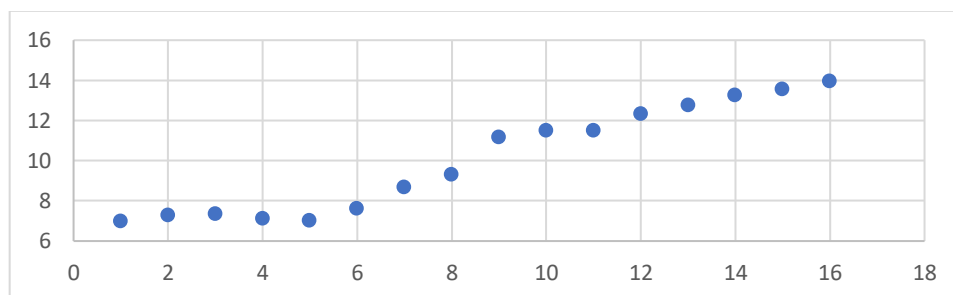
№3

Постановка задачи:

Задача 3. загрязнение региона в баллах было определено за последние 16 месяцев. Данные о загрязнении приведены в таблице. Разработать трендовую модель для определения будущей экологической обстановки в районе в регионе в последние пять месяцев.

| месяцы | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 |
|----------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| загрязне | 6.9 | 7.2 | 7.3 | 7.1 | 6.9 | 7.6 | 8.6 | 9.2 | 11. | 11. | 11. | 12. | 12. | 13. | 13. | 13. |
| ние | 6 | 7 | 3 | 1 | 9 | 0 | 6 | 8 | 15 | 48 | 49 | 33 | 74 | 26 | 54 | 95 |

1. Построить график зависимости между переменными, по которому необходимо подобрать модель регрессии:



По графику можно предположить наличие линейной зависимости.

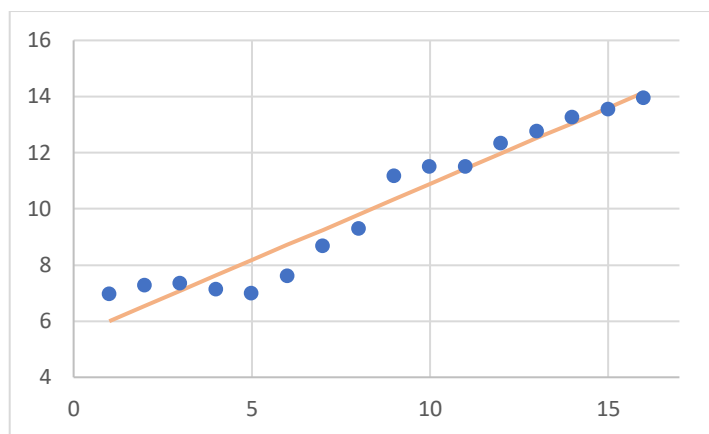
2. Рассчитать параметры (коэффициентов) уравнения методом наименьших квадратов:

$$b_1 = \frac{\overline{xy} - \bar{x} \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{97,15 - 8,5 * 10,0713}{93,5 - 8,5^2} = 0,54326$$

$$b_0 = \bar{y} - a\bar{x} = 10,0713 - 0,54326 * 8,5 = 5,4535$$

$$y' = 0,54326x + 5,4535$$

3. С учетом вычисленных параметров построить модель, описывающую зависимость загрязнения в зависимости от месяца:



4. Критерии для оценки полученной регрессионной модели:

- 1) средняя ошибка аппроксимации:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \tilde{y}_i}{y_i} * 100\%,$$

где n — количество измерений;

y_i — заданные значения;

\tilde{y}_i — рассчитанные значения;

$$\bar{A} = 6,2185\%$$

- 2) коэффициент эластичности:

$$\varepsilon = a \frac{\bar{x}}{\bar{y}} = 0,46$$

- 3) коэффициент корреляции:

$$\sigma_x = \sqrt{\overline{x^2} - \bar{x}^2},$$

где σ_x – среднее квадратическое отклонение по x ;

$$\sigma_y = \sqrt{\overline{y^2} - \bar{y}^2},$$

где σ_y – среднее квадратическое отклонение по y ;

$$r = \frac{\overline{xy} - \bar{x} \bar{y}}{\sigma_x \sigma_y} = \frac{97,15 - 8,5 * 10,0713}{4,60977 * 2,5819} = 0,969953748$$

4) коэффициент детерминации:

$$r^2 = r^2 * 100\% = 94,08\%$$

5) t – критерий Стьюдента:

Т.к. исходные данные выборочные, то необходимо оценить значимость величины коэффициента корреляции.

Выдвигаем гипотезу H_0 :

коэффициент корреляции генеральной совокупности равен 0, и течение времени не оказывает существенного влияния на загрязнение.

Тогда

$$H_0: r = 0$$

$$H_1: r \neq 0$$

t – критерий Стьюдента применяется для проверки гипотезы H_0 .

Расчетное значение t – критерия Стьюдента:

$$t_p = \frac{|r|}{\sqrt{\frac{1 - r^2}{n - 2}}},$$

где $(n - 2)$ – количество степеней свободы;

$$t_p = 14,9174$$

Критическое значение t находится из таблицы “ t – распределение Стьюдента” (уровень значимости $\alpha = 0,05$):

$$t_{кр} = 2,14$$

Т.к. $t_p = 14,9174 > t_{кр} = 2,14$, то H_0 опровергается, а следовательно верна гипотеза H_1 .

6) Выводы:

Коэффициент регрессии показывает, что изменение времени на 1% загрязнение в среднем увеличивается на 0,54326.

Т.к. $\bar{A} = 6,2185\% < 10\%$, то уравнение регрессии считается качественным.

Коэффициент эластичности показывает, что в среднем при изменении времени на 1% загрязнение увеличивается в среднем на 46%.

Т.к. $0,9 < r = 0,969953748 < 1$, то взаимосвязь между признаками высокая.

Из значения $r^2 = 94,08\%$ можно сделать вывод, что загрязнение на 94,08% напрямую зависит от течения времени и загрязнение только на 5,92% зависит от других факторов.

Т.к. подтвердилась гипотеза H_1 , значит течение времени оказывает существенное влияние на загрязнение.

5. Результаты:

| № | x | y | xy | x ² | y ² | y' | A' | ср(xy) | 97,15 | A | 6,2185% |
|---------|-----|---------|--------|----------------|----------------|---------|---------|----------------------|---------|----------------|---------|
| 1 | 1 | 6,96 | 6,96 | 1 | 48,4416 | 5,99676 | 0,1384 | ср(x)*ср(y) | 85,6056 | Э | 46% |
| 2 | 2 | 7,27 | 14,54 | 4 | 52,8529 | 6,54003 | 0,10041 | ср(x ²) | 93,5 | | |
| 3 | 3 | 7,33 | 21,99 | 9 | 53,7289 | 7,08329 | 0,03366 | (ср(x)) ² | 72,25 | ср кв откл X | 4,60977 |
| 4 | 4 | 7,11 | 28,44 | 16 | 50,5521 | 7,62656 | 0,07265 | b1 | 0,54326 | ср кв откл Y | 2,5819 |
| 5 | 5 | 6,99 | 34,95 | 25 | 48,8601 | 8,16982 | 0,16879 | | | r | 0,96995 |
| 6 | 6 | 7,6 | 45,6 | 36 | 57,76 | 8,71309 | 0,14646 | ср(x) | 8,5 | r ² | 94,08% |
| 7 | 7 | 8,66 | 60,62 | 49 | 74,9956 | 9,25635 | 0,06886 | ср(y) | 10,0713 | t p | 14,9174 |
| 8 | 8 | 9,28 | 74,24 | 64 | 86,1184 | 9,79962 | 0,05599 | b0 | 5,4535 | t кр | 2,14 |
| 9 | 9 | 11,15 | 100,35 | 81 | 124,323 | 10,3429 | 0,07239 | | | | |
| 10 | 10 | 11,48 | 114,8 | 100 | 131,79 | 10,8861 | 0,05173 | | | | |
| 11 | 11 | 11,49 | 126,39 | 121 | 132,02 | 11,4294 | 0,00527 | | | | |
| 12 | 12 | 12,33 | 147,96 | 144 | 152,029 | 11,9727 | 0,02898 | | | | |
| 13 | 13 | 12,74 | 165,62 | 169 | 162,308 | 12,5159 | 0,01759 | | | | |
| 14 | 14 | 13,26 | 185,64 | 196 | 175,828 | 13,0592 | 0,01514 | | | | |
| 15 | 15 | 13,54 | 203,1 | 225 | 183,332 | 13,6025 | 0,00461 | | | | |
| 16 | 16 | 13,95 | 223,2 | 256 | 194,603 | 14,1457 | 0,01403 | | | | |
| Итого | 136 | 161,14 | 1554,4 | 1496 | 1729,54 | 161,14 | 0,99496 | | | | |
| Среднее | 8,5 | 10,0713 | 97,15 | 93,5 | 108,096 | 10,7427 | 0,06219 | | | | |

