**Задание 1:** Простая линейная регрессия по сгруппированным данным

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -0,15 | 0,67 | 1,49 | 2,31 | 3,13 | 3,95 | 4,77 | 5,59 | 6,41 | 7,23 | 8,05 | 8,87 |
| -7,09 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 |
| -5,91 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 |
| -4,73 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 5 | 2 | 0 |
| -3,55 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 8 | 2 | 0 | 0 |
| -2,37 | 0 | 0 | 0 | 0 | 0 | 1 | 8 | 11 | 4 | 0 | 0 | 0 |
| -1,19 | 0 | 0 | 0 | 0 | 0 | 7 | 10 | 6 | 0 | 0 | 0 | 0 |
| -0,01 | 0 | 0 | 0 | 0 | 10 | 19 | 13 | 1 | 0 | 0 | 0 | 0 |
| 1,17 | 0 | 0 | 0 | 0 | 8 | 7 | 4 | 0 | 0 | 0 | 0 | 0 |
| 2,35 | 0 | 0 | 0 | 9 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3,53 | 0 | 0 | 7 | 6 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4,71 | 0 | 4 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5,89 | 1 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

Найти уравнение регрессии y на x и x на y.

Решение.

Выборочные средние - это величины, которые определяются по формулам:

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

Вычислим все необходимые суммы:

В результате получаем выборочные средние:

Дисперсией называется средняя арифметическая квадратов отклонений вариантов от их средней арифметической.

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

Вычислим все необходимые суммы:

В результате получаем выборочные дисперсии:

Выборочный коэффициент корреляции - это показатель тесноты линейной связи между X и Y, который определяется по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

Вычислим все необходимые суммы:

В результате получаем выборочный коэффициент корреляции:

Связь между переменными Y и X можно описать линейным уравнением, которое называется линейным уравнением регрессии. Оно определяется по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

Аналогично определяется уравнение регрессии X по Y:

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

Коэффициент регрессии показывает, на сколько единиц в среднем изменяется переменная Y при увеличении переменной X на одну единицу.

Коэффициент регрессии показывает, на сколько единиц в среднем изменяется переменная X при увеличении переменной Y на одну единицу.

Найдем значения коэффициентов регрессии по формулам:

|  |  |  |
| --- | --- | --- |
|  |  | (6) |
|  |  | (7) |

В результате получаем:

По формулам (4) - (5) находим уравнения линейной регрессии:

Для каждого значения т.е. для каждой строки корреляционной таблицы вычислим групповые средние

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

где частоты пар и число интервалов по переменной Y.

Вычисленные групповые средние поместим в последнем столбце корреляционной таблицы и изобразим графически в виде ломаной, называемой эмпирической линией регрессии Y по X.

Аналогично для каждого значения по формуле

|  |  |  |
| --- | --- | --- |
|  |  | (9) |

вычислим групповые средние , где число интервалов по переменной X.

Вычисленные групповые средние поместим в последнюю строку корреляционной таблицы 1.

и т.д.

Аналогично и для X:

и т.д.

Поместим все необходимые данные в таблицу 1.

Таблица 1

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | -0,15 | 0,67 | 1,49 | 2,31 | 3,13 | 3,95 | 4,77 | 5,59 | 6,41 | 7,23 | 8,05 | 8,87 | Всего, |  |
| -7,09 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 0 | 3 | 7,50 |
| -5,91 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1 | 1 | 4 | 7,85 |
| -4,73 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 5 | 2 | 0 | 11 | 7,08 |
| -3,55 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 8 | 2 | 0 | 0 | 16 | 6,21 |
| -2,37 | 0 | 0 | 0 | 0 | 0 | 1 | 8 | 11 | 4 | 0 | 0 | 0 | 24 | 5,39 |
| -1,19 | 0 | 0 | 0 | 0 | 0 | 7 | 10 | 6 | 0 | 0 | 0 | 0 | 23 | 4,73 |
| -0,01 | 0 | 0 | 0 | 0 | 10 | 19 | 13 | 1 | 0 | 0 | 0 | 0 | 43 | 4,05 |
| 1,17 | 0 | 0 | 0 | 0 | 8 | 7 | 4 | 0 | 0 | 0 | 0 | 0 | 19 | 3,78 |
| 2,35 | 0 | 0 | 0 | 9 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 19 | 2,74 |
| 3,53 | 0 | 0 | 7 | 6 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 20 | 2,31 |
| 4,71 | 0 | 4 | 2 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 | 1,49 |
| 5,89 | 1 | 2 | 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 | 0,94 |
| Всего, | 1 | 6 | 12 | 19 | 35 | 34 | 35 | 24 | 16 | 11 | 4 | 1 | 198 |  |
|  | 5,89 | 5,10 | 4,32 | 3,22 | 1,64 | -0,08 | -0,75 | -2,27 | -3,55 | -5,16 | -5,62 | -5,91 |  |  |

Изобразим полученную зависимость графически точками координатной плоскости. Такая зависимость называется полем корреляции.

Построим групповые средние и прямые линии регрессии на поле корреляции.

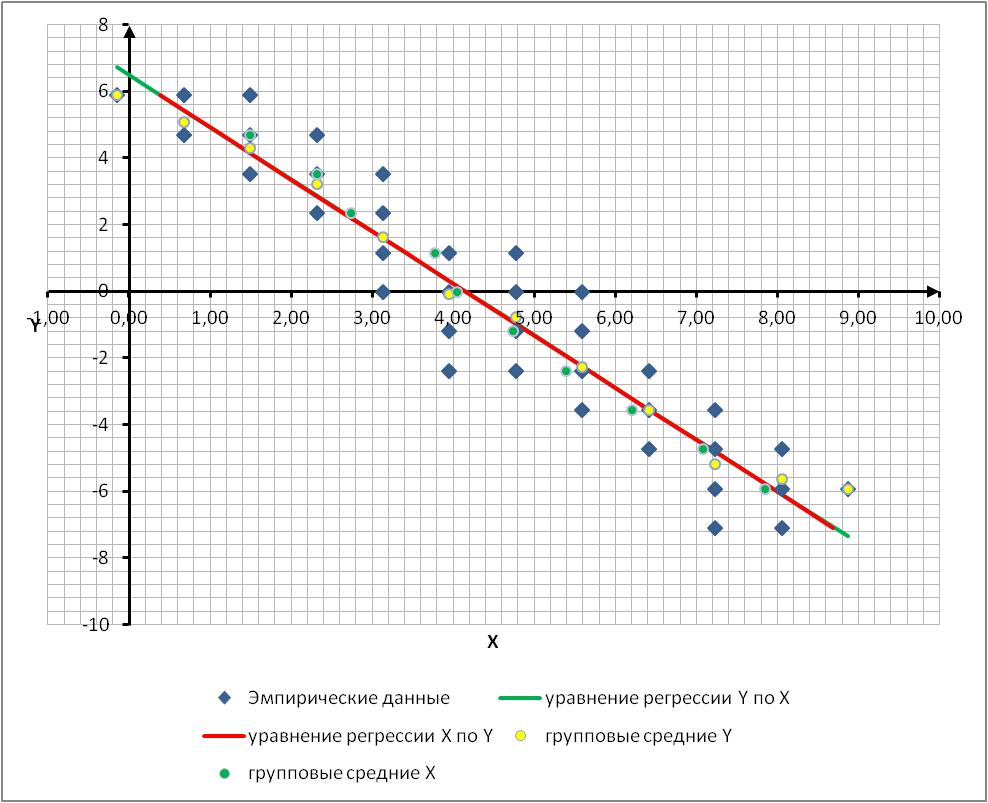


Рисунок 1. Поле корреляции.

Оценим статистическую значимость параметров регрессии с помощью критерия Стьюдента. Оценку статистической значимости параметров регрессии проведем с помощью t-статистики Стьюдента и путем расчета доверительного интервала каждого из показателей.

Выдвигаем гипотезу о статистической незначимости коэффициента корреляции.

Гипотеза отвергается, т.е. выборочный коэффициент корреляции значимо отличается от нуля, если

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

где табличное значение t-критерия Стьюдента, определяемое при уровне значимости при числе степеней свободы

Статистика критерия по формуле (10) составит:

Для уровня значимости и числа степеней свободы находим критическое значение статистики Поскольку коэффициент корреляции между Y и X значимо отличается от нуля.

Для значимого коэффициента корреляции целесообразно найти доверительный интервал (интервальную оценку), который с заданной надежностью содержит неизвестный генеральный коэффициент корреляции . Для построения такого интервала необходимо знать выборочное распределение коэффициента корреляции , которое при несимметрично и очень медленно (с ростом n) сходится к нормальному распределению. Поэтому прибегают к специально подобранным функциям от r, которые сходятся к хорошо изученным распределениям. Чаще всего для подбора функции применяют z-преобразование Фишера:

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

Так как коэффициент корреляции значим, то построим доверительный интервал для генерального коэффициента корреляции , применяя z - преобразование Фишера.

По формуле (11) получаем:

Распределение z является нормальным с математическим ожиданием

|  |  |  |
| --- | --- | --- |
|  |  | (12) |

и дисперсией

|  |  |  |
| --- | --- | --- |
|  |  | (13) |

Сначала строим доверительный интервал для

|  |  |  |
| --- | --- | --- |
|  |  | (14) |

где нормированное отклонение z, определяемое с помощью функции Лапласа:

|  |  |  |
| --- | --- | --- |
|  |  | (15) |

По формуле (15) из условия по таблице функции Лапласа находим

По формуле (14) построим доверительный интервал для M(z).

или

При определении границ доверительного интервала для т.е. для перехода от z к , существует специальная таблица. При ее отсутствии переход может быть осуществлен по формуле:

|  |  |  |
| --- | --- | --- |
|  |  | (16) |

где гиперболический тангенс z.

По формуле (16) находим границы доверительного интервала для

Если коэффициент корреляции значим, то коэффициенты регрессии и также значимо отличаются от нуля, а интервальные оценки для генеральных коэффициентов регрессии и могут быть получены по формулам, основанным на том, что статистики имеют t-распределение Стьюдента с n-2 степенями свободы:

|  |  |
| --- | --- |
|  | (17) |
|  | (18) |

По формулам (17) - (18) находим доверительные интервалы для и :

|  |
| --- |
|  |
|  |

Оценим статистическую значимость полученного уравнения регрессии . Выдвигаем гипотезу о статистически незначимом отличии показателей от нуля:

для числа степеней свободы и составит

Определим случайные ошибки

Находим фактические значения t-статистики:

и

поэтому гипотеза не принимается, т.е. параметры уравнения регрессии и статистически значимы.

Аналогично оценим статистическую значимость полученного уравнения регрессии . Выдвигаем гипотезу о статистически незначимом отличии показателей от нуля:

для числа степеней свободы и составит

Определим случайные ошибки

Находим фактические значения t-статистики:

и

поэтому гипотеза не принимается, т.е. параметры уравнения регрессии и статистически значимы.

Для построения гистограммы по оси абсцисс указывают значения границ интервалов и на их основании строят прямоугольники, высота которых пропорциональна частотам.

Для направляющих X и Y построим гистограммы.

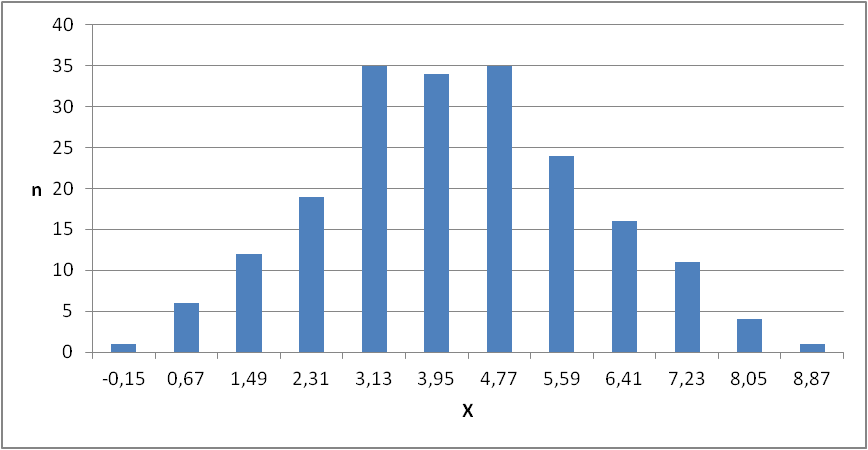


Рисунок 2. Гистограмма распределения X.

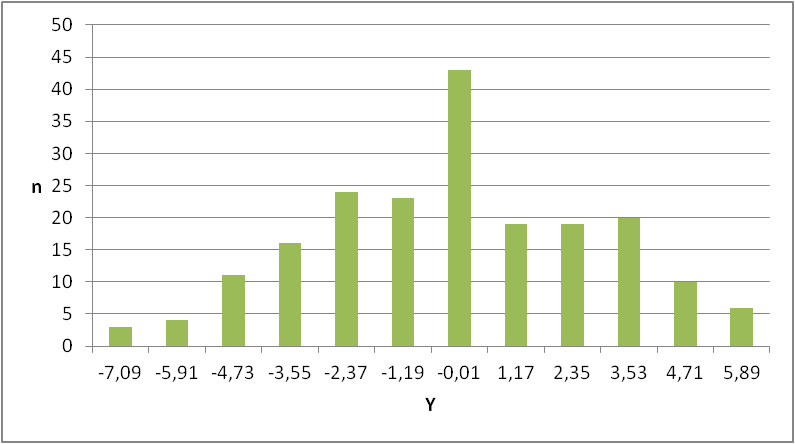


Рисунок 3. Гистограмма распределения Y.

По виду гистограммы распределения (рис. 2) можно предположить нормальный

закон распределения признака. Итак, выдвигаем гипотезу случайная величина X распределена нормально с параметрами

Проверку гипотезы осуществляем с помощью критерия Пирсона.

Для расчета вероятностей попадания величины X в интервал используем формулу Лапласа в соответствии со свойством нормального распределения:

Для определения статистики удобно составить таблицу:

Таблица 2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| i |  | Эмпирические частоты, | Вероятности, | Теоретические частоты, |  |  |
| 1 | -0,15 | 1 |  |  |  |  |
| 2 | 0,67 | 6 | 0,016 | 3,07 | 8,563 | 2,786 |
| 3 | 1,49 | 12 | 0,039 | 7,72 | 18,344 | 2,377 |
| 4 | 2,31 | 19 | 0,079 | 15,65 | 11,204 | 0,716 |
| 5 | 3,13 | 35 | 0,130 | 25,65 | 87,385 | 3,407 |
| 6 | 3,95 | 34 | 0,172 | 33,97 | 0,001 | 0,000 |
| 7 | 4,77 | 35 | 0,184 | 36,34 | 1,802 | 0,050 |
| 8 | 5,59 | 24 | 0,159 | 31,42 | 55,037 | 1,752 |
| 9 | 6,41 | 16 | 0,111 | 21,95 | 35,367 | 1,611 |
| 10 | 7,23 | 11 | 0,063 | 12,39 | 1,924 | 0,155 |
| 11 | 8,05 | 4 | 0,029 | 5,65 | 2,718 | 0,481 |
| 12 | 8,87 | 1 | 0,011 | 2,08 | 1,168 | 0,561 |
|  |  | 198 | 0,99 | 195,89 | 223,513 | 13,896 |

Итак, наблюдаемое значение критерия равно 13.896.

Число интервалов равно , а нормальный закон распределения определяется параметрами, то число степеней свободы равно . Соответствующее критическое значение по таблице распределения равно

Так как , то гипотеза о выбранном теоретическом нормальном законе согласуется с опытными данными.

Аналогично, выдвигаем гипотезу о нормальном распределении величины Y. Итак, выдвигаем гипотезу случайная величина Y распределена нормально с параметрами

Для определения статистики удобно составить таблицу:

Таблица 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| j |  | Эмпирические частоты, | Вероятности, | Теоретические частоты, |  |  |
| 1 | -7,09 | 3 |  |  |  |  |
| 2 | -5,91 | 4 | 0,015 | 3,018 | 0,965 | 0,320 |
| 3 | -4,73 | 11 | 0,033 | 6,610 | 19,276 | 2,916 |
| 4 | -3,55 | 16 | 0,062 | 12,349 | 13,332 | 1,080 |
| 5 | -2,37 | 24 | 0,099 | 19,680 | 18,660 | 0,948 |
| 6 | -1,19 | 23 | 0,135 | 26,755 | 14,102 | 0,527 |
| 7 | -0,01 | 43 | 0,157 | 31,028 | 143,326 | 4,619 |
| 8 | 1,17 | 19 | 0,155 | 30,696 | 136,787 | 4,456 |
| 9 | 2,35 | 19 | 0,131 | 25,904 | 47,667 | 1,840 |
| 10 | 3,53 | 20 | 0,094 | 18,648 | 1,828 | 0,098 |
| 11 | 4,71 | 10 | 0,068 | 13,464 | 11,999 | 0,891 |
| 12 | 5,89 | 6 | 0,040 | 7,920 | 3,686 | 0,465 |
|  |  | 198 | 1,00 | 196,071 | 411,627 | 18,161 |

Итак, наблюдаемое значение критерия равно 18.161

Число интервалов равно , а нормальный закон распределения определяется параметрами, то число степеней свободы равно . Соответствующее критическое значение по таблице распределения равно

Так как , то гипотеза о выбранном теоретическом нормальном законе согласуется с опытными данными.

**Задание 2:** Зависимость вида

Найти уравнение регрессии

Таблица 4

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 0,2 | 0,3 | 0,4 | 0,5 | 0,6 | 0,7 | 0,8 | 0,9 | 1 | 1,1 |
| y | 5,74 | 4,03 | 6,31 | 4,96 | 7,33 | 6,08 | 8,51 | 7,83 | 10,32 | 9,18 |
| x | 1,2 | 1,3 | 1,4 | 1,5 | 1,6 | 1,7 | 1,8 | 1,9 | 2 | 2,1 |
| y | 12,05 | 11,37 | 14,46 | 14,15 | 17,2 | 16,99 | 20,47 | 20,38 | 23,72 | 24,59 |

Построению модели вида предшествует процедура линеаризации переменных. Линеаризация производится путем логарифмирования обеих частей уравнения:

где

Для расчета параметров a и b нелинейной регрессии решаем систему нормальных уравнений относительно a и b:

Для расчетов используем данные таблицы 5.

Таблица 5

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| n | x | y | X | Y |  | YX |
| 1 | 0,20 | 5,74 | 0,2 | 1,747459 | 0,04 | 0,349492 |
| 2 | 0,30 | 4,03 | 0,3 | 1,393766 | 0,09 | 0,41813 |
| 3 | 0,40 | 6,31 | 0,4 | 1,842136 | 0,16 | 0,736854 |
| 4 | 0,50 | 4,96 | 0,5 | 1,601406 | 0,25 | 0,800703 |
| 5 | 0,60 | 7,33 | 0,6 | 1,991976 | 0,36 | 1,195185 |
| 6 | 0,70 | 6,08 | 0,7 | 1,805005 | 0,49 | 1,263503 |
| 7 | 0,80 | 8,51 | 0,8 | 2,141242 | 0,64 | 1,712994 |
| 8 | 0,90 | 7,83 | 0,9 | 2,057963 | 0,81 | 1,852166 |
| 9 | 1,00 | 10,32 | 1 | 2,334084 | 1 | 2,334084 |
| 10 | 1,10 | 9,18 | 1,1 | 2,217027 | 1,21 | 2,43873 |
| 11 | 1,20 | 12,05 | 1,2 | 2,489065 | 1,44 | 2,986878 |
| 12 | 1,30 | 11,37 | 1,3 | 2,430978 | 1,69 | 3,160272 |
| 13 | 1,40 | 14,46 | 1,4 | 2,671386 | 1,96 | 3,739941 |
| 14 | 1,50 | 14,15 | 1,5 | 2,649715 | 2,25 | 3,974572 |
| 15 | 1,60 | 17,2 | 1,6 | 2,844909 | 2,56 | 4,551855 |
| 16 | 1,70 | 16,99 | 1,7 | 2,832625 | 2,89 | 4,815462 |
| 17 | 1,80 | 20,47 | 1,8 | 3,01896 | 3,24 | 5,434129 |
| 18 | 1,90 | 20,38 | 1,9 | 3,014554 | 3,61 | 5,727653 |
| 19 | 2,00 | 23,72 | 2 | 3,166319 | 4 | 6,332637 |
| 20 | 2,10 | 24,59 | 2,1 | 3,20234 | 4,41 | 6,724914 |
| Сумма | 23 | 245,67 | 23 | 47,45291 | 33,1 | 60,55015 |

Получаем систему уравнений:

Получено линейное уравнение:

Произведем потенцирование полученного уравнения и запишем его в обычной форме:

Ответ:

**Задание 3**: Зависимость вида

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 1,05 | 1,1 | 1,15 | 1,2 | 1,25 | 1,3 | 1,35 | 1,4 | 1,45 | 1,5 |
| y | 11 | 10,99 | 10,96 | 10,86 | 10,82 | 10,75 | 10,64 | 10,54 | 10,4 | 10,25 |
| x | 1,55 | 1,6 | 1,65 | 1,7 | 1,75 | 1,8 | 1,85 | 1,9 | 1,95 | 2 |
| y | 10,1 | 9,94 | 9,74 | 9,54 | 9,32 | 9,1 | 8,84 | 8,58 | 8,32 | 8 |

Найти уравнение регрессии

Решение.

Для расчета параметров нелинейной регрессии решаем систему нормальных уравнений:

Для расчетов используем данные таблицы 6.

Таблица 6

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| n | x | y |  |  |  |  |  |
| 1 | 1,05 | 1,01 | 1,1025 | 1,157625 | 1,215506 | 1,0605 | 1,113525 |
| 2 | 1,1 | 1,04 | 1,21 | 1,331 | 1,4641 | 1,144 | 1,2584 |
| 3 | 1,15 | 1,07 | 1,3225 | 1,520875 | 1,749006 | 1,2305 | 1,415075 |
| 4 | 1,2 | 1,14 | 1,44 | 1,728 | 2,0736 | 1,368 | 1,6416 |
| 5 | 1,25 | 1,19 | 1,5625 | 1,953125 | 2,441406 | 1,4875 | 1,859375 |
| 6 | 1,3 | 1,29 | 1,69 | 2,197 | 2,8561 | 1,677 | 2,1801 |
| 7 | 1,35 | 1,38 | 1,8225 | 2,460375 | 3,321506 | 1,863 | 2,51505 |
| 8 | 1,4 | 1,49 | 1,96 | 2,744 | 3,8416 | 2,086 | 2,9204 |
| 9 | 1,45 | 1,62 | 2,1025 | 3,048625 | 4,420506 | 2,349 | 3,40605 |
| 10 | 1,5 | 1,77 | 2,25 | 3,375 | 5,0625 | 2,655 | 3,9825 |
| 11 | 1,55 | 1,91 | 2,4025 | 3,723875 | 5,772006 | 2,9605 | 4,588775 |
| 12 | 1,6 | 2,1 | 2,56 | 4,096 | 6,5536 | 3,36 | 5,376 |
| 13 | 1,65 | 2,28 | 2,7225 | 4,492125 | 7,412006 | 3,762 | 6,2073 |
| 14 | 1,7 | 2,49 | 2,89 | 4,913 | 8,3521 | 4,233 | 7,1961 |
| 15 | 1,75 | 2,7 | 3,0625 | 5,359375 | 9,378906 | 4,725 | 8,26875 |
| 16 | 1,8 | 2,93 | 3,24 | 5,832 | 10,4976 | 5,274 | 9,4932 |
| 17 | 1,85 | 3,18 | 3,4225 | 6,331625 | 11,71351 | 5,883 | 10,88355 |
| 18 | 1,9 | 3,44 | 3,61 | 6,859 | 13,0321 | 6,536 | 12,4184 |
| 19 | 1,95 | 3,72 | 3,8025 | 7,414875 | 14,45901 | 7,254 | 14,1453 |
| 20 | 2 | 4,02 | 4 | 8 | 16 | 8,04 | 16,08 |
| Сумма | 30,5 | 41,77 | 48,175 | 78,5375 | 131,6167 | 68,948 | 116,9495 |

Получаем систему уравнений:

Решаем систему по формулам Крамера:

Рассчитываем

Получаем уравнение регрессии

Ответ: