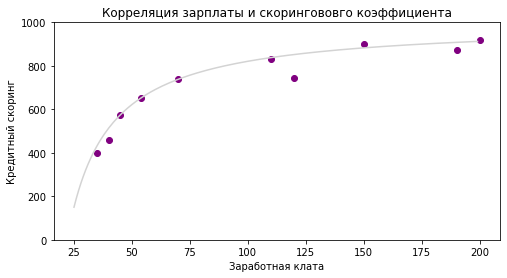
1. Даны значения величины заработной платы заемщиков банка (zp) и значения их поведенческого кредитного скоринга (ks): zp = [35, 45, 190, 200, 40, 70, 54, 150, 120, 110], ks = [401, 574, 874, 919, 459, 739, 653, 902, 746, 832]. Используя математические операции, посчитать коэффициенты линейной регрессии, приняв за X заработную плату (то есть, zp - признак), а за y - значения скорингового балла (то есть, ks - целевая переменная). Произвести расчет как с использованием intercept, так и без.

Для более адекватного анализа сперва построю график распределения и попробую сразу оценить зависимость, хотя точек и не очень много.



Проведя несколько «визуальных» экспериментов и ручную подгонку коэффициентов, нашел аналитическую функцию, которая на отрезке [40; 200] похожа на это распределение. Эта функция



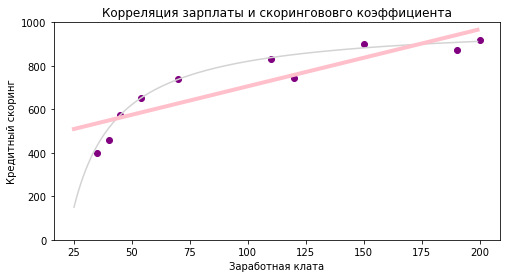
Но в задаче необходимо рассчитать коэффициенты линейной регрессии

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **№** |  |  |  |  |  |
| 1 | 35 | 401 | 14 035 | 4 408,96 | 95 419,21 |
| 2 | 40 | 459 | 18 360 | 3 769,96 | 62 950,81 |
| 3 | 45 | 574 | 25 830 | 3 180,96 | 18 468,81 |
| 4 | 54 | 653 | 35 262 | 2 246,76 | 3 237,61 |
| 5 | 70 | 739 | 51 730 | 985,96 | 846,81 |
| 6 | 110 | 832 | 91 520 | 73,96 | 14 908,41 |
| 7 | 120 | 746 | 89 520 | 345,96 | 1 303,21 |
| 8 | 150 | 902 | 135 300 | 2 361,96 | 36 902,41 |
| 9 | 190 | 874 | 166 060 | 7 849,96 | 26 928,81 |
| 10 | 200 | 919 | 183 800 | 9 721,96 | 43 722,81 |
|  |  |  |  |  |  |
| **Σ** | **1 014** | **7 099** | **811 417** | **34 946,40** | **304 688,90** |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
| **Ср. зн.** | 101,4 | 709,9 | 81 141,7 | 3 494,64 | 30 468,89 |
|  |  |  |  |  |  |
|  |  |  |  | 59,1154802 | 174,5534016 |

Выборочный коэффициент корреляции

Уравнение линейной регрессии имеет вид

= 2,6205389 \* x + 444,17736



Использование инструментов EXEL дает следующий результат:

Функция ОТРЕЗОК(ks; zp) = 444,1774  
Функция НАКЛОН(ks; zp) = 2,620539

Идеально совпадает с рассчитанными значениями.

Ответ: y = 2,6205389 \* x + 444,17736

1. В каких случаях для вычисления доверительных интервалов и проверки статистических гипотез используется таблица значений функции Лапласа, а в каких - таблица критических точек распределения Стьюдента?

Если имеет место нормальный закон распределения и число измерений n достаточно велико, доверительный интервал легко отыскать при помощи удвоенной функции Лапласа при условии, что ИЗВЕСТНА генеральная дисперсия

В том случае, когда случайная величина распределена по нормальному закону, но генеральная дисперсия НЕИЗВЕСТНА, доверительные интервалы можно отыскать при помощи распределения Стьюдента.