ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

Пусть некоторый объект характеризуется двумя признаками Х и Y.

Между X и Y могут существовать различные виды зависимостей:

**Функциональная** – каждому значению признака X соответствует единственное значение признака Y. 

**Статистическая** – каждому значению признака Х соответствует статистическое распределение признака Y ( корреляционная таблица).

**Корреляционная**– частный случай статистической, когда каждому значению признака х соответствует среднее значение признака Y: ,связь между ними достаточно хорошо описывается функцией , называется уравнением регрессии у по х.

Основные задачи теории корреляции:

* 1. Оценить силу (тесноту) связи между признаками X и Y.
  2. Найти вид этой связи в виде уравнения прямой регрессии 

1. Пример 3.1: По данным корреляционной таблицы найти условные средние . Оценить тесноту связи и составить уравнение регрессии у по х. Сделать чертёж.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| х у | 5 | 7 | 9 | 11 | nx |
| 1,8 | - | - | 2 | 5 | 7 |
| 4,4 | - | 5 | 5 | - | 10 |
| 7 | - | 8 | 3 | 2 | 13 |
| 9,6 | 3 | 11 | 1 | 2 | 17 |
| 12,2 | 2 | 1 | - | - | 3 |
| nу | 5 | 25 | 11 | 9 | 50 |

Частота nху показывает сколько раз наблюдалась каждая пара значений.

Например: пара значений (9,6; 7) наблюдалась 11 раз.

В таблице каждому значению х соответствует статистическое распределение признака у.

Например для х=7

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| у | 5 | 7 | 9 | 11 |
| nху | 0 | 8 | 3 | 2 |

Σnxy=13  
Среднее значение у при х=7 или условная средняя 

Аналогично, каждому значению у можно найти статистическое распределение по х.

Найдём условные варианты:

ui=(xi-с1)/h1

νi=(уi-c2)/h2

c1=7, h1=4,4-1,8=2,6 c2=7, h2=7-5=2

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Ui  Vi | -1 | 0 | 1 | 2 | nu | nu∙u |
| -2 | - | - | 2 | 5 | 7 | -14 |
| -1 | - | 5 | 5 | - | 10 | -10 |
| 0 | - | 8 | 3 | 2 | 13 | 0 |
| 1 | 3 | 11 | 1 | 2 | 17 | +17 |
| 2 | 2 | 1 | - | - | 3 | 6 |
| nv | 5 | 25 | 11 | 9 | 50 | ∑nu∙u=-1 |
| nv∙v | -5 | 0 | 11 | 18 | ∑nu∙u=24 | |

Найдём вспомогательные величины





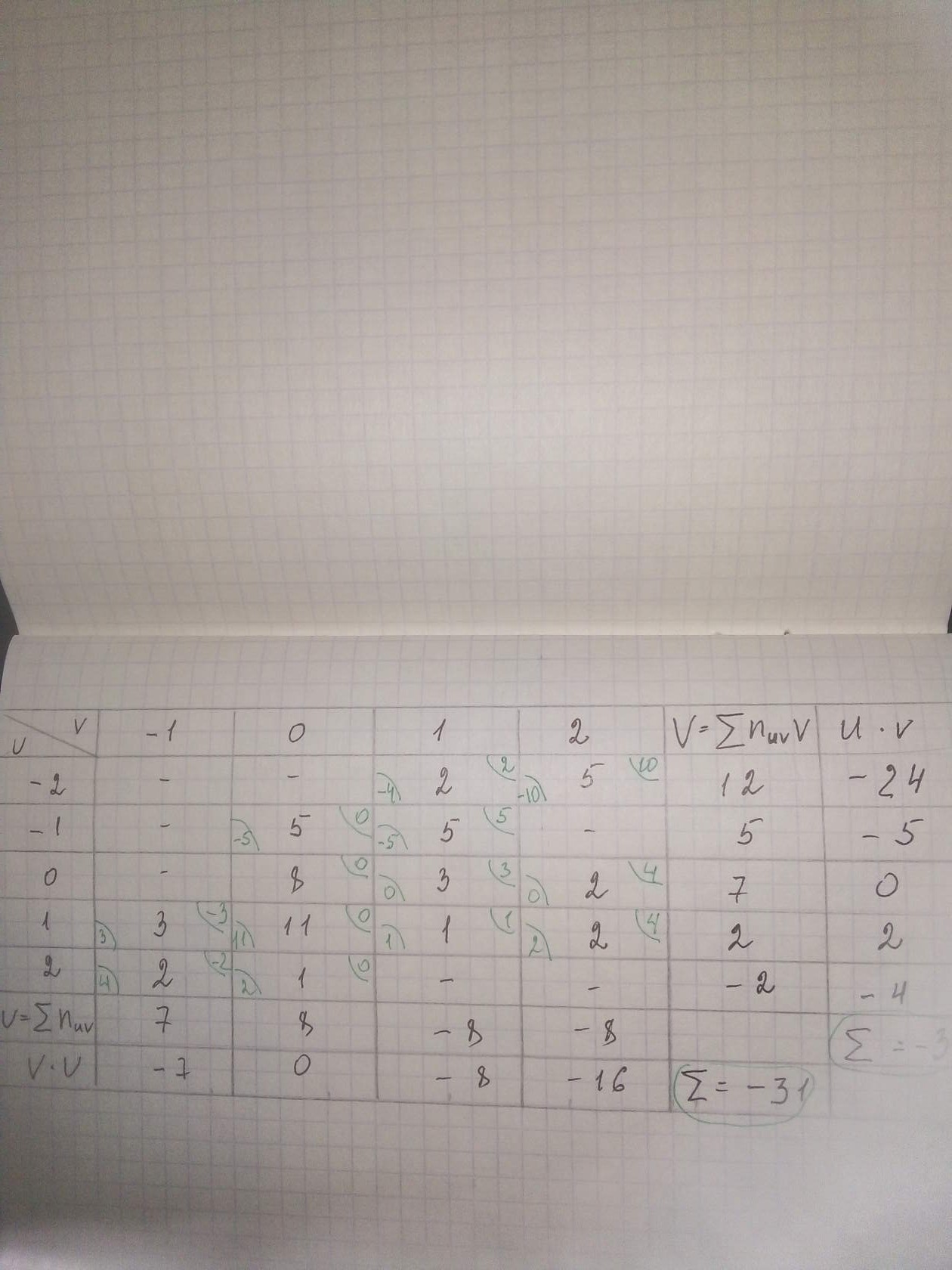




∑nnvu∙v найдём из таблицы:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| u v | -1 | 0 | 1 | 2 | V=∑nnvv | u∙v |
| -2 | - | - | 2 | 5 | 12 | -24 |
| -1 | - | 5 | 5 | - | 5 | -5 |
| 0 | - | 8 | 3 | 2 | 7 | 0 |
| 1 | 3 | 11 | 1 | 2 | 2 | 2 |
| 2 | 2 | 1 | - | - | -2 | -4 |
| u=∑nnvu | 7 | 8 | -8 | -8 |  | ∑=-31 |
| v∙u | -7 | 0 | -8 | -16 | ∑=-31 |  |

∑u∙v=∑v∙u=-31=∑nuvuv



Искомый выборочный коэффициент корреляции



Пояснения к расчётной таблице

1. В каждой строчке, где частота nuv≠0 в правом верхнем углу стоит произведение частоты nuv на варианту v.
2. Складываются все числа, полученные в правых углах одной строки, записываются в столбец v.
3. Варианту u умножают на v.
4. Складывают все числа столбца uv, получают

∑uv=∑nuvuv

Аналогично (для контроля) вычисляют ∑vu=∑nuvuv по столбцам.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Оценка тесноты связи | | | | | | | |
| Значение ׀rв׀ | 0-0,1 | 0,1-0,3 | 0,3-0,5 | 0,5-0,7 | 0,7-0,9 | 0,9-0,99 | 1 |
| Теснота линейной  связи | нет | слабая | умеренная | заметная | высокая | очень высокая | функциональная |



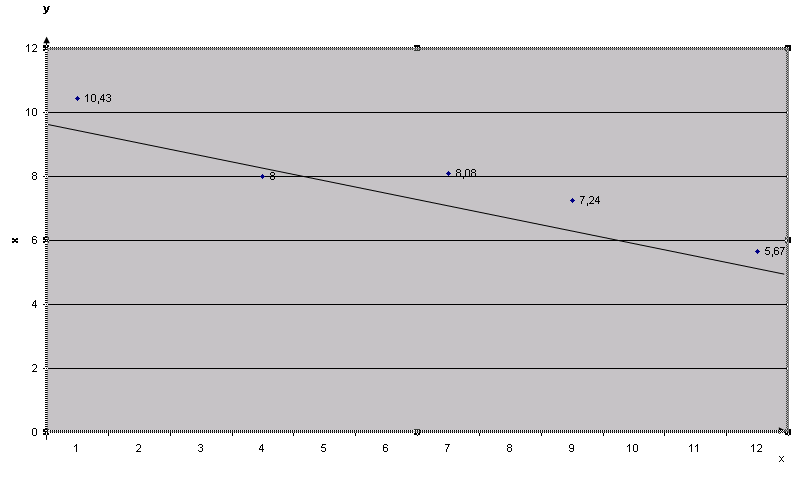
Т.к. rв=-0,58<0, связь обратная, т.е с ростом х убывает y.

׀rв׀Є(0,5-0,7)=>связь заметная.

Строим график функции  и отмечаем условные средние:



Замечание: Если расчёты и построения выполнены верно, то прямая регрессии y по х должна пройти вблизи всех точек , отмеченных \*



**Задания к самостоятельной работе**

По данной таблице найти выборочное уравнение прямой линии регрессии Y на X. Оценить тесноту связи. Построить эмпирическую и теоретическую линии регрессии.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 4 | 9 | 14 | 19 | 24 | 29 | ny |
| 10 | 2 | 3 |  |  |  |  | 5 |
| 20 |  | 7 | 3 |  |  |  | 10 |
| 30 |  |  | 2 | 50 | 2 |  | 54 |
| 40 |  |  | 1 | 10 | 6 |  | 17 |
| 50 |  |  |  | 4 | 7 | 3 | 14 |
| nx | 2 | 10 | 6 | 64 | 15 | 3 | n = 100 |

1)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 10 | 15 | 20 | 25 | 30 | 35 | ny |
| 30 | 2 | 6 |  |  |  |  | 8 |
| 40 |  | 4 | 4 |  |  |  | 8 |
| 50 |  |  | 7 | 35 | 8 |  | 50 |
| 60 |  |  | 2 | 10 | 8 |  | 20 |
| 70 |  |  |  | 5 | 6 | 3 | 14 |
| nx | 2 | 10 | 13 | 50 | 22 | 3 | n = 100 |

2)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 15 | 20 | 25 | 30 | 35 | 40 | ny |
| 5 | 4 | 2 |  |  |  |  | 6 |
| 10 |  | 6 | 4 |  |  |  | 10 |
| 15 |  |  | 6 | 45 | 2 |  | 53 |
| 20 |  |  | 2 | 8 | 6 |  | 16 |
| 25 |  |  |  | 4 | 7 | 4 | 15 |
| nx | 4 | 8 | 12 | 57 | 15 | 4 | n = 100 |

3)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 5 | 10 | 15 | 20 | 25 | 30 | ny |
| 20 | 1 | 5 |  |  |  |  | 6 |
| 30 |  | 5 | 3 |  |  |  | 8 |
| 40 |  |  | 9 | 40 | 2 |  | 51 |
| 50 |  |  | 4 | 11 | 6 |  | 21 |
| 60 |  |  |  | 4 | 7 | 3 | 14 |
| nx | 1 | 10 | 16 | 55 | 15 | 3 | n = 100 |

4)

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 10 | 15 | 20 | 25 | 30 | 35 | ny |
| 6 | 4 | 2 |  |  |  |  | 6 |
| 12 |  | 6 | 2 |  |  |  | 8 |
| 18 |  |  | 5 | 40 | 5 |  | 50 |
| 24 |  |  | 2 | 8 | 7 |  | 17 |
| 30 |  |  |  | 4 | 7 | 8 | 19 |
| nx | 4 | 8 | 9 | 52 | 19 | 8 | n = 100 |

**5)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 5 | 10 | 15 | 20 | 25 | 30 | ny |
| 8 | 2 | 4 |  |  |  |  | 6 |
| 12 |  | 3 | 7 |  |  |  | 10 |
| 16 |  |  | 5 | 30 | 10 |  | 45 |
| 20 |  |  | 7 | 10 | 8 |  | 25 |
| 24 |  |  |  | 5 | 6 | 3 | 14 |
| nx | 2 | 7 | 19 | 45 | 24 | 3 | n = 100 |

**6)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 2 | 7 | 12 | 17 | 22 | 27 | ny |
| 30 | 2 | 4 |  |  |  |  | 6 |
| 40 |  | 6 | 2 |  |  |  | 8 |
| 50 |  |  | 3 | 50 | 2 |  | 55 |
| 60 |  |  | 1 | 10 | 6 |  | 17 |
| 70 |  |  |  | 4 | 7 | 3 | 14 |
| nx | 2 | 10 | 6 | 64 | 15 | 3 | n = 100 |

**7)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 11 | 16 | 21 | 26 | 31 | 36 | ny |
| 35 | 2 | 4 |  |  |  |  | 6 |
| 45 |  | 6 | 3 |  |  |  | 9 |
| 55 |  |  | 6 | 45 | 4 |  | 55 |
| 65 |  |  | 2 | 8 | 6 |  | 16 |
| 75 |  |  |  | 4 | 7 | 3 | 14 |
| nx | 2 | 10 | 11 | 57 | 17 | 3 | n = 100 |

**8)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 4 | 9 | 14 | 19 | 24 | 29 | ny |
| 8 | 3 | 3 |  |  |  |  | 6 |
| 18 |  | 5 | 4 |  |  |  | 9 |
| 28 |  |  | 40 | 2 | 8 |  | 50 |
| 38 |  |  | 5 | 10 | 6 |  | 21 |
| 48 |  |  |  | 4 | 7 | 3 | 14 |
| nx | 3 | 8 | 49 | 16 | 21 | 3 | n = 100 |

**9)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 10 | 15 | 20 | 25 | 30 | 35 | ny |
| 30 | 4 | 2 |  |  |  |  | 6 |
| 40 |  | 5 | 3 |  |  |  | 8 |
| 50 |  |  | 5 | 45 | 5 |  | 55 |
| 60 |  |  | 2 | 8 | 7 |  | 17 |
| 70 |  |  |  | 4 | 7 | 3 | 14 |
| nx | 4 | 7 | 10 | 57 | 19 | 3 | n = 100 |

**10)**

**Пример.** Среди владельцев иномарок было выбрано 100 человек. Из стоимости автомашин в тыс. у.е.(*X*) и годового дохода владельцев также в тыс. у.е.(*Y*) составлена корреляционная таблица:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *Y* | *X* | | | | | https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-rBzVbN.png |
| 5 | 10 | 15 | 20 | 25 |
| 10 | 10 | 5 | - | - | - | 15 |
| 20 | 5 | 10 | 5 | - | - | 20 |
| 30 | 5 | 5 | 10 | 5 | - | 25 |
| 40 | - | 5 | 5 | 10 | - | 20 |
| *50* | - | 5 | 5 | 5 | 5 | 20 |
| https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-68AQWf.png | 20 | 30 | 25 | 20 | 5 | https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-nJrOqT.png |

Найти коэффициент корреляции величин *X* и *Y* и уравнение прямой линии регрессии *Y* на *X*.

По аналогии с предыдущим примером, находим

https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-8WJj9U.png

https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-tsbn9S.png

Проверяя значимость коэффициента корреляции по указанной выше схеме, получим:

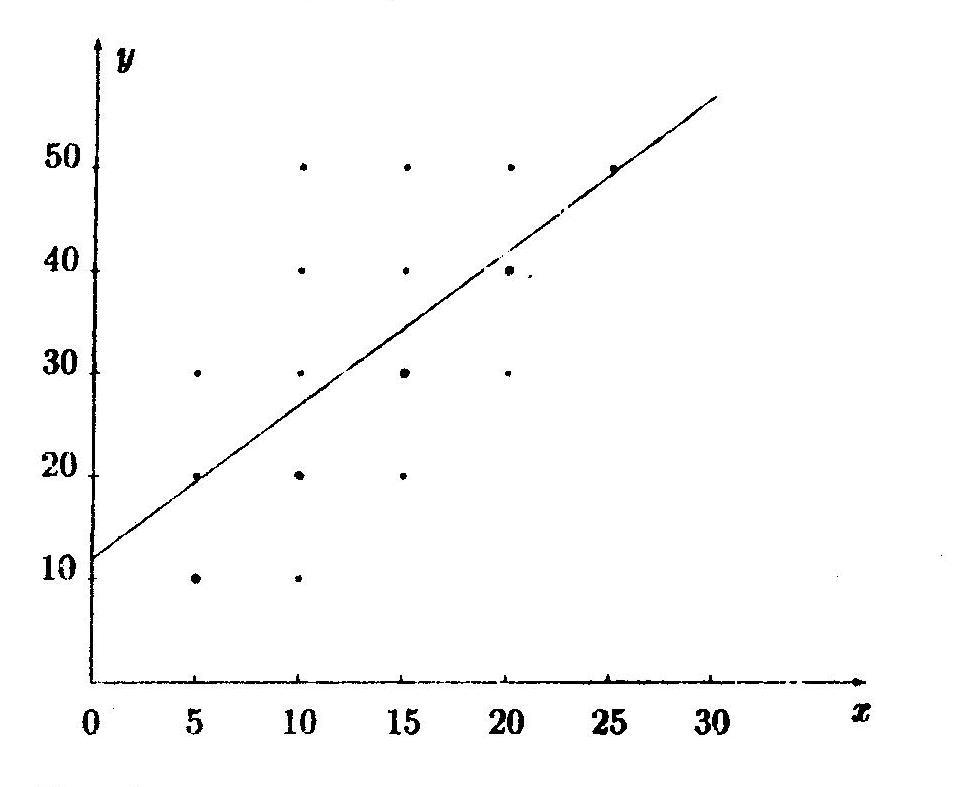
https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-FbnEyM.png

В предположении нормальности и независимости величин *X* и *Y*, критическое значение этой величины при уровне значимости https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-O6leIV.png найденное по таблице с  https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-pjQQOz.png равно https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-4xJtnF.png т.е. имеет место https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-lqtPMa.png, что свидетельствует о существовании линейной зависимости между *X* и *Y*.

Уравнение прямой линии регрессии*Y* на *X*запишется в виде:

https://studfile.net/html/2706/568/html_FanUVYhZId.tAkL/img-LVM5zx.png

Или у(х) = 1,5х + 12



То есть на графике отмечены все значения х и у из таблицы. Можно отмечать только выборочные средние.

**Самостоятельно**. По данной таблице найти выборочное уравнение прямой линии регрессии Y на Х. Оценить тесноту связи. Построить эмпирическую и теоретическую линии регрессии.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Y | X | | | | | | |
| 15 | 20 | 25 | 30 | 35 | 40 | ny |
| 30 | 1 | 6 |  |  |  |  | 7 |
| 40 |  |  | 4 |  |  | 5 | 9 |
| 50 |  | 4 | 7 | 30 | 9 |  | 50 |
| 60 |  |  | 2 | 10 | 8 |  | 20 |
| 70 | 5 |  |  |  | 6 | 3 | 14 |
| nx | 6 | 10 | 13 | 40 | 23 | 8 | n = 100 |

Проверить гипотезу о равенстве 0 генерального коэффициента

H0: rг = 0

Вычисления:

.

Если промежуточные вычисления округлять, то может получиться немного другой ответ.

Для построения графика-прямой, можно взять х=0, тогда у=10,36-можно округлить до 10. Второе значение может х= 10, тогда у= - 3,5 + 10,36 = 6,86 – можно взять у = 7.

Строим прямую по точкам (0; 10) и (10; 7)