

Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»**

**Департамент анализа данных,
принятия решений и финансовых технологий**

**Практическая работа №2
«Нахождение коэффициента корреляции»
(по дисциплине «Эконометрика»)**

Выполнила:

студентка группы ПИ19-1

факультета

«Информационных технологий и анализа больших данных»

Воронина К. М.

_____ (Подпись)

Дата выполнения: _____

Дата защиты: _____

Преподаватель:

доцент, к.т.н., Петросов Д. А.

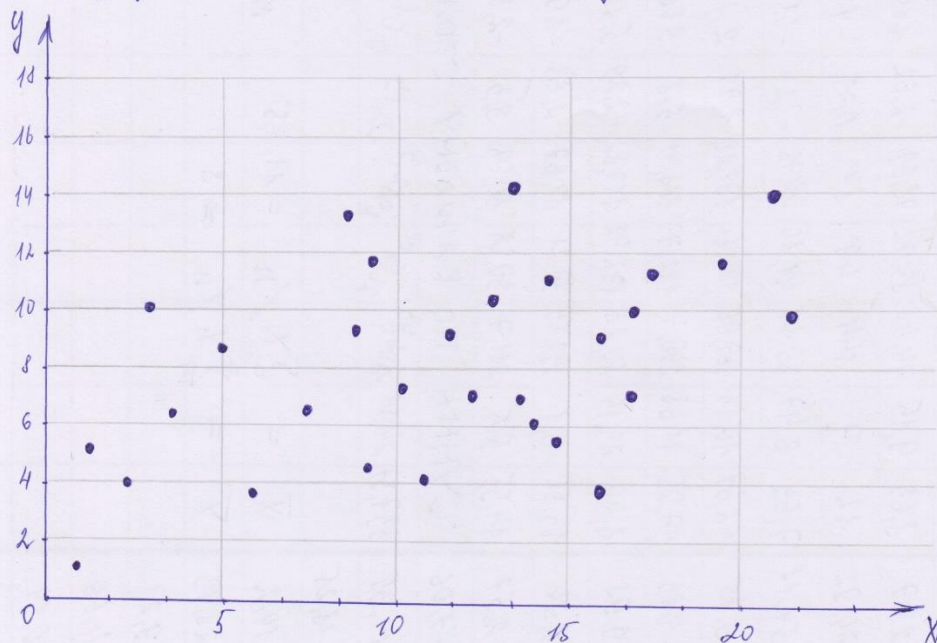
_____ (Подпись)

Нахождение коэффициента корреляции

Практическая работа №2
«Нахождение коэффициента корреляции»

Воронина Ксения, ФМ19-1
Лиз

1) Зададим 30 точек на координатной плоскости:



2) Для расчета коэффициента корреляции воспользуемся двумя формулами:

$$\textcircled{1} \quad r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

$$\textcircled{2} \quad r = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{(n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2)(n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2)}}$$

X_i	Y_i	X_i^2	Y_i^2	$X_i Y_i$	$X_i - \bar{X}$	$Y_i - \bar{Y}$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$	$(X_i - \bar{X})(Y_i - \bar{Y})$										
1,33	0,92	1,77	0,85	1,22	-9,92	-7,04	98,32	50,13	70,21	9,19	4,54	84,46	20,61	41,72	-2,06	-3,46	4,23	11,92	7,11
1,55	5,05	2,4	25,5	7,83	-9,7	-2,95	94,01	8,7	28,61	3,77	6,55	13,76	42,9	24,30	-7,54	-1,45	56,79	2,1	10,93
10,61	4,18	112,57	17,47	44,35	-0,64	-3,82	0,4	14,59	2,43	3,2	9,88	10,24	97,61	31,62	-8,05	-1,88	64,73	3,53	-15,12
5,92	3,55	35,05	12,6	21,02	-5,33	-4,45	28,36	19,81	23,7	8,63	9,15	74,48	83,72	78,69	-2,62	-1,15	6,84	1,32	-3,01
2,77	3,88	7,67	15,05	10,75	-8,48	-4,12	71,84	16,98	34,92	12	7	144	49	84	0,75	-1	0,57	1	-0,75
20,62	13,78	425,13	189,89	284,14	9,37	5,75	87,88	33,4	54,18	7,69	6,47	59,14	41,86	49,75	-3,56	-1,53	12,64	2,34	5,44
5,03	8,62	25,3	74,3	43,36	-6,22	0,62	38,63	0,38	-3,85	12,69	10,13	167,96	102,6	131,28	1,71	2,13	2,94	4,54	3,65
16,73	7,12	279,9	50,7	119,12	5,48	-0,88	30,08	0,77	-4,83	17,05	11,08	290,7	122,77	188,91	5,8	3,08	33,69	9,48	17,88
14,7	10,89	216,09	118,6	160,08	3,45	2,89	11,93	8,35	9,98	8,96	13,15	80,28	172,92	117,82	-2,29	5,15	5,22	26,52	-11,77
10,13	7,16	102,62	51,27	72,53	-1,12	-0,84	1,24	0,71	0,94	13,88	6,1	192,65	37,21	84,67	2,63	-1,9	6,94	3,61	-5,01
13,46	6,84	181,17	46,79	92,07	2,21	-1,16	4,9	1,35	-2,57	14,55	5,5	211,7	30,28	30,03	3,3	-2,5	10,92	6,25	-8,26
9,25	11,54	85,56	133,17	106,57	-2	3,54	3,98	12,53	-7,06	СУММА ПО КАЖДОМУ СТОЛБЦУ									
20,94	9,79	438,48	95,84	205	9,64	1,79	93,98	3,2	17,35	337,37	240,01	4706,43	2236,14	2962,3	0,13	0,01	912,45	315,98	263,23
15,55	3,76	241,8	14,14	58,47	4,3	-4,24	18,53	17,98	-18,25										
13,57	14,11	184,15	199,1	191,47	2,32	6,11	5,4	37,33	14,2	$\bar{X} = \frac{\sum X_i}{n} = 11,25$	$n = 30$								
19,4	11,45	376,36	131,1	222,13	8,15	3,45	66,49	11,9	25,13	$\bar{Y} = \frac{\sum Y_i}{n} \approx 8$									
16,57	9,73	274,57	94,67	161,23	5,32	1,73	28,35	2,99	9,21										
16	8,98	256	80,64	143,68	4,75	0,98	22,6	0,96	4,66										
11,42	9,11	130,42	82,99	104,04	0,17	1,11	0,03	1,23	0,19										

3) По первой формуле получаем:

$$r_1 = \frac{263,23}{\sqrt{912,45 \cdot 315,98}} = \frac{263,23}{536,959} \approx 0,490223648$$

По второй формуле получаем:

$$r_2 = \frac{30 \cdot 2962,3 - 337,37 \cdot 240,01}{\sqrt{(30 \cdot 4706,43 - (337,37)^2)(30 \cdot 240,01 - (2236,14)^2)}} =$$

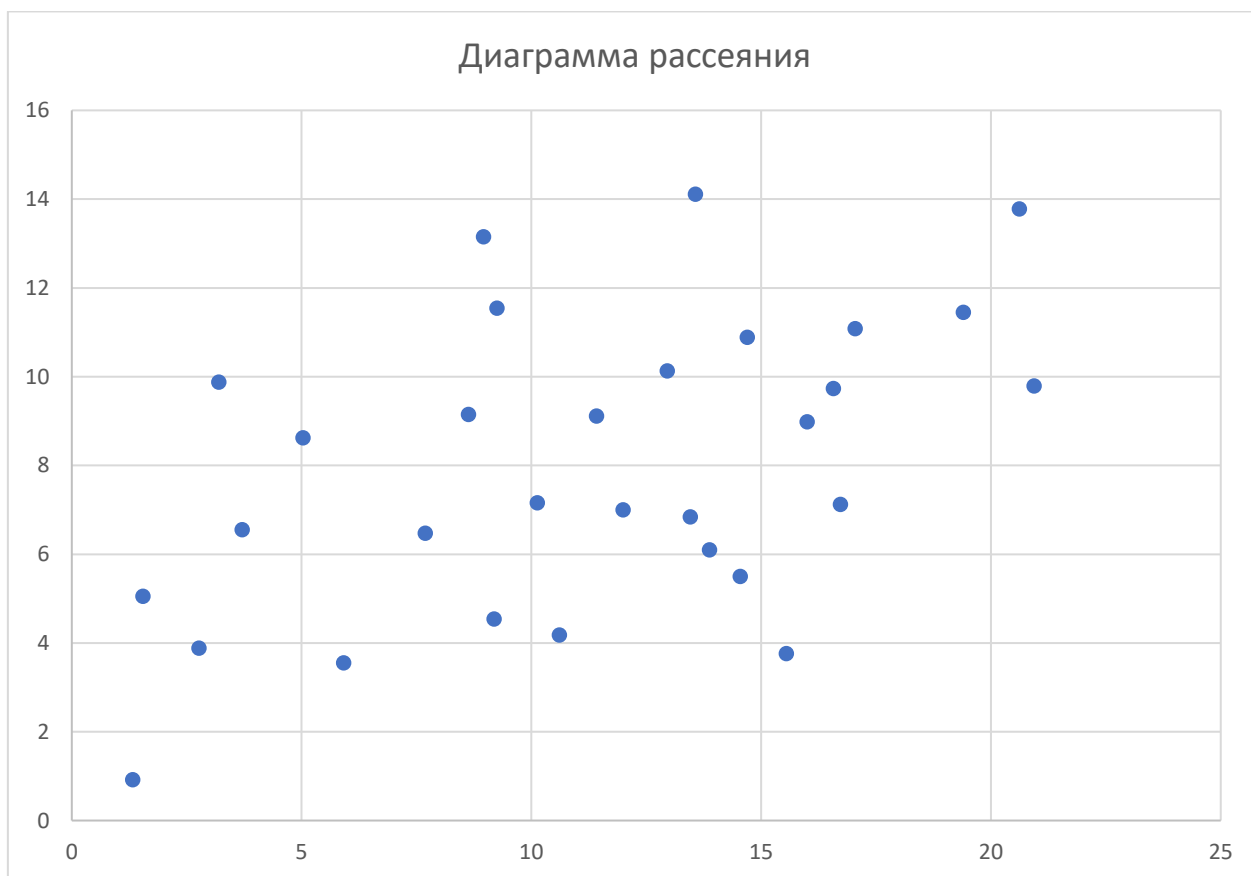
$$= \frac{7896,826}{\sqrt{27374,38 \cdot 9479,399}} = \frac{7896,826}{16108,746} = 0,490218522$$

Нахождение коэффициента корреляции в Excel

1) Формулы для расчета:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2 \sum (y - \bar{y})^2}}$$
$$r = \frac{n \sum xy - \sum x \sum y}{\sqrt{\left(n \sum x^2 - (\sum x)^2 \right) \left(n \sum y^2 - (\sum y)^2 \right)}}$$

2) Набор точек на диаграмме рассеяния:



3) Таблица найденных значений для удобства расчета:

X	Y	X^2	Y^2	Xi*Yi	Xi-Xср	Yi-Yср	(Xi-Xср)^2	(Yi-Yср)^2	(Xi-Xср)*(Yi-Yср)
1,33	0,92	1,7689	0,8464	1,2236	-9,9157	-7,080333333	98,32044544	50,13112011	70,20622522
1,55	5,05	2,4025	25,5025	7,8275	-9,6957	-2,950333333	94,00595211	8,704466778	28,60544856
10,61	4,18	112,572	17,4724	44,3498	-0,6357	-3,820333333	0,404072111	14,59494678	2,428458556
5,92	3,55	35,0464	12,6025	21,016	-5,3257	-4,450333333	28,36272544	19,80546678	23,70099189
2,77	3,88	7,6729	15,0544	10,7476	-8,4757	-4,120333333	71,83692544	16,97714678	34,92257189
20,62	13,78	425,184	189,888	284,1436	9,37433	5,779666667	87,87812544	33,40454678	54,18052189
5,03	8,62	25,3009	74,3044	43,3586	-6,2157	0,619666667	38,63451211	0,383986778	-3,851641444
16,73	7,12	279,893	50,6944	119,1176	5,48433	-0,880333333	30,07791211	0,774986778	-4,828041444
14,7	10,89	216,09	118,592	160,083	3,45433	2,889666667	11,93241878	8,350173444	9,981871889
10,13	7,16	102,617	51,2656	72,5308	-1,1157	-0,840333333	1,244712111	0,706160111	0,937531889
13,46	6,84	181,172	46,7856	92,0664	2,21433	-1,160333333	4,903272111	1,346373444	-2,569364778
9,25	11,54	85,5625	133,172	106,745	-1,9957	3,539666667	3,982685444	12,52924011	-7,063994778
20,94	9,79	438,484	95,8441	205,0026	9,69433	1,789666667	93,98009878	3,202906778	17,34962522
15,55	3,76	241,803	14,1376	58,468	4,30433	-4,240333333	18,52728544	17,98042678	-18,25180811
13,57	14,11	184,145	199,092	191,4727	2,32433	6,109666667	5,402525444	37,32802678	14,20090189
19,4	11,45	376,36	131,103	222,13	8,15433	3,449666667	66,49315211	11,90020011	28,12973189
16,57	9,73	274,565	94,6729	161,2261	5,32433	1,729666667	28,34852544	2,991746778	9,209321889
16	8,98	256	80,6404	143,68	4,75433	0,979666667	22,60368544	0,959746778	4,657661889
11,42	9,11	130,416	82,9921	104,0362	0,17433	1,109666667	0,030392111	1,231360111	0,193451889
9,19	4,54	84,4561	20,6116	41,7226	-2,0557	-3,460333333	4,225765444	11,97390678	7,113291889
3,71	6,55	13,7641	42,9025	24,3005	-7,5357	-1,450333333	56,78627211	2,103466778	10,92922856
3,2	9,88	10,24	97,6144	31,616	-8,0457	1,879666667	64,73275211	3,533146778	-15,12317144
8,63	9,15	74,4769	83,7225	78,9645	-2,6157	1,149666667	6,841712111	1,321733444	-3,007144778
12	7	144	49	84	0,75433	-1,000333333	0,569018778	1,000666778	-0,754584778
7,69	6,47	59,1361	41,8609	49,7543	-3,5557	-1,530333333	12,64276544	2,341920111	5,441355222
12,96	10,13	167,962	102,617	131,2848	1,71433	2,129666667	2,938938778	4,535480111	3,650958556
17,05	11,08	290,703	122,766	188,914	5,80433	3,079666667	33,69028544	9,484346778	17,87541189
8,96	13,15	80,2816	172,923	117,824	-2,2857	5,149666667	5,224272111	26,51906678	-11,77042144
13,88	6,1	192,654	37,21	84,668	2,63433	-1,900333333	6,939712111	3,611266778	-5,006111444
14,55	5,5	211,703	30,25	80,025	3,30433	-2,500333333	10,91861878	6,251666778	-8,261934778
337,37	240,01	4706,43	2236,14	2962,299	6,6E-14	-1,15463E-14	912,4795367	315,9796967	263,2263433

4) Результаты:

Третий расчет проводился с помощью встроенной функции Excel
 =КОРРЕЛ(Таблица1[X];Таблица1[Y])

ПО ПЕРВОЙ ФОРМУЛЕ		
0,49021680480095		
ПО ВТОРОЙ ФОРМУЛЕ		
0,49021680480095		
ПО ВСТРОЕННОЙ ФУНКЦИИ		
0,490216805		

Нахождение коэффициента корреляции в Python

1) Импортируем библиотеки и загружаем лист с X и Y в виде DataFrame pandas;

```
import numpy as np
import pandas as pd
df = pd.read_excel('Dannye.xlsx', sheet_name='Данные для python')
```

2) Вычисляем табличные значения сумм;

```
n = df.shape[0]
xmean=df['X'].mean()
ymean=df['Y'].mean()
#сумма x
x = df['X'].sum()
#сумма y
y = df['Y'].sum()
#сумма x^2
x2 = (df['X']*df['X']).sum()
#сумма y^2
y2 = (df['Y']*df['Y']).sum()
#сумма x*y
xy = (df['X']*df['Y']).sum()
#сумма x-xmean
x_xmean = (df['X'] - xmean).sum()
#сумма y-ymean
y_ymean = (df['Y'] - ymean).sum()
#сумма (x-xmean)^2
x_xmean2 = ((df['X'] - xmean)*(df['X'] - xmean)).sum()
#сумма (y-ymean)^2
y_ymean2 = ((df['Y'] - ymean)*(df['Y'] - ymean)).sum()
#сумма (x-xmean)*(y-ymean)
sumxymean = ((df['X'] - xmean)*(df['Y'] - ymean)).sum()
```

3) Находим значение корреляции по формулам и встроенной функции NumPy.

```
#1
k1 = sumxymean / (x_xmean2*y_ymean2)**(1/2)
k1
```

0.49021680480095176

```
#2
```

```
k2 = (n*xy-x*y)/((n*x2-x*x)*(n*y2-y*y))**(1/2)  
k2
```

```
0.4902168048009515
```

```
#встроенная функция
```

```
df['X'].corr(df['Y'])
```

```
0.49021680480095153
```