

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ
УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ПЕДАГОГИЧЕСКИЙ УНИВЕРСИТЕТ им. А. И. ГЕРЦЕНА**

Институт информационных технологий и технологического образования

Кафедра компьютерные технологии и электронного обучения

Основная профессиональная образовательная программа

Направление подготовки 09.03.01 Информатика и вычислительная техника

Направленность (профиль) «Технологии разработки программного
обеспечения»

форма обучения – очная

ЛАБОРАТОРНАЯ РАБОТА №7.3

по дисциплине: «Анализ данных и основы Data science»

**КОРРЕЛЯЦИОННЫЙ АНАЛИЗ. ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ
КОРРЕЛЯЦИИ**

Руководитель:

кандидат педагогических наук, доцент,

Светлана Викторовна Гончарова

Автор работы студент 2 курса

1 группы 1 подгруппы

Чирцов Тимофей Александрович

Санкт-Петербург
2023

Цель работы: провести вычисления коэффициентов корреляции, ранговой корреляции и линейной корреляции.

Оборудование: ПК, Excel

Математические модели:

$$R_i = \sqrt{1 - \frac{|A|}{A_{ii}}},$$

$$R_i = \sqrt{\frac{r_{ij}^2 + r_{ik}^2 - 2r_{ij} r_{ik} r_{jk}}{1 - r_{jk}^2}}$$

$$t = \frac{R^2(n-p)}{(1-R^2)(p-1)} > F_{кр},$$

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

$$R_z = R_3 \sqrt{\frac{r_{13}^2 + r_{23}^2 - 2r_{13} r_{23} r_{12}}{1 - r_{12}^2}}$$

Задание 2.1

Задача 2.1. Проведено исследование шести фермерских хозяйств для изучения зависимости урожайности зерновых культур (Z, ц/га) от качества пашни (X, в баллах) и количества внесенных минеральных удобрений на 1 га пашни (Y, кг/га).

Результаты приведены в таблице 1.

Таблица 1

X	26	35	36	40	41	45
Y	2.1	2.3	2.4	2.6	2.9	3.0
Z	18.0	21.0	22.1	25.3	28	28.5

Определить тесноту связи между величиной Z и величинами X и Y, используя множественный коэффициент корреляции. Проверить на уровне $\alpha = 0.05$ значимость R_z .

Результат работы:

i	x _i	y _i	z _i	x ²	y ²	z ²	x _i y _i	x _i z _i	y _i z _i
1	26,00	2,10	18,00	676,00	4,41	324,00	54,60	468,00	37,80
2	35,00	2,30	21,00	1225,00	5,29	441,00	80,50	735,00	48,30
3	36,00	2,40	22,10	1296,00	5,76	488,41	86,40	795,60	53,04
4	40,00	2,60	25,30	1600,00	6,76	640,09	104,00	1012,00	65,78
5	41,00	2,90	28,00	1681,00	8,41	784,00	118,90	1148,00	81,20
6,00	45,00	3,00	28,50	2025,00	9,00	812,25	135,00	1282,50	85,50
Ср. знач.	37,17	2,55	23,82	1417,17	6,61	581,63	96,57	906,85	61,94
Сумм	223,00	15,30	142,90	8503,00	39,63	3489,75	579,40	5441,10	371,62

(x _i - x _{ср}) ²	(y _i - y _{ср}) ²	(z _i - z _{ср}) ²	(x _i - x _{ср}) * (y _i - y _{ср})	(x _i - x _{ср}) * (z _i - z _{ср})	(y _i - y _{ср}) * (z _i - z _{ср})
124,69	0,20	33,83	5,025	64,95278	2,6175
4,69	0,06	7,93	0,5416667	6,102778	0,704167
1,36	0,02	2,95	0,175	2,002778	0,2575
8,03	0,00	2,20	0,1416667	4,202778	0,074167
14,69	0,12	17,50	1,3416667	16,03611	1,464167
61,36	0,20	21,93	3,525	36,68611	2,1075
214,83	0,62	86,35	10,75	129,98	7,23

г _{xy}	0,935234		
г _{xz}	0,954355		
г _{zy}	0,991456		

	1	0,935234	0,954355
A	0,935234	1	0,954355
	0,954355	0,991456	1

R ₃	0,999301
R _z	0,993714
p	3
t	118,1844
k ₁	2
k ₂	3
крит	0,05
F	9,55

Вывод: $t = 118,1844 > F_{кр} = 9,55 \Rightarrow$ из этого следует, что коэффициент R значимый

Задание 3.1

Задание 3. Вычислите вычисления.

Задача 3.1 При изучении финансовой деятельности компании в течение некоторого времени был собран статистический материал. В таблице содержатся данные о ежемесячной прибыли Z (тыс. усл. ед.), расходах на рекламу X (тыс. усл. ед.) и вложении капитала в ценные бумаги Y (тыс. усл. ед.).

Z	10	12	12	14	16	17	18
X	0.2	0.5	0.3	0.5	0.5	0.6	0.8
Y	0.8	0.2	1	1.2	0.9	1	1.1

Определить тесноту связи между переменной Z и переменными X и Y с помощью выборочного множественного коэффициента корреляции R и определить его значимость на уровне $\alpha = 0.05$.

Результат работы:

i	x	y	z	x ²	y ²	z ²	xy	xz	yz
1	0,20	0,80	10,00	0,04	0,64	100,00	0,16	2,00	8,00
2	0,50	0,20	12,00	0,25	0,04	144,00	0,10	6,00	2,40
3	0,30	1,00	12,00	0,09	1,00	144,00	0,30	3,60	12,00
4	0,50	1,20	14,00	0,25	1,44	196,00	0,60	7,00	16,80
5	0,50	0,90	16,00	0,25	0,81	256,00	0,45	8,00	14,40
6	0,60	1,00	17,00	0,36	1,00	289,00	0,60	10,20	17,00
7	0,80	1,10	18,00	0,64	1,21	324,00	0,88	14,40	19,80
Сумм	3,40	6,20	99,00	1,88	6,14	1453,00	3,09	51,20	90,40
Ср знач	0,485714	0,88571	14,1429	0,26857	0,87714	207,571	0,44143	7,31429	12,9143

(xi - xcp) ²	(yi - ycp) ²	(zi - zcp) ²	(xi - xcp) * (yi - ycp)	(xi - xcp) * (zi - zcp)	(zi - zcp) * (yi - ycp)
0,08	0,01	17,16	0,0244898	1,18367	0,3551
0,00	0,47	4,59	-0,009796	-0,0306	1,46939
0,03	0,01	4,59	-0,021224	0,39796	-0,2449
0,00	0,10	0,02	0,0044898	-0,002	-0,0449
0,00	0,00	3,45	0,0002041	0,02653	0,02653
0,01	0,01	8,16	0,0130612	0,32653	0,32653
0,10	0,05	14,88	0,0673469	1,21224	0,82653
0,23	0,65	52,86	0,08	3,11	2,71

r _{xy}	0,204068		
r _{xz}	0,895974		
r _{yz}	0,46358		
	1	0,20407	0,89597
A	0,204068	1	0,46358
	0,895974	0,46358	y

R3	0,943292
RZ	0,97
t	31,84095
k1	2
k2	4
α	0.05
F _{кр}	6,94

Вывод: $R = 0.97$, статистика $t = 31.8$, R – значим

Вывод по лабораторной работе: с помощью электронных таблиц Excel нам удалось провести вычисления коэффициентов корреляции.