

**Федеральное государственное бюджетное образовательное учреждение  
высшего образования**

**"Уфимский государственный авиационный технический университет"**

**Кафедра** Высокопроизводительных вычислительных технологий и систем

**Дисциплина:** Математическая статистика

**Отчет по лабораторной работе № 1**

**Тема:** «Линейный корреляционный и регрессионный анализ  
многомерного временного сигнала»

Группа ПМ-453	Фамилия И.О.	Подпись	Дата	Оценка
Студент	Шамаев И.Р.			
Принял	Маякова С.А.			

**Уфа 2022**

**Цель работы:** изучить методы линейного регрессионного корреляционного анализа многомерного временного сигнала.

### Практическая часть

#### Задача 1.

Результаты наблюдений за характеристиками канала представлены в табл.1.

№ пп	Пропускная способность канала,  кбит/с	Соотношение сигнал/шум,  дБ	Остаточное затухание, дБ, на частоте, Гц		
			1020	1800	2400
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
1	26,37	41,98	17,66	16,05	22,85
2	28,00	43,83	17,15	15,47	23,25
3	27,83	42,83	15,38	17,59	24,55
4	31,67	47,28	18,39	16,92	26,59
5	23,50	38,75	18,32	15,66	26,22
6	21,04	35,12	17,81	17,00	27,52
7	16,94	32,07	21,42	16,77	25,76
8	37,56	54,25	26,42	15,68	23,10
9	18,84	32,70	17,23	15,92	23,41
10	25,77	40,51	30,43	15,29	25,17
11	33,52	49,78	21,71	15,61	25,39
12	28,21	43,84	28,33	15,70	24,56
13	28,76	44,03	30,42	16,87	24,45

14	24,60	39,46	21,66	15,25	23,81
15	24,51	38,78	25,77	16,05	24,48

Найдем выборочное мат. ожидание  $\mu_1$ , выборочную дисперсию  $\mu_2$  и среднеквадратическое отклонение  $\sigma$ .

**Решение:**

Вычислим для каждого столбца выборочное среднее и дисперсию

$$\mu_1(x_j) = \frac{1}{n} \sum_{i=1}^n x_{ij}$$

$$\mu_2(x_j) = \sigma^2(x_j) = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \mu_1(x_j))^2$$

Оценка параметра распределения	Варианта				
	$X_1$	$X_2$	$X_3$	$X_4$	$X_5$
1	26,47	41,68	21,87	16,12	24,74
2	29,10	36,47	26,37	0,52	1,88
	5,39	6,04	5,13	0,72	1,37

Вычислим новую матрицу  $U (u_{ij} = \frac{(x_{ij} - \mu_1(x_j))}{\sigma(x_j)})$

№	Пропускная способность	Соотношение сигнал/шум,	Остаточное затухание, дБ
---	------------------------	-------------------------	--------------------------

пп			на частоте, Гц		
	канала, кбит/с	дБ	1020	1800	2400
	$U_1$	$U_2$	$U_3$	$U_4$	$U_5$
1	-0,019	0,050	-0,821	-0,100	-1,379
2	0,283	0,356	-0,920	-0,902	-1,087
3	0,251	0,190	-1,265	2,030	-0,139
4	0,963	0,927	-0,678	1,104	1,349
5	-0,551	-0,485	-0,692	-0,639	1,079
6	-1,007	-1,086	-0,791	1,214	2,027
7	-1,768	-1,592	-0,088	0,896	0,743
8	2,055	2,081	0,885	-0,611	-1,196
9	-1,415	-1,487	-0,904	-0,279	-0,970
10	-0,131	-0,194	1,666	-1,151	0,313
11	1,306	1,341	-0,032	-0,708	0,474
12	0,322	0,358	1,257	-0,584	-0,132
13	0,424	0,389	1,664	1,034	-0,212
14	-0,348	-0,368	-0,042	-1,206	-0,679
15	-0,364	-0,480	0,759	-0,100	-0,190

Теперь вычислим корреляции

$$\rho_{1k} = \frac{1}{15} \sum_{i=1}^{15} u_{i1} u_{ik}, k=2,3,4,5$$

и значения статистик критерия Стьюдента

$$t = \frac{\hat{\rho}_{1k} \sqrt{n-2}}{\sqrt{1-\rho_{1k}^2}}.$$

	$X_2$	$X_3$	$X_4$	$X_5$
$r_{1j}$	0,931	0,254	-0,138	-0,220
$t$	9,161	0,946	0,501	0,814

Проверка гипотезы о значимости оценок коэффициентов корреляции с уровнем значимости  $\alpha$ , равным 0,1. Т.к.  $t_{кр}(13,0.1)=1,77$ , то при условии  $t < t_{кр}(13,0.1)$  можно отбросить третий, четвертый и пятые столбцы. И только с  $X_2$   $X_1$  имеет линейную положительную взаимосвязь, так как  $\rho_{12} \approx 1$ .

## Задача 2.

Построить уравнение регрессии для пропускной способности канала по выборке, заданной в табл. 1.

Применительно к указанной выборке построение аналитической зависимости в основной своей части выполнено в рамках примера 1. Осталось лишь выбрать вид регрессии, представив графически зависимость  $X_1$  от  $X_2$ , найти коэффициенты регрессии.

## Решение:

Решение будем искать в виде

$$\hat{y} = \mu_1(x_1) - a_2 \mu_1(x_2) \frac{\sigma(x_1)}{\sigma(x_2)} + a_2 \frac{\sigma(x_1)}{\sigma(x_2)} x_2.$$

Коэффициент регрессии  $a_2$  найдем из уравнения  $\rho_{y2} - a_2 \rho_{22} = 0$ ,

$\rho_{y2} = \rho_{12}$ ,  $\rho_{22} = 0,9333$ , тогда  $a_2 = 0,99699$ .

$$\hat{y} = -10,6463 + 0,890604 x_2.$$

№ пп	Пропускная способность	Соотношени е сигнал/шум,	Значение функции, кбит/с	Погрешность, кбит/с
	канала, кбит/с	дБ		
	$Y$	$X$	$\hat{y}$	$e$
1	26,37	41,98	26,74125	0,371254
2	28,00	43,83	28,38887	0,388871
3	27,83	42,83	27,49827	0,331733
4	31,67	47,28	31,46145	0,208547
5	23,50	38,75	23,8646	0,364604
6	21,04	35,12	20,63171	0,408287
7	16,94	32,07	17,91537	0,975372
8	37,56	54,25	37,66896	0,10896
9	18,84	32,70	18,47645	0,363547

10	25,77	40,51	25,43207	0,337933
11	33,52	49,78	33,68796	0,167962
12	28,21	43,84	28,39778	0,187777
13	28,76	44,03	28,56699	0,193009
14	24,60	39,46	24,49693	0,103067
15	24,51	38,78	23,89132	0,618678

Построим графически зависимость  $X_1$  от  $X_2$

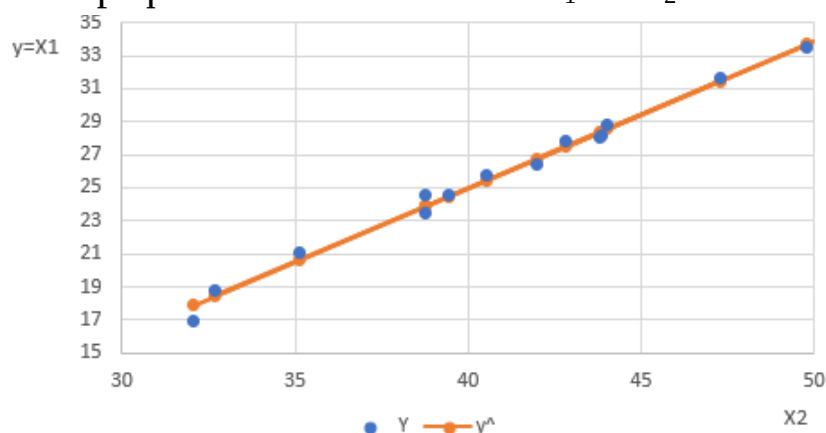


Рисунок 1. График зависимости

По графику видим, что  $X_1$  и  $X_2$  имеют положительную линейную связь.

### Задача 3.

Выяснить коррелированность курсов акций компаний «Газпром», «Роснефть», ТНК, «РУСАЛ». Определить, есть ли существенные связи при уровне значимости  $\alpha=0,1; 0,5$ . Если существенные связи есть, оставить 3 курса акций с наиболее сильными связями. Построить уравнение двумерной линейной регрессии  $Y=aX_1+bX_2+c$ , рассчитав неизвестные коэффициенты  $a$ ,  $b$ ,  $c$  методом наименьших квадратов. Построить график плоскости  $Y$ , указав на том же графике экспериментальные точки выборки (по которым был реализован метод наименьших квадратов). Оценить максимальное отклонение экспериментальных точек от плоскости.

**Решение:**

В качестве выборки возьмем 15 значений курсов акций за последние 3 месяца (в одни и те же даты).

№	«Газпром»	«Роснефть»	ТНК	«РУСАЛ»
	$X_1$	$X_2$	$X_3$	$X_4$
1	191,65	371,15	15,47	30,69
2	189,81	387,25	14,11	30,7
3	186,99	388,5	12,9	32,15
4	183,51	372,3	12,23	30,59
5	178,36	371,45	11,98	31,68
6	177,94	379,1	12,19	32,16
7	184,5	373,2	10,42	32,4
8	177,96	384,2	10,84	32,38
9	170,95	390,7	11,25	31,07
10	167,78	383,15	11,05	32,05
11	168,62	382,45	10,9	32,52
12	163,92	380,4	10,23	32,84
13	164,66	351,8	9,35	30,63
14	155,0	384,35	10,4	31,89
15	165,27	425,1	11,09	32,3

Заполним необходимые таблицы



Оценка параметра распределения	Варианты			
	$X_1$	$X_2$	$X_3$	$X_4$
$\mu_1$	175,128	381,673	11,6273	31,736667
$\mu_2$	118,930574	10086,3	2,52641	0,6165238
$\sigma$	10,9055295	100,43	1,58947	0,7851903

№	$U_1$	$U_2$	$U_3$	$U_4$
1	1,515011257	-0,10478	2,417581	-1,33301
2	1,34628951	0,055528	1,561949	-1,32027
3	1,087705092	0,067974	0,800687	0,526412
4	0,768600917	-0,09333	0,379163	-1,46037
5	0,296363417	-0,1018	0,221877	-0,07217
6	0,257850845	-0,02562	0,353997	0,539147
7	0,859380553	-0,08437	-0,75958	0,844806
8	0,259684777	0,025158	-0,49534	0,819334
9	-0,383108403	0,08988	-0,2374	-0,84905
10	-0,673786631	0,014703	-0,36322	0,399054
11	-0,596761485	0,007733	-0,4576	0,997635
12	-1,027735514	-0,01268	-0,87912	1,40518
13	-0,959880029	-0,29745	-1,43277	-1,40942
14	-1,845669204	0,026652	-0,77217	0,195282
15	-0,903945102	0,432406	-0,33806	0,717448

Теперь вычислим матрицу парных корреляций

$$\rho_{jk} = \frac{1}{15} \sum_{i=1}^{15} u_{ij} u_{ik}.$$

$\rho_{1k}$	-0,025	0,728117	-0,32864
$\rho_{2k}$	-0,025	0,007518	0,055696

$\rho_{3k}$	0,728117	0,007518	-0,41993
$\rho_{4k}$	-0,32864	0,055696	-0,41993

и значения статистик критерия Стьюдента

$$t = \rho_{jk} \sqrt{\frac{n-2}{1-\rho_{jk}^2}}.$$

$t_1$	0,090168	3,829973	1,254624
$t_2$	0,090168	0,027107	0,201127
$t_3$	3,829973	0,027107	1,86831
$t_4$	1,254624	0,201127	1,86831

Критическое значение  $t_{кр}(n-2; \alpha) = (13; 0,1) = 1,77$

Критическое значение  $t_{кр}(n-2; \alpha) = (13; 0,5) = 0,69$ .

Сравнив значения  $t$ , приходим к выводу, что столбец  $X_2$  можно отбросить. Оставим 3 курса акций с наиболее сильными связями, это  $X_1, X_3$  и  $X_4$ .  $X_3$  и  $X_1$  имеют линейную положительную взаимосвязь, так как  $t_{13} > t_{кр}$  и  $\rho_{13} \approx 1$ .

Далее построим уравнение двумерной линейной регрессии. Решение будем искать в виде

$$\hat{y}(X_3) = c + aX_1 + bX_4.$$

Коэффициенты регрессии найдем методом наименьших квадратов. из уравнения

$$\sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - c - aX_{1i} - bX_{4i})^2 \rightarrow \min.$$

Частные производные функции приравниваем к нулю

и

Из системы получаем  $c = 6,364, a = 0,103, b = -0,405$ , тогда

$$\hat{y}(X_3) = 6,364 + 0,103X_1 - 0,405X_4.$$

№	Y	$X_1$	$X_4$	$\hat{y}$	e
1	15,47	91,65	30,69	13,6745	1,7955

2	14,11	189,81	30,7	13,48093	0,62907
3	12,9	186,99	32,15	12,60322	0,29678
4	12,23	183,51	30,59	12,87658	0,64658
5	11,98	178,36	31,68	11,90468	0,07532
6	12,19	177,94	32,16	11,66702	0,52298
7	10,42	184,5	32,4	12,2455	1,8255
8	10,84	177,96	32,38	11,57998	0,73998
9	11,25	170, 195	31,07	11,3885	0,1385
10	11,05	167,78	32,05	10,66509	0,38491
11	10,9	168,62	32,52	10,56126	0,33874
12	10,23	163,92	32,84	9,94756	0,28244
13	9,35	164,66	30,63	10,91883	1,56883
14	10,4	155,0	31,89	9,41355	0,98645
15	11,09	165,27	32,3	10,30531	0,78469

Построим график плоскости  $Y$

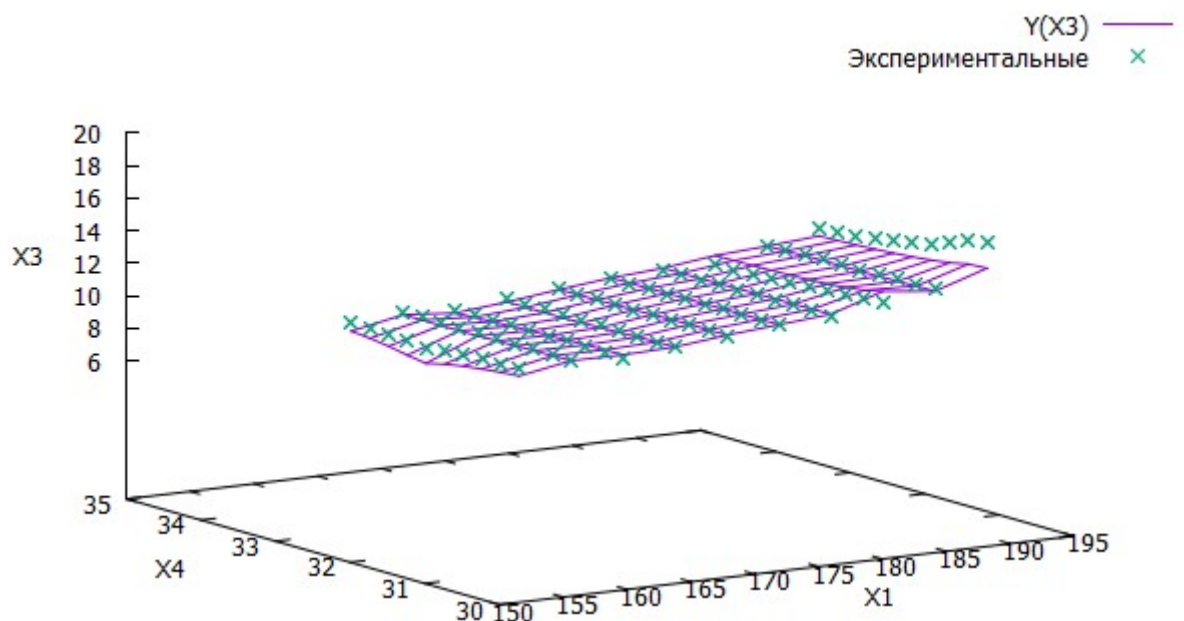


Рисунок 2. График зависимости

Максимальная погрешность равна 1,7955.

## **Вывод**

Таким образом, в ходе лабораторной работы был изучен линейный корреляционный и регрессионный анализ многомерного временного сигнала.