

Лабораторная работа №7.
Преобразование данных по методу главных компонент.

Имеются данные измерений двух признаков X_1 и X_2 :

X_1	X_2	X_1	X_2
-4,36	-5,64	12,25	6,27
6,36	-3,64	5,66	18,22
8,36	7,36	14,28	5,64
6,36	3,36	13,27	7,27
-7,44	6,9	-9,4	9,64
13,86	13,12	-6,23	8,61
4,81	-4,82	-6,4	8,67
5,82	-2,83		
3,85	-1,58		
5,48	9,81		
1,8	9,88		
-7,32	8,22		
-7,25	-9,14		
-2,27	-9,25		
-8,82	-9,41		
-5,32	8,22		
-11,83	12,88		
-10,28	3,78		
-10,86	14,89		
-13,88	19,12		
-11,12	11,19		
-7,15	8,45		
12,62	5,12		
17,18	11,85		
10,12	8,53		
5,18	16,22		
-5,29	7,81		
-3,92	11,82		
-6,26	10,38		
-6,72	4,28		
5,38	2,84		
3,81	4,55		
4,22	9,62		
5,32	2,92		
6,28	12,28		

1. Рассчитаем выборочное среднее для каждого признака:

$$\bar{x}_B = \sum_{i=1}^m \frac{n_i * x_i}{N}$$

$$\overline{x_{B1}} = 0,62, \overline{x_{B2}} = 6,71;$$

2. Рассчитаем дисперсию для каждого признака:

$$D(x_i) = \frac{\sum_{i=1}^m (x_i - \bar{x}_{Bi})^2}{N - 1}$$

$$D(x_1) = 67, D(x_2) = 18;$$

3. Рассчитаем корреляцию по формуле: $K(x_1; x_2) = \frac{\sum_{i=1}^m (x_{1i} - x_{1B})(x_{2i} - x_{2B})}{N - 1}$

$$K(x_1; x_2) = 11$$

4. Составим ковариационную матрицу:

$$[\Sigma] = \begin{bmatrix} 67 & 11 \\ 11 & 18 \end{bmatrix}$$

5. Вычислим собственные значения и соответствующие им собственные векторы ковариационной матрицы:

$$\begin{bmatrix} 67 - \lambda & 11 \\ 11 & 18 - \lambda \end{bmatrix} = (67 - \lambda) * (18 - \lambda) - (11) * (11) =$$

$$\lambda_1 = 69,5; \quad \lambda_2 = 15,5;$$

при λ_1 имеем:

$$\begin{bmatrix} -2,5 & 11 \\ 11 & -51,5 \end{bmatrix} * \begin{bmatrix} B1_1 \\ B1_2 \end{bmatrix} = 0 \Rightarrow \begin{cases} -2,5B1_1 + 11B1_2 = 0 \\ 11B1_1 - 51,5B1_2 = 0 \end{cases}$$

$$B1_2 = C; \quad B1_1 = 4,68B1_2; \quad |B1| = (C; 4,68C)$$

$$|B1| = \sqrt{C^2 + 21,92C^2} = 4,79C;$$

$$B01_2 = \frac{B1_2}{|B1|} = \frac{C}{4,79C} = 0,21;$$

$$B01_1 = \frac{B1_1}{|B1|} = \frac{4,68C}{4,79C} = 0,98;$$

$$B01 = \begin{pmatrix} 0,98 \\ 0,21 \end{pmatrix};$$

При λ_2 имеем:

$$\begin{bmatrix} 51,5 & 11 \\ 11 & 2,5 \end{bmatrix} * \begin{bmatrix} B2_1 \\ B2_2 \end{bmatrix} = 0 \Rightarrow \begin{cases} 51,5B2_1 + 11B2_2 = 0 \\ 11B2_1 + 2,5B2_2 = 0 \end{cases}$$

$$B2_2 = -0,48B2_1; B2_1 = C; |B2| = (-4,68C; C)$$

$$|B2| = \sqrt{C^2 + 21,90C^2} = 4,79C;$$

$$B02_1 = \frac{B2_1}{|B2|} = \frac{C}{4,79C} = 0,21;$$

$$B02_2 = \frac{B2_2}{|B2|} = \frac{-4,68C}{4,79C} = -0,98;$$

$$B02 = \begin{pmatrix} 0,2 \\ -0,98 \end{pmatrix};$$

$$tg\alpha = \frac{0,21}{0,98}; \quad tg\beta = \frac{-0,98}{0,21};$$

$$\alpha = 12; \quad \beta = 78;$$

6. Сумма собственных значений равна следу матрицы: $\lambda_1 + \lambda_2 = 69,5 + 15,5 = 85$;
 $D(x_1) + D(x_2) = 67 + 18 = 85$

Таким образом, вклад переменной x_1 в общую дисперсию $(69,5/85)100 = 81,8 \%$, а вклад второй переменной $x_2 - (15,5/85)100 = 18,2 \%$.

С другой стороны, главные оси эллипса также характеризуют суммарную дисперсию, следовательно: первая главная ось составляет $(67/85)100 = 78,8\%$ суммарной дисперсии, а вторая главная ось эллипса $(18/85)100 = 21,2\%$.

Таким образом, главная ось эллипса учитывает чуть большую часть изменчивости множества данных.

Угол поворота новых осей координат по отношению к старой определяется из координат собственных векторов:

$$B1_1 = \sin\alpha = 0,21/1 = 0,21$$

$$B1_2 = \cos\alpha = 0,98;$$

$$B2_1 = \sin\beta = -0,98;$$

$$B2_2 = \cos\beta = 0,21;$$

Если сделаем преобразование вида:

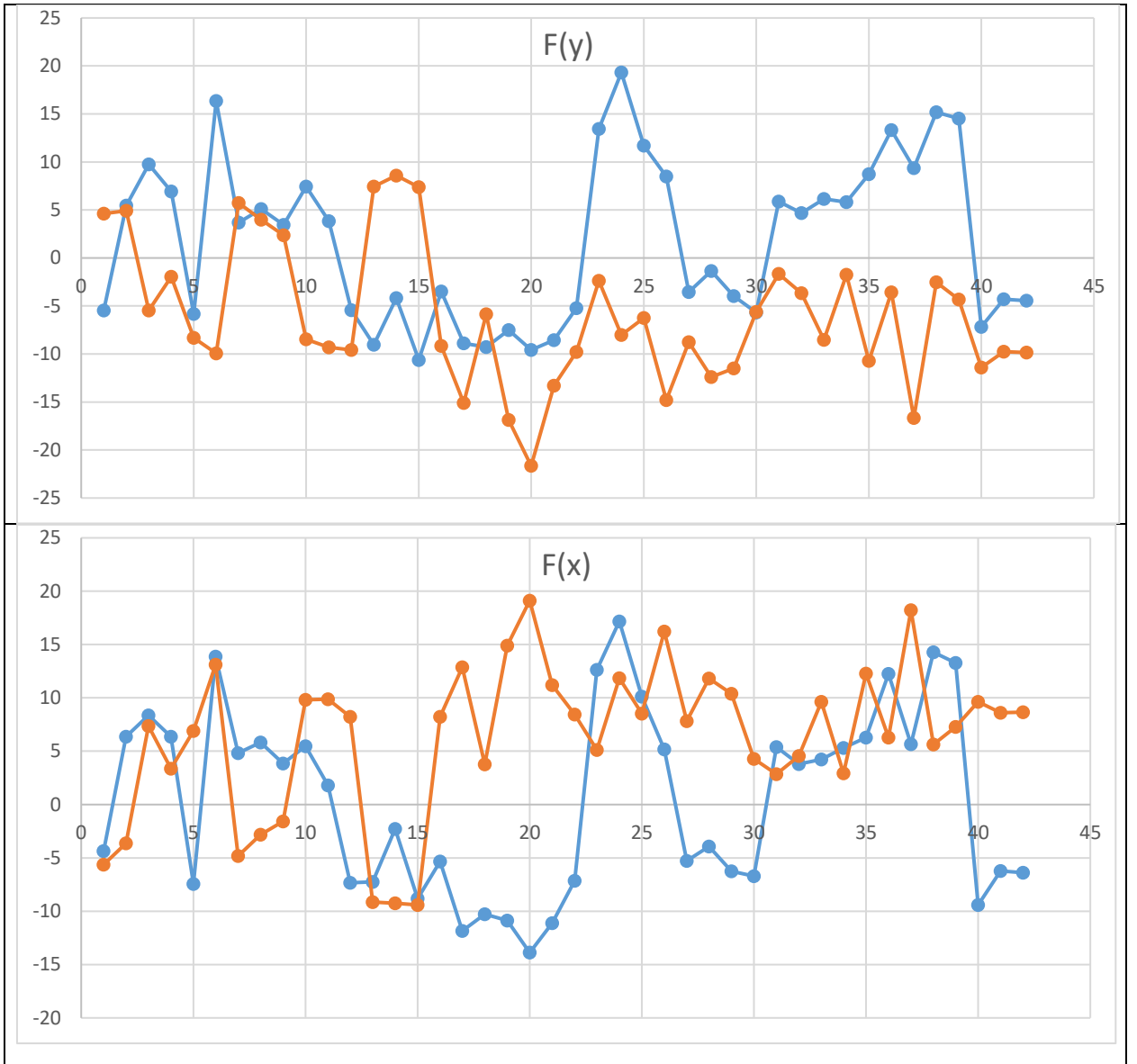
$$Y_1 = B1_1x_1 + B1_2x_2;$$

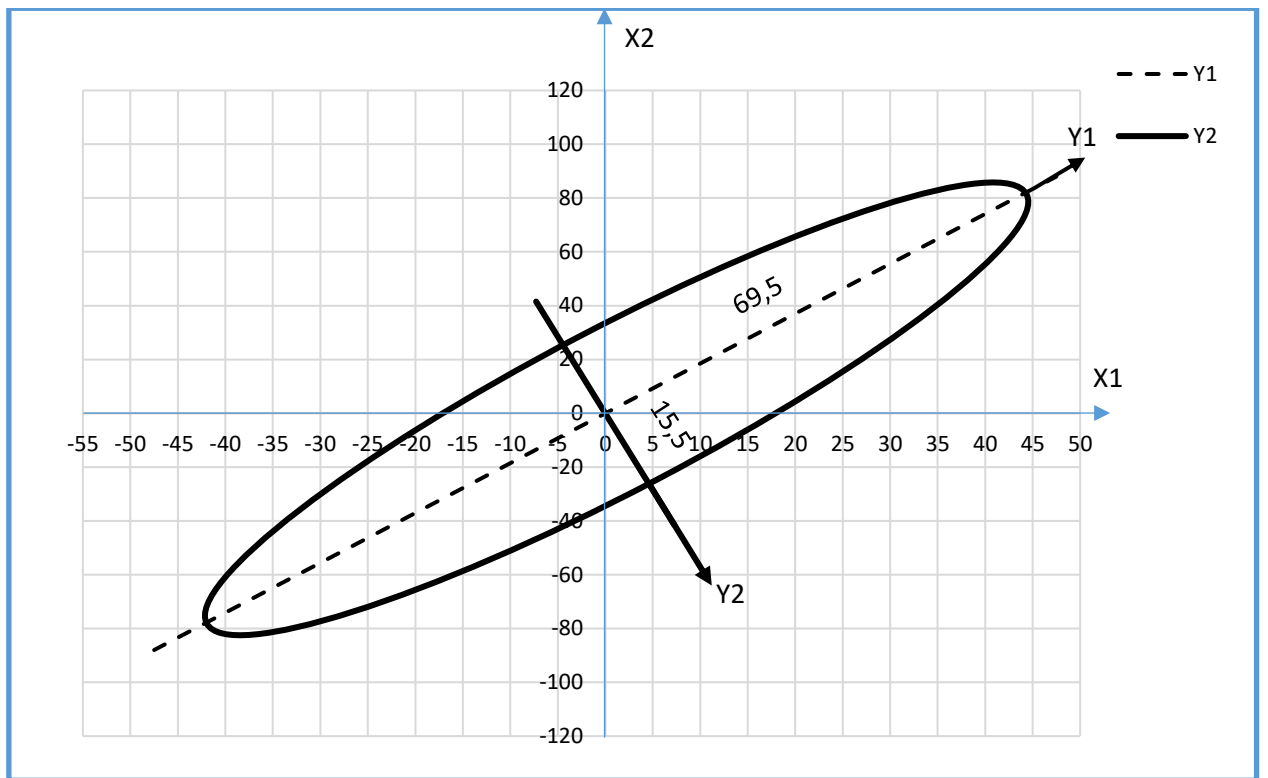
$$Y_1 = 0,98x_1 + 0,21x_2;$$

$$Y_2 = B2_1x_1 + B2_2x_2;$$

$$Y_2 = 0,21x_1 - 0,98x_2;$$

Y1	Y2
-5,4572	4,6116
5,4684	4,9028
9,7384	-5,4572
6,9384	-1,9572
-5,8422	-8,3244
16,338	-9,947
3,7016	5,7337
5,1093	3,9956
3,4412	2,3569
7,4305	-8,463
3,8388	-9,3044
-5,4474	-9,5928
-9,0244	7,4347
-4,1671	8,5883
-10,6197	7,3696
-3,4874	-9,1728
-8,8886	-15,1067
-9,2806	-5,8632
-7,5159	-16,8728
-9,5872	-21,6524
-8,5477	-13,3014
-5,2325	-9,7825
13,4428	-2,3674
19,3249	-8,0052
11,7089	-6,2342
8,4826	-14,8078
-3,5441	-8,7647
-1,3594	-12,4068
-3,955	-11,487
-5,6868	-5,6056
5,8688	-1,6534
4,6893	-3,6589
6,1558	-8,5414
5,8268	-1,7444
8,7332	-10,7156
13,3217	-3,5721
9,373	-16,667
15,1788	-2,5284
14,5313	-4,3379
-7,1876	-11,4212
-4,2973	-9,7461
-4,4513	-9,8406





Оставляем одну главную компоненту, так как она описывает значительную часть данных