

Федеральное государственное бюджетное образовательное учреждение  
высшего профессионального образования  
«Новосибирский государственный технический университет»  
Кафедра Автоматизированных систем управления

Отчет по лабораторной работе №4  
по дисциплине «Методы анализа данных»

**Факторный анализ**

Факультет: АВТФ

Группа: АВТ-412

Выполнили: Лазаревич М.М.

Евтушенко Н.С.

Проверила: Ганелина Н.Д.

Новосибирск

2017 г.

## **Цель работы:**

Познакомиться с теорией и практикой факторного анализа.

## **Постановка задачи:**

Для выбранного массива данных провести факторный анализ.

Проанализировать полученные результаты.

## **Описание исходного массива данных:**

Представленные данные были получены путём исследования изображений различных видов листов. Исследуемый массив содержит 16 атрибутов:

1. Класс
2. Номер листа
3. Эксцентриситет эллипса, описывающего лист
4. Соотношение сторон описывающего лист прямоугольника
5. Удлинение, характеризует закруглённость контура листа
6. Твёрдость: отношение площади поверхности листа к площади оптимального выпуклого многоугольника, описывающего листа.
7. Стохастическая выпуклость: вероятность случайно выбранной точки быть внутри границ листа.
8. Изопериметрический фактор: отношение площади листа к квадрату длины контура.
9. Максимальная глубина отступа: максимум отношение разности расстояний по  $x$  и  $y$  точки внутри листа до центроида внешнего выпуклого многоугольника к длине дуги
10. Дольчатость: мера, характеризующая деление листа на отдельные доли(лопасти), зависит от параметра 9.
11. Средняя интенсивность
12. Средняя контрастность: среднеквадратичное отклонение интенсивности
13. Гладкость: мера относительной гладкости интенсивности, зависит от параметра 12
14. Третий момент: мера асимметрии гистограммы интенсивности
15. Однородность: сумма квадратов частот интенсивностей по всем уровням, максимальное значение 1.
16. Энтропия: мера случайности интенсивности

Атрибуты 10 и 13 не используются в анализе, так как они зависят от параметров 9 и 12 соответственно.

## Корреляционная матрица переменных, используемых в дальнейшем.

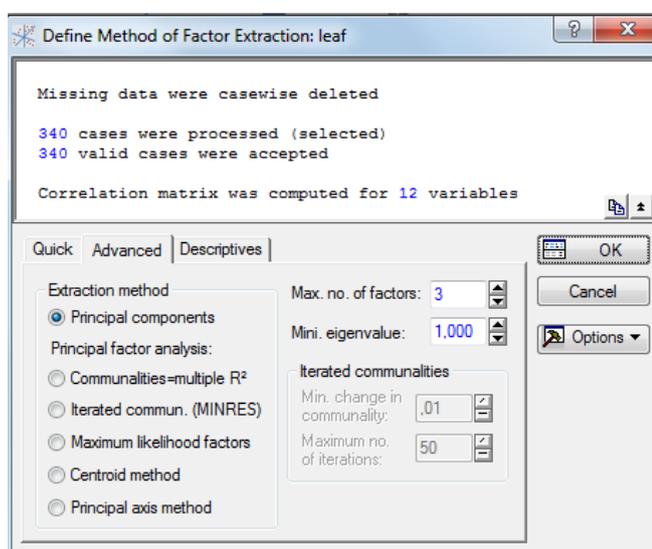
Корреляционная матрица

	Эксцентриситет	Соотношение_сторон	Удлинение	Твёрдость	Выпуклость	Изопериметрический_фактор	Максимальная_глубина_отступа	Средняя_интенсивность	Средняя_контрастность	Третий_момент	Однородность	Энтропия
Корреляция	1.000	.551	.554	.374	.386	-.036	-.274	-.226	-.195	-.150	-.250	-.240
Эксцентриситет		1.000	.678	.005	.107	-.471	.092	-.282	-.298	-.234	-.229	-.313
Соотношение_сторон	.551		1.000	-.411	-.379	-.793	.437	-.205	-.189	-.151	-.251	-.211
Удлинение	.554	.678		1.000	.863	.755	-.886	.085	.083	.059	.126	.056
Твёрдость	.374	.005	-.411		1.000	.656	-.769	.063	.078	.037	.107	.053
Выпуклость	.386	.107	-.379	.863		1.000	-.736	.081	.078	.067	.133	.070
Изопериметрический_фактор	-.036	-.471	-.793	.755	.656		1.000	-.090	-.089	-.023	-.138	-.112
Максимальная_глубина_отступа	-.274	.092	.437	-.886	-.769	-.736		1.000	.955	.918	1.000	.860
Средняя_интенсивность	-.226	-.282	-.205	.085	.063	.081	-.090		1.000	.822	.804	.936
Средняя_контрастность	-.195	-.298	-.189	.083	.058	.078	-.089	.955		1.000	.653	.860
Третий_момент	-.150	-.234	-.151	.059	.037	.067	-.023	.822	.918		1.000	.645
Однородность	-.250	-.229	-.251	.126	.107	.133	-.138	.804	.653	.409		1.000
Энтропия	-.240	-.313	-.211	.056	.053	.070	-.112	.936	.860	.645	.409	

### Ход работы:

### Факторный анализ в Statistica:

Проведём факторный анализ методом главных компонент с параметрами, изображёнными ниже:



Результатирующие нагрузки факторов при отсутствии вращения:

Variable	Factor Loadings (Unrotated) (leaf) Extraction: Principal components (Marked loadings are >.700000)		
	Factor 1	Factor 2	Factor 3
Эксцентриситет	0,287622	-0,402820	0,762182
Соотношение сторон	0,512802	-0,052131	0,715329
Удлинение	0,600519	0,382658	0,656337
Твёрдость	-0,420319	-0,841023	0,236155
Выпуклость	-0,365345	-0,800964	0,288024
Изопериметрический фактор	-0,524131	-0,738599	-0,314167
Максимальная глубина отступа	0,439575	0,799708	-0,144306
Средняя интенсивность	-0,872640	0,407500	0,250779
Средняя контрастность	-0,846464	0,398772	0,255569
Третий момент	-0,717781	0,366762	0,244020
Однородность	-0,745146	0,262660	0,139936
Энтропия	-0,829463	0,385623	0,195709
Expl.Var	4,720863	3,522618	2,027282
Prp.Totl	0,393405	0,293552	0,168940

В результате проведения факторного анализа было выделено три фактора. Первый фактор имеет большие нагрузки у переменных, являющихся

статистическими характеристиками интенсивности изображения листа. Второй фактор соответствует переменным, относящимся к характеристикам, относящимся к площади листа. Третий фактор имеет большие нагрузки у переменных “Эксцентриситет” и “Соотношение сторон”, эти переменные характеризуют форму листа, при этом нагрузка, соответствующая переменной “Удлинение” имеет значение 0.65 для фактора 3 и 0.6 для фактора 1, эту переменную нельзя явным образом отнести к тому или иному фактору.

Проведём факторный анализ тем же методом (главных компонент), только на этот раз используем метод вращения кватримакс. Данный метод должен минимизировать число факторов для объяснения каждой переменной.

Результаты проведённого анализа:

Variable	Factor Loadings (Quartimax normalized) (leaf) Extraction: Principal components (Marked loadings are >.700000)		
	Factor 1	Factor 2	Factor 3
Эксцентриситет	-0,176346	0,363540	-0,814034
Соотношение сторон	-0,226613	-0,051219	-0,850531
Удлинение	-0,128290	-0,487109	-0,827100
Твёрдость	0,059264	0,965015	-0,070649
Выпуклость	0,047116	0,914617	-0,138691
Изопериметрический фактор	0,016609	0,829270	0,480601
Максимальная глубина отступа	-0,064395	-0,921386	-0,022002
Средняя интенсивность	0,991976	0,030813	0,074040
Средняя контрастность	0,967666	0,028678	0,060300
Третий момент	0,841789	0,002888	0,025557
Однородность	0,786036	0,091332	0,132742
Энтропия	0,928586	0,024143	0,110274
Expl.Var	4,218190	3,687184	2,365390
Prp.Totl	0,351516	0,307265	0,197116

Графики нагрузок:

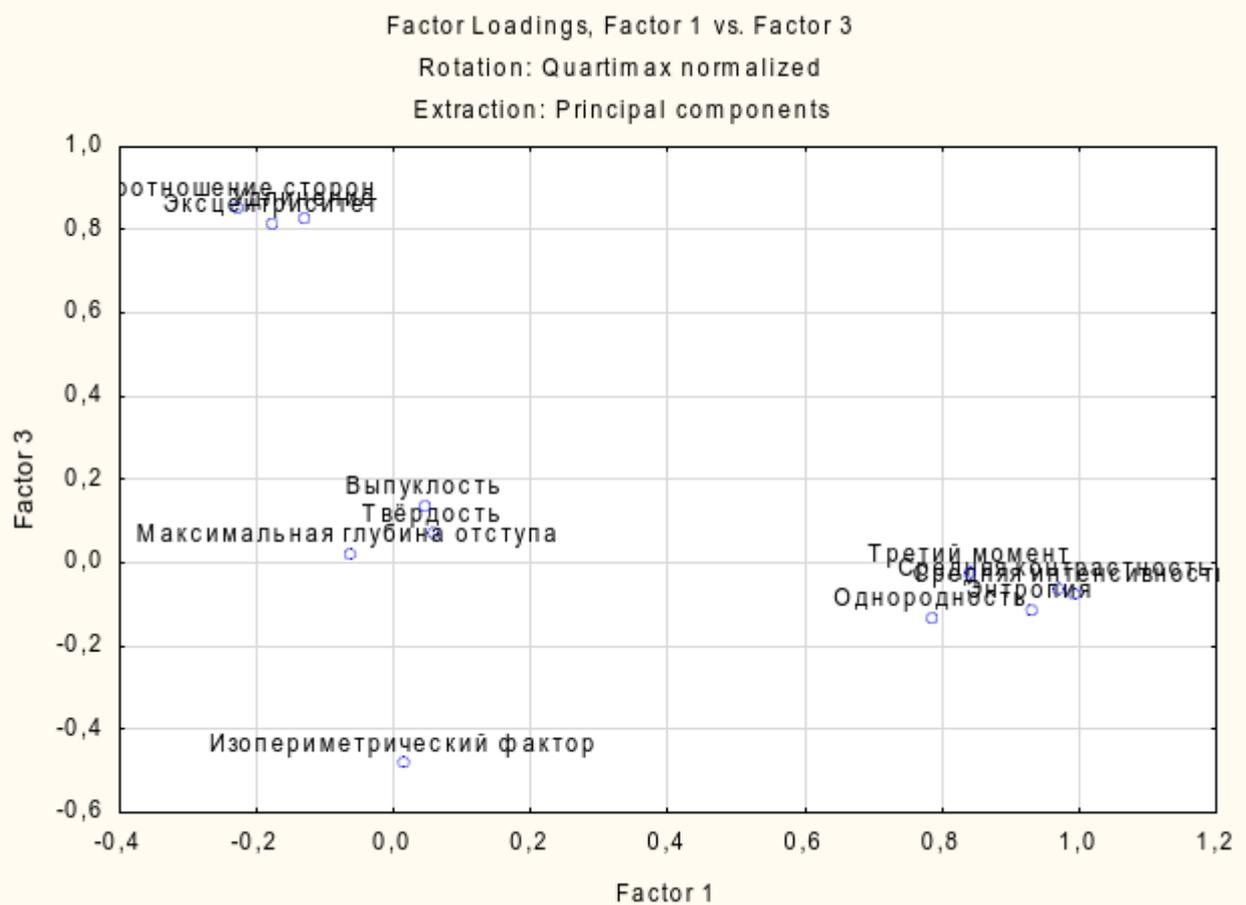
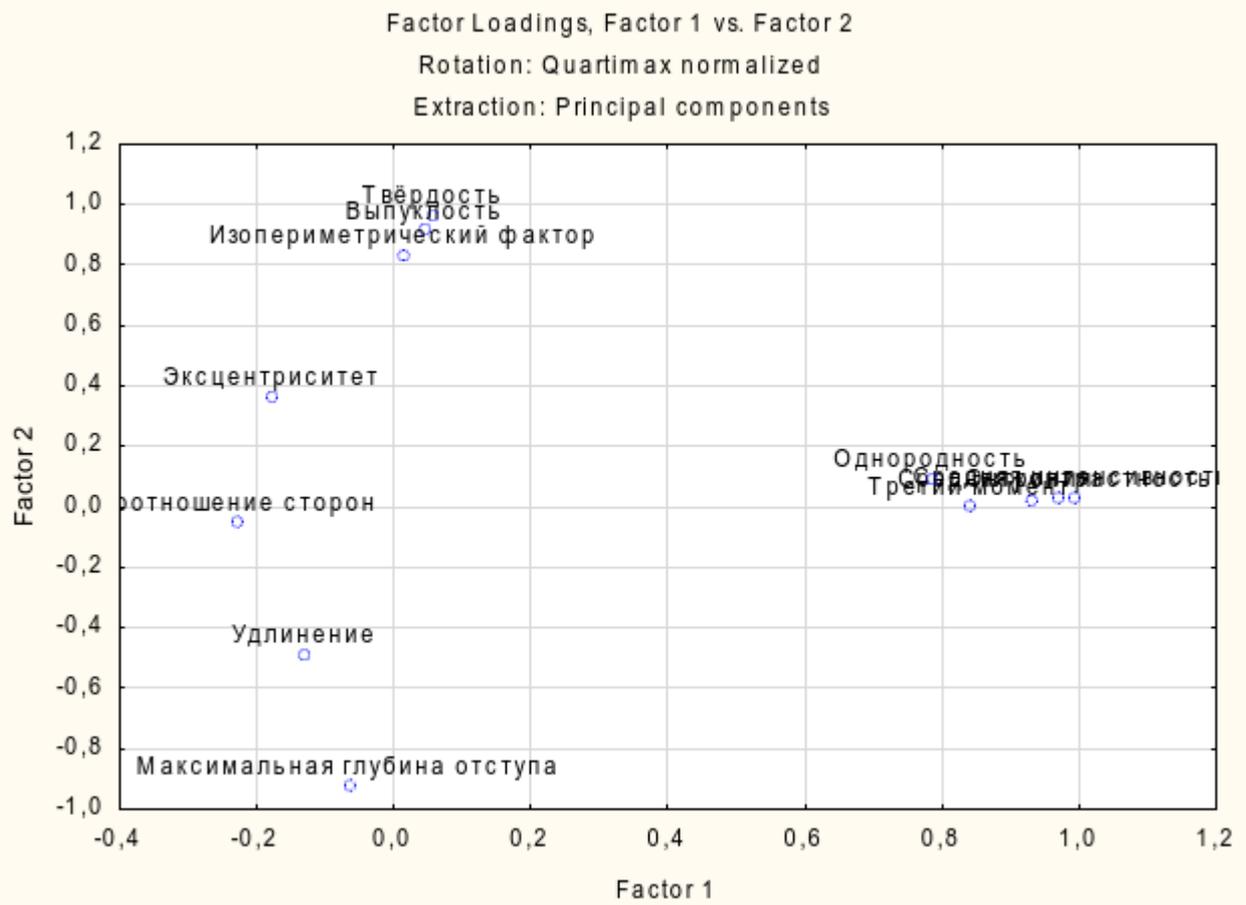
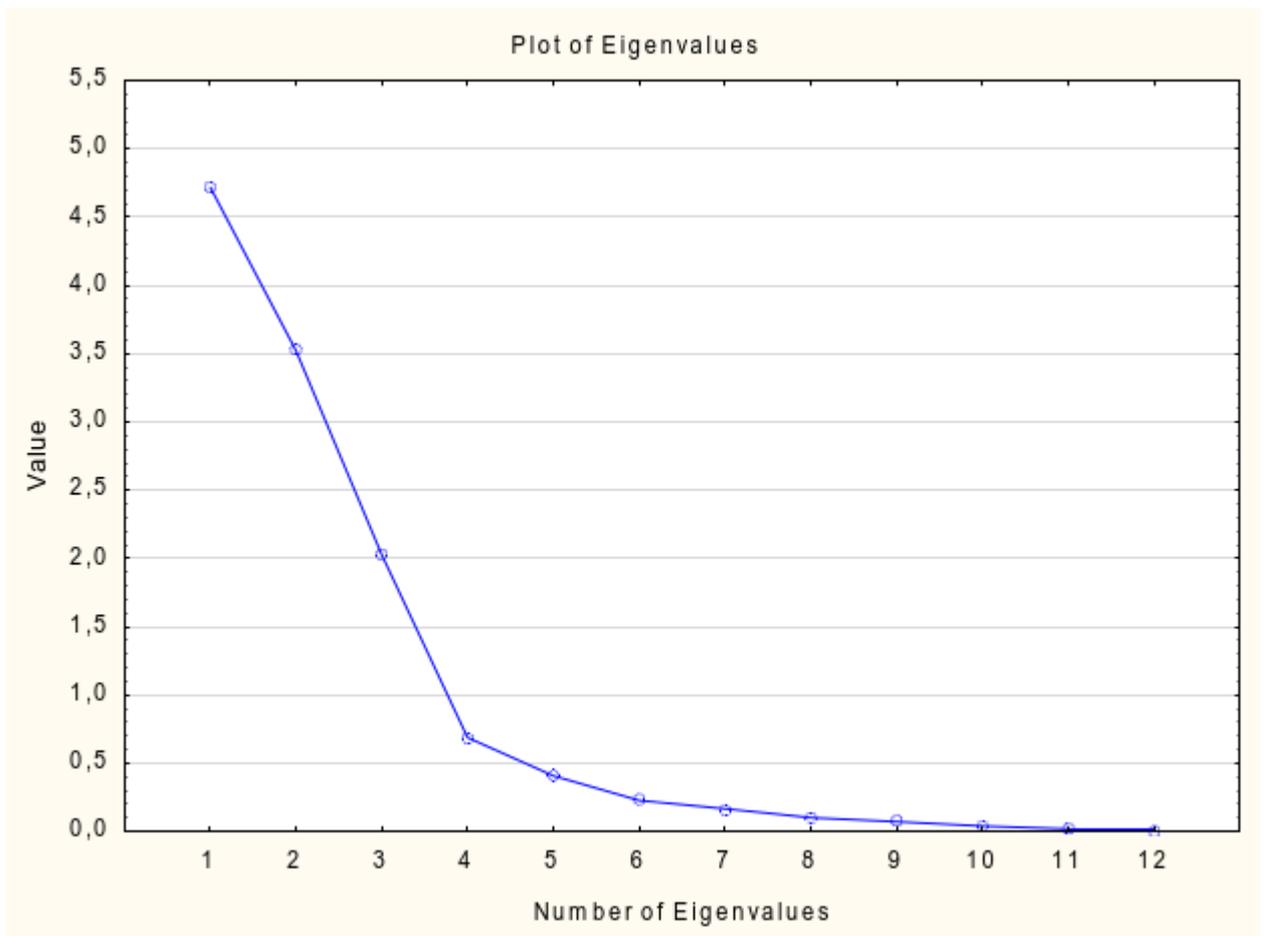


График собственных значений:



Судя по виду графика собственных значений можно сделать вывод, что число факторов в данном случае точно не превышает 4.

Как видно из таблицы нагрузок, нагрузки, который раньше превосходили значени 0.7 для всех факторов увеличили своё значение, при этом фактор 3 теперь имеет существенную нагрузку для переменной “Удлинение”. Также стоит отметить, что большая часть не существенных в прошлом анализе нагрузок теперь имеют меньшие значения.

Для проверки полученных результатов проведём анализ другим методом, на этот раз испльзуется метод "общности как множественный R-квадрат" с методом вращения варимакс.

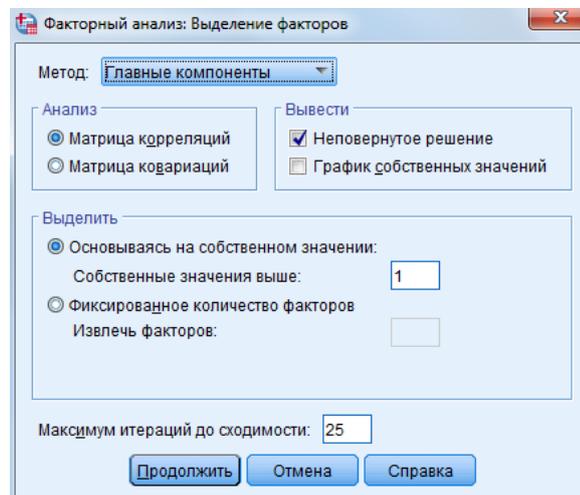
Результаты:

Variable	Factor Loadings (Varimax normalized) (leaf) Extraction: Principal factors (comm.=multiple R-square) (Marked loadings are >,700000)		
	Factor 1	Factor 2	Factor 3
Эксцентриситет	-0,151705	0,368382	-0,788110
Соотношение сторон	-0,203543	-0,040543	-0,783860
Удлинение	-0,099156	-0,468947	-0,843519
Твёрдость	0,057353	0,958465	-0,052042
Выпуклость	0,045951	0,889350	-0,110678
Изопериметрический фактор	0,000609	0,813854	0,489578
Максимальная глубина отступа	-0,059924	-0,892432	-0,040836
Средняя интенсивность	0,990658	0,033125	0,108673
Средняя контрастность	0,964590	0,031495	0,092529
Третий момент	0,837722	0,006157	0,052534
Однородность	0,760855	0,090984	0,166801
Энтропия	0,917211	0,025345	0,144769
Expl. Var	4,117060	3,536698	2,275303
Prp. Totl	0,343088	0,294725	0,189609

Полученные нагрузки факторов соответствуют результатам предыдущего анализа.

### Факторный анализ в SPSS Statistics:

Используем метод главных компонент с методом вращения варимакс, в качестве критерия определения числа факторов используется критерий Кайзера (отбираются факторы с собственными значениями не меньше единицы).



Факторный анализ: Вращение

Метод

Нет
  Квартимакс
  Варимакс
  Эквимакс
  Прямой облимин
  Промакс

Дельта:  Каппа

Вывести

Повернутое решение
  График(и) нагрузок

Максимум итераций до сходимости:

Результаты:

Матрица компонент<sup>а</sup>

	Компонент		
	1	2	3
Эксцентриситет	-.288	-.403	.762
Соотношение_сторон	-.513	-.052	.715
Удлинение	-.601	.383	.656
Твердость	.420	-.841	.236
Выпуклость	.365	-.801	.288
Изопериметрический_фактор	.524	-.739	-.314
Максимальная_глубина_отсу́па	-.440	.800	-.144
Средняя_интенсивность	.873	.408	.251
Средняя_контрастность	.846	.399	.256
Третий_момент	.718	.367	.244
Однородность	.745	.263	.140
Энтропия	.829	.386	.196

Метод выделения факторов: метод главных компонент.

а. Извлечено компонентов - 3.

**Повернутая матрица компонентов<sup>а</sup>**

	Компонент		
	1	2	3
Эксцентриситет	-.153	.376	.813
Соотношение_сторон	-.201	-.038	.858
Удлинение	-.101	-.474	.838
Твердость	.058	.966	.053
Выпуклость	.048	.917	.122
Изопериметрический_фактор	-.001	.821	-.494
Максимальная_глубина_отсуха	-.060	-.921	.039
Средняя_интенсивность	.989	.033	-.105
Средняя_контрастность	.965	.031	-.090
Третий_момент	.841	.005	-.051
Однородность	.781	.092	-.158
Энтропия	.925	.025	-.139

Метод выделения факторов: метод главных компонент.

Метод вращения: варимакс с нормализацией Кайзера.

а. Вращение сошлось за 4 итераций.

**Объясненная совокупная дисперсия**

Компонент	Начальные собственные значения			Извлечение суммы квадратов нагрузок			Ротация суммы квадратов нагрузок		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	4.721	39.341	39.341	4.721	39.341	39.341	4.165	34.712	34.712
2	3.523	29.355	68.696	3.523	29.355	68.696	3.676	30.635	65.347
3	2.027	16.894	85.590	2.027	16.894	85.590	2.429	20.243	85.590
4	.687	5.726	91.315						
5	.407	3.394	94.709						
6	.227	1.892	96.601						
7	.166	1.387	97.988						
8	.101	.845	98.833						
9	.072	.598	99.431						
10	.039	.328	99.759						
11	.019	.158	99.917						
12	.010	.083	100.000						

Метод выделения факторов: метод главных компонент.

В результате анализа было выделено 3 фактора, они объясняют 86 процентов дисперсии. В целом выводы на основе проведённого анализа совпадают с предыдущими.

Вторым используемым методом является невзвешенный метод наименьших квадратов. Метод вращения – варимакс.

Результаты:

**Факторная матрица<sup>а</sup>**

	Факторный анализ		
	1	2	3
Эксцентриситет	-.266	-.372	.708
Соотношение_сторон	-.477	-.049	.616
Удлинение	-.617	.387	.687
Твердость	.441	-.835	.268
Выпуклость	.371	-.757	.290
Изопериметрический_фактор	.534	-.724	-.294
Максимальная_глубина_отсуха	-.442	.755	-.160
Средняя_интенсивность	.881	.442	.244
Средняя_контрастность	.848	.428	.246
Третий_момент	.668	.355	.197
Однородность	.683	.253	.104
Энтропия	.810	.400	.174

Метод выделения факторов: метод невзвешенных наименьш

а. Извлечено факторов - 3. Потребовалось итераций - 6.

### Повернутая факторная матрица<sup>а</sup>

	Факторный анализ		
	1	2	3
Эксцентриситет	-.158	.361	.745
Соотношение_сторон	-.210	-.032	.751
Удлинение	-.097	-.462	.883
Твердость	.059	.978	.052
Выпуклость	.046	.884	.102
Изопериметрический_фактор	.001	.804	-.499
Максимальная_глубина_отсу́па	-.062	-.886	.046
Средняя_интенсивность	1.009	.032	-.105
Средняя_контрастность	.976	.030	-.091
Третий_момент	.778	.009	-.068
Однородность	.712	.086	-.162
Энтропия	.908	.024	-.143

Метод выделения факторов: метод невзвешенных наименьш

Метод вращения: варимакс с нормализацией Кайзера.

а. Вращение сошлось за 4 итераций.

### Объясненная совокупная дисперсия

Факторный анализ	Начальные собственные значения			Извлечение суммы квадратов нагрузок			Ротация суммы квадратов нагрузок		
	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %	Всего	% дисперсии	Суммарный %
1	4.721	39.341	39.341	4.554	37.948	37.948	3.997	33.307	33.307
2	3.523	29.355	68.696	3.381	28.173	66.121	3.524	29.368	62.675
3	2.027	16.894	85.590	1.821	15.173	81.295	2.234	18.619	81.295
4	.687	5.726	91.315						
5	.407	3.394	94.709						
6	.227	1.892	96.601						
7	.166	1.387	97.988						
8	.101	.845	98.833						
9	.072	.598	99.431						
10	.039	.328	99.759						
11	.019	.158	99.917						
12	.010	.083	100.000						

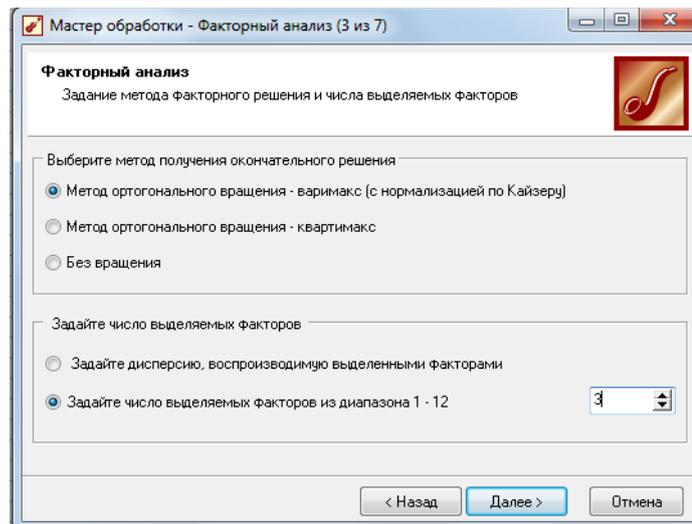
Метод выделения факторов: метод невзвешенных наименьших квадратов.

Данный метод также позволил выделить 3 фактора, но доля объяснённой дисперсии стала ниже, как и распределение нагрузок.

Также в рамках данной лабораторной была сделана попытка использовать обобщённый метод наименьших квадратов, однако в процессе анализа не удалось выделить ни одного фактора.

### Факторный анализ в Deductor:

Используем метод ортогонального вращения варимакс с нормализацией по методу Кайзера.



В результате оставим три главные компоненты.

Главные компоненты	Собственное значение	Вклад в результат	Суммарный вклад
<input checked="" type="checkbox"/> Значение 1	4,721	39,3405 %	39,3405 %
<input checked="" type="checkbox"/> Значение 2	3,523	29,3552 %	68,6957 %
<input checked="" type="checkbox"/> Значение 3	2,027	16,8940 %	85,5897 %
<input type="checkbox"/> Значение 4	0,687	05,7255 %	
<input type="checkbox"/> Значение 5	0,407	03,3940 %	
<input type="checkbox"/> Значение 6	0,227	01,8920 %	
<input type="checkbox"/> Значение 7	0,166	01,3872 %	
<input type="checkbox"/> Значение 8	0,101	00,8447 %	
<input type="checkbox"/> Значение 9	0,072	00,5976 %	
<input type="checkbox"/> Значение 10	0,039	00,3284 %	
<input type="checkbox"/> Значение 11	0,019	00,1584 %	
<input type="checkbox"/> Значение 12	0,010	00,0825 %	

Как и следовало ожидать, полученные результаты совпадают с результатами применения метода главных компонент с методом вращения варимакс в пакете SPSS Statistics.

Переменные	Окончательные факторы (Варимакс метод)		
	Фактор 1	Фактор 2	Фактор 3
Эксцентриситет			0,8130
Соотношение_сторон			0,8577
Удлинение			0,8384
Твердость		0,9662	
Выпуклость		0,9169	
Изопериметрический факт...		0,8214	
Максимальная глубина от...		-0,9211	
Средняя интенсивность	0,9891		
Средняя контрастность	0,9653		
Третий момент	0,8406		
Однородность	0,7813		
Энтропия	0,9247		

**Выводы:**

Для выбранного массива данных был проведён факторный анализ в программах Statistica, SPSS Statistics, Deductor. По результатам проведённых различными методами анализов было выделено 3 факторы, первый фактор можно назвать “Текстура листа”, второй фактор – “Форма поверхности листа”, третий фактор – “Внешний контур листа”.