

**МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ
ФГБОУ ВПО БАШКИРСКИЙ ГАУ**

Т.Н. Лубова

УЧЕБНОЕ ПОСОБИЕ
Многомерные статистические методы

Уфа

Издательство БГАУ

2015

УДК 519.21(075)

ББК 22.171

Л 82

Рекомендовано к изданию Редакционно-издательским советом
БГАУ

Автор Т.Н. Лубова

Рецензент: к.т.н., доцент кафедры вычислительной математики и
кибернетики ГОУ ВПО «Уфимский государственный авиационный
технический университет» Васильева Л.И.;

к.т.н., доцент кафедры математики ФГОУ ВПО «Башкирский
государственный аграрный университет» Анасова Т.А.

Л 82 Учебное пособие по многомерным статистическим
методам. – Уфа: Издательство БГАУ, 2015. - 64 с.

Учебное пособие предназначено для студентов обучающихся по
направлениям подготовки бакалавров 38.03.01 (080100.62)
Экономика, 38.03.05 (080500.62) Бизнес-информатика, 38.03.04
(081100.62) Государственное и муниципальное управление, 09.03.03
(230700) Прикладная информатика.

Учебное пособие включает рекомендации по изучению
основных тем по многомерным статистическим методам, список
рекомендуемой литературы, вопросы для самопроверки, контрольные
и индивидуальные задачи.

© Башкирский государственный
аграрный университет, 2015

© Т.Н. Лубова, 2015

ОГЛАВЛЕНИЕ

Многомерные статистические методы	4
Тема 1 Многомерный статистический анализ	5
Тема 2 Множественный корреляционно-регрессионный анализ	6
Тема 3 Компонентный анализ	24
Тема 4 Факторный анализ	34
Тема 5 Методы многомерной классификации. Классификация без обучения. Кластерный анализ	37
Тема 6 Дискриминантный анализ. Классификации с обучением	42
Тема 7 Канонические корреляции	49
Тема 8 Множественный ковариационный анализ	53
Индивидуальные задания	54
Библиографический список	59
Приложение	60

МНОГОМЕРНЫЕ СТАТИСТИЧЕСКИЕ МЕТОДЫ

Социально-экономические, технические, технологические, природные процессы и явления зависят от большого числа факторов, их характеризующих, что обуславливает трудности, связанные с выявлением структуры взаимосвязи изучаемых параметров. Поэтому, чтобы принимать верные решения при управлении или прогнозировании на основании анализа неполной информации (в условиях неопределённости), возникает необходимость в использовании методов многомерного статистического анализа для построения моделей, адекватных изучаемым процессам.

Методы многомерного статистического анализа следует рассматривать как логическое развитие методов теории вероятностей и математической статистики. Принципиальное отличие состоит в том, что объекты, социальные и экономические явления рассматриваются здесь с учётом не одного - двух, а одновременно некоторого множества признаков.

Развитие многомерного статистического анализа как науки начинается с 1901-1904 гг. - в это время появились статьи К. Пирсона и Ч. Спирмена, посвященные теории факторного анализа.

Методы многомерного статистического анализа базируются на представлении исходной информации в многомерном признаковом пространстве и позволяют определить неявные, но объективно существующие закономерности в организационной структуре и тенденциях развития изучаемых явлений и процессов. Практически все методы многомерного статистического анализа можно разбить на

две большие группы: 1) методы вероятностного анализа многомерных данных; 2) методы логико-алгебро-геометрического направления. Следует отметить, что любая классификация статистических методов является условной - многие методы основываются и на вероятностной, и на алгебро-геометрической идеологии.

В многомерном статистическом анализе обобщаются методы проверки статистических гипотез, статистического оценивания, дисперсионный, корреляционно-регрессионный и другие разделы математической статистики на многомерный случай. Кроме того, в многомерном статистическом анализе рассматриваются методы сжатия информации, классификации, снижения размерности исследуемого признакового пространства и отбора наиболее информативных показателей.

ТЕМА 1 МНОГОМЕРНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ

Данная тема включает в себя:

- области применения многомерного статистического анализа;
- характеристика основных методов.

Вопросы для самопроверки

1 Сформулируйте основные области применения многомерного статистического анализа.

2 Дайте характеристику основных методов многомерного статистического анализа.

ТЕМА 2 МНОЖЕСТВЕННЫЙ КОРРЕЛЯЦИОННО-РЕГРЕССИОННЫЙ АНАЛИЗ

Данная тема включает в себя:

- цели и задачи множественного корреляционно-регрессионного анализа;
- основные понятия;
- двумерная модель:
 - точечные оценки параметров,
 - приемы вычисления выборочных характеристик,
 - проверка значимости параметров связи,
 - интервальные оценки параметров связи,
 - задачи, решаемы при помощи статистики Фишера;
- трехмерная модель:
 - основные параметры модели,
 - оценивание и проверка значимости параметров;
- вопросы корреляционного анализа многомерной модели;
- простейшее линейное уравнение регрессии:
 - нахождение оценки уравнения,
 - определение интервальных оценок для коэффициентов регрессии, проверка их значимости,

- определение интервальной оценки для условного математического ожидания;
- множественное линейное уравнение регрессии:
 - нахождение и анализ свойств оценки уравнения регрессии,
 - проверка значимости уравнения регрессии,
 - доверительные интервалы для параметров линейной модели,
 - взвешенный метод наименьших квадратов.

Контрольные задачи

1 По результатам выборки объемом $n=20$ единицам из двумерной нормальной генеральной совокупности получены выборочные характеристики: $b_{yx} = -0,50$; $s_x^2 = 31,16$; $s_y^2 = 9,21$. Проверить при $\alpha=0,05$ значимость генерального коэффициента корреляции.

2 Из двумерной нормальной генеральной совокупности взята выборка объемом $n=32$ единицы и найдены выборочные коэффициенты регрессии $b_{yx} = -0,50$ и $b_{xy} = -1,71$. Проверить значимость генерального коэффициента корреляции при $\alpha=0,01$,

3 На основании шестнадцати выборочных данных получены следующие характеристики: $s_x^2 = 10,5$ и $s_{x/y}^2 = 4,5$. Требуется при $\alpha=0,05$ проверить значимость генерального коэффициента корреляции.

4 По результатам трех выборок из совокупности (x,y) имеем:

$n_1=20$; $r_1=0,80$; $n_2=30$; $r_2=0,85$ и $n_3=30$; $r_3=0,9$. Требуется при $\alpha=0,05$ проверить гипотезу однородности коэффициентов корреляции.

5 На основе ста выборочных данных об объеме и весе одного вида кондитерских изделий было выяснено, что при увеличении веса на 1г объем увеличивается в среднем на $0,3\text{см}^3$, увеличению объема на 1см^3 соответствует увеличение веса в среднем на 2,7г. Необходимо:

а) при $\alpha=0,05$ проверить значимость генерального коэффициента корреляции;

б) с надежностью $\gamma=0,954$ найти доверительный интервал для генерального коэффициента корреляции между объемом и весом кондитерских изделий;

в) с надежностью $\gamma=0,95$ найти длину интервальной оценки генерального коэффициента регрессии объема на вес кондитерских изделий;

г) с надежностью $\gamma=0,9$ найти интервальную оценку коэффициента регрессии веса на объем кондитерских изделий.

6 На основании пятидесяти выборочных данных о выручке и массе реализованной сельскохозяйственной продукции было выяснено, что выборочная доля дисперсии выручки, вызванная случайной вариацией выручки без учета вариации массы, составляет 9%. Требуется;

а) при $\alpha=0,01$ проверить значимость генерального коэффициента корреляции;

б) с надежностью $\gamma=0,95$ найти интервальную оценку ρ ,

в) при $\gamma=0,9$ определить интервальную оценку коэффициента регрессии выручки от массы реализованной продукции.

7 По данным ста выборочных наблюдений об объеме реализованной продукции некоторого вида и себестоимости единицы продукции было выяснено, что при увеличении объема на единицу его измерения себестоимость единицы продукции уменьшается в среднем на 0,4 единицы своего измерения, а при увеличении себестоимости на единицу объем реализованной продукции уменьшается в среднем на 1,6 единицы. С надежностью $\gamma =0,99$ найти:

а) интервальную оценку генерального коэффициента корреляции между объемом реализации и себестоимостью единицы продукции;

б) интервальную оценку коэффициента регрессии себестоимости единицы продукции от объема реализации.

8 На основании двадцати наблюдений было выяснено, что выборочная доля дисперсии случайной величины y , вызванная вариацией x , составляет 64%. Найти интервальную оценку коэффициента корреляции ρ с надежностью $\gamma=0,543$.

9 По данным обследования тридцати хозяйств для анализа зависимости между средней урожайностью (x) и средней себестоимостью (y) хлопка были получены следующие выборочные характеристики: $r=-0,56$; $S_x=5,3$ т/га; $S_y=2,5$ тыс. руб. С надежностью $\gamma=0,90$ найти нижнюю границу оценки генерального коэффициента регрессии себестоимости на урожайность.

10 По результатам шестнадцати выборочных наблюдений

получены выборочные коэффициенты регрессии: $b_{yx} = -0,50$, $b_{xy} = -1,72$. С надежностью $\gamma=0,95$ найти доверительные интервалы для коэффициента корреляции ρ и коэффициента регрессии β_{yx} .

11 Выборочное среднее квадратическое отклонение и условное среднее квадратическое отклонение случайной величины y при фиксированной случайной величине x , полученные на основании выборки объемом 30 единиц, равны соответственно 6,25; 2,25. Проверить с $\alpha=0,05$ гипотезу о том, что генеральный коэффициент корреляции равен 0,95, если выборочный коэффициент корреляции положителен.

12 На основании пяти выборочных точек (2,5), (3,5), (1,3), (2,4), (3,4) из генеральной совокупности (x, y) проверить с $\alpha=0,02$ гипотезу $H_0: \rho=0,6$.

13 По результатам пяти наблюдений (1,5), (2,4), (3,1), (4,1), (5,2) проверить с $\alpha=0,03$ гипотезу $H_0: \rho=-0,6$.

14 Проверить с $\alpha=0,01$ гипотезу $H_0: \rho=-0,7$, если известны: $b_{xy} = -0,3$; $s_x^2 = 2,25$; $s_y^2 = 6,25$, $n=40$.

15 Для вычисления взаимозависимости между себестоимостью 1т песка (z), сменной добычей песка (y) и фондоотдачей (x) было обследовано восемь карьеров. В результате получены следующие данные:

X	30	20	40	35	45	25	50	30
$Y, \text{ т}$	20	30	50	70	80	20	90	25
$Z, \text{ тыс. руб.}$	20	25	20	15	10	30	10	20

Вычислить:

- а) матрицу выборочных парных коэффициентов корреляции;
- б) выборочный частный коэффициент корреляции между фондоотдачей и сменной добычей песка;
- в) выборочный частный коэффициент корреляции между фондоотдачей и себестоимостью 1 т песка;
- г) выборочный множественный коэффициент корреляции себестоимости 1 т песка на остальные показатели;
- д) выборочный множественный коэффициент корреляции фондоотдачи на остальные показатели;
- е) выборочный множественный коэффициент корреляции сменной добычи песка на остальные показатели.

16 На основании годовых отчетных данных пяти строительномонтажных предприятий были получены значения следующих показателей (таблица).

Найти:

- а) матрицу выборочных парных коэффициентов корреляции;
- б) выборочный частный коэффициент корреляции между объемом выполненных работ и численностью рабочих;

№ предприятия	Объем выполненных работ, млрд. руб.	Численность рабочих, чел.	Фонд заработной платы, млрд. руб.
1	13	320	3,2
2	14	570	5,5
3	16	780	8,0
4	12	200	2,5
5	15	700	7,2

в) выборочный частный коэффициент корреляции между объемом выполненных работ и фондом заработной платы;

г) выборочный частный коэффициент корреляции между численностью рабочих и фондом заработной платы;

д) выборочный множественный коэффициент корреляции между объемом выполненных работ и остальными показателями;

е) выборочный множественный коэффициент корреляции между численностью рабочих и остальными показателями.

17 По результатам $n=100$ наблюдений найдены выборочные характеристики трехмерной генеральной совокупности:

$$\bar{x} = 4; \quad s_x = 2; \quad r_{xy} = -0,6$$

$$\bar{y} = 5; \quad s_y = 2; \quad r_{xz} = 0,8$$

$$\bar{z} = 7; \quad s_z = 3; \quad r_{zy} = -0,6$$

Требуется:

а) проверить значимость:

частных коэффициентов корреляции при $\rho_{xy/z}$, $\rho_{xz/y}$ и $\rho_{yz/x}$ при $\alpha=0,05$; множественного коэффициента корреляции $\rho_{x/yz}$ при $\alpha=0,01$;

б) найти интервальную оценку для $\rho_{xy/z}$ с надежностью $\gamma=0,925$.

18 По результатам $n=30$ наблюдений найдено $s_y^2 = 4,1$; $r_{yz} = 0,81$; $s_{y/yz}^2 = 0,4$. Требуется проверить при $\alpha=0,05$ значимость частного коэффициента корреляции $\rho_{xy/z}$.

19 На основании $n=20$ наблюдений найдено $s_z^2 = 9,2$; $r_{yz}^2 = 0,36$ и $s_{z/xy}^2 = 1$.

Требуется:

а) при $\alpha=0,05$ проверить значимость частного коэффициента корреляции $\rho_{xz/y}$

б) с надежностью $\gamma=0,92$ найти интервальную оценку $\rho_{xz/y}$.

20 По результатам $n=100$ наблюдений найдены частные коэффициенты регрессии $b_{xy/z} = -0,4$, $b_{yx/z} = -0,9$.

Требуется:

а) проверить $\alpha=0,05$ значимость частного коэффициента корреляции $\rho_{xy/z}$;

б) с надежностью $\gamma=0,95$ найти интервальную оценку для $\rho_{yz/x}$.

21 На основании $n=50$ наблюдений найдены выборочные характеристики $b_{yz/x} = -10$, $b_{zy/x} = -0,081$. Проверить при $\alpha=0,01$ значимость частного коэффициента корреляции $\rho_{yz/x}$ и с надежностью $\gamma=0,95$ найти интервальную оценку для $\rho_{yz/x}$.

22 По результатам $n=50$ наблюдений найдены выборочные характеристики $s_y^2 = 4,1$, $r_{xz}^2 = 0,36$ и $s_{yz/x}^2 = 0,81$. Определить значимость множественного коэффициента корреляции $\rho_{y/xz}$ при $\alpha=0,05$,

23 На основании $n=30$ наблюдений найдено $s_z^2 = 4,1$, $r_{yz}^2 = 0,36$, $r_{zx/y}^2 = 0,49$. Проверить при $\alpha=0,01$ значимость множественного коэффициента корреляции $\rho_{z/xy}$.

24 По данным задачи 17 найти интервальные оценки для:

а) частного коэффициента регрессии $\beta_{yx/z}$ с надежностью $\gamma=0,95$;

б) частного коэффициента регрессии $\beta_{xz/y}$ при $\gamma=0,9$.

Определить точечные оценки условного математического ожидания x при $y=5$ и $z=7$, а также условного математического ожидания y при $x=4$, $z=7$.

25 По отчетным данным $n=14$ машиностроительных предприятий:

№ предприятия	x	y	z	№ предприятия	x	y	z
1	9,4	62	11	8	6,6	77	19
2	9,9	53	9	9	5,5	32	7
3	9,1	56	23	10	9,4	200	14
4	5,6	30	10	11	5,7	91	8
5	6,7	18	9	12	5,2	82	18
6	4,3	14	5	13	10,0	76	17
7	7,4	90	10	14	6,7	37	13

провести корреляционный анализ взаимосвязи следующих показателей эффективности: x - производительность труда (млн. руб./чел.); y - индекс снижения себестоимости продукции (%); r - рентабельность (%).

26 По отчетным данным $n=12$ машиностроительных предприятий, представленных в таблице, провести корреляционный анализ взаимосвязи следующих показателей эффективности: x - фондоотдача активной части на 1 руб. ОПФ (руб.); y - рентабельность (%); z - производительность труда (млн. руб./чел.).

Таблица Отчетные данные машиностроительных предприятий

№ предприятия	x	y	z	№ предприятия	x	y	z
1	2,0	13,2	9,4	7	0,7	6,6	5,5
2	1,2	12,9	6,7	8	1,3	19,1	6,6
3	2,6	17,2	10,0	9	1,1	9,9	7,4
4	1,8	17,5	5,2	10	0,6	5,4	4,3
5	1,4	8,0	5,7	11	0,9	9,1	6,6
6	2,3	14,2	9,4	12	1,0	9,7	5,5

27 По данным $n=14$ цементных заводов, представленных в таблице, провести корреляционный анализ взаимосвязи следующих показателей эффективности: x - фондоотдачи, y - выработки натурального цемента на одного работающего (млн.т); z - среднезаводской себестоимости 1т цемента.

№ предприятия	x	y	z	№ предприятия	x	y	z
1	26,2	1,2	19,0	8	35,2	1,7	11,9
2	38,4	1,6	10,3	9	62,5	1,8	10,9
3	31,6	1,1	11,9	10	29,2	1,2	12,8
4	42,7	2,2	10,5	11	38,7	1,5	9,2
5	32,9	1,7	11,6	12	56,7	1,8	6,7
6	58,2	1,6	7,8	13	23,1	0,7	13,2
7	44,8	0,9	9,7	14	48,4	1,9	9,3

28 По данным годовой отчетности $n=30$ угольных шахт найдены следующие выборочные характеристики:

$$\bar{x}_1 = 34,3; \quad s_1 = 12,9; \quad r_{12} = -0,427;$$

$$\bar{x}_2 = 0,02; \quad s_2 = 0,0053; \quad r_{13} = -0,750$$

$$\bar{x}_3 = 568,8; \quad s_3 = 167,2; \quad r_{23} = 0,102$$

Требуется провести корреляционный анализ, если x_1 - среднемесячная производительность труда рабочего по добыче угля (т); x_2 - фондоотдача; x_3 - трудоемкость работ по добыче (человеко-дни на 1000 т).

29 На основании данных годовых отчетов $n=20$ цементных заводов вычислены:

$$\begin{array}{lll} \bar{x}_1 = 35,4; & s_1 = 12,1; & r_{12} = -0,429; \\ \bar{x}_2 = 1417; & s_2 = 633; & r_{13} = -0,520 \\ \bar{x}_3 = 12,5; & s_3 = 3,2; & r_{23} = -0,706 \end{array}$$

Требуется провести корреляционный анализ, если x_1 - фондоотдача на 1 руб. ОПФ (кг); x_2 - выработка натурального цемента на одного работающего (т); x_3 - среднезаводская себестоимость 1 т цемента (тыс. руб.).

30 По данным годовых отчетов $n=30$ угольных шахт найдены следующие выборочные характеристики:

$$\begin{array}{lll} \bar{x}_1 = 34,3; & s_1 = 12,9; & r_{12} = -0,229; \\ \bar{x}_2 = 59,5; & s_2 = 21,83; & r_{13} = -0,750 \\ \bar{x}_3 = 568,8; & s_3 = 167,2; & r_{23} = 0,361 \end{array}$$

Провести корреляционный анализ, если x_1 - среднемесячная производительность труда рабочего по добыче угля (т); x_2 - фондоемкость (тыс. руб./т.); x_3 - фондоемкость работ по добыче (человеко-дни на 100 т).

31 На основании данных о зависимости усушки формового хлеба от продолжительности хранения требуется:

а) найти оценки b_0 , b_1 и остаточной дисперсии s^2 в предположении, что генеральное сравнение регрессии имеет вид $\tilde{y} = \beta_0 + \beta_1 x$;

б) проверить при $\alpha = 0,05$ значимости уравнения регрессии, т.е. гипотезу $H_0: \beta_1 = 0$;

в) с надежностью $\gamma = 0,9$ определить интервальные оценки параметров β_0, β_1 ;

г) с надежностью $\gamma = 0,95$ определить интервальную оценку условного математического ожидания \tilde{y} при $x^0 = 6$;

д) определить при $\gamma = 0,95$ доверительный интервал предсказания y_{n+1} в точке $x^0 = 12$.

Продолжительность хранения (ч)	1	3	6	8	10
Усушка (% к весу горячего хлеба)	1,6	2,4	2,8	3,2	3,3

32 На основании данных о динамике процента хронических больных на 1000 жителей, приведенных в таблице.

Годы, x	0	1	2	3	4
%, y	10	8	5	3	4

и предположения, что генеральное уравнение регрессии имеет вид $\tilde{y} = \beta_0 + \beta_1 x$:

а) определить оценки b_0 и b_1 параметров уравнения регрессии и остаточной дисперсии s^2 ;

б) проверить при $\alpha = 0,01$ значимость уравнения регрессии, т.е. гипотезу $H_0: \beta_1 = 0$;

в) с надежностью $\gamma = 0,95$ определить интервальные оценки параметров β_0 и β_1 ;

г) с надежностью $\gamma=0,9$ определить интервальную оценку условного математического ожидания \tilde{y} при $x^0=4$;

д) определить при $\gamma=0,9$ доверительный интервал предсказания \tilde{y}_{n+1} в точке $x=5$.

33 Результаты исследования динамики веса поросят приведены в таблице.

Возраст (недели) (x)	0	1	2	3	4	5	6
Вес (кг) (y)	1,2	2,5	3,9	5,2	6,4	7,7	9,2

Предполагая, что генеральное уравнение регрессии - линейное:

а) определить оценки b_0 и b_1 параметров уравнения регрессии и остаточной дисперсии s^2 ;

б) проверить при уровне значимости $\alpha=0,05$ значимость уравнения регрессии, т.е. гипотезу $H_0: \beta_1=0$;

в) с надежностью $\gamma=0,8$ определить интервальную оценку параметров β_0 и β_1 ;

г) с надежностью $\gamma=0,98$ определить интервальную оценку условного математического ожидания \tilde{y} при $x^0=3$ и $x^0=6$;

д) определить при $\gamma=0,98$ доверительный интервал предсказания y_{n+1} в точке $x=8$.

34 При анализе зависимости удельного объема валовой продукции совхозов области (y) от средней мощности тракторов (x), приходящихся на одного работника, получены следующие результаты:

Таблица Зависимость удельного объема валовой продукции совхозов от факторов

Средняя мощность тракторов, л.с.	Число совхозов	Валовая продукция на одного работника, тыс. руб.
4,1	10	1,4
6,1	27	2,0
7,8	34	2,5
9,9	22	2,6
12,4	7	3,2

Требуется:

- определить оценки b_0 и b_1 параметров линейного уравнения регрессии и остаточной дисперсии s^2 ;
- проверить при $\alpha=0,01$ значимость уравнения регрессии, т.е. гипотезу $H_0: \beta_1=0$;
- с надежностью $\gamma=0,9$ определить интервальную оценку параметров b_0 , b_1 и условного математического ожидания \tilde{y} при $x=10$.

35 Себестоимость y (тыс.руб.) одного экземпляра книги в зависимости от тиража x (тыс. экз.) характеризуется данными, собранными издательством (таблица).

x (тыс. экз.)	1	2	3	5	10	20	30	50
y (тыс.руб.)	9,10	5,30	4,11	2,83	2,11	1,62	1,41	1,30

Определить оценки b_0 и b_1 параметров уравнения регрессии гиперболического вида $\tilde{y} = \beta_0 + \beta_1 \frac{1}{x}$.

С надежностью $\gamma=0,9$ построить доверительные интервалы для параметров b_0 и b_1 , а также условного математического ожидания y при $x=10$.

36 Данные о расходе электроэнергии (кВт.-ч) на изготовление 1 т цемента (y) в зависимости от объема выпуска (x) продукции (тыс. т) цементными заводами приводятся в таблице.

Выпуск продукции (тыс. т) x	5	10	15	20	25	30
Расход электроэнергии (кВт.-ч) y	10,0	8,2	7,3	6,3	6,4	5,2

Определить оценки b_0 и b_1 параметров уравнения регрессии вида $\tilde{y} = \beta_0 + \beta_1 \frac{1}{x}$, проверив при $\alpha=0,05$ гипотезу $H_0: \beta_1=0$, и построить с надежностью $\gamma=0,9$ доверительные интервалы для параметров b_0 и b_1 и условного математического y ожидания при $x=20$.

37 Измерение некоторой величины y через равные интервалы x дали результаты, приведенные в таблице.

X	-3	-2	-1	0	1	2	3
Y	-1,70	-1,01	-0,21	0,52	0,73	1,30	1,50

Считая, что генеральное уравнение регрессии имеет вид $\tilde{y} = \beta_0 + \beta_1 x + \beta_2 x^2$, определить:

- оценки b_0 , b_1 и b_2 параметров уравнения;
- оценку остаточной дисперсии s^2 ;
- оценки дисперсий $\sigma_{b_j}^2$ выборочных характеристик b_j , где $j=0,1,2$.

Проверить:

- значимость при $\alpha=0,05$ уравнения регрессии, т.е. гипотезу $H_0: \beta_0 = \beta_1 = \beta_2 = 0$;
- при $\alpha = 0,05$ значимость отдельных коэффициентов регрессии.

38 Осредненные данные по группе хозяйств, характеризующие связь между выходом продукции на 1 га сельскохозяйственных угодий (y) и основными средствами совхозов, также приходящимися на 1 га угодий (x), приведены в таблице.

x	11,3	12,9	13,6	16,8	18,8	22,0	22,2	23,7
y	13,2	15,6	17,2	18,8	20,2	23,3	22,4	23,0

Требуется:

- а) найти оценки параметров линейного уравнения регрессии, т.е. $\tilde{y} = \beta_0 + \beta_1 x$;
- б) вычислить несмещенную оценку s^2 остаточной дисперсии;
- б) при $\alpha = 0,05$ проверить значимость линейного уравнения регрессии;
- в) построить с надежностью $\gamma = 0,95$ доверительные интервалы параметров β_0 , β_1 и условного математического ожидания \tilde{y} при $x = 22$.

39 По данным задачи 38:

- а) определить оценки b_0 , b_1 и b_2 параметров уравнения регрессии вида $\tilde{y} = \beta_0 + \beta_1 x + \beta_2 x^2$;
- б) вычислить несмещенную оценку s^2 остаточной дисперсии и сравнить ее с соответствующей оценкой в задаче 65;
- в) проверить при $\alpha = 0,05$ значимость уравнения регрессии, т.е. гипотезы $H_0: \beta_0 = \beta_1 = \beta_2 = 0$;
- г) проверить при $\alpha = 0,05$ значимость коэффициентов регрессии, т.е. гипотезу $H_0: \beta_j$, где $j = 0, 1, 2$.

д) построить с надежностью $\gamma=0,9$ доверительные интервалы параметров $\beta_0, \beta_1, \beta_2$, и условного математического ожидания \tilde{y} при $x=22$.

40 Данные анализа зависимости средней себестоимости хлопка (y) от урожайности приведены в таблице.

Средняя урожайность (ц/га) x	13	16	22	28	33
Средняя себестоимость 1 ц (тыс.руб.) y	26	22	18	15	14

Считая, что генеральное уравнение регрессии имеет вид

$$\tilde{y} = \beta_0 + \beta_1 \frac{1}{x}, \text{ определить:}$$

- оценки b_0, b_1 параметров уравнения регрессии;
- оценки остаточной дисперсии s^2 ;
- точечные оценки дисперсий выборочных характеристик $b_0,$

b_1 и условной средней $\tilde{y} = \beta_0 + \beta_1 \frac{1}{x}$ при $x=16$.

41 Результаты лабораторных испытаний прочности стальных проволок различных диаметров приведены в таблице.

Диаметр проволоки (мм) x	1	1,2	1,4	1,5	1,6
Разрывное усилие (т) y	0,15	0,20	0,3	0,4	0,6

Приняв, что генеральное уравнение регрессии имеет вид

$$\tilde{y} = \beta_0 + \beta_1 x^2, \text{ определить:}$$

- оценки b_0, b_1 параметров β_0, β_1 ;
- несмещенную оценку s^2 остаточной дисперсии;
- интервальные оценки неизвестных параметров β_0, β_1 и

условного математического ожидания \tilde{y} при $x^0=1$, приняв $\gamma=0,9$.

42 По приведенным в таблице данным $n=14$ машиностроительных предприятий провести регрессионный анализ зависимости индекса снижения себестоимости продукции (y) от трудоемкости единицы продукции (x_1) и удельного веса покупных изделий (x_2).

№ п/п	y	x_1	x_2	№ п/п	y	x_1	x_2
1	204	0,23	0,40	8	56	0,26	0,44
2	209	0,24	0,26	9	52	0,49	0,17
3	222	0,19	0,40	10	46	0,36	0,39
4	236	0,17	0,50	11	53	0,37	0,33
5	62	0,23	0,40	12	31	0,43	0,25
6	53	0,43	0,19	13	146	0,35	0,32
7	172	0,31	0,25	14	18	0,38	0,02

43 По данным $n=12$ угольных шахт провести регрессионный анализ зависимости полной себестоимости добычи 1 т угля y (тыс. руб.) от среднесуточной добычи угля на шахте - x_1 (т) и удельного веса комбайновой проходки выработки - x_2 (%).

№ п/п	y	x_1	x_2	№ п/п	y	x_1	x_2
1	12,2	4795	69	7	12,7	4888	55
2	7,6	6962	82	8	10,5	6237	81
3	10,4	6571	87	9	15,1	2997	65
4	9,9	4249	92	10	10,6	2990	98
5	15,7	9540	23	11	15,2	1748	100
6	14,0	3488	31	12	17,2	2128	69

44 На основании данных $n=16$ цементных заводов провести регрессионный анализ зависимости выработки натурального цемента на одного работающего y (т/чел.) от среднечасовой

производительности вращающихся печей x_1 (т/ч) и среднечасовой производительности цементных мельниц x_2 (т/ч).

№ п/п	y	x_1	x_2	№ п/п	y	x_1	x_2
1	996	37	46	9	1620	44	38
2	1362	23	44	10	3018	34	32
3	759	15	26	11	1831	63	50
4	1216	36	34	12	1167	8	23
5	1350	26	26	13	1424	44	38
6	1026	24	31	14	2387	43	35
7,	1099	15	20	15	1438	30	29
8	1726	33	32	16	1199	19	38

Вопросы для самопроверки

1 Сформулируйте цели и задачи множественного корреляционно-регрессионного анализа.

2 Приведите способы оценки параметров двумерной корреляционно-регрессионной модели.

3 Приведите способы оценки параметров трехмерной корреляционно-регрессионной модели.

4 Приведите способы оценки параметров многомерной корреляционно-регрессионной модели.

ТЕМА 3 КОМПОНЕНТНЫЙ АНАЛИЗ

Данная тема включает в себя:

- статистический подход в методе главных компонент:

- многомерное нормальное распределение как модель,

- геометрическая интерпретация плотности вероятности двумерного нормального распределения;
- линейная модель метода главных компонент:
 - общие положения,
 - получение матрицы весовых коэффициентов по алгоритму метода главных компонент,
 - линейный оператор и отвечающая ему матрица,
 - характеристический многочлен подобных матриц,
 - собственные векторы при одном и том же характеристическом числе,
 - характеристическая матрица и характеристический многочлен,
 - выражение присоединенной матрицы через характеристический многочлен,
 - метод Фаддеева – одновременное вычисление коэффициентов характеристического многочлена и присоединенной матрицы;
- квадратичные формы и главные компоненты:
 - главные компоненты трехмерного и конечно-мерного пространства,
 - дисперсия исследуемых признаков в методе главных компонент.

Контрольные задачи

1 Хозяйственная деятельность предприятий приборостроения ($N=10$) характеризуется следующими показателями ($n=2$):

Таблица 1 Показатели деятельности предприятий приборостроения

№ п/п	Общие затраты на 1 руб. продукции, x_1	Фондоотдача, x_2
1	0,92	0,51
2	0,72	0,59
3	0,83	1,03
4	0,81	1,21
5	0,82	0,63
6	0,93	0,68
7	0,84	0,57
8	0,89	1,52
9	0,89	1,04
10	0,95	0,99

Требуется:

- а) ранжировать предприятия по первой главной компоненте;
- б) графически изобразить предприятия в пространстве двух главных компонент. Дать экономическую интерпретацию полученным результатам.

2 Деятельность предприятий приборостроения ($N=10$) характеризуется показателями ($n=2$).

Приняв за результативный показатель y производительность труда работника, построить уравнение регрессии на главной компоненте, наиболее тесно связанной с y , где

$$y^T = (3,5; 4,4; 3,9; 3,8; 3,6; 3,5; 3,0; 4,1; 4,0; 3,1).$$

Дать интерпретацию полученным результатам.

Таблица 2 Показатели деятельности предприятий приборостроения

№ п/п	Общие затраты на единицу товарной продукции, x_1	Среднегодовая заработная плата на единицу ППП, x_2
1	0,92	2,33
2	0,72	2,31
3	0,83	2,39
4	0,81	2,34
5	0,82	2,38
6	0,93	2,43
7	0,84	2,20
8	0,79	2,33
9	0,89	2,23
10	0,95	2,13

3 Значения экономических показателей, отражающих работу предприятий приборостроения ($N=10$), приводятся в таблице 3.

Таблица 3 Показатели деятельности предприятий приборостроения

№ п/п	Фондоотдача, x_1	Среднегодовая заработная плата на единицу ППП, x_2
1	0,98	2,24
2	0,73	2,29
3	0,73	2,40
4	1,47	2,46
5	0,75	2,58
6	0,81	2,18
7	0,83	2,12
8	0,92	2,10
9	0,46	2,29
10	1,11	2,36

Требуется:

а) определить собственные значения главных компонент и их вклад в суммарную дисперсию исходных показателей;

б) определить ортогональную матрицу собственных векторов корреляционной матрицы R ;

в) дать экономическую интерпретацию главных компонент.

4 Найдены значения главных компонент для i -го предприятия $f_{i1} = 0,661$; $f_{i2} = -2,151$ и матрица факторных нагрузок

$$A = \begin{pmatrix} -0,756 & 0,654 \\ 0,756 & 0,654 \end{pmatrix}$$

Требуется найти значения исходных показателей для этого предприятия (x_{i1}, x_{i2}) , если выборка характеризуется векторами средних $\bar{x} = \begin{pmatrix} 0,850 \\ 0,877 \end{pmatrix}$ и средних квадратических отклонений $S = \begin{pmatrix} 0,072 \\ 0,333 \end{pmatrix}$

5 В матрице значений главных компонент F оказалось пропущенным последнее десятое наблюдение ($i=10$)

$$F = \begin{pmatrix} -0,484 & 1,053 \\ 1,227 & -1,572 \\ 0,773 & 0,527 \\ 0,606 & -0,175 \\ 0,799 & 0,314 \\ 0,136 & 2,097 \\ -0,671 & -1,109 \\ 0,720 & -0,507 \\ -0,921 & -0,231 \end{pmatrix}$$

Матрица факторных нагрузок

$$A = \begin{pmatrix} -0,791 & 0,611 \\ 0,791 & 0,611 \end{pmatrix}$$

Определить:

а) значения главных компонент f_{i1}, f_{i2} для десятого наблюдения ($i=10$);

б) нормированные значения исходных показателей для десятого наблюдения (y_{i1}, y_{i2}) , где $i=10$;

в) значения исходных показателей для $i=10$ наблюдения (x_{i1} , x_{i2}), если выборка характеризуется векторами средних $\bar{x} = \begin{pmatrix} 0,850 \\ 2,307 \end{pmatrix}$;

$$S = \begin{pmatrix} 0,072 \\ 0,093 \end{pmatrix}$$

г) значения исходных показателей для шестого наблюдения.

6 Хозяйственная деятельность машиностроительных предприятий ($N=10$) характеризуется следующими показателями ($n=3$).

Таблица 4 Показатели деятельности машиностроительных предприятий

№п/п	Трудоемкость единицы продукции, x_1	Удельный вес рабочих в составе ППП, x_2	Коэффициент сменности оборудования, x_3
1	0,38	0,72	1,40
2	0,24	0,70	1,20
3	0,31	0,66	1,15
4	0,42	0,69	1,09
5	0,51	0,71	5,00
6	0,31	0,73	1,36
7	0,37	0,65	1,15
8	0,16	0,82	1,87
9	0,18	0,80	2,17
10	0,43	0,83	1,61

Требуется:

- а) ранжировать предприятия по первой главной компоненте;
- б) графически изобразить предприятия в пространстве двух первых компонент;
- в) сопоставить и дать интерпретацию полученным результатам.

7 Экономическая деятельность предприятий машиностроения ($N=10$) характеризуется показателями ($n=3$).

Таблица 5 Показатели деятельности машиностроительных предприятий

№п/п	Удельный вес покупных изделий, x_1	Коэффициент сменности оборудования, x_2	Премии и вознаграждения на одного работника, x_3	Y
1	0,30	1,40	0,67	9,8
2	0,56	1,20	0,98	13,2
3	0,42	1,15	1,16	17,3
4	0,26	1,09	0,54	7,1
5	0,16	1,26	1,23	11,5
6	0,45	1,36	0,78	12,1
7	0,31	1,15	1,16	15,2
8	0,08	1,87	2,44	31,3
9	0,63	2,17	1,06	11,6
10	0,03	1,61	2,13	30,1

Приняв за результативный показатель y - рентабельность, построить уравнение регрессии на главной компоненте, наиболее тесно связанной с y . Дать экономическую интерпретацию полученным результатам.

8 Значения экономических показателей предприятий машиностроения ($N=10$) характеризуются следующими показателями (таблица).

Требуется:

- а) определить собственные значения главных компонент и их вклад в суммарную дисперсию исходных признаков;
- б) определить ортогональную матрицу собственных векторов V ;
- в) дать экономическую интерпретацию главных компонент.

Таблица 6 Показатели деятельности предприятий машиностроения

№ п/п	Удельный вес рабочих в составе ППП, x_1	Удельный вес покупных изделий, x_2	Премии и вознаграждения на одного работника, x_3
1	0,68	0,33	0,82
2	0,74	0,25	0,84
3	0,66	0,32	0,67
4	0,72	0,02	1,04
5	0,68	0,06	0,66
6	0,77	0,15	0,86
7	0,78	0,08	0,79
8	0,78	0,20	0,34
9	0,81	0,20	1,60
10	0,79	0,30	1,46

9 Деятельность предприятий ($N=10$) характеризуется следующими показателями ($n=3$).

Таблица 7 Показатели деятельности предприятий

№ п/п	Трудоемкость единицы продукции, x_1	Удельный вес покупных изделий, x_2	Коэффициент сменности оборудования, x_3	Y
1	0,51	0,20	1,47	21,9
2	0,36	0,64	1,27	48,4
3	0,23	0,42	1,51	173,5
4	0,26	0,27	1,46	74,1
5	0,27	0,37	1,27	68,6
6	0,29	0,38	1,43	60,8
7	0,01	0,35	1,50	355,6
8	0,02	0,42	1,35	264,8
9	0,18	0,32	1,41	526,6
10	0,25	0,33	1,47	118,6

Приняв за результативный показатель Y - индекс снижения себестоимости продукции, построить уравнение регрессии на главные компоненты, наиболее тесно связанные с y . Дать экономическую интерпретацию полученным результатам.

10 Заданы значения главных компонент для i -го объекта $f_{i1} = -0,252$; $f_{i2} = 0,335$; $f_{i3} = -0,542$ и матрица факторных нагрузок

$$A = \begin{pmatrix} -0,718 & 0,691 & -0,081 \\ 0,879 & 0,436 & 0,191 \\ 0,964 & 0,116 & -0,235 \end{pmatrix}$$

Найти значения исходных показателей для этого объекта (x_{i1} , x_{i2} , x_{i3}), если выборка характеризуется векторами средних

$$\bar{x} = \begin{pmatrix} 0,331 \\ 0,731 \\ 1,426 \end{pmatrix}$$

и средних квадратических отклонений

$$s = \begin{pmatrix} 0,107 \\ 0,051 \\ 0,337 \end{pmatrix}$$

11 В матрице значений главных компонент F оказалось пропущенным восьмое наблюдение ($i=8$):

Таблица 8 Матрица значений главных компонент

1	-0,25	0,34	-0,54
2	-0,22	-1,30	1,31
3	-0,75	-1,08	-0,14
4	-0,97	0,24	0,39
5	-0,89	1,39	-0,87
6	-0,03	-0,24	0,60
7	-1,00	-0,62	-0,92
8			
9	1,86	-0,33	-1,91
10	0,58	2,07	1,09

Матрица факторных нагрузок равна:

$$A = \begin{pmatrix} -0,72 & 0,69 & -0,08 \\ 0,88 & 0,44 & 0,19 \\ 0,96 & 0,12 & -0,24 \end{pmatrix}$$

Определить:

а) значения главных компонент f_{i1}, f_{i2}, f_{i3} для $i=8$;

б) нормированные значения исходных показателей восьмого наблюдения (Y_{81}, Y_{82}, Y_{83});

в) значения исходных показателей для восьмого наблюдения (X_{81}, X_{82}, X_{83}), если выборка характеризуется векторами средних

$$\bar{x} = \begin{pmatrix} 0,33 \\ 0,73 \\ 1,43 \end{pmatrix}$$

и средних квадратических отклонений

$$s = \begin{pmatrix} 0,11 \\ 0,06 \\ 0,34 \end{pmatrix}$$

г) значения исходных показателей для третьего наблюдения.

Вопросы для самопроверки

- 1 Сформулируйте цели и задачи компонентного анализа.
- 2 Приведите схему алгоритма метода главных компонент.
- 3 Как определить характеристическую матрицу, ее определитель и присоединенную матрицу?
- 4 Сформулируйте особенности метода Фаддеева.
- 5 Как определить главные компоненты двумерного пространства?

6 Как определить главные компоненты трехмерного пространства?

ТЕМА 4 ФАКТОРНЫЙ АНАЛИЗ

Данная тема включает в себя:

- основные понятия факторного анализа:
 - основные проблемы факторного анализа,
 - основная модель факторного анализа,
 - компоненты дисперсии в факторном анализе,
 - получение матрицы коэффициентов парной корреляции и ее преобразование в факторном анализе,
 - корреляционная матрица с общностями на главной диагонали,
 - факторное отображение и факторная структура,
 - пространство общих факторов и полное факторное пространство,
 - связь факторных решений, полученных разными методами;
- метод главных факторов и его алгоритм;
- проблема вращения:
 - понятие ортогонального и косоугольного вращения,
 - вращение при помощи ортогональных матриц,
 - обобщенные факторы;
- проблема оценки факторов и задачи классификации:

- измерение факторов,
- факторный анализ и методы классификации многомерных наблюдений;
- классификация задач факторного анализа и метода главных компонент.

Контрольные задачи

1 Составить уравнения для факторного анализа при $n=5$, $m=2$.

2 Чему равны вклады общих факторов и характерного фактора в дисперсию y_1 :

$$y_1 = 0,5f_1 + 0,8f_2 + 0,33v_1;$$

3 Найти общность в задаче 2.

4 Найти общность и характерность параметра y_1 , если его надежность $r_1^2 = 0,95$, а специфичность равна $0,15$.

5 По таблице, содержащей две составляющие дисперсии, найти три другие составляющие.

Компоненты дисперсии признака	№ заданий							
	1	2	3	4	5	6	7	8
Специфичность – b_j^2	0,10	0,15	0,20	0,25				
Дисперсия ошибок - c_j^2	0,20				0,05	0,10		
Характерность - d_j^n			0,35			0,45		0,30
Общность - h_j^2		0,75			0,60		0,50	
Надежность - r_j^2				0,85			0,75	0,90

6 Заданы два общих фактора: $a_{11}=0,6$; $a_{12}=0,5$, а также надежность $r_1^2 = 0,84$.

Составить полное уравнение для параметра y_1 .

7 Найти долю суммарной общности, вносимую двумя общими факторами и каждым в отдельности (в %), по заданной матрице A (знаки весовых коэффициентов не приводятся):

$$A = \begin{pmatrix} 0,7 & 0,3 \\ 0,8 & 0,0 \\ 0,7 & 0,0 \\ 0,8 & 0,6 \\ 0,6 & 0,5 \\ 0,5 & 0,0 \\ 0,6 & 0,4 \\ 0,7 & 0,6 \end{pmatrix}$$

8 Вычислить долю суммарной дисперсии, объясняемую каждым из общих факторов в задаче 7.

9 Вычислить коэффициенты корреляции, редуцированные на основе задачи : $r'_{12}, r'_{14}, r'_{25}, r'_{26}$.

10 По данным задачи 7 вычислить наибольшее и наименьшее значения характерности.

Вопросы для самопроверки

- 1 Сформулируйте цели, задачи и проблемы факторного анализа.
- 2 Приведите основную модель факторного анализа.
- 3 Приведите схему алгоритма метода главных компонент.
- 4 Покажите преобразование матрицы коэффициентов парной корреляции в факторном анализе.
- 5 Приведите методы определения и оценки общностей.

6 Проиллюстрируйте связь факторных решений, полученных разными методами.

7 Охарактеризуйте проблемы вращения осей.

ТЕМА 5 МЕТОДЫ МНОГОМЕРНОЙ КЛАССИФИКАЦИИ. КЛАССИФИКАЦИЯ БЕЗ ОБУЧЕНИЯ. КЛАСТЕРНЫЙ АНАЛИЗ

Данная тема включает в себя:

- основные понятия,
- расстояние между объектами и мера близости,
- расстояние между кластерами,
- функционалы качества разбиения,
- иерархические кластер-процедуры.

Контрольные задачи

1 По иерархическому агломеративному алгоритму провести классификацию $n=6$ предприятий машиностроения, деятельность которых характеризуется показателями: x_1 - рентабельности (%); и x_2 - производительности труда.

№ предприятия	1	2	3	4	5	6
x_1	23,4	17,5	9,7	18,2	6,6	8,0
x_2	9,1	5,2	5,5	9,4	7,5	5,7

В качестве расстояния между объектами принять:

- а) обычное евклидово расстояние;

б) взвешенное евклидово расстояние с весами $\omega_1=0,3$ и $\omega_2=0,7$.

Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий.

Расстояние между кластерами определить по принципу "ближайшего соседа". Данные не нормализовывать.

2 Решить задачу 1, предварительно нормировав исходные данные.

3 Решить задачу 1, измеряя расстояние между кластерами по принципу "дальнего соседа", нормируя предварительно исходные данные.

4 Решить задачу 1, измеряя расстояние между кластерами по "центрам тяжести" групп, не нормируя предварительно исходные данные.

5 По агломеративному алгоритму провести классификацию $n=5$ фермерских хозяйств, работа которых характеризуется показателями объема реализованной продукции: x_1 - растениеводства и x_2 - животноводства с 1 га пашни.

№ колхоза	1	2	3	4	5
x_1	2,49	1,51	1,17	1,67	2,73
x_2	0,38	0,51	0,28	0,29	0,34

В качестве расстояния между объектами принять обычное евклидово расстояние, а расстояния между кластерами измерять по принципу:

- а) "ближайшего соседа";
- б) "дальнего соседа".

Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий. Исходные данные не нормализовывать.

6 Решить задачу 5, предварительно нормировав исходные данные.

7 Решить задачу 5 при условии, что расстояния между кластерами измеряются по принципу "центра тяжести" групп, нормируя предварительные исходные данные.

8 Решить задачу 5 при условии, что расстояние между кластерами измеряется по принципу:

а) "средней связи";

б) "центра тяжести" групп, не нормируя предварительно исходные данные.

9 По агломеративному алгоритму провести классификацию $n=6$ регионов по уровню медицинского обслуживания населения, который характеризуется показателями: x_1 - число врачей на 10 тыс. жителей и x_2 - число больничных коек на 10 тыс. жителей.

№ региона	1	2	3	4	5	6
x_1	34,8	31,2	32,1	35,7	30,2	34,2
x_2	126	112	123	128	115	123

За расстояние между объектами принять взвешенное евклидово расстояние с $\omega_1=0,4$ и $\omega_2=0,6$, а расстояние между кластерами измерять по принципу:

а) "ближайшего соседа";

б) "дальнего соседа".

Сравнить разбиение на кластеры по критерию минимума суммы внутриклассовых дисперсий. Исходные данные не нормализовывать.

10 Решить задачу 9, предварительно нормировав исходные данные.

11 Решить задачу 9 при условии, что расстояние между кластерами измеряются по принципу "центра тяжести" групп.

12 Решить задачу 9 при условии, что расстояние между кластерами измеряются по принципу "средней связи".

13 По агломеративному алгоритму провести классификацию $n=5$ районов, работа которых оценивалась по показателям: x_1 - производительности труда, объема реализованной продукции: x_2 - растениеводства и x_3 - животноводства с 1га посевной площади.

В качестве расстояния между объектами принять:

а) обычное евклидово расстояние;

б) взвешенное евклидово расстояние с весами: $\omega_1=0,2$; $\omega_2=0,4$; $\omega_3=0,4$. Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий.

Расстояние между кластерами определять по принципу "ближайшего соседа". Исходные данные не нормировать.

№ колхоза	1	2	3	4	5
x_1	8,22	4,33	6,43	6,39	4,92
x_2	0,25	0,49	0,51	0,27	0,32
x_3	0,41	0,38	0,51	0,42	0,55

14 Решить задачу 13, предварительно нормировав исходные данные.

15 Решить задачу 13, измеряя расстояние между кластерами по принципу "дальнего соседа". Исходные данные не нормировать.

16 Решить задачу 13, измеряя расстояние между кластерами по принципу "центра тяжести", не нормируя исходные данные.

17 По агломеративному алгоритму провести классификацию $n=6$ машиностроительных предприятий, работа которых оценивалась по показателям: x_1 - производительности труда, x_2 - удельного веса потерь от брака (%) и x_3 - фондоотдачи активной части производственных фондов.

№ колхозов	1	2	3	4	5	6
x_1	9,4	6,6	7,4	10,0	6,6	9,1
x_2	0,15	0,48	0,62	0,32	0,50	0,90
x_3	1,91	0,88	1,09	2,62	1,32	1,89

В качестве расстояния между объектами принять обычное евклидово расстояние, а расстояние между кластерами измерить по принципу:

- а) "ближайшего соседа";
- б) "дальнего соседа".

Сравнить разбиение на два кластера по критерию минимума суммы внутриклассовых дисперсий. Исходные данные не нормировать.

18 Решить задачу 17, предварительно пронормировав исходные данные.

19 Решить задачу 17 при условии, что расстояние между кластерами измеряется по принципу "центра тяжести" групп, не нормируя предварительно исходные данные.

20 Решить задачу 17 при условии, что расстояние между кластерами измеряется по принципу "средней связи".

Вопросы для самопроверки

1 Сформулируйте цели и задачи кластерного анализа.

2 Приведите наиболее часто используемые расстояния между объектами и кластерами, меры близости в задачах кластерного анализа.

3 Охарактеризуйте иерархические кластер-процедуры.

ТЕМА 6 ДИСКРИМИНАНТНЫЙ АНАЛИЗ. КЛАССИФИКАЦИИ С ОБУЧЕНИЕМ

Данная тема включает в себя:

- методы классификации с обучением,
- линейный дискриминантный анализ,
- дискриминантный анализ при нормальном законе распределения показателей

Контрольные задачи

1 В таблице представлены группы машиностроительных предприятий с высоким и низким уровнями организации управления производством. Характеризуя деятельность предприятий показателями рентабельности и производительности труда с помощью дискриминантного анализа, классифицируйте три последних предприятия.

Таблица Результаты деятельности предприятий

№ предприятия	Показатель	Рентабельность, %	Производительность труда, млн. руб./чел.
	Группа предприятий		
1	Высокий уровень, X	23,4	9,1
2		19,1	6,6
3		17,5	5,2
4		17,2	10,1
5	Низкий уровень, Y	5,4	4,3
6		6,6	5,5
7		8,0	5,7
8		9,7	5,5
9		9,1	6,6
10	Подлежат дискриминации, Z	9,9	7,4
11		14,2	9,4
12		12,9	6,7

2 При оценке эффективности деятельности предприятий приборостроения были получены два класса предприятий *A* и *B* (таблица). Используя их как обучающие выборки провести дискриминацию трех последних предприятий по показателям фондоотдачи и материалоемкости.

Таблица Результаты деятельности предприятий

№ предприятия	Показатель		Фондоотдача	Материалоемкость
	Предприятия			
1	Класс А, Х		1,47	3,90
2			1,52	6,18
3			1,38	13,80
4			1,27	8,18
5	Класс В, У		0,59	9,08
6			0,57	2,10
7			0,46	6,5
8			0,53	13,3
9	Подлежат дискриминации, Z		0,92	13,80
10			1,21	4,80
11			0,63	12,33

3 Анализ эффективности использования земельных угодий в сельскохозяйственных районах области позволил выделить регионы с низким *A* и высоким *B* уровнями использования земли (таблица).

Таблица Результаты деятельности предприятий

№ района	Показатель		Объем реализованной продукции	
	Уровень использования земли		Растениеводства	Животноводства
1	Низкий А, Х		0,25	0,41
2			0,51	0,51
3			0,27	0,42
4			0,33	0,56
5	Высокий В, У		1,17	0,28
6			4,99	0,67
7			5,18	0,45
8			2,49	0,38
9			2,73	0,33
10	Подлежат дискриминации, Z		0,32	0,45
11			0,67	0,32
12			4,6	0,56

С помощью дискриминантного анализа провести классификацию трех последних районов по показателям объема реализованной продукции растениеводства и животноводства с 1га посевной площади.

4 По эффективности работы выделены две группы, состоящие из $n_x=4$ и $n_y=5$ ферм. Для этих групп по показателям производительности труда и объема реализованной продукции растениеводства с 1га получены оценки векторов средних

$$\bar{X} = \begin{pmatrix} 6,72 \\ 0,340 \end{pmatrix}; \quad \bar{Y} = \begin{pmatrix} 4,05 \\ 2,51 \end{pmatrix}$$

и ковариационных матриц

$$S_x = \begin{pmatrix} 1,07 & 0,23 \\ 0,23 & 0,08 \end{pmatrix}; \quad S_y = \begin{pmatrix} 0,30 & 0,16 \\ 0,16 & 0,24 \end{pmatrix}$$

Вычислить значения дискриминантной функции для ферм, показатели которых представлены в матрице

$$Z = \begin{pmatrix} 4,92 & 0,32 \\ 5,22 & 0,67 \\ 5,62 & 4,60 \end{pmatrix}$$

5 При анализе уровня медицинского обслуживания населения выделены две группы регионов. При этом первая группа включает $n_x=4$, а вторая $n_y=5$ регионов.

Для этих групп по двум показателям - число врачей на 10 тыс. жителей; число больничных коек на 10 тыс. человек - получены оценки векторов средних

$$\bar{X} = \begin{pmatrix} 34,47 \\ 124,75 \end{pmatrix}; \quad \bar{Y} = \begin{pmatrix} 31,08 \\ 112,00 \end{pmatrix}$$

и ковариационных матриц

$$S_x = \begin{pmatrix} 1,10 & 2,79 \\ 2,79 & 7,58 \end{pmatrix}; \quad S_y = \begin{pmatrix} 2,41 & 3,45 \\ 3,45 & 5,02 \end{pmatrix}$$

Вычислить оценки значений дискриминантной функции для регионов, показатели которых представлены в матрице

$$Z = \begin{pmatrix} 31,3 & 123 \\ 31,8 & 110 \\ 34,6 & 117 \end{pmatrix}$$

6 По уровню механизации добычи выделены две группы, включающие $n_x=5$ и $n_y=4$ угольных шахт. Для этих групп по двум показателям - удельный вес добычи угля из комплексно-механизированных забоев по отношению ко всей добыче (%); удельный вес комбайновой проходки выработок (%) - получены оценки векторов средних и ковариационных матриц

$$\bar{X} = \begin{pmatrix} 43,26 \\ 28,54 \end{pmatrix}; \quad \bar{Y} = \begin{pmatrix} 14,98 \\ 38,95 \end{pmatrix}$$

и ковариационных матриц

$$S_x = \begin{pmatrix} 84,61 & 29,54 \\ 29,54 & 57,91 \end{pmatrix}; \quad S_y = \begin{pmatrix} 79,88 & 26,98 \\ 26,98 & 54,77 \end{pmatrix}$$

Определить оценки значений дискриминантной функции для угольных шахт, показатели которых представлены в матрице

$$Z = \begin{pmatrix} 34,2 & 24,8 \\ 17,0 & 60,0 \\ 38,0 & 9,7 \end{pmatrix}$$

7 Эффективность деятельности фермерских хозяйств оценивалась по трем показателям: производительность труда (млн. руб./чел.); объем реализованной продукции растениеводства и животноводства с 1га посевной площади (млн. руб./га). В результате были выделены хозяйства с высоким A и низким B уровнями использования земли (таблица).

Используя данные таблицы, с помощью дискриминантного анализа произвести классификацию трех последних хозяйств.

Таблица Результаты деятельности хозяйств

№ район а	Показатель Уровень использования земли	Производительность труда	Объем реализованной продукции	
			Растениеводство	Животноводства
1	Высокий А, Х	8,22	0,25	0,41
2		6,43	0,51	0,51
3		6,39	0,27	0,42
4		5,84	0,33	0,56
5	Низкий В, У	3,14	1,17	0,28
6		4,02	4,99	0,67
7		4,19	5,18	0,45
8		4,33	2,49	0,38
9		4,58	2,73	0,33
10	Подлежат дискриминации, Z	4,92	0,32	0,45
11		5,22	0,67	0,32
12		5,62	4,60	0,56

8 Имеется 12 предприятий, характеризуемых тремя экономическими показателями: производительность труда; удельный вес потерь от брака (%) и фондоотдача активной части основных производственных фондов. Из этих предприятий выделены две обучающие выборки, первая из которых включает $n_x=4$ предприятий группы А, а вторая - $n_y=4$ предприятий группы В. Используя данные таблицы требуется:

- найти оценки векторов средних \bar{X} , \bar{Y} и ковариационных матриц S_x и S_y ;
- определить несмещенную оценку суммарной ковариационной

матрицы \hat{S} и обратной матрицы \hat{S}^{-1} ;

- получить вектор оценок коэффициентов дискриминантной функции;
- найти оценки значений дискриминантной функции \hat{U}_x и \hat{U}_y для матриц исходных данных X и Y ;
- определить оценку константы \hat{C} ;
- вычислить оценки значений дискриминантной функции для группы оставшихся предприятий и провести их классификацию.

Таблица Результаты деятельности предприятий

№ предприятия	Показатель	Производительность труда	Удельный вес потерь от брака	Фондоотдача
	Предприятия			
1	Группы А, X	9,4	0,15	1,91
2		9,9	0,34	1,68
3		9,1	0,09	1,89
4		9,4	0,21	2,30
5	Группы В, Y	6,6	0,48	0,88
6		4,3	0,41	0,62
7		7,4	0,62	1,09
8		6,6	0,50	1,32
9		5,5	1,20	0,68
10	Подлежат дискриминации, Z	5,5	0,05	1,02
11		5,7	0,66	1,43
12		10,0	0,32	2,62

Вопросы для самопроверки

- 1 Сформулируйте цели и задачи дискриминантного анализа.
- 2 Приведите алгоритм дискриминантного анализа.

ТЕМА 7 КАНОНИЧЕСКИЕ КОРРЕЛЯЦИИ

Данная тема включает в себя:

- основные понятия;
- канонические корреляции и канонические величины генеральной совокупности;
- канонические корреляции и их интерпретация;
- оценка канонических корреляций и канонических величин.

Контрольные задачи

Для статистического анализа эффективности подземной угледобычи была собрана информация по 30 угольным шахтам. В качестве исходных параметров были отобраны 8 показателей эффективности угледобычи и 19 признаков, ее определяющих.

Показатели эффективности угледобычи:

y_1 - выработка валовой продукции на одного работающего, т;

y_2 - среднемесячная производительность труда рабочего по добыче угля, т;

y_3 - среднемесячная производительность труда промышленно-производственного персонала по добыче угля, т;

y_4 - среднемесячная производительность труда одного рабочего на очистных работах, т.;

y_5 - полная себестоимость добычи 1т угля;

y_6 - фондоотдача;

y_7 - фондоемкость;

y_8 - трудоемкость работ по добыче, чел.-дни на 1000 т.

Показатели, определяющие эффективность угледобычи:

x_1 - среднесуточная добыча угля (нагрузка на шахту), т;

x_2 - среднединамическая мощность пласта;

x_3 - максимальная глубина разработки, м;

x_4 - приток воды в шахту, м³/час;

x_5 - количество метана на 1 т среднемесячной добычи, м³;

x_6 - среднесуточная нагрузка на забой, т;

x_7 - среднедействующая длина линии очистных забоев, м;

x_8 - среднемесячное продвижение линии действующих очистных забоев, м;

x_9 - удельный вес добычи угля из комплексно-механизированных забоев по отношению ко всей добыче из очистных забоев, м;

x_{10} - удельный вес комбайновой проходки выработок, %;

x_{11} - удельный объем проведения подготовительных выработок на 1000 т добычи, м³;

x_{12} - удельный объем выработок, закрепленных металлической и железобетонной крепью, %;

x_{13} - протяженность выработок на 1000 т добычи, м;

x_{14} - среднемесячная производительность работающего электровоза, тыс. км;

x_{15} - протяженность действующих рельсовых путей, км;

x_{16} - удельный вес забалластированных рельсовых путей, %;

x_{17} - удельная емкость вагонеточного парка на 1000 т среднесуточной выработки, т;

x_{18} - объем породы на 1000 т добычи, т;

x_{19} - зольность угля, %.

Провести исследование информации по совокупности шахт, приведенной в таблицах, методом канонических корреляций для заданных наборов показателей эффективности подземной угледобычи y и признаков x , ее определяющих.

Вычислите первую пару канонических величин и каноническую корреляцию по показателям эффективности и показателям, определяющим эффективность.

Проведите сокращение числа показателей, определяющих эффективность угледобычи при помощи канонического анализа.

Варианты заданий

№ варианта	№ переменной 1 группы, y	№ переменной 2 группы, x
1	1,2	x_1-x_5
2	1,3	x_1-x_5
3	2,3	x_1-x_5
4	1,2,3	x_1-x_6
5	1,2,3	x_1-x_7
6	1,2	x_3-x_8
7	1,3	x_3-x_8
8	2,3	x_3-x_8
9	1,2	x_2-x_7
10	1,3	x_2-x_7
11	2,3	x_2-x_7
12	1,2	x_4-x_8
13	1,3	x_4-x_8
14	2,3	x_1-x_8
15	1,2	x_1-x_8
16	1,3	x_1-x_8
17	2,3	x_1-x_8
18	1,2	x_1-x_3 И x_5-x_8
19	1,3	x_1-x_3 И x_3-x_8
20	2,3	x_1-x_3 И x_3-x_8

Исходные данные по совокупности шахт

№ шахты	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	y_1	y_2	y_3
1	879	2,80	311	250	146	198	114	34,7	3090	23,0	19,0
2	1533	3,24	250	420	57	212	212	30,9	3388	23,9	20,0
3	2526	3,66	220	400	18,7	260	103	41,4	7525	56,9	46,5
4	1646	1,55	307	200	62	243	582	33,8	2795	27,4	23,4
5	1997	3,26	360	400	123	115	2978	3,60	0,40	22,9	25,3
6	6005	2,60	275	420	21,0	166	940	20,2	0,00	18,7	50,9
7	1824	2,96	343	120	19,68	160	585	10,4	0,00	44,3	36,5
8	2693	2,53	280	420	9,92	128	206	37,8	5360	40,2	33,1
9	3416	3,0	190	800	14,0	168	509	34,2	6514	47,5	40,2
10	1386	2,44	258	439	27,0	93	194	24,7	4995	37,8	30,7
11	2547	3,00	330	335	39,5	222	164	67,0	4889	35,9	30,1
12	4217	2,75	250	520	16,0	190	439	55,8	7361	54,7	45,4
13	3506	3,17	247	320	15,0	195	409	40,3	4878	36,1	30,0
14	2531	3,63	280	380	29,5	179	222	37,1	5344	39,4	32,8
15	2229	3,09	260	430	10,0	28	177	45,8	5938	43,4	36,2
16	2161	2,99	240	90,0	27,0	251	58	68,7	5268	39,0	32,4
17	3169	2,97	126	400	4,50	313	325	43,8	8299	62,3	51,1
18	3552	2,10	88,0	475	29,0	178	93	99,8	7041	52,0	43,1
19	1332	2,72	310	92,0	12,0	208	117	57,2	5155	40,6	31,8
20	3419	3,18	145	250	5,30	445	241	56,6	9011	66,8	55,5
21	1194	3,03	290	165	58,0	184	150	31,9	4697	35,8	28,7
22	2828	2,86	330	260	29,0	257	219	48,0	7008	51,5	42,7
23	2748	2,17	502	180	17,0	329	613	36,4	5467	36,3	31,5
24	1196	3,21	263	485	7,00	193	127	39,6	3535	25,7	21,7
25	2265	3,69	100	128	1,00	177	11	105,1	10644	76,7	65,7
26	1241	1,80	720	320	43,0	294	432	23,6	2768	19,2	16,7
27	3217	1,77	420	400	7,00	465	920	36,3	5225	37,0	32,1
28	1997	1,82	360	220	8,00	311	643	31,0	5167	38,1	31,5
29	1983	1,66	230	474	1,00	414	742	31,9	8393	45,8	39,1
30	842	3,47	289	250	21,0	165	175	23,6	3810	28,4	23,2

Вопросы для самопроверки

- 1 Сформулируйте задачи, решаемые с помощью метода канонических корреляций.
- 2 Дайте основные понятия метода канонических корреляций.
- 3 Приведите интерпретацию канонических корреляций

**ТЕМА 8 МНОЖЕСТВЕННЫЙ КОВАРИАЦИОННЫЙ
АНАЛИЗ**

Данная тема включает в себя:

- основные понятия;
- фиксированные коварианты;
- переменные коварианты.

Вопросы для самопроверки

- 1 Сформулируйте задачи, решаемые с помощью множественного ковариационного анализа.
- 2 Дайте основные понятия множественного ковариационного анализа.
- 3 Охарактеризуйте различия между фиксированными и переменными ковариантами.

ИНДИВИДУАЛЬНЫЕ ЗАДАНИЯ

Варианты заданий по корреляционному, регрессионному, компонентному и кластерному анализу даны в таблице 1, а варианты задания по дискриминантному анализу приведены в таблице 2.

Таблица 1 Варианты заданий по корреляционному, регрессионному, компонентному и кластерному анализу

№ варианта	Результативный признак, Y	Факторные признаки, X
1	1	5, 7, 10, 11, 16
2	1	5, 7, 10, 12, 16
3	1	7, 10, 11, 12, 16
4	1	5, 7, 12, 13, 16
5	1	7, 10, 12, 13, 16
6	1	5, 7, 11, 12, 16
7	1	6, 10, 11, 12, 16
8	1	6, 8, 11, 12, 16
9	1	7, 10, 11, 12, 16
10	1	7, 8, 12, 13, 16
11	1	4, 5, 6, 8, 16
12	1	4, 6, 8, 10, 16
13	1	4, 5, 11, 12, 16
14	1	4, 6, 9, 13, 16
15	1	4, 5, 9, 13, 16
16	3	7, 9, 14, 15, 16
17	3	4, 5, 9, 14, 16
18	3	4, 5, 6, 10, 11
19	3	7, 8, 9, 10, 16
20	3	7, 8, 9, 11, 16
21	2	4, 5, 7, 8, 9
22	2	4, 5, 6, 8, 9
23	2	4, 5, 7, 9, 13
24	2	4, 7, 8, 9, 16
25	2	4, 6, 8, 9, 16

Таблица 2 Варианты заданий по дискриминантному анализу

№ варианта	Номера организаций	
	Группа А	Группа В
1	1, 2, 3, 4, 5, 6, 8, 24, 29	12, 14, 15, 16, 21, 23, 27, 28
2	1, 2, 3, 4, 5, 6, 8, 24, 43	12, 14, 15, 16, 21, 23, 27, 41
3	1, 2, 3, 4, 5, 6, 8, 29, 43	12, 14, 15, 16, 21, 23, 28, 41
4	1, 2, 3, 4, 5, 6, 24, 29, 43	12, 14, 15, 16, 21, 27, 28, 41
5	1, 2, 3, 4, 5, 8, 24, 29, 43	12, 14, 15, 16, 23, 27, 28, 41
6	1, 2, 3, 4, 6, 8, 24, 29, 43	12, 14, 15, 21, 23, 27, 28, 41
7	1, 2, 3, 5, 6, 8, 24, 29, 43	12, 14, 16, 21, 23, 27, 28, 41
8	1, 2, 4, 5, 6, 8, 24, 29, 43	12, 15, 16, 21, 23, 27, 28, 41
9	1, 3, 4, 5, 6, 8, 24, 29, 43	14, 15, 16, 21, 23, 27, 28, 41
10	2, 3, 4, 5, 6, 8, 24, 29, 43	12, 14, 15, 16, 21, 23, 27, 28
11	1, 2, 3, 4, 5, 6, 8, 24, 29	12, 14, 15, 16, 21, 23, 27, 41
12	1, 2, 3, 4, 5, 6, 8, 24, 43	12, 14, 15, 16, 21, 23, 28, 41
13	1, 2, 3, 4, 5, 6, 8, 29, 43	12, 14, 15, 16, 21, 27, 28, 41
14	1, 2, 3, 4, 5, 6, 24, 29, 43	12, 14, 15, 16, 23, 27, 28, 41
15	1, 2, 3, 4, 5, 8, 24, 29, 43	12, 14, 15, 21, 23, 27, 28, 41
16	1, 2, 3, 4, 6, 8, 24, 29, 43	12, 14, 16, 21, 23, 27, 28, 41
17	1, 2, 3, 5, 6, 8, 24, 29, 43	12, 15, 16, 21, 23, 27, 28, 41
18	1, 2, 4, 5, 6, 8, 24, 29, 43	14, 15, 16, 21, 23, 27, 28, 41
19	1, 3, 4, 5, 6, 8, 24, 29, 43	12, 14, 15, 16, 21, 23, 27, 28
20	2, 3, 4, 5, 6, 8, 24, 29, 43	12, 14, 15, 16, 21, 23, 27, 41
21	1, 2, 3, 4, 5, 6, 8, 24, 29	12, 14, 15, 16, 21, 23, 28, 41
22	1, 2, 3, 4, 5, 6, 8, 24, 43	12, 14, 15, 16, 21, 27, 28, 41
23	1, 2, 3, 4, 5, 6, 8, 29, 43	12, 14, 15, 16, 23, 27, 28, 41
24	1, 2, 3, 4, 5, 6, 24, 29, 43	12, 14, 15, 21, 23, 27, 28, 41
25	1, 2, 3, 4, 5, 8, 24, 29, 43	12, 14, 16, 21, 23, 27, 28, 41

1 Корреляционный анализ

Задание. На основании данных приложения А проведите корреляционный анализ:

1) определите оценки параметров шестимерного нормального закона распределения (векторы средних арифметических и среднего

квадратического отклонения, матрица парных коэффициентов корреляции);

2) получите оценку матрицы частных коэффициентов корреляции. Проверьте значимость и найдите интервальные оценки частных коэффициентов корреляции;

3) найдите оценки шести множественных коэффициентов корреляции (детерминации). Проверьте их значимость, предварительно выбрав уровень α .

4) постройте уравнение регрессии, выбрав в качестве результативного показатель, которому соответствует наибольший множественный коэффициент корреляции;

5) дайте интерпретацию полученным результатам корреляционного анализа.

2 Регрессионный анализ

Задание. На основании данных приложения А проведите регрессионный анализ:

1) найдите оценки уравнений регрессии линейного и степенного вида;

2) проверьте значимость уравнения и коэффициентов регрессии;

3) пошаговым регрессионным анализом получите уравнения регрессии с максимальным числом значимых коэффициентов регрессии;

4) для полученного уравнения регрессии:

- а) дайте экономическую интерпретацию коэффициентам регрессии и множественному коэффициенту детерминации;
- б) оцените адекватность полученной модели по величине абсолютных и относительных отклонений и средней относительной ошибки;
- в) постройте интервальные оценки для коэффициентов регрессии.

3 Компонентный анализ

Задание. На основании данных приложения А проведите компонентный анализ:

- 1) для факторных признаков x найдите оценку матрицы R парных коэффициентов корреляции;
- 2) на основании матрицы собственных значений Λ определите вклад компонент в суммарную дисперсию. Отберите и укажите вклад ($m < k$) первых главных компонент;
- 3) используя матрицу факторных нагрузок A , дайте экономическую интерпретацию полученным главным компонентам;
- 4) по матрице значений главных компонент F проведите классификацию объектов по двум первым главным компонентам. Дайте интерпретацию полученным результатам;
- 5) используя вектор значений результативного признака y и матрицу F , постройте уравнение регрессии на главные компоненты с максимальным числом значимых коэффициентов регрессии.

4 Факторный анализ

Задание. На основании данных приложения А проведите факторный анализ:

- 1) получите матрицу парных коэффициентов корреляции R ;
- 2) Используя метод наибольшего элемента в строке, преобразуйте матрицу R в редуцированную матрицу R_h ;
- 3) получите первые три общие факторы и дайте им экономическую интерпретацию по матрице факторных нагрузок A ;
- 4) на основании матрицы индивидуальных значений F проведите графически классификацию объектов по двум первым общим факторам;
- 5) проведите графически классификацию признаков по двум первым общим факторам.

5 Дискриминантный анализ

Задание. Предприятия характеризуются пятью экономическими показателями. Значения данных показателей приведены в приложении А. В каждом варианте (приложение А) даны две обучающие выборки (предприятия группы А и предприятия группы В).

Вычислите оценки значений дискриминантной функции для оставшихся предприятий и проведите их классификацию. Дайте экономическую интерпретацию результатов дискриминации.

6 Кластерный анализ

Задание. На основании данных приложения А по алгоритму кластерного анализа проведите классификацию предприятий и постройте дендрограмму.

По дендрограмме выберите окончательный вариант разбиения предприятий на кластеры и дайте интерпретацию полученным результатам, используя статистические характеристики кластеров (векторы средних и средних квадратических отклонений).

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1 Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. – М.: Финансы и статистика, 2000. – 352 с.

2 Айвазян С.А. Прикладная статистика и основы эконометрики: Учебник. – М.: ЮНИТИ, 1998. – 640 с.

3 Салин В.Н., Чурилова Э.Ю. Практикум по курсу «Статистика» (в системе STATISTICA). – М.: Издательский дом «Социальные отношения». Издательство «Перспектива», 2002. – 188 с.

4 Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шеффер М. Многомерный статистический анализ в экономике: Учебное пособие для вузов/под. ред. проф. В.Н. Тамашевича. – М.: ЮНИТИ-ДАНА, 1999. – 598 с.

Приложение А

Рассматриваются следующие показатели:

Y_1 - производительность труда, тыс. руб.;

Y_2 - индекс снижения себестоимости продукции, %;

Y_3 - рентабельность, %;

X_4 - удельный вес рабочих в составе промышленно-производственного персонала;

X_5 - удельный вес покупных изделий;

X_6 - коэффициент сменности оборудования;

X_7 - премии и вознаграждения на одного работника, тыс. руб.;

X_8 - удельный вес потерь от брака;

X_9 - фондоотдача;

X_{10} - среднегодовая численность промышленно-производственного персонала, чел.;

X_{11} - среднегодовая стоимость основных производственных фондов, тыс. руб.;

X_{12} - среднегодовой фонд заработной платы промышленно-производственного персонала, руб.;

X_{13} - фондовооруженность труда;

X_{14} - оборачиваемость нормируемых оборотных средств;

X_{15} - оборачиваемость ненормируемых оборотных средств;

X_{16} - непроизводственные расходы, тыс. руб.

Таблица А.1 Таблица исходных данных

№ организации	Y_1	Y_2	Y_3	X_4	X_5	X_6	X_7	X_8
1	2	3	4	5	6	7	8	9
1	9,26	204,20	13,26	0,78	0,40	1,37	1,23	0,23
2	9,38	209,60	10,16	0,75	0,26	1,49	1,04	0,39
3	12,11	222,60	13,72	0,68	0,40	1,44	1,80	0,43
4	10,81	236,70	12,85	0,70	0,50	1,42	0,43	0,18
5	9,35	62,00	10,63	0,62	0,40	1,35	0,88	0,15
6	9,87	53,10	9,12	0,76	0,19	1,39	0,57	0,34
7	8,17	172,10	25,83	0,73	0,25	1,16	1,72	0,38
8	9,12	56,50	23,39	0,71	0,44	1,27	1,70	0,09
9	5,88	52,60	14,68	0,69	0,17	1,16	0,84	0,14
10	6,30	46,60	10,05	0,73	0,39	1,25	0,60	0,21
11	6,22	53,20	13,99	0,68	0,33	1,13	0,82	0,42
12	5,49	30,10	9,68	0,74	0,25	1,10	0,84	0,05
13	6,50	146,40	10,03	0,66	0,32	1,15	0,67	0,29
14	6,61	18,10	9,13	0,72	0,02	1,23	1,04	0,48
15	4,32	13,60	5,37	0,68	0,06	1,39	0,66	0,41
16	7,37	89,80	9,86	0,77	0,15	1,38	0,86	0,62
17	7,02	62,50	12,62	0,78	0,08	1,35	0,79	0,56
18	8,25	46,30	5,02	0,78	0,20	1,42	0,34	1,76
19	8,15	103,50	21,18	0,81	0,20	1,37	1,60	1,31
20	8,72	73,30	25,17	0,79	0,30	1,41	1,46	0,45
21	6,64	76,60	19,40	0,77	0,24	1,35	1,27	0,50
22	8,10	73,10	21,00	0,78	0,10	1,48	1,58	0,77
23	5,52	32,30	6,57	0,72	0,11	1,24	0,68	1,20
24	9,37	199,60	14,19	0,79	0,47	1,40	0,86	0,21
25	13,17	598,10	15,81	0,77	0,53	1,45	1,98	0,25
26	6,67	71,20	5,23	0,80	0,34	1,40	0,33	0,15
27	5,68	90,80	7,99	0,71	0,20	1,28	0,45	0,66
28	5,22	82,10	17,50	0,79	0,24	1,33	0,74	0,74
29	10,02	76,20	17,16	0,76	0,54	1,22	0,03	0,32
30	8,16	119,50	14,54	0,78	0,40	1,28	0,99	0,89
31	3,78	21,90	6,24	0,62	0,20	1,47	0,24	0,23
32	6,48	48,40	12,08	0,75	0,64	1,27	0,57	0,32
33	10,44	173,50	9,49	0,71	0,42	1,51	1,22	0,54
34	7,65	74,10	9,28	0,74	0,27	1,46	0,68	0,75

1	2	3	4	5	6	7	8	9
35	8,77	68,60	11,42	0,65	0,37	1,27	1,00	0,16
36	7,00	60,80	10,31	0,66	0,38	1,43	0,81	0,24
37	11,06	355,60	8,65	0,84	0,35	1,50	1,27	0,59
38	9,02	264,80	10,94	0,74	0,42	1,35	1,14	0,56
39	13,28	526,60	9,87	0,75	0,32	1,41	1,89	0,63
40	9,27	118,60	6,14	0,75	0,33	1,47	0,67	1,10
41	6,70	37,10	12,93	0,79	0,29	1,35	0,96	0,39
42	6,69	57,70	9,78	0,72	0,30	1,40	0,67	0,73
43	9,42	51,60	13,22	0,70	0,56	1,20	0,98	0,28
44	7,24	64,70	17,29	0,66	0,42	1,15	1,16	0,10
45	5,39	48,30	7,11	0,69	0,26	1,9	0,54	0,68
46	5,61	15,00	22,49	0,71	0,16	1,26	1,23	0,87
47	5,59	87,50	12,14	0,73	0,45	1,36	0,78	0,49
48	6,57	108,40	15,25	0,65	0,31	1,15	1,16	0,16
49	6,54	267,30	31,34	0,82	0,08	1,87	4,44	0,85
50	4,23	34,20	11,56	0,80	0,68	1,17	1,06	0,13

Таблица А.1 Таблица исходных данных

№ организации	X_9	X_{10}	X_{11}	X_{12}	X_{13}	X_{14}	X_{15}	X_{16}
1	2	3	4	5	6	7	8	9
1	1,45	26006	167,69	47750	6,40	166,32	10,08	17,72
2	1,30	23935	186,10	50391	7,80	92,88	14,76	18,39
3	1,37	22589	220,45	43149	9,76	158,04	6,48	26,46
4	1,65	21220	169,30	41089	7,90	93,96	21,96	22,37
5	1,91	7394	39,53	14257	5,35	173,88	11,88	28,13
6	1,68	11586	40,41	22661	9,90	162,30	12,60	17,55
7	1,94	26609	102,96	52509	4,50	88,56	11,52	21,92
8	1,89	7801	37,02	14903	4,88	101,16	8,28	19,52
9	1,94	11587	45,74	25587	3,46	166,32	11,52	23,99
10	2,06	9475	40,07	16821	3,60	140,76	32,40	21,76
11	1,96	10811	45,44	19459	3,56	128,52	11,52	25,68
12	1,02	6371	41,08	12973	5,65	177,84	17,28	18,13
13	1,85	26761	136,14	50907	4,28	114,48	16,20	25,74
14	0,88	4210	42,39	6920	8,85	93,24	13,32	21,21
15	0,62	3557	37,39	5736	8,52	126,72	17,28	22,97

1	2	3	4	5	6	7	8	9
16	1,09	14148	101,78	26705	7,19	91,80	9,72	16,38
17	1,60	9872	47,55	20068	4,82	69,12	16,20	13,21
18	1,53	5975	32,61	11487	5,46	66,24	24,84	14,48
19	1,40	16662	103,25	32029	6,20	67,68	14,76	13,38
20	2,22	9166	38,95	18946	4,25	50,40	7,56	13,69
21	1,32	15118	81,32	28025	5,38	70,56	8,64	16,66
22	1,48	11429	67,26	20968	5,88	72,00	8,64	15,06
23	0,68	6462	59,92	11049	9,27	97,20	9,00	20,09
24	2,30	24628	107,34	45893	4,36	80,28	14,76	15,98
25	1,37	49727	512,60	99400	10,31	51,48	10,08	18,27
26	1,51	11470	53,81	20719	4,69	105,12	14,76	14,42
27	1,43	19448	80,83	36813	4,16	128,52	10,44	22,76
28	1,82	18963	59,42	33956	3,13	94,68	14,76	15,41
29	2,62	9185	36,96	17016	4,02	85,32	20,52	19,35
30	1,75	17478	91,43	34873	5,23	76,32	14,40	16,83
31	1,54	6265	17,16	11237	2,74	153,00	24,84	30,53
32	2,25	8810	27,29	17306	3,10	107,64	11,16	17,98
33	1,07	17659	184,33	39250	10,44	90,72	6,48	22,09
34	1,44	10342	58,42	19074	5,65	82,44	9,72	18,29
35	1,40	8901	59,40	18452	6,67	79,92	3,24	26,05
36	1,31	8402	49,63	17500	5,91	120,96	6,48	26,20
37	1,12	32625	391,27	7888	11,99	84,60	5,40	17,26
38	1,16	31160	258,62	58947	8,30	85,32	6,12	18,83
39	0,88	46461	75,66	94697	1,63	101,52	8,64	19,70
40	1,07	13833	123,68	29626	8,94	107,64	11,88	16,87
41	1,24	6391	37,21	11688	5,82	85,32	7,92	14,63
42	1,49	11115	53,37	21955	4,80	131,76	10,08	22,17
43	2,03	6555	32,87	12243	5,01	116,64	18,72	22,62
44	1,84	11085	45,63	20193	4,12	138,24	13,68	26,44
45	1,22	9484	48,41	20122	5,10	156,96	16,56	22,26
46	1,72	3967	13,58	7612	3,49	137,52	14,76	19,13
47	1,75	15283	63,99	27404	4,19	135,72	7,92	18,28
48	1,46	20874	104,55	39648	5,01	155,52	18,36	28,23
49	1,60	19418	222,11	43799	11,44	48,60	8,28	12,39
50	1,47	3351	25,76	6235	7,67	42,84	14,04	11,64

Учебное пособие

Лубова Татьяна Николаевна

Учебное пособие по многомерным статистическим методам

Печатается в авторской редакции

Подписано в печать _____ 2015 г. Формат бумаги

Усл. печ. л. ____ Усл. изд. л. _____ Бумага типографская.

Гарнитура Таймс. Печать офсетная. Заказ _____. Тираж 50 экз.

Издательство Башкирского государственного аграрного университета

Типография Башкирского государственного аграрного университета

Адрес издательства и типографии: 450001, г. Уфа, ул. 50 лет Октября, 34