

**Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное
учреждение высшего образования
Санкт-Петербургский горный университет**

Кафедра системного анализа и управления

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
ИССЛЕДОВАНИЯ ЭКОНОМИЧЕСКИХ И
СОЦИАЛЬНЫХ СИСТЕМ**

*Методические указания к практическим занятиям
для студентов бакалавриата направления 27.03.03*

**САНКТ-ПЕТЕРБУРГ
2021**

УДК 519.237(073)

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ
ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ:** Методические указания к
практическим занятиям / Санкт-Петербургский горный университет. Сост.:
О.В. Афанасьева, Д.А. Первухин. СПб, 2021. 30 с.

Содержатся краткие теоретические сведения и задания для проведения
практических занятий по учебной дисциплине «Математические методы
исследования экономических и социальных систем».

Предназначены для студентов бакалавриата направления 27.03.03
«Системный анализ и управление».

Научный редактор проф. *В.Е. Трушников*

Рецензент доц. *Г.А. Митрофанов* (ФГКВОУ ВО «Михайловская военная
артиллерийская академия»)

© Санкт-Петербургский
горный университет, 2021

МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

*Методические указания к практическим занятиям
для студентов бакалавриата направления 27.03.03*

Сост. *О.В. Афанасьева, Д.А. Первухин*

Печатается с оригинал-макета, подготовленного кафедрой
системного анализа и управления

Ответственный за выпуск *О.В. Афанасьева*

Лицензия ИД № 06517 от 09.01.2002

Подписано к печати 08.06.2021. Формат 60×84/16.
Усл. печ. л. 1,7. Усл.кр.-отт. 1,7. Уч.-изд.л. 1,5. Тираж 75 экз. Заказ 562.

Санкт-Петербургский горный университет
РИЦ Санкт-Петербургского горного университета
Адрес университета и РИЦ: 199106 Санкт-Петербург, 21-я линия, 2

Введение

Цель изучения дисциплины «Математические методы исследования экономических и социальных систем»: системный анализ и обобщение научно-технической информации, отечественного и зарубежного опыта, на базе системно-аналитического исследования, принципов и технологий управления; системно-аналитическая постановка задач математического, физического и других видов моделирования процессов и объектов исследования и управления ими, формулировка задач исследования на базе системного анализа и управления, включая модели, методы, технологии и алгоритмы программного обеспечения автоматизированного проектирования и системных исследований; применение методов системного анализа, управления и современных инструментальных проектных и технологических методов при разработке аппаратных и программных средств.

Основными задачами дисциплины «Математические методы исследования экономических и социальных систем» являются: сбор и системный анализ данных для проектирования и конструирования; проведение предварительного технико-экономического обоснования и системно-аналитических проектных и конструкторских решений; проведение натуральных вычислительных, имитационных и других типов исследований по заданной методике и системный анализ их результатов; выполнение измерений и описание исследований, подготовка данных для составления отчетов по результатам исследований и научных публикаций; формирование отчета по теме исследований, участие во внедрении результатов исследований и разработок; применение Web-технологий при удаленном доступе в системах и распределенных вычислениях при выполнении проектно-технологических работ; использование проектно-технологических стандартов и типовых методов контроля и оценки качества продукции; проектирование и конструирование систем, устройств и баз данных в соответствии с техническим заданием с использованием современных технологий проектирования; разработка и оформление проектно-конструкторской и рабочей технической документации; контроль соответствия проектов и

технической документации стандартам, техническим условиям и другим нормативным документам; участие в работах по проектированию и автоматизации технологических процессов при подготовке производства новой продукции; освоение и применение современных проектно-технологических комплексов исследования и автоматизированного проектирования объектов деятельности.

Главным содержанием этого вида учебных занятий является работа каждого обучающегося по овладению практическими умениями и навыками профессиональной деятельности.

1. Общая характеристика

Целью практических занятий является совершенствование умений и навыков решения практических задач.

Главным содержанием этого вида учебных занятий является работа каждого обучающегося по овладению практическими умениями и навыками профессиональной деятельности. В таблице 1. приведен перечень практических работ по дисциплине «Математические методы исследования экономических и социальных систем»

Таблица 1

Перечень практических работ по дисциплине «Математические методы исследования экономических и социальных систем»

№ п/п	Наименование работ	Трудоемкость (час.)
1.	Применение методов многомерного регрессионного анализа при исследовании технических, экономических и социальных систем	6
2.	Применение факторного анализа для исследования технических, экономических и социальных систем	6
3.	Применение кластерного анализа для исследования технических, экономических и социальных систем	6
4.	Применение дискриминантного анализа для исследования технических, экономических и социальных систем	9
	Итого:	27

2. Контрольные точки и виды отчетности по ним

№ п/п	Вид деятельности студентов	Сроки выполнения	Кол. баллов
1.	Сдача отчета по практической работе 1 «Применение методов многомерного регрессионного анализа при исследовании технических, экономических и социальных систем»	3-ая неделя	6
2.	Сдача отчета по практической работе 2 «Применение факторного анализа для исследования технических, экономических и социальных систем»	6-ая неделя	6
3.	Сдача отчета по практической работе 3 «Применение кластерного анализа для исследования технических, экономических и социальных систем»	9-ая неделя	6
4.	Сдача отчета по практической работе 4 «Применение дискриминантного анализа для исследования технических, экономических и социальных систем»	12-ая неделя	9
	Итого за семестр		27
	Итого		27

3. Методические указания для проведения практических занятий по учебной дисциплине «Математические методы исследования экономических и социальных систем»

Практическая работа 1

ПРИМЕНЕНИЕ МЕТОДОВ МНОГОМЕРНОГО РЕГРЕССИОННОГО АНАЛИЗА ПРИ ИССЛЕДОВАНИИ ТЕХНИЧЕСКИХ, ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

В зависимости от двух фундаментальных признаков построения математической модели – вида описания причинно-следственных связей и изменений их во времени – различают детерминистические и стохастические, статические и динамические модели.

Математическая модель называется детерминистической (детерминированной), если все ее параметры и переменные являются однозначно определяемыми величинами, а также выполняется условие полной определенности информации. В противном случае, в условиях неопределенности информации, когда параметры и переменные модели – случайные величины, модель называется стохастической (вероятностной).

Модель называется динамической, если как минимум одна переменная изменяется по периодам времени, и статической, если принимается гипотеза, что переменные не изменяются по периодам времени.

В простейшем случае балансовые модели выступают в виде уравнения баланса, где в левой части располагается сумма каких-либо поступлений, а в правой – расходная часть также в виде суммы. Например, в таком виде представляется годовой бюджет организации.

На основе статистических данных могут строиться не только балансовые, но и корреляционно-регрессионные модели.

Условие. Проведённый анализ причин количества срывов сроков выполнения работ по нанесению огнезащитных покрытий показал, что все срывы (в основном) происходили либо из-за

поломки оборудования (фактор X_1), либо из-за человеческого фактора (фактор X_2), либо из-за погодных условий. В табл. 1 приведены статистические данные срывов сроков за последний год.

Таблица 3.1

Статистические данные количества срывов сроков

Период	поломка оборудования	человеческий фактор	Процент срывов сроков выполнения работ
	x1	x2	y
июл.19	10	12	23
авг.19	13	7	22
сен.19	11	8	22
окт.19	14	6	23
ноя.19	12	10	24
дек.19	13	11	26
январь.20	18	4	25
фев.20	17	8	27
мар.20	13	11	26
апр.20	11	15	27
май.20	12	9	29

Установим зависимость между количествами срывов работ и их причинами, для этого построим уравнение множественной регрессии. Для построения уравнения множественной регрессии чаще всего используются следующие функции: линейную, степенную, экспоненту и гиперболу.

Теоретическая часть

Уравнение множественной регрессии в естественной форме имеет вид:

$$\hat{Y} = a + b_1X_1 + b_2X_2$$

Для решения целесообразно использовать процедуру «Регрессия» из пакета «Анализ данных» электронных таблиц Excel, описание которой приведено в учебном пособии.

В результате получили, что (рис.1) срывы сроков выполнения работ по нанесению огнезащитных покрытий на 50 процентов связано с нехваткой людей, а технический фактор

(поломка или нехватка оборудования) был причиной только в 15% случаев.

	поломка оборудования	человеческий фактор	процент срывов сроков выполнения работ
	y	x_1	x_2
y	1		
x_1	0,156083634	1	
x_2	0,504617138	-0,604355659	1

Рис. 1. Степень зависимости количества срывов сроков выполнения работ от технического и человеческого факторов

Расчёт зависимость количества срывов сроков выполнения работ от поломки оборудования

Воспользовавшись процедурой «Регрессия» из пакета «Анализ данных» электронных таблиц Excel были установлены параметры (рис.2), позволяющие установить и оценить зависимость количества срывов сроков выполнения работ от поломки оборудования ($Y=f(X_1)$).

Регрессионная статистика									
Множественный R	0,156084								
R-квадрат	0,024362								
Нормированный R-квадрат	-0,0732								
Стандартная ошибка	2,734939								
Наблюдения	12								
Дисперсионный анализ									
	df	SS	MS	F	Значимость F				
Регрессия	1	1,867761	1,867761	0,24970433	0,628094557				
Остаток	10	74,79891	7,479891						
Итого	11	76,66667							
	Коэффи- циенты	Станда- ртная ошибка	t- статис- тика	P- Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%	
Y-пересечение	23,04241	4,652047	4,953176	0,00057582	12,67700044	33,40781	12,677	33,40781	
x_1	0,175103	0,350412	0,499704	0,62809456	-0,605665042	0,95587	-0,60567	0,95587	

Рис. 2. Зависимость количества срывов сроков выполнения работ от поломки оборудования

Коэффициент множественной корреляции (показывают степень влияния одного фактора X_1 на результирующий фактор Y при фиксированном значении другого фактора X_2 . В данном случае он равен 0,16.

Анализ данных решения показывает слабую зависимость между количеством срывов сроков выполнения работ и техническим фактором (поломки или нехватка оборудования).

Зависимость количества срывов сроков выполнения работ от человеческого фактора

Воспользовавшись процедурой «Регрессия» из пакета «Анализ данных» электронных таблиц Excel были установлены параметры (рис.3), позволяющие установить и оценить зависимость количества срывов сроков выполнения работ от человеческого фактора (нехватка работников) ($Y=f(X_2)$).

Регрессионная статистика									
Множественный R	0,504617138								
R-квадрат	0,254638456								
Нормированный R-квадрат	0,180102302								
Стандартная ошибка	2,390489176								
Наблюдения	12								
Дисперсионный анализ									
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимость F</i>				
Регрессия	1	19,52228164	19,52228	3,41630794	0,094303577				
Остаток	10	57,14438503	5,714439						
Итого	11	76,66666667							
	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>	<i>P-значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>	<i>Нижние 95,0%</i>	<i>Верхние 95,0%</i>	
Y-пересечение	21,50802139	2,18162452	9,858718	1,8107E-06	16,64705906	26,36898	16,64706	26,36898	
x2	0,395721925	0,214097506	1,848326	0,09430358	-0,081317044	0,872761	-0,08132	0,872761	

Рис. 3. Зависимость количества срывов сроков выполнения работ от человеческого фактора

Коэффициент множественной корреляции (показывают

степень влияния фактора X_2 на результирующий фактор Y при фиксированном значении фактора X_1 . В данном случае он равен 0,5.

Анализ данных решения показывает имеющуюся зависимость между количеством срывов сроков выполнения работ и человеческим фактором (нехватка сотрудников).

Результаты использования методов многомерного корреляционно – регрессионного анализа

Результаты применения многомерного корреляционно-регрессионного анализа, проведённого при помощи процедуры «Регрессия» из пакета «Анализ данных» электронных таблиц Excel, приведены на рис. 4.

Регрессионная статистика								
Множественный R	0,767803428							
R-квадрат	0,589522104							
Нормированный R-ква	0,498304794							
Стандартная ошибка	1,869936778							
Наблюдения	12							
Дисперсионный анализ								
	df	SS	MS	F	Значимость F			
Регрессия	2	45,19669467	22,59835	6,46283149	0,018188798			
Остаток	9	31,46997199	3,496664					
Итого	11	76,66666667						
	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%	Нижние 95,0%	Верхние 95,0%
Y-пересечение	7,519344466	5,437177971	1,38295	0,20002336	-4,780406601	19,8191	-4,78041	19,8191
x1	0,814853231	0,300715666	2,709713	0,02400877	0,134587134	1,495119	0,134587	1,495119
x2	0,739964734	0,210207857	3,520157	0,00651383	0,264441525	1,215488	0,264442	1,215488

Рис. 4. Зависимость количества срывов сроков выполнения работ от технического и человеческого факторов

На основе проведенных вычислений получается уравнение множественной регрессии в естественной форме:

$$\hat{Y} = a + b_1 X_1 + b_2 X_2 .$$

То есть уравнение регрессии будет иметь вид:

$$\hat{Y} = 7,52 + 0,82 X_1 + 0,74 X_2 .$$

Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции в данном случае он равен 0,77.

С целью оценки адекватности модели также используют общий критерий Фишера (F -критерий), проверяющий основную гипотезу (H_0) о статической значимости уравнения множественной регрессии и показателя тесноты связи ($R_{yx1x2\dots xn}$). Для этого используется формула (4.14):

$$F_{расч} = \frac{R_{yx1x2}^2}{1 - R_{yx1x2}^2} \cdot \frac{n - m - 1}{m},$$

где m - число факторов переменных (в нашем случае $m = 2$ (X_1, X_2);

По таблицам F – критерия находится значение $F_{табл}$

$$F_{табл} = \alpha = 1 - p / \begin{array}{l} v_1 = m' - 1; \\ v_2 = n - m', \end{array}$$

где m' - число коэффициентов (параметров) уравнения регрессии (в данном случае $m' = 3$ (a, b_1, b_2)).

Основная гипотеза(H_0) принимается, если $F_{табл} < F_{расч}$; в этом случае с вероятностью $p = 1 - \alpha = 0,95$ делается заключение о статической значимости уравнения в целом и показателя тесноты связи $R_{yx1x2\dots xn}$, которые сформировались под воздействием факторов X_1, X_2 .

Частные F критерии - F_{X1} (оно равно 0,32) и (оно равно 3,42) оценивают статическую значимость присутствия факторов X_1 и X_2 в уравнении множественной регрессии. Кроме того, F_{x1} оценивает целесообразность включения в уравнение регрессии первого фактора X_1 после того, как в него был включен фактор X_2 .

В данном случае основная гипотеза (H_0) принимается (является статистически значимой) так как $F_{табл} < F_{расч}$.

Выводы Данные анализа решения задачи наглядно свидетельствуют о корректности применения процедур упрощения

моделей множественной регрессии. Так, расчетное значение критерия Фишера ($F_{расч}$) при двухфакторной форме модели имеет значение (6,46), в то время как при исследовании однофакторной линейной зависимости $Y = f(X_2)$, $F_{расч} = 3,42$, а для $Y = f(X_1)$, $F_{расч} = 0,32$, указывая тем самым на слабую статистическую значимость фактора X_1 . Тесноту совместного влияния факторов на результат оценивает индекс множественной корреляции в данном случае он равен 0,77.

Таким образом, несмотря на то, что фактор X_2 , в ходе одномерного корреляционно – регрессионного анализа оказался значимым, а фактор X_1 оказался незначимый, в целом в ходе многомерного корреляционно – регрессионного анализа подтверждена гипотеза о статической значимости линейной модели множественной корреляции.

$$\hat{Y} = 7,52 + 0,82 X_1 + 0,74 X_2 .$$

Таким образом, влияние на исследуемый процесс (срывы сроков выполнения работ) имеет как человеческий фактор, так и поломка оборудования, но влияние человеческого фактора сильнее.

Практическая работа 2

ПРИМЕНЕНИЕ ФАКТОРНОГО АНАЛИЗА ДЛЯ ИССЛЕДОВАНИЯ ТЕХНИЧЕСКИХ, ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

Математическое моделирование основано на явлении изоморфизма - сходстве форм при качественном различии явлений. Благодаря изоморфизму мы можем моделировать одну систему с помощью другой, вместо одного явления изучать другое. При математическом моделировании вместо изучения и исследования оригинала исследуются математические зависимости, описывающие оригинал.

Математической моделью объекта называют его описание математическими средствами, позволяющее выводить суждение о некоторых свойствах объекта при помощи формальных процедур. Использование математического языка предопределяет необходимость все операции и преобразования в математических моделях осуществлять над математическими объектами: числами, векторами, множествами, матрицами, функциями и т. д. В наиболее общем виде математическая модель объекта представляется уравнением

$$F(X, Y) = \text{const},$$

где X , Y - векторы управляемых и неуправляемых параметров модели.

При изучении сложных объектов приходится учитывать большое число взаимосвязанных факторов, при этом реально доступная информация может иметь любую степень детерминированности, быть плохо формализуемой, поступать в произвольной форме и эволюционировать во времени. При математическом моделировании факторы отображаются в виде математических конструкций - таких, как параметры состояния объекта и среды, ограничения по области совместного и согласованного изменения этих параметров и т.д. Многие факторы при этом могут оказаться плохо формализуемыми, т.е. их отражение математическими конструкциями оказывается либо невозможным, либо просто плохим.

Условие Совокупность из четырех промышленных предприятий оценена по трем характерным признакам: выработке на одного среднегодового работника X_1 , уровню рентабельности X_2 и уровню фондоотдачи X_3 .

В результате предварительных аналитических расчетов по исходным данным X получена матрица парных корреляций:

$$R = \begin{pmatrix} 1 & 0,581 & 0,154 \\ 0,581 & 1 & 0,439 \\ 0,154 & 0,439 & 1 \end{pmatrix}$$

Используя алгоритм метода главных компонент, найдем собственные числа и собственные векторы матрицы R и построим матрицы с аналитическими результатами (A и F).

Решение.

- По рекуррентным соотношениям Фаддеева исчислим определитель матрицы парных корреляций $|R|$.

Первый шаг.

$$R = A \text{ и } A = A_1, \text{ тогда } P_1 = t_r A_1 = 1 + 1 + 1 = 3,$$

$$B_1 = A_1 - P_1 E = \begin{pmatrix} -2 & 0,581 & 0,154 \\ 0,581 & -2 & 0,439 \\ 0,154 & 0,439 & -2 \end{pmatrix}$$

Второй шаг:

$$\begin{aligned} A_2 = AB_1 &= \begin{pmatrix} 1 & 0,581 & 0,154 \\ 0,581 & 1 & 0,439 \\ 0,154 & 0,439 & 1 \end{pmatrix} \cdot \begin{pmatrix} -2 & 0,581 & 0,154 \\ 0,581 & -2 & 0,439 \\ 0,154 & 0,439 & -2 \end{pmatrix} \\ &= \begin{pmatrix} -1,638 & -0,513 & 0,101 \\ -0,513 & -1,469 & -0,35 \\ 0,101 & -0,35 & -1,783 \end{pmatrix}; \end{aligned}$$

$$P_2 = (1/2)t_r A_2 = (1/2)(-1,638) + (-1,469) + (-1,783) = -2,445,$$

$$B_2 = A_2 - P_2 E = \begin{pmatrix} 0,807 & 0,581 & 0,154 \\ -0,513 & 0,976 & -0,350 \\ 0,101 & -0,350 & 0,662 \end{pmatrix}$$

Третий шаг:

$$A_3 = AB_2 = \begin{pmatrix} 1 & 0,581 & 0,154 \\ 0,581 & 1 & 0,439 \\ 0,154 & 0,439 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0,807 & 0,581 & 0,154 \\ -0,513 & 0,976 & -0,350 \\ 0,101 & -0,350 & 0,662 \end{pmatrix} \\ = \begin{pmatrix} 0,524 & 0 & 0 \\ 0 & 0,524 & 0 \\ 0 & 0 & 0,524 \end{pmatrix}; \\ P_3 = \left(\frac{1}{3}\right) (3 \cdot 0,524) = 0,524, \quad B_3 = A_3 - P_3 E = 0.$$

В итоге $|R| = 0,524$ и $B_3 = 0$. Обратим внимание, что в ходерасчетов все промежуточные матрицы A_j, uB_j — симметрические.

2. Построим характеристическое уравнение:

$$\lambda^3 - 3\lambda^2 + 2,445\lambda - 0,524 = 0,$$

Откуда найдём:

$$\lambda_1 = 1,798; \lambda_2 = 0,875 \text{ и } \lambda_3 = 0,327.$$

Таким образом, наши исходные элементарные признаки X_1, X_2, X_3 могут быть обобщены значениями трех главных компонент, причем первая главная компонента F_1 объяснит примерно 60% всей вариации X_j ($1,798/3 = 0,599$), вторая главная компонента F_2 объяснит 29,2% — меньшую часть по сравнению с F_1 общей дисперсии ($0,875/3 = 0,292$), наконец, третья главная компонента F_3 охватывает оставшуюся, еще не объясненную вариацию входных признаков — 10,9% ($0,327/3 = 0,109$). Все главные компоненты F_1, F_2, F_3 объясняют вариацию X_1, X_2, X_3 полностью, на 100% ($59,9 + 29,2 + 10,9$).

Собственные векторы матрицы парных корреляций R найдем решением трех систем линейных уравнений соответственно для

$$\lambda_1 = 1,798, \lambda_3 = 0,875 \text{ и } \lambda_3 = 0,327.$$

Для определения области решений в каждой системе будем задавать одному из неизвестных признаков u_{3j} значение, равное единице.

Первая система уравнений, для $\lambda_1 = 1,798$:

$$(1 - 1,798) \cdot u_{11} + 0,581 \cdot u_{21} + 0,154 \cdot u_{31} = 0, \quad u_{11} = 1,262$$

$$\begin{aligned} 0,581 \cdot u_{11} + (1 - 1,798) \cdot u_{21} + 0,439 u_{31} &= 0, & u_{11} &= 1,469 \\ 0,154 \cdot u_{11} + 0,581 \cdot u_{21} + (1 - 1,798) \cdot u_{31} &= 0, & u_{11} &= 1,000 \end{aligned}$$

Вторая система уравнений, $\lambda_2 = 0,875$:

$$\begin{aligned} (1 - 0,875) \cdot u_{12} + 0,581 \cdot u_{22} + 0,154 \cdot u_{32} &= 0, & u_{12} &= 0,144 \\ 0,581 \cdot u_{12} + (1 - 0,875) \cdot u_{22} + 0,439 \cdot u_{32} &= 0, & u_{12} &= 0,234 \\ 0,154 \cdot u_{12} + 0,439 \cdot u_{22} + (1 - 0,875) \cdot u_{32} &= 0, & u_{12} &= 1,000 \end{aligned}$$

Третья система уравнений, $\lambda_3 = 0,327$:

$$\begin{aligned} (1 - 0,327) \cdot u_{13} + 0,581 \cdot u_{23} + 0,154 \cdot u_{33} &= 0, & u_{12} &= 1,307 \\ 0,581 \cdot u_{13} + (1 - 0,327) \cdot u_{23} + 0,439 \cdot u_{33} &= 0, & u_{12} &= 1,779 \\ 0,154 \cdot u_{13} + 0,439 \cdot u_{23} + (1 - 0,327) \cdot u_{33} &= 0, & u_{12} &= 1,000 \end{aligned}$$

Матрица собственных векторов принимает вид:

$$U = \begin{pmatrix} 1,262 & -0,144 & 1,307 \\ 1,469 & -0,234 & -1,779 \\ 1,000 & 1,000 & 1,000 \end{pmatrix}$$

Пронормируем векторы U_j , т.е. найдем $V_j = U_j/|U_j|$ и получим матрицу нормированных значений собственных векторов:

$$V = \begin{pmatrix} 0,579 & -1,139 & 0,539 \\ 0,674 & -0,225 & -0,734 \\ 0,459 & 0,964 & 0,413 \end{pmatrix}$$

так как V — матрица, отображающая ортонормированное пространство, в общем должно выполняться условие: $VV = E$.

Матрицу факторного отображения (A) получим из матричного уравнения $A = V\Lambda^{1/2}$

$$A = \begin{pmatrix} 0,579 & -0,139 & 0,539 \\ 0,674 & -0,225 & -0,734 \\ 0,459 & 0,964 & 0,413 \end{pmatrix} \cdot \begin{pmatrix} \sqrt{1,798} & & \\ & \sqrt{0,875} & \\ & & \sqrt{0,327} \end{pmatrix} =$$

	F_1	F_2	F_3
X_1	0,776	-0,130	0,308
X_2	0,904	-0,210	-0,420
X_3	0,616	0,902	0,236

Матрица A содержит частные коэффициенты корреляции, представляющие связи исходных признаков X_j и главных компонент F_r . Соответственно все элементы a_{ij} могут варьировать в пределах от -1 до +1.

Из равенства $A'A = \Lambda$ следует условие $\sum_j a_{jr}^2 = \lambda_r$. Проверим, как оно выполняется на исчисленных данных матрицы A :

$$\sum_j a_{j1}^2 = 0,776^2 + 0,904^2 + 0,616^2 = 1,798;$$

$$\sum_j a_{j2}^2 = 0,875; \quad \sum_j a_{j3}^2 = 0,327$$

Теперь запишем системы линейных уравнений зависимости элементарных признаков Z_j и главных компонент, или обобщенных признаков F_r ; максимально $r = j = 3$:

$$Z_1 = 0,776 \cdot F_1 - 0,130 \cdot F_2 + 0,308 \cdot F_3,$$

$$Z_2 = 0,904 \cdot F_1 - 0,210 \cdot F_2 - 0,420 \cdot F_3,$$

$$Z_3 = 0,616 \cdot F_1 + 0,902 \cdot F_2 + 0,236 \cdot F_3, u$$

$$F_1 = \frac{1}{1,798} \cdot (0,776 \cdot Z_1 + 0,904 \cdot Z_2 + 0,616Z \cdot_3),$$

$$F_2 = \frac{1}{0,875} (-0,130 \cdot Z_1 - 0,210 \cdot Z_2 + 0,902 \cdot Z_3),$$

$$F_3 = \frac{1}{0,375} (0,308 \cdot Z_1 - 0,420Z \cdot_2 + 0,236 \cdot Z_3).$$

На завершающем шаге алгоритма исчислим значения главных компонент для всех наблюдаемых объектов и построим матрицу F : $F = A^{-1}Z'$ и матрица Z известна из условия задачи, тогда

$$F = \begin{pmatrix} 0,542 & 0,507 & 0,196 \\ -0,776 & -0,010 & 0,994 \\ 1,554 & -1,283 & -0,075 \end{pmatrix} \cdot \begin{pmatrix} -0,971 & -0,868 & 1,478 & 0,361 \\ 0,549 & -1,684 & 0,882 & 0,253 \\ 0,076 & -1,069 & -0,534 & 1,527 \end{pmatrix}$$

$$= \begin{pmatrix} -0,233 & -1,533 & 1,143 & 0,623 \\ 0,823 & -0,372 & -1,687 & 1,236 \\ -2,218 & 0,892 & 1,205 & 0,121 \end{pmatrix}$$

Более привычной формой записи $n \times r$ значений главных компонент является транспонированная матрица F .

		F_r - главные компоненты		
		→		
		F_1	F_2	F_3
$F' =$	n_1	-0,233	0,823	-0,218
	n_2	-1,533	-0,372	0,892
	n_3	1,143	-1,687	1,205
	n_4	0,623	1,236	0,121
$\sum f_{ri}$		0	0	0

n_i - объёкты

Центр распределения значений главных компонент F_2 находится в точке $(0, 0, \dots, 0)$. Отсюда следует правило равенства суммы элементов каждого столбца матрицы F' нулю. Далее аналитические выводы по результатам расчетов следуют уже после принятия решения о числе значащих признаков Z_j и главных компонент F_r и определения названий главным компонентам.

Практическая работа 3

ПРИМЕНЕНИЕ КЛАСТЕРНОГО АНАЛИЗА ДЛЯ ИССЛЕДОВАНИЯ ТЕХНИЧЕСКИХ, ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

Условие. Определим сходство между предприятиями, если каждое из них характеризуется тремя признаками: X_1 - производство продукции, млрд.руб., X_2 - стоимость основных производственных фондов, млрд.руб.; X_3 - фонд заработной платы промышленно-производственного персонала, млрд.руб. (табл. 3.1, 3.2).

Таблица 3.1.

Матрица исходных данных

№ п/п	X_1	X_2	X_3
1	32,5	40,3	3,5
2	38,4	46,8	4,3
3	16,7	25,7	2,0
4	42,3	44,0	4,5

Таблица 3.2.

Матрица евклидовых расстояний

№ п/п	1	2	3	4
1	0	8,81	21,55	10,36
2		0	30,35	30,48
3			0	31,57
4				0

Решение. Оценка сходства между объектами сильно зависит от абсолютного значения признака и от степени его вариации в совокупности. Чтобы устранить подобное влияние на процедуру классификации, можно значения исходных переменных нормировать одним из следующих способов:

$$1. z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j}, \quad 2. z_{ij} = \frac{x_{ij}}{x_{maxj}}, \quad 3. z_{ij} = \frac{x_{ij}}{\bar{x}_j}, \quad 4. z_{ij} = \frac{x_{ij}}{x_{minj}}.$$

Продемонстрируем на нашем примере, как скажется нормирование исходных переменных на мерах сходства между объектами. Заменяем x_{ij} новыми значениями z_{ij} , полученными по формуле:

$$z_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_j},$$

и построим матрицу стандартизованных значений признаков и новую матрицу расстояний (табл. 3.3, 3.4).

В первой матрице расстояний (табл. 3.3) самыми «близкими» были объекты n_1 и n_2 ($d_{12} = 8,81$), а самыми «дальними» - объекты

n_3 и n_4 ($d_{34} = 31,57$).

После нормирования значений исходных переменных самыми «близкими» стали объекты n_2 и n_4 ($d_{24} = 0,56$), а самыми «дальними» - объекты n_2 и n_3 ($d_{23} = 13,2$) (табл. 4).

Таблица 3.3.

Матрица стандартизированных значений признаков

№ п/п	z_1	z_2	z_3
1	0,00205	0,13530	-0,10204
2	0,60718	0,93481	0,71429
3	-1,61846	-1,66052	-1,63215
4	1,00718	0,59041	0,91837

Таблица 3.4.

Матрица евклидовых расстояний

№ п/п	1	2	3	4
1	0	1,29	2,86	1,50
2		0	13,20	0,56
3			0	4,30
4				0

В качестве меры сходства отдельных переменных используются парные коэффициенты корреляции Пирсона. Если исходные переменные являются альтернативными признаками, т.е. принимают только два значения, то в качестве меры сходства можно использовать коэффициенты ассоциативности.

Вопрос о придании переменным соответствующих весов должен решаться после проведения исследователем тщательного анализа изучаемой совокупности и социально-экономической сущности классифицирующих переменных. Например, если для классификации предприятия используются переменные: X_1 - прибыль предприятия, X_2 - выработка продукции на одного работающего, X_3 - среднегодовая стоимость основных производственных фондов, то можно переменным задать веса пропорционально их степени важности для эффективности работы

предприятия:

$$w_{x_1} = 0,6; \quad w_{x_2} = 0,3; \quad w_{x_3} = 0,1.$$

Тогда евклидово расстояние будет определяться по формуле:

$$d_{ij} = \sqrt{0,6 \cdot (x_{i1} - \bar{x}_{j1})^2 + 0,3 \cdot (x_{i2} - x_{j2})^2 + 0,1 \cdot (x_{i3} - x_{j3})^2}.$$

Выбор меры расстояния и весов для классифицирующих переменных – очень важный этап кластерного анализа, так как от этих процедур зависят состав и количество формируемых кластеров, а также степень сходства объектов внутри кластеров.

Практическая работа 4.

ПРИМЕНЕНИЕ ДИСКРИМИНАНТНОГО АНАЛИЗА ДЛЯ ИССЛЕДОВАНИЯ ТЕХНИЧЕСКИХ, ЭКОНОМИЧЕСКИХ И СОЦИАЛЬНЫХ СИСТЕМ

Задание 1. Использование дискриминантного анализа для проведения многомерной классификации объектов.

В качестве обучающих будем использовать сначала две выборки, принадлежащие двум классам, а затем обобщим алгоритм классификации на случай k классов.

Условие. Имеются данные по двум группам промышленных предприятий машиностроительного комплекса:

X_1 — фондоотдача основных производственных фондов, руб.;

X_2 — затраты на рубль произведенной продукции, коп.;

X_3 — затраты сырья и материалов на один рубль продукции, коп.

Таблица 4.1

Исходные данные

	Номер предприятия	X_1	X_2	X_3
1-я группа	1	0,50	94,0	8,50
	2	0,67	75,4	8,79
	3	0,68	85,2	9,10
	4	0,55	98,8	8,47
2-я группа	5	1,52	81,5	4,95
	6	1,20	93,8	6,95
	7	1,46	86,5	4,70

Необходимо провести классификацию четырех новых предприятий, имеющих следующие значения исходных переменных:

1-е предприятие: $x_1 = 1,07$, $x_2 = 93,5$, $x_3 = 5,30$,

2-е предприятие: $x_1 = 0,99$, $x_2 = 84,0$, $x_3 = 4,85$,

3-е предприятие: $x_1 = 0,70$, $x_2 = 76,8$, $x_3 = 3,50$,

4-е предприятие: $x_1 = 1,24$, $x_2 = 88,0$, $x_3 = 4,95$.

Решение.

Для удобства запишем значения исходных переменных для каждой группы предприятий в виде матриц X_1 и X_2 :

$$X_1 = \begin{pmatrix} 0,50 & 94,0 & 8,50 \\ 0,67 & 75,4 & 8,79 \\ 0,68 & 85,2 & 9,10 \\ 0,55 & 98,8 & 8,47 \end{pmatrix},$$

$$X_2 = \begin{pmatrix} 1,52 & 81,5 & 4,95 \\ 1,20 & 93,8 & 6,95 \\ 1,46 & 86,5 & 4,70 \end{pmatrix}.$$

Рассчитаем среднее значение каждой переменной в отдельных группах для определения положения центров этих групп:

I группа: $\bar{x}_{11}=0,60$,

$\bar{x}_{21}=88,4$,

$\bar{x}_{31}=8,72$.

II группа: $\bar{x}_{12}=1,39$,

$\bar{x}_{22}=87,3$,

$\bar{x}_{32}=5,53$.

Дискриминантная функция $f(x)$ в данном случае имеет вид:

$$f(x) = a_1x_1 + a_2x_2 + a_3x_3$$

Коэффициенты a_1 , a_2 и a_3 вычисляются по формуле:

$$A = S_*^{-1}(\bar{X}_1 - \bar{X}_2).$$

где \bar{X}_1, \bar{X}_2 - векторы средних в первой и второй группах; A - вектор коэффициентов; S_* - матрица, обратная совместной ковариационной матрице.

Для определения совместной ковариационной матрицы

S_* нужно рассчитать матрицы S_1 и S_2 . Каждый элемент этих матриц представляет собой разность между соответствующим значением исходной переменной x_{ij} и средним значением этой переменной в данной группе \bar{x}_{ik} (k - номер группы):

$$S_1 = \begin{pmatrix} 0,0238 & -2,2460 & 0,0698 \\ -2,2460 & 318,76 & -5,958 \\ 0,0698 & -5,958 & 0,2602 \end{pmatrix};$$

$$S_2 = \begin{pmatrix} 0,00579 & -2,0450 & -0,4033 \\ -2,0450 & 76,530 & 13,2580 \\ -0,4033 & 13,258 & 3,0417 \end{pmatrix}.$$

Тогда совместная ковариационная матрица S_* будет равна:

$$S_* = \frac{1}{n_1 + n_2 = 2} (S_1 + S_2),$$

где n_1, n_2 - число объектов 1-й и 2-й группы;

$$S_* = \frac{1}{(4 + 3 - 2)} \begin{pmatrix} 0,0817 & -4,291 & -0,3335 \\ -4,291 & 395,290 & 7,300 \\ -0,3335 & 7,300 & 3,3019 \end{pmatrix} =$$

$$= \begin{pmatrix} 0,01634 & -0,8582 & -0,0667 \\ -0,8582 & 79,058 & 1,460 \\ -0,0667 & 1,460 & 0,6604 \end{pmatrix}$$

Отсюда находим вектор коэффициентов дискриминантной функции по формуле:

$$A = S_*^{-1}(\bar{X}_1 - \bar{X}_2) = \begin{pmatrix} 330,970 & -3,190 & 27,290 \\ -3,190 & 0,043 & -0,227 \\ 27,290 & -0,227 & 8,380 \end{pmatrix} \cdot \begin{pmatrix} -0,79 \\ 1,10 \\ 3,19 \end{pmatrix}$$

$$= \begin{pmatrix} -185,03 \\ 1,84 \\ 4,92 \end{pmatrix}$$

т.е. $a_1 = -185,03$, $a_2 = 1,84$, $a_3 = 4,92$.

Рассчитаем значения дискриминантной функции для каждого объекта с учётом полученных значения коэффициентов:

Для 1 -го множества

$$\begin{cases} f_{11} = 0,5(-185,03) + 94 \times 1,84 + 6,5 \times 4,92 = 122,265. \\ f_{12} = 0,67(-185,03) + 75,4 \times 1,84 + 8,79 \times 4,92 = 58,0127. \\ f_{13} = 0,68(-185,03) + 85,2 \times 1,84 + 9,1 \times 4,92 = 75,7196. \\ f_{13} = 0,55(-185,03) + 98,8 \times 1,84 + 8,47 \times 4,92 = 121,6979. \\ \bar{f}_1 = 94,4238 \end{cases}$$

Для 2 -го множества

$$\begin{cases} f_{21} = 1,52 - 185,03 + 81,5 \times 1,84_{4,95} \cdot 4,92 = -106,9316, \\ f_{22} = 1,20 - 185,03 + 93,8 \times 1,84 + 6,95 \times 4,92 = -15,25, \\ f_{23} = 1,46 - 185,03 + 86,5 \times 1,84 + 4,7 \times 4,92 = -83,8508, \\ \bar{f}_2 = -70,0138 \end{cases}$$

Тогда константа дискриминации C будет равна:

$$C = \frac{1}{2}(94,4238 - 70,0138) = 12,205.$$

После получения константы дискриминации можно проверить правильность распределения объектов в уже существующих двух классах, а также провести классификацию новых объектов.

Рассмотрим, например, объекты с номерами 1, 2, 3, 4. Для того чтобы отнести эти объекты к одному из двух множеств, рассчитаем для них значения дискриминантных функций (по трем переменным):

$$\begin{aligned} f_1 &= -185,03 \times 1,07 + 1,84 \times 93,5 + 4,92 \times 5,30 = 0,1339, \\ f_2 &= -185,03 \times 0,99 + 1,84 \times 84,0 + 4,92 \times 4,85 = -4,7577, \\ f_3 &= -185,03 \times 0,70 + 1,84 \times 76,8 + 4,92 \times 3,50 = 29,0110, \\ f_4 &= -185,03 \times 1,24 + 1,84 \times 88,0 + 4,92 \times 4,95 = -43,1632. \end{aligned}$$

Таким образом, объекты 1, 2 и 4 относятся ко второму классу, а объект 3 относится к первому классу, так как

$$f_1 < C, f_2 < C, f_3 > C, f_4 < C.$$

Задание 2.

Условие. Рассмотрим случай, когда существует три класса (множества) объектов.

Для этого к двум классам из предыдущего примера добавим еще один.

В этом случае будем иметь уже три матрицы исходных данных:

$$X_1 = \begin{pmatrix} 0,5 & 94,0 & 8,50 & 6707 \\ 0,67 & 75,4 & 8,79 & 5037 \\ 0,68 & 85,2 & 9,10 & 3695 \\ 0,55 & 98,8 & 8,47 & 6815 \end{pmatrix},$$

$$X_2 = \begin{pmatrix} 1,52 & 81,5 & 4,95 & 3211 \\ 1,20 & 93,8 & 6,95 & 2890 \\ 1,46 & 86,5 & 4,70 & 2935 \end{pmatrix},$$

$$X_3 = \begin{pmatrix} 1,70 & 80,0 & 4,5 & 3510 \\ 1,65 & 85,0 & 4,8 & 2900 \\ 1,49 & 78,5 & 4,1 & 2850 \end{pmatrix}.$$

Решение.

Если в процессе дискриминации используются все четыре переменные (X_1, X_2, X_3, X_4), то для каждого класса дискриминантные функции имеют вид:

$$f_1 = 613,6X_1 + 5,482X_2 + 37,53X_3 - 8,286X_4 + 22460,$$

$$f_2 = 657,2X_1 + 6,11X_2 + 33,82X_3 - 8,377X_4 + 11800,$$

$$f_3 = 625,0X_1 + 5,778X_2 + 31,53X_3 - 7,692X_4 + 11060.$$

Определим теперь, к какому классу можно отнести каждое из четырех наблюдений, приведенных в табл. 4.2

Таблица 4.2

Значения переменных				
Номер наблюдения	X_1	X_2	X_3	X_4
1	1,07	93,5	5,30	5385
2	0,99	84,0	4,85	5225
3	0,70	76,8	3,50	5190
4	1,24	88,0	4,95	6280

Вычислим разности:

$$f_1 - f_2 = -20792,082 + 31856,41 = 11064,328 > 0,$$

$$f_1 - f_2 = -20792,082 + 40016,428 = 19224,346 > 0.$$

Следовательно, наблюдение 1 в табл. 4.2 относится к первому классу. Аналогичные расчеты показывают, что и остальные три наблюдения следует отнести тоже к первому классу.

Чтобы показать влияние числа дискриминантных переменных на результаты классификации, изменим условие последнего примера. Будем использовать для расчета дискриминантных функций только три переменные: X_1 , X_2 и X_3 . В этом случае выражения для дискриминантных функций будут иметь вид:

$$f_1 = 36,93X_1 + 1,288X_2 + 8,644X_3 - 105,6,$$

$$f_2 = 29,29X_1 + 2,043X_2 + 5,617X_3 - 125,1,$$

$$f_3 = 25,65X_1 + 2,211X_2 + 4,579X_3 - 120,6.$$

Подставив в эти выражения значения исходных переменных для классифицируемых объектов, нетрудно убедиться, что все они попадают в третий класс, так как

$$f_1 - f_2 = -26,87 < 0,$$

$$f_1 - f_3 = -37,68 < 0,$$

$$f_2 - f_3 = -10,809 < 0.$$

Таким образом, мы видим, что изменение числа переменных сильно влияет на результат дискриминантного анализа. Чтобы судить о целесообразности включения (удаления) дискриминантной переменной, обычно используют специальные статистические критерии, позволяющие оценить значимость ухудшения или улучшения разбиения после включения (удаления) каждой из отобранных переменных.

Библиографический список

Основная литература:

1. Качала В.В. Основы теории систем и системного анализа: учеб. пособие. 2-е изд. М.: Гор. линия - Телеком, 2012. 210 с.

Режим доступа: <http://znanium.com/bookread2.php?book=351396>

2. Моделирование систем и процессов: учебник для академического бакалавриата / В. Н. Волкова [и др.]; под ред. В. Н. Волковой, В. Н. Козлова. - М.: Издательство Юрайт, 2017. - 450 с.

Режим доступа: <https://www.biblio-online.ru/viewer/E7D370B9-3C64-4A0F-AF1B-F6BD0EEEEBCD0#page/1>

3. Моделирование систем [Электронный ресурс] : учеб. пособие — Электрон. дан. — Москва : Горная книга, 2006. — 295 с.

Режим доступа: <https://e.lanbook.com/reader/book/3511/#1>

4. Рубан А.И. Адаптивные системы управления с идентификацией. Красноярск: СФУ, 2015. 140 с. [Электронный ресурс] – <http://znanium.com/bookread2.php?book=550540#>

5. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов и др. - М.: НИЦ ИНФРА-М, 2015. - 890 с.

Режим доступа: <http://znanium.com/bookread2.php?book=515227>

6. Теория систем и системный анализ : учебник и практикум для академического бакалавриата / М. Б. Алексеева, П. П. Ветренко. — М. : Издательство Юрайт, 2017. — 304 с.

Режим доступа: <https://www.biblio-online.ru/viewer/B791EB3D-7CD9-48A7-B7DD-BEB4670DB29E#page/1>

7. Экономико-математические методы и прикладные модели : учебник для бакалавриата и магистратуры / А. Н. Гармаш, И. В. Орлова, В. В. Федосеев. — 4-е изд., перераб. и доп. — М. : Издательство Юрайт, 2017. — 328 с.

Режим доступа: <https://www.biblio-online.ru/viewer/62CA472C-1C3E-48F7-B963-6762D5A89A50#/>

Дополнительная литература

1. *Ивченко Б.П., Мартыщенко Л.А., Табухов М.Е.* Управление в экономических и социальных системах. Системный

анализ. Принятие решений в условиях неопределенности. – СПб.: «Нордмед-Издат», 2001. – 248 с.

2. *Ивченко, Б.П.* Информационная микроэкономика. Ч. 2 /Б.П. Ивченко, Л.А. Мартыщенко, Г.С. Губин. – СПб., 1998.

3. *Ивченко, Б.П.* Информационная микроэкономика. Ч. 1/Б.П. Ивченко, Л.А. Мартыщенко, И.Б. Иванцов. – СПб., 1997.

4. *Клавдиев А.А., Пасевич В.* Адаптивные технологии информационно-вероятностного анализа транспортных систем.- СПб.: Издат-во СЗГУ, 2009.-305с.

5. *Мартыщенко Л.А.* и др. Теоретические основы информационно-статистического анализа сложных систем.-Спб.: Лань, 1997.-320с.

6. *Покровский А.К.* Исследование систем управления. Транспортная отрасль. Учебное пособие. Издат-во «КноРус», 2010.- 368 с.

7. *Мищенко А.В.* Методы и модели управления ограниченными ресурсами в логистических системах: учеб. пособие. 2-е изд. М.: ИНФРА-М, 2011. 185 с. [Электронный ресурс] – <http://znanium.com/bookread2.php?book=911255>

СОДЕРЖАНИЕ

Введение	3
1. Общая характеристика.....	5
2. Контрольные точки и виды отчетности по ним	6
3. Методические указания для проведения практических занятий по учебной дисциплине «Математические методы исследования экономических и социальных систем»	7
Практическая работа 1_Применение методов многомерного регрессионного анализа при исследовании технических, экономических и социальных систем	7
Расчёт зависимость количества срывов сроков выполнения работ от поломки оборудования	9
Практическая работа 2_Применение факторного анализа для исследования технических, экономических и социальных систем	14
Практическая работа 3_Применение кластерного анализа для исследования технических, экономических и социальных систем	19
Практическая работа 4._Применение дискриминантного анализа для исследования технических, экономических и социальных систем	22
Библиографический список	28