Министерство науки и высшего образования Российской Федерации

**Федеральное государственное бюджетное образовательное учреждение высшего образования
Санкт-Петербургский горный университет**

Кафедра системного анализа и управления

ВЕРОЯТНОСТНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ СЛОЖНЫХ СИСТЕМ

Методические указания к практическим занятиям для студентов бакалавриата направления 27.03.03

**САНКТ-ПЕТЕРБУРГ
2021**

УДК 519.237(073)

ВЕРОЯТНОСТНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ СЛОЖНЫХ СИСТЕМ: Методические указания к практическим занятиям / Санкт-Петербургский горный университет. Сост. *О.В. Афанасьева*. СПб, 2021. 27 с.

Содержатся краткие теоретические сведения и задания для проведения практических занятий по учебной дисциплине «Вероятностные методы прогнозирования сложных систем».

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

Научный редактор проф. *Д.А. Первухин*

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

ВЕРОЯТНОСТНЫЕ МЕТОДЫ ПРОГНОЗИРОВАНИЯ СЛОЖНЫХ СИСТЕМ

Методические указания к практическим занятиям для студентов бакалавриата направления 27.03.03

Сост. *О.В. Афанасьева*

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. 1,6. Усл.кр.-отт. 1,6. Уч.-изд.л. 1,3. Тираж 75 экз. Заказ 560.

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

Введение

Цель изучения дисциплины «Вероятностные методы прогнозирования сложных систем» изучение вероятностных методов прогнозирования и их применение при решении задач управления; знакомство с применением методов прогнозирования при решении задач управления сложными системами; овладение в комплексе научно-методическим аппаратом прогнозирования, основываясь на статистических данных при решении задач управления.

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 «Вероятностные методы прогнозирования сложных систем»

Таблица 1.

Перечень практических работ по дисциплине «Вероятностные методы прогнозирования сложных систем»

№ п/п	Наименование работ	Трудоемкость (час.)
1.	Математическое описание доставки сырья автомобильным транспортом на предприятие	4
2.	Математическое описание разгрузки автомобильного транспорта на промышленном предприятии	4
3.	Имитация работы грузового терминала	6
4.	Имитационное моделирование времени разгрузки автомобильного транспорта	6
5.	Оценка технического уровня сложной системы	8
6.	Имитация работы специализированного участка разгрузки автомобилей на предприятии	6
	Итого:	34

2. Контрольные точки и виды отчетности по ним

№ п/п	Вид деятельности студентов	Сроки выполнения
1.	Сдача отчета по практической работе 1 «Математическое описание доставки сырья автомобильным транспортом на предприятие»	2-ая неделя
2.	Сдача отчета по практической работе 2 «Математическое описание разгрузки автомобильного транспорта на промышленном предприятии»	4-ая неделя
3.	Сдача отчета по практической работе 3 «Имитация работы грузового терминала»	6-ая неделя
4.	Сдача отчета по практической работе 4 «Имитационное моделирование времени разгрузки автомобильного транспорта»	8-ая неделя
5.	Сдача отчета по практической работе 5 «Оценка технического уровня сложной системы»	10-ая неделя
6.	Сдача отчета по практической работе 6 «Имитация работы специализированного участка разгрузки автомобилей на предприятии»	12-ая неделя

3. Методические указания для проведения практических занятий по учебной дисциплине

«Вероятностные методы прогнозирования сложных систем»

Практическая работа 1. МАТЕМАТИЧЕСКОЕ ОПИСАНИЕ ДОСТАВКИ СЫРЬЯ АВТОМОБИЛЬНЫМ ТРАНСПОРТОМ НА ПРЕДПРИЯТИЕ

Условие задачи

Имеются сведения о ежесуточном прибытии автомобилей с сырьем на промышленное предприятие в течение шести месяцев (рис. 1). Методами математической статистики необходимо установить аналитические выражения, описывающие этот случайный процесс.

Последовательность выполнения

1) Определить среднее значение и дисперсию количества прибывающих на предприятие автомобилей за сутки.

2) С помощью принятого критерия согласия проверить возможность замены эмпирического распределения числа пребывающих на предприятие автомобилей теоретическим законом распределения Пуассона.

3) Построить полигон эмпирического и теоретического распределений суточного поступления автомобилей с сырьем на предприятие.

4) Определить вероятность, что

- на предприятие в течение суток не прибудет ни одного автомобиля;

- на предприятие в течение суток прибудет хотя бы один автомобиль;

- на предприятие в течение суток поступит автомобилей больше, чем имеется разгрузочных терминалов.

Количество разгрузочных терминалов принимается равным ближайшему большему целому относительно среднесуточного прибытия автомобилей.

Май	Июнь	Июль	Август	Сентябрь	Октябрь
	1	1	2	2	1
	2	2	4	2	3
	0	1	1	1	2
	1	1	3	3	2
1	2	3	5	5	1
0	0	0	4	3	1
0	2	1	0	2	0
0	1	2	2	1	0
0	2	1	3	2	3
0	0	1	5	0	1
1	2	1	3	1	1
0	3	1	2	1	2
1	1	2	4	2	1
1	2	0	1	3	3
2	1	1	5	1	0
0	0	3	2	1	2
4	3	1	2	2	1
0	1	0	3	2	2
4	1	2	4	1	0
0	2	0	1	3	3
2	2	1	5	5	4
0	2	1	2	4	5
0	1	1	3	3	1
0	2	1	2	1	2
0	2	0	2	2	3
0	0	0	3	2	1
0	2	1	2	4	0
0	1	1	2	3	1
1	0	2	4	2	2
2	2	1	4	2	3
2		1	0		3

Рис.1. Исходные данные

Для каждого месяца составим таблицу и построим графики. Для выполнения работы необходимо воспользоваться следующими формулами:

1) $P_i = \dfrac{m_i}{\sum m_i} = \dfrac{m_i}{n}$, где m_i - частота i-го признака;

P_i - эмпирическая вероятность появления i-го признака;

2) $P_i' = e^{-\lambda} \dfrac{\lambda^{x_i}}{x_i!}$, где P_i' - теоретическое значение вероятности появления i-го признака для закона Пуассона в зависимости от значений x_i и λ-параметра закона Пуассона, равного $M(x)$;

3) $\chi^2 = \sum \dfrac{(m_i - m_i')^2}{m_i'}$, где

m' - теоретическое значение частоты по закону Пуассона;

χ^2 - значение критерия Пирсона

4) $M(x) = \sum P_i x_i = \lambda$, где x_i - i-е значение количества поступивших на предприятие автомобилей;

$M(x)$ - математическое ожидание;

5) $D(x) = M(x^2) - [M(x)]^2 = \sum P_i x_i^2 - (\sum P_i x_i)^2$, где $D(x)$ - дисперсия.

По рассчитанному значению критерия Пирсона χ^2 делается вывод о возможности замены эмпирического распределения числа поступающих на предприятие автомобилей теоретическим законом Пуассона.

6) $k = l - 1 - r$, где k - число степеней свободы; l - число различающихся значений x_i, $l = N + 1$; r - количество параметров теоретического закона. Для закона Пуассона используется один параметр λ, то есть в этом случае r=1.

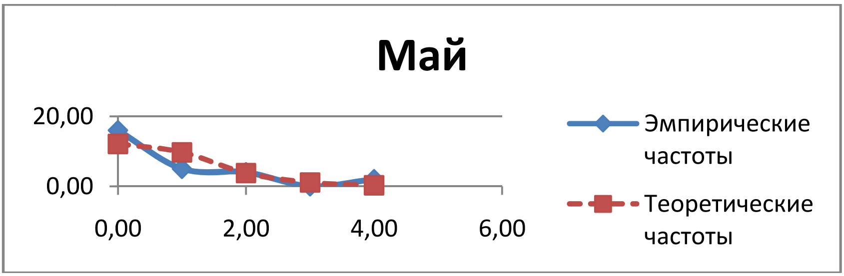
По значению критерия χ^2 и параметру k (на их пересечении) считывается значение функции распределения Пирсона, которое представляет собой «вероятностную степень недоверия» к гипотезе о возможности замены опытного распределения соответствующим теоретическим. Иначе говоря, чем ближе это значение к нулю, тем лучше согласуется опытное распределение и теоретическое, тем больше подтверждается гипотеза.

Решение задачи

1) Май: $M(x)$ =0,8; $D(x)$ =1,36;

Столбец x_i - это число автомобилей, прибывших на предприятие в мае. Его значение от 1 до 4. Столбец m_i показывает, как часто в мае месяце приходило x_i количество автомобилей.

Май									
xi	mi	Pi	Pixi	Pi(xi)^2	P'i	m'i	(mi-m'i)^2	((mi-m'i)^2)/m'i	
0,00	16,00	0,59	0,00	0,00	0,45	12,15	14,82	1,22	0,09
1,00	5,00	0,19	0,19	0,19	0,36	9,72	22,28	2,29	0,26
2,00	4,00	0,15	0,30	0,59	0,14	3,78	0,05	0,01	0,00
3,00	0,00	0,00	0,00	0,00	0,04	1,08	1,17	1,08	0,09
4,00	2,00	0,07	0,30	1,19	0,01	0,27	2,99	11,08	0,97



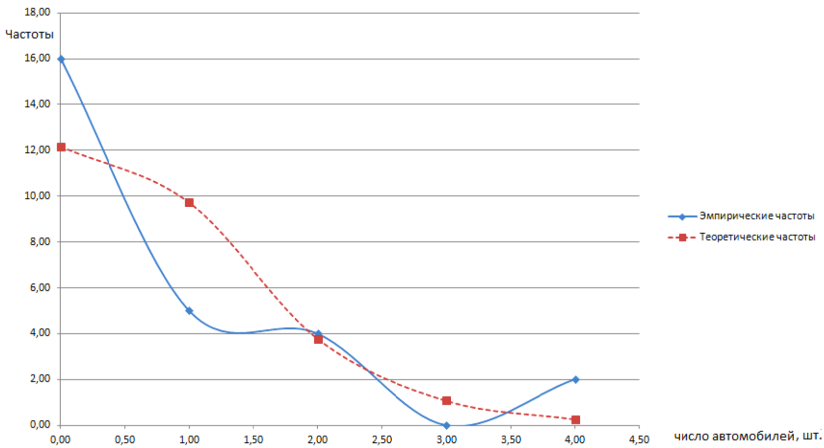


Рисунок 2. Результаты расчёта для месяца: май

Число разгрузочных терминалов – 1

Вероятность, что

- на предприятие в течение суток не прибудет ни одного автомобиля: 0,45;
- на предприятие в течение суток прибудет хотя бы один автомобиль: 0,55;
- на предприятие в течение суток поступит автомобилей больше, чем имеется разгрузочных терминалов: 0,19.

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

$$F(X, Y) = const,$$

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 па область 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.

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

Первый шаг.

$R = A$ и $A = A_1$, 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:

$$A_2 = AB_1 = \begin{pmatrix} 1 & 0{,}581 & 0{,}154 \\ 0{,}581 & 1 & 0{,}439 \\ 0{,}154 & 0{,}439 & 1 \end{pmatrix} \cdot \begin{pmatrix} -2 & 0{,}581 & 0{,}154 \\ 0{,}581 & -2 & 0{,}439 \\ 0{,}154 & 0{,}439 & -2 \end{pmatrix}$$

$$= \begin{pmatrix} -1{,}638 & -0{,}513 & 0{,}101 \\ -0{,}513 & -1{,}469 & -0{,}35 \\ 0{,}101 & -0{,}35 & -1{,}783 \end{pmatrix};$$

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

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

Третий шаг:

$$A_3 = AB_2 = \begin{pmatrix} 1 & 0{,}581 & 0{,}154 \\ 0{,}581 & 1 & 0{,}439 \\ 0{,}154 & 0{,}439 & 1 \end{pmatrix} \cdot \begin{pmatrix} 0{,}807 & 0{,}581 & 0{,}154 \\ -0{,}513 & 0{,}976 & -0{,}350 \\ 0{,}101 & -0{,}350 & 0{,}662 \end{pmatrix}$$

$$= \begin{pmatrix} 0{,}524 & 0 & 0 \\ 0 & 0{,}524 & 0 \\ 0 & 0 & 0{,}524 \end{pmatrix};$$

$$P_3 = \left(\frac{1}{3}\right)(3 \cdot 0{,}524) = 0{,}524, \qquad B_3 = A_3 - P_3 E = 0.$$

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

2. Построим характеристическое уравнение:

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

Откуда найдём:

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

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 примерно60% 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

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

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

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

$$(1 - 1{,}798) \cdot u_{11} + 0{,}581 \cdot u_{21} + 0{,}154 \cdot u_{31} = 0, \qquad u_{11} = 1{,}262$$
$$0{,}581 \cdot u_{11} + (1 - 1{,}798) \cdot u_{21} + 0{,}439u \cdot_{31} = 0, \qquad u_{11} = 1{,}469$$
$$0{,}154 \cdot u_{11} + 0{,}581 \cdot u_{21} + (1 - 1{,}798) \cdot u_{31} = 0, \qquad u_{11} = 1{,}000$$

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

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

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

$$(1 - 0{,}327) \cdot u_{13} + 0{,}581 \cdot u_{23} + 0{,}154 \cdot u_{33} = 0, \qquad u_{12} = 1{,}307$$

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

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: $V'V = E$.

Матрицу факторного отображения (А) 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 А 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 А:

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

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

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 матрицуF: $F = A^{-1}Z'$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.

Fr - главные компоненты

$$F' = \begin{array}{c|ccc} & F_1 & F_2 & F_3 \\ n_1 & -0,233 & 0,823 & -0,218 \\ n_2 & -1,533 & -0,372 & 0,892 \\ n_3 & 1,143 & -1,687 & 1,205 \\ n_4 & 0,623 & 1,236 & 0,121 \\ \sum f_{ri} & 0 & 0 & 0 \end{array}$$

n_i - объекты

Центр распределения значений главных компонент F_2находится 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_jи главных компонент F_rи 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. $z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j}$, 2. $z_{ij} = \frac{x_{ij}}{x_{maxj}}$, 3. $z_{ij} = \frac{x_{ij}}{\overline{x_j}}$, 4. $z_{ij} = \frac{x_{ij}}{x_{minj}}$.

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

$$z_{ij} = \frac{x_{ij} - \overline{x_j}}{\sigma_j},$$

и построим матрицу стандартизированных 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 - прибыть 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:

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

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: $x_1 = 1{,}07,\ x_1 = 93{,}5,\ x_1 = 5{,}30,$
2-е предприятие: $x_1 = 0{,}99,\ x_1 = 84{,}0,\ x_1 = 4{,}85,$
3-е предприятие: $x_1 = 0{,}70,\ x_1 = 76{,}8,\ x_1 = 3{,}50,$
4-е предприятие: $x_1 = 1{,}24,\ x_1 = 88{,}0,\ x_1 = 4{,}95.$

Решение.

Для удобства запишем значения исходных переменных для каждой группы предприятий в виде матриц X_1и 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: $\bar{\boldsymbol{x}}_{\mathbf{11}}$=0,60,
$\bar{\boldsymbol{x}}_{\mathbf{21}}$=88,4,
$\bar{\boldsymbol{x}}_{\mathbf{31}}$ =8,72.

II группа: $\bar{\boldsymbol{x}}_{\mathbf{11}}$=1,39,
$\bar{\boldsymbol{x}}_{\mathbf{22}}$ =87,3,
$\bar{\boldsymbol{x}}_{\mathbf{32}}$= 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_3вычисляются по формуле:

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

где $\overline{X}_1, \overline{X}_2$ - 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_1и 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:

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

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

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

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

Отсюда находим вектор коэффициентов дискриминантной функции по формуле:

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

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

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

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

$$\overline{f}_1 = 94{,}4238$$

Для 2 -го множества

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

$$\overline{f}_2 = -70{,}0138$$

Тогда константа дискриминации *С* будет равна:

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

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):

$$f_1 = -185{,}03 \times 1{,}07 + 1{,}84 \times 93{,}5 + 4{,}92 \times 5{,}30 = 0{,}1339,$$

$$f_2 = -185{,}03 \times 0{,}99 + 1{,}84 \times 84{,}0 + 4{,}92 \times 4{,}85 = -4{,}7577,$$

$$f_3 = -185{,}03 \times 0{,}70 + 1{,}84 \times 76{,}8 + 4{,}92 \times 3{,}50 = 29{,}0110,$$

$$f_4 = -185{,}03 \times 1{,}24 + 1{,}84 \times 88{,}0 + 4{,}92 \times 4{,}95 = -43{,}1632.$$

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

$$f_1 < \text{C},\ f_2 < \text{C},\ f_3 > \text{C},\ f_4 < \text{C}.$$

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-F6BD0EEEBCD0#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 СЗТУ, 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-еизд. М.: ИНФРА-М, 2011. 185 с. [Электронный ресурс] – http://znanium.com/bookread2.php?book=911255

СОДЕРЖАНИЕ