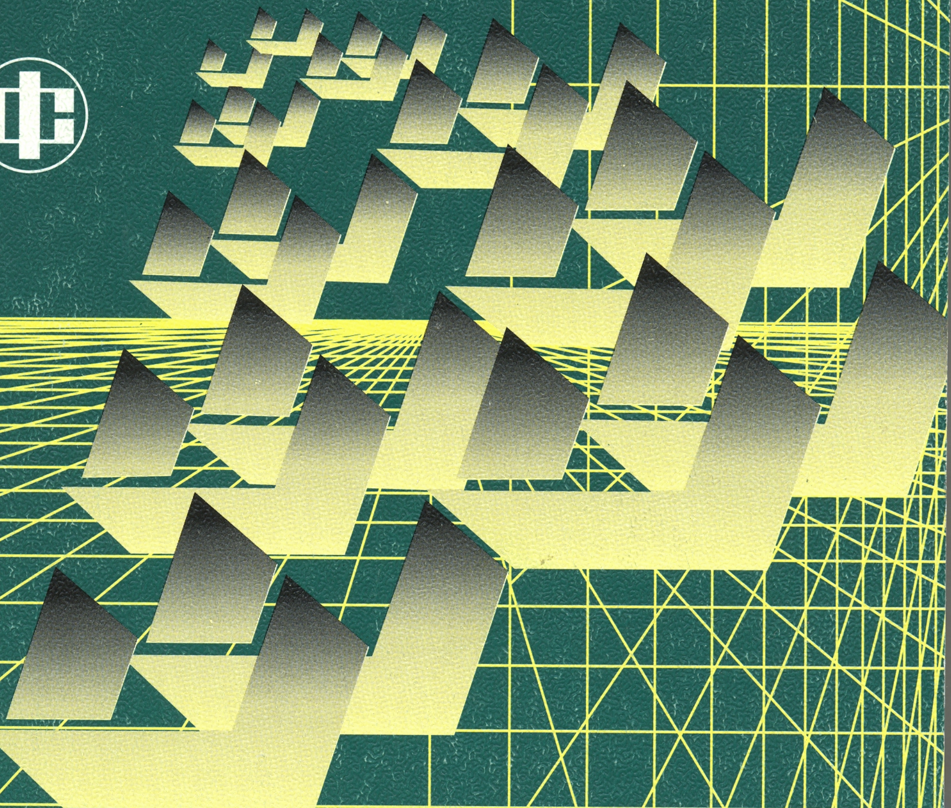
А.М. ДУБРОВ

КОМПОНЕНТНЫЙ АНАЛИЗ И ЭФФЕКТИВНОСТЬ В ЭКОНОМИКЕ



А.М. ДУБРОВ

КОМПОНЕНТНЫЙ АНАЛИЗ И ЭФФЕКТИВНОСТЬ В ЭКОНОМИКЕ

Рекомендовано
Учебно-методическим объединением
по образованию в области экономики,
статистики, информационных систем
и математических методов в экономике
в качестве учебного пособия
для студентов высших учебных заведений,
обучающихся по специальности 061700 "Статистика"

Москва
"Финансы и статистика"
2002

УДК 519.237(075.8)+330.131.52(075.8)
ББК 65в6я73
Д79

РЕЦЕНЗЕНТЫ:

кафедра экономической кибернетики
Московской сельскохозяйственной академии
имени К. А. Тимирязева;
В. А. Колемаев,
доктор экономических наук, профессор

Дубров А. М.

Д79 Компонентный анализ и эффективность в экономике: Учеб. пособие. — М.: Финансы и статистика, 2002. — 352 с.: ил.

ISBN 5-279-02430-9

Излагаются метод многомерного статистического анализа при помощи главных компонент, а также методы оценки эффективности экономических организаций, экономических систем и систем управления. Рассматриваются задачи обработки многомерных наблюдений в экономике и проблемы совершенствования метода главных компонент и расширения области его приложения. Изучаются основные принципы исследования операций, используемые в теории эффективности; дается оценка эффективности на основе критериев — игровых, информационных, теории массового обслуживания.

Для студентов экономических специальностей вузов, обучающихся по специальности 061700 «Статистика», а также для аспирантов, преподавателей вузов, экономистов, менеджеров.

Д $\frac{1602090000-045}{010(01)-2002}$ 217–2002

УДК 519.237(075.8)+330.131.52(075.8)
ББК 65в6я73

ISBN 5-279-02430-9

ПРЕДИСЛОВИЕ

В рыночных условиях хозяйствования большое внимание уделяется развитию исследований с применением методов прикладной математики и ЭВМ в управлении народным хозяйством, повышению качества и научного уровня работ по изучению развития народного хозяйства и внедрению наиболее совершенных математических моделей управления экономикой.

При административно-командном управлении экономикой анализ деятельности предприятий в основном ограничивался контролем выполнения планов, который осуществлялся традиционными методами. Традиционные методы исследования не позволяли глубоко проанализировать процессы производственно-хозяйственной деятельности.

В рыночных условиях хозяйствования управление экономикой проводится на основе системы управления (СУ), которая не ограничивается традиционными методами управления. Современный анализ производственно-хозяйственной деятельности в условиях СУ превращается в подсистему управления, позволяющую проводить исследования, охватывающие все области деятельности предприятий.

В настоящее время в экономике России проводится большое число исследований с применением теории вероятностей и математической статистики. Особое внимание уделяется корреляционному и регрессионному анализу, позволяющим прогнозировать функционирование и развитие отраслей народного хозяйства, предприятий, фирм. Это вполне естественно, так как в современных условиях многочисленные технико-экономические связи между отраслями экономики, объединениями, предприятиями, фирмами многообразны и весьма сложны.

Исследование и классификация экономических показателей и связей между фирмами, объединениями, торговыми организациями требуют от управления планированием всестороннего анализа условий производства и конечных результатов деятельности предприятий.

Расширяются исследования с применением более сложного математического аппарата — многомерного статистического анализа. Методы многомерного статистического анализа позволяют определять неявные закономерности, объективно существующие в изучаемых экономических явлениях. Наиболее перспективными в этом плане являются методы распознавания образов, факторный, кластерный и компонентный анализ. Данные компонентного анализа позволяют выделить и обосновать систему признаков, наиболее существенно влияющих на исследуемый процесс, а также отделить группы взаимозависимых признаков от независимых, существенных признаков от несущественных.

В книге делается попытка ввести читателя в круг основ этих математических методов, так как ранее опубликованная литература была рассчитана на готовых специалистов в области многомерного статистического анализа.

Учебное пособие написано на основе курсов, прочитанных автором в техническом и экономическом вузах, и более чем тридцатилетнего опыта работы по применению методов многомерного статистического анализа.

Математический аппарат, представленный в книге для облегчения овладения методами, снабжен характерными примерами. В приложение вынесены элементы линейной алгебры, связанные с методом главных компонент. Это позволило рассмотреть упрощенный вариант алгоритма решения, его математическое обоснование и привести новые, недостаточно освещенные ранее в литературе разделы по уточнению оценок, построению динамических моделей и др.

В настоящее время следует не только проводить анализ функционирования системы в целом, необходимо оценивать эффективность отдельных звеньев системы. Данная задача часто выполняется на основе показателей с использованием различных математических и статистических методов.

В связи с вышеизложенным учебное пособие разделено на две части.

Первая часть — **«Компонентный анализ»** — посвящена многомерному статистическому анализу. В ней изучается многомерное нормальное распределение. Приводятся понятие

системы случайных величин, характеристики многомерного нормального распределения, двумерное нормальное распределение, критерии значимости коэффициента корреляции.

Показана значимость метода главных компонент по сравнению с другими методами многомерного статистического анализа, обратные и ортогональные преобразования в методе главных компонент, собственные векторы и собственные значения матрицы, преобразование корреляционной матрицы, квадратичные формы, главные компоненты в трехмерном и конечномерном пространстве, ортогональная регрессия, математическая модель, блок-схема алгоритма.

Приведены практические задачи, которые исследуются методом главных компонент. Большое внимание уделяется дальнейшему совершенствованию метода главных компонент и расширению области его приложения. Это уточнение числовых характеристик закона распределения исходных параметров модели, динамическая модель, метод главных факторов.

Во второй части — **«Оценка эффективности систем управления»** — рассматриваются основы теории эффективности сложных систем, экономическая эффективность СУ, принципы исследования операций, используемые в теории эффективности, оценка эффективности на основе информационных критериев, критериев теории массового обслуживания, а также игровых критериев.

Основная теорема минимакса, ставшая краеугольным камнем теории игр, была доказана Джоном фон Нейманом. В 1947 г. в США вышла книга Дж. Неймана и О. Моргенштерна «Теория игр и экономическое поведение»[1], которая была переведена в России в 1970 г. Однако при плановой экономике она не вызвала такого интереса, как при рыночной экономике.

Оценки эффективности делятся на два класса:

- оценки эффективности самих СУ;
- оценки эффективности функционирования управляемых объектов, получаемые с помощью СУ.

Системы управления являются сложными системами, поэтому в книге представлены подходы А. Н. Колмогорова, Ю. Б. Гермейера, Н. П. Бусленко. Общий подход к оценке

[1] *Neumann G., Morgenstern O.* Teory of Games and Economic Behavior. — New York: Viley, 1947.

эффективности должен учитывать, что современные СУ — это сложные человеко-машинные системы.

В учебном пособии приведены практические примеры, показаны роль коэффициента готовности и возможности графических методов исследования. Рассматриваются как простые, так и более сложные примеры, решаемые на основе игровых критериев при помощи рисковых ситуаций.

Приводится оценка эффективности организации работы вычислительного центра на основе критериев теории массового обслуживания. Поднимается вопрос об оценке эффективности функционирования технических средств СУ в зависимости от уровня обученности обслуживающего персонала.

Таким образом, практические проблемы управления могут быть решены на основе анализа его фактических результатов различными математическими критериями, а при наличии вероятности событий или процессов — методами математической статистики при помощи современных пакетов прикладных программ и вычислительной техники.

Автор книги выражает благодарность рецензентам — *кафедре экономической кибернетики Московской сельскохозяйственной академии имени К. А. Тимирязева и ее заведующему члену-корреспонденту РАСХН, доктору экономических наук, профессору А. М. Гатаулину, доктору экономических наук В. А. Колемаеву* за кропотливую работу по рецензированию рукописи и ценные замечания, позволившие улучшить качество учебного пособия.

ВВЕДЕНИЕ

Объект изучения в экономике, например фирма, банк, предприятие, может быть всесторонне охарактеризован только при помощи целого набора признаков (параметров, показателей). При характеристике объекта исследования многомерными случайными признаками строится корреляционная матрица, элементы которой учитывают тесноту линейной стохастической связи. Однако при большом числе признаков охарактеризовать выявленные связи довольно сложно. Возникает потребность в сжатии информации, т. е. описании объектов меньшим числом обобщенных показателей, например факторами или главными компонентами. Главные компоненты являются более удобными укрупненными показателями. Они отражают внутренние, объективно существующие закономерности, которые не поддаются непосредственному наблюдению.

При корреляционном или регрессионном анализе на основе полученной корреляционной матрицы строятся, например, уравнения регрессии, связывающие факторные признаки с результативным признаком. Сами уравнения регрессии являются конечной целью исследования. По ним проводится содержательная экономическая интерпретация полученных результатов и управлением предприятия вырабатываются соответствующие решения.

При использовании метода главных компонент корреляционная матрица является исходной ступенью для дальнейшего анализа наблюдавшихся ранее значений признаков. Появляется возможность извлечения дополнительной информации об изучаемом объекте или процессе. При этом весьма ценную новую информацию можно получить на основе статистических данных, ранее собранных для проведения классического регрессионного анализа по главным компонентам.

Какие же задачи можно решить методом главных компонент?

По мнению автора, можно выделить четыре основных типа решаемых задач.

Первая задача — отыскание скрытых, но объективно существующих закономерностей, определяемых воздействием внутренних и внешних причин.

Вторая задача — описание изучаемого процесса числом главных компонент m, значительно меньшим, чем число первоначально взятых признаков n. Эта задача обусловлена наличием большого числа признаков и связей между ними. Главные компоненты адекватно, в более компактной форме отражают исходную информацию. Выделенные главные компоненты содержат в среднем больше информации, чем непосредственно замеряемые отдельные признаки.

Третья задача — выявление и изучение стохастической связи признаков с главными компонентами. Выявление признаков, наиболее тесно связанных с данной главной компонентой, позволяет руководителю разработать и принять научно обоснованное управляющее решение, способствующее повышению эффективности функционирования изучаемого процесса.

Четвертая задача предусматривает возможность использования полученных результатов для прогнозирования хода развития процесса на основе уравнения регрессии, построенного по полученным главным компонентам. Такой метод прогнозирования обладает определенными преимуществами перед классическим регрессионным анализом на основе корреляционной матрицы.

Так, при классическом регрессионном анализе стремление исследователя наиболее полно отразить влияние факторных признаков на результирующий заставляет его включать в математическую модель как можно большее число исходных показателей. Исходные показатели в экономических задачах часто обладают мультиколлинеарностью, т. е. существенной коррелированностью входных параметров. Это затрудняет проведение анализа, интерпретацию и определение коэффициентов регрессионной модели. Если же по полученным уравнениям регрессии строить прогноз, то он не будет удовлетворять существующим требованиям по точности. Появляется задача замены исходных взаимосвязанных признаков на некоторую совокупность некоррелированных параметров,

которая всю информацию об изменчивости изучаемого процесса должна сохранить без искажения. Как известно, таким математическим аппаратом является метод главных компонент. Действительно, главные компоненты — это характеристики, построенные на основе наблюденных и измеренных признаков.

Регрессионная модель, построенная по главным компонентам, позволяет устранить отмеченные выше недостатки. Практика показала, что классификация объектов исследования по полученным главным компонентам оказывается более объективной, чем их разделение при помощи отдельных исходных признаков.

Это один из результатов, полученных на практике.

Практические возможности решения ранее упомянутых четырех задач реализованы в следующих семи направлениях [66]:

- причинный анализ взаимосвязей показателей и определение их стохастической связи с главными компонентами;
- построение обобщенных технико-экономических показателей;
- ранжирование объектов или наблюдений по главным компонентам;
- классификация объектов наблюдений;
- ортогонализация исходных показателей;
- сжатие исходной информации;
- построение уравнений регрессии по обобщенным технико-экономическим показателям.

Негативной стороной метода является сложность математического аппарата, требующего знания как теории вероятностей и математической статистики, так и линейной алгебры и математического обеспечения ЭВМ.

Однако развитие вычислительной техники, внедрение в экономическую практику современных пакетов прикладных программ по статистической обработке данных снимают указанные трудности, но создают иллюзию «легкости» проведения исследований без глубокой математической проработки в области метода главных компонент.

Практика подтвердила, что применение готовых программ без глубокого понимания математической сущности используемых процедур при изменении условий эксперимента или явления может привести к необоснованным выводам.

Следует отметить, что при решении практических задач нельзя формально пользоваться методами многомерного статистического анализа без профессионального знания моделируемого процесса. Успеха можно добиться только на пути профессионального анализа, подкрепленного хорошим знанием математических методов. Рассматриваемый метод главных компонент является мощным математическим аппаратом, описывающим самые разнообразные реальные процессы в различных областях исследования.

Однако в условиях рыночной экономики на этом останавливаться нельзя. Обычно представляет интерес прибыль, которую желательно не только заранее оценить, но и определить вероятность ее получения.

Технической базой систем управления служит современная электронная вычислительная техника, обеспеченная соответствующими пакетами с математической основой и сетевыми пакетами, которые позволяют оценить эффективность.

Автор не ограничивается рассмотрением чисто экономической эффективности, являющейся важнейшей составляющей общей теории эффективности. Проблема рассмотрена шире, когда исследователь не может определить общий эффект, но может найти выигрыш во времени, оценить выигрыш в объеме обрабатываемой информации.

ЧАСТЬ I

КОМПОНЕНТНЫЙ АНАЛИЗ

Глава 1
МНОГОМЕРНОЕ НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

1.1. Понятие системы случайных величин

В экономических исследованиях часто встречаются задачи, в которых конечный результат зависит от многих случайных величин. Так, выполнение финансового плана предприятиями зависит от производительности труда, регулярности поступления материалов, уровня квалификации работников, руководства на предприятиях, состояния оборудования, рационализаторской и изобретательской работы и других признаков. Следовательно, интересуясь конечным результатом выполнения финансового плана или степенью взаимосвязи признаков, мы сталкиваемся с системой признаков. Если учесть, что на проявление каждого признака влияет большое число случайных факторов, то можно считать, что признаки носят случайный характер. Если это подтверждается практикой, то можно заключить, что необходимо исследовать систему случайных признаков (величин).

Свойства системы нескольких случайных величин не представляют собой сумму свойств отдельных величин. Они в значительной степени зависят от взаимных связей случайных величин. О системе n случайных величин удобнее говорить как о точке в пространстве n измерений или как о случайном n-мерном векторе. Системы случайных величин полностью характеризуются многомерными законами распределения. Их часто характеризуют при помощи числовых характеристик этих законов. Наиболее широкое распространение при изучении реальных случайных процессов получило многомерное нормальное распределение. Из центральной предельной теоремы следует, что предельным распределением одномерных независимых случайных величин является одномерный нормальный закон. Из обобщенной центральной предельной теоремы получаем, что предельным распределением в случае n измерений ($n > 1$) является многомерное нормальное распределение.

Практика использования многомерного нормального распределения в изучении экономики показала, что оно является хорошим приближением к реальным распределениям. Широкое распространение получил статистический анализ, основанный на модели нормального распределения. Это связано с тем, что одним из замечательных свойств многомерного нормального распределения является нормальность частных и условных распределений.

Рассмотрим многомерное нормальное распределение, характеристики которого широко изложены в отечественной и зарубежной литературе, например в работе [6].

1.2. Характеристики многомерного нормального распределения

Плотность хорошо знакомого из курса теории вероятностей одномерного нормального распределения может быть представлена следующим выражением:

$$f(x) = k \exp\left[-\frac{1}{2} a (x - \beta)^2\right].$$

Выражение в показателе степени может быть переписано так:

$$a(x - \beta)^2 = (x - \beta)\, a\, (x - \beta), \tag{1.1}$$

тогда

$$f(x) = k \exp\left[-\frac{1}{2} (x - \beta)\, a\, (x - \beta)\right], \tag{1.2}$$

где $a = 1/\sigma^2$, $a > 0$;

σ — среднее квадратическое отклонение случайной величины;

x — значение случайной величины;

$\beta = m_x$ — математическое ожидание случайной величины;

k — постоянный коэффициент.

Этот коэффициент выбирается из следующего условия:

$$\int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Обратимся к многомерному нормальному распределению величин $X_1, X_1, ..., X_n$. Его форма соответствует выражению (1.2), но при этом смысл составляющих форму меняется. Скалярная величина x заменяется конечномерным вектором:

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{pmatrix}.$$

Скалярная постоянная β заменяется n-мерным вектором:

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix}.$$

Положительная постоянная величина a заменяется положительно определенной симметрической матрицей:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{pmatrix}.$$

Выражение (1.1) заменяется квадратичной формой:

$$(\mathbf{x} - \mathbf{b})'\mathbf{A}(\mathbf{x} - \mathbf{b}) = \sum_{i,j=1}^{n} a_{ij}(x_i - b_i)(x_j - b_j),$$

где a_{ij} — компоненты матрицы **A**;

b_i — математическое ожидание случайной величины x_i; $(i = \overline{1, n})$;

b_i, b_j — компоненты вектора **b**;

$(\mathbf{x} - \mathbf{b})$ — вектор, компоненты которого являются центрированными случайными величинами.

$$(\mathbf{x}-\mathbf{b}) = \begin{pmatrix} x_1 - b_1 \\ x_2 - b_2 \\ \cdots \\ x_n - b_n \end{pmatrix}.$$

Транспонированный вектор $(\mathbf{x}-\mathbf{b})'$ — вектор-строка.

$$(\mathbf{x}-\mathbf{b})' = ((x_1 - b_1), (x_2 - b_2), \ldots, (x_n - b_n)).$$

Переход от показателя степени при одномерном распределении к квадратичной форме полезно пояснить простейшим примером.

Пример 1.1

Задана квадратичная функция

$$f(\mathbf{x}) = 2x_1^2 + 2x_1x_2 + 2x_2^2.$$

Запишите эту функцию в матричной форме.

Решение

1. Этой функции соответствует матрица

$$\mathbf{A} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}.$$

У матрицы $\mathbf{A}$ диагональные элементы являются коэффициентами при x^2, а недиагональные элементы равны половине коэффициента при $x_k x_j$ $(k \neq j)$.

2. Тогда следующее произведение будет представлять собой функцию

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = (x_1, x_2)\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = ((2x_1 + x_2), (x_1 + 2x_2))\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$= 2x_1^2 + x_2x_1 + x_1x_2 + 2x_2^2 = 2x_1^2 + 2x_1x_2 + 2x_2^2,$$

которая соответствует исходной функции.

Если известны значения x_1 и x_2, то в результате можно получить некоторое число.

Теперь запишем выражение для плотности n-мерного нормального распределения вероятностей (в матричной форме):

$$f(x_1, x_2, \ldots, x_n) = \mathrm{K} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{b})'\mathbf{A}(\mathbf{x}-\mathbf{b})\right).$$

Коэффициент K выбирается бóльшим нуля, при этом интеграл по всему n-мерному евклидову пространству переменных $x_1, x_2, \ldots, x_n$ равен единице.

В данной форме мы видим сходство с записью (1.2) для одномерной плотности распределения. Можно показать, что плотность многомерного нормального распределения вероятностей неотрицательна и ограничена, т. е.

$$f(x_1, x_2, \ldots, x_n) \geqslant 0;$$

$$f(x_1, x_2, \ldots, x_n) \leqslant \mathrm{K}.$$

Это следует из определения $\mathbf{A}$ как положительно определенной матрицы, что влечет за собой

$$(\mathbf{x}-\mathbf{b})'\mathbf{A}(\mathbf{x}-\mathbf{b}) \geqslant 0.$$

Определим значение K, при котором интеграл по всему n-мерному евклидову пространству n переменных был бы равен единице. Для этого обозначим

$$\mathbf{L} = \int_{-\infty}^{\infty} \ldots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}(\mathbf{x}-\mathbf{b})'\mathbf{A}(\mathbf{x}-\mathbf{b})\right) dx_n dx_{n-1} \ldots dx_1. \qquad (1.3)$$

Воспользуемся известной теоремой и ее следствием из работы [6].

Теорема 1.1. Для любой симметрической матрицы $\mathbf{A}$ существует ортогональная матрица $\mathbf{C}$, такая, что

$$\mathbf{C}'\mathbf{A}\mathbf{C} = \mathbf{D} = \begin{pmatrix} d_1 & 0 & \ldots & 0 \\ 0 & d_2 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & d_n \end{pmatrix}.$$

Если при этом $\mathbf{A}$ — положительно определенная матрица, то все $d_j > 0$. Отсюда вытекает следствие.

Следствие. Если матрица **A** положительно определена, то существует невырожденная матрица **C**, такая, что

$$\mathbf{C'AC} = \mathbf{E}, \tag{1.4}$$

где **E**—единичная матрица;

C'—матрица, транспонированная по отношению к ортогональной матрице **C**.

Теперь примем

$$\mathbf{x} - \mathbf{b} = \mathbf{Cy}, \tag{1.5}$$

где

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Перепишем квадратичную форму:

$$(\mathbf{x} - \mathbf{b})'\mathbf{A}(\mathbf{x} - \mathbf{b}) = \mathbf{y}'\mathbf{C}'\mathbf{ACy} = \mathbf{y}'\mathbf{y}.$$

После перехода к новым переменным, когда в качестве линейного оператора преобразования использована матрица **C**, можно матрицу (1.3) записать так:

$$\mathbf{L} = \mathrm{mod}\,|\mathbf{C}| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\mathbf{y}'\mathbf{y}\right) dy_n dy_{n-1} \dots dy_1, \tag{1.6}$$

где mod $|\mathbf{C}|$—определитель преобразования **L** от переменной x к переменной y, абсолютная величина определителя матрицы **C**.

Можно представить

$$e^{-\frac{1}{2}\mathbf{y}'\mathbf{y}} = e^{-\frac{1}{2}\sum_{i=1}^{n} y_i^2} = \prod_{i=1}^{n} e^{-\frac{1}{2}y_i^2}. \tag{1.7}$$

Тогда выражение (1.6) запишется так:

$$\mathbf{L} = \mathrm{mod}\,|\mathbf{C}| \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y_1^2\right) \exp\left(-\frac{1}{2}y_2^2\right) \dots$$

$$\dots \exp\left(-\frac{1}{2}y_n^2\right) dy_n \dots dy_1 = \mathrm{mod}\,|\mathbf{C}| \prod_{i=1}^{n} \left\{\int_{-\infty}^{\infty} e^{-\frac{1}{2}y_i^2} dy_i\right\}.$$

Из теории вероятностей известно, что

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y_i^2\right) dy_i = 1,$$

следовательно,

$$\int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}y_i^2\right) dy_i = \sqrt{2\pi}.$$

Значит,

$$\mathbf{L} = \mathrm{mod}\,|\mathbf{C}| \prod_{i=1}^{n} \{\sqrt{2\pi}\} = \mathrm{mod}\,|\mathbf{C}|\,(2\pi)^{\frac{1}{2}n}.$$

Из линейной алгебры известно, что определитель произведения нескольких квадратных матриц равен произведению определителей этих же матриц, а определители матрицы и ее транспонированной матрицы равны между собой. В данном случае из выражения (1.4) заключаем:

$$|\mathbf{C}'||\mathbf{A}||\mathbf{C}| = |\mathbf{E}| \tag{1.8}$$

и, используя $|\mathbf{C}'| = |\mathbf{C}|$, $|\mathbf{E}| = 1$, перепишем матрицу (1.8) так:

$$|\mathbf{C}|^2|\mathbf{A}| = 1; \ [\mathrm{mod}\,|\mathbf{C}|]^2|\mathbf{A}| = 1; \ \mathrm{mod}\,|\mathbf{C}| = 1/\sqrt{|\mathbf{A}|}.$$

Коэффициент K определяется из условия

$$K \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{b})'\mathbf{A}(\mathbf{x} - \mathbf{b})\right] dx_n dx_{n-1} \dots dx_1 = 1,$$

что соответствует, согласно матрице (1.3), $K\mathbf{L} = 1$, откуда

$$K = 1/\mathbf{L} = \sqrt{|\mathbf{A}|}\,(2\pi)^{-\frac{1}{2}n}.$$

Теперь можно записать выражение для плотности многомерного нормального распределения вероятностей:

$$f(x_1, x_2, \dots, x_n) = \frac{\sqrt{|\mathbf{A}|}}{(2\pi)^{\frac{1}{2}n}} \exp\left[-\frac{1}{2}(\mathbf{x} - \mathbf{b})'\mathbf{A}(\mathbf{x} - \mathbf{b})\right].$$

Выясним вероятностный смысл матрицы **A** и вектора **b**. Для этого надо определить первый и второй моменты случайных величин $X_1, X_2, ..., X_n$, а случайные величины будем считать компонентами случайного вектора

$$\mathbf{x} = \begin{pmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{pmatrix}.$$

Случайный вектор является частным случаем случайной матрицы, когда число ее столбцов равно единице.

Определение 1.1. *Случайная матрица* $\mathbf{Z} = (z_{ij})$; $(i = \overline{1, m}$; $j = \overline{1, n})$ *есть матрица случайных величин* $z_{11}, z_{12}, ..., z_{mn}$.

Определение 1.2. *Математическое ожидание случайной матрицы* $\mathbf{Z}$ есть $\mathbf{MZ} = (\mathbf{M}z_{ij})$.

Свойства математического ожидания случайной матрицы (или вектора) выражаются леммой [6].

Лемма 1.1. Если $\mathbf{Z}$ является случайной матрицей порядка $m \times n$, а действительные матрицы $\mathbf{D}, \mathbf{G}, \mathbf{F}$ имеют порядок $l \times m$, $n \times q$, $l \times q$ соответственно, то

$$\mathbf{M(DZG + F) = D(MZ)G + F}.$$

Доказательство. Элемент h-й строки и p-го столбца матрицы $\mathbf{M(DZG + F)}$ равен:

$$\mathbf{M}(\Sigma d_{hj} z_{ji} g_{ip} + f_{hp}) = \Sigma d_{hj} (\mathbf{M} z_{ji}) g_{ip} + f_{hp},$$

а это является элементом h-й строки и p-го столбца матрицы $\mathbf{D(MZ)G + F}$, что и требовалось доказать.

Из выражения (1.5) следует, что

$$\mathbf{x = Cy + b;\ Mx = C(My) + b}. \tag{1.9}$$

Плотность распределения вероятностей вектора $\mathbf{y}$ пропорциональна формуле (1.7):

$$(2\pi)^{\frac{1}{2}n} \exp\left(-\frac{1}{2}\mathbf{y'y}\right) = \prod_{j=1}^{n} \left\{(2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}y_j^2\right)\right\}. \tag{1.10}$$

Математическое ожидание i-й компоненты вектора $\mathbf{y}$:

$$\mathbf{M}y_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i \exp\left(-\frac{1}{2}y_i^2\right) dy_i \times$$

$$\times \prod_{\substack{j=1 \\ j \neq i}}^{n} \left\{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_j^2\right) dy_j\right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i e^{-\frac{1}{2}y_i^2} dy_i = 0,$$

а

$$\mathbf{M}y_i = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} y_i \prod_{j=1}^{n} \left\{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_j^2\right)\right\} dy_1 dy_2 \dots dy_{n-1} dy_n.$$

Равенство нулю естественно, так как $y_i \exp\left(-\frac{1}{2}y_i^2\right)$ является нечетной функцией y_i, поэтому первый (нечетный) момент ее равен нулю.

Если любая компонента случайного вектора $\mathbf{y}$ равна нулю, то и $\mathbf{My} = 0$. Таким образом, в формуле (1.9) $\mathbf{C(My)} = 0$, следовательно, $\mu = \mathbf{Mx} = \mathbf{b}$. Выяснив, что вектор $\mathbf{b}$ представляет собой среднее значение случайного вектора $\mathbf{x}$, которое мы обозначили μ, можем приступить к определению статистического смысла матрицы $\mathbf{A}$.

Предположим, что имеется случайный вектор $\mathbf{x}$. Ковариационная матрица этого вектора определится как математическое ожидание произведения центрированного многомерного вектора $\mathbf{x}^0$ на его транспозицию, т. е.

$$\mathbf{M}(\mathbf{x} - \mu)(\mathbf{x} - \mu)' = (\mathbf{M}(X_i - \mu_i)(X_j - \mu_j)). \tag{1.11}$$

Значит, элементом ковариационной матрицы является математическое ожидание произведения центрированных случайных величин X_i^0 и X_j^0 $(i, j = \overline{1, n})$. В этой матрице диагональный $(i = j)$ элемент $\mathbf{M}(X_i - \mu_i)^2$ является дисперсией случайной величины X_i. Недиагональные $(i \neq j)$ элементы этой матрицы $\mathbf{M}(X_i - \mu_i)(X_j - \mu_j)$ являются ковариацией [6] случайных величин X_i и X_j.

К выражению (1.11) применим преобразование (1.5) и получим

$$\mathbf{M}(\mathbf{x} - \mu)(\mathbf{x} - \mu)' = \mathbf{MCyy'C'} = \mathbf{C(Myy')C'}. \tag{1.12}$$

Далее необходимо выяснить значение матрицы ($\mathbf{Myy'}$). Элементом i-й строки и j-го столбца этой матрицы является

$$\mathbf{M}y_iy_j = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} y_i\, y_j \prod_{m=1}^{n} \left\{\frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_m^2\right)\right\} dy_1 \dots dy_{n-1} dy_n,$$

так как плотность распределения вероятностей случайного вектора $\mathbf{y}$ равна выражению (1.10). Рассмотрим случаи, когда $i = j$ и $i \neq j$.

Если $i = j$, то

$$\mathbf{M}y_i^2 = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i^2 \exp\left(-\frac{1}{2}\, y_i^2\right) dy_i \prod_{\substack{m=1\\ m \neq i}}^{n} \left\{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y_m^2\right) dy_m\right\} =$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i^2 \exp\left(-\frac{1}{2}\, y_i^2\right) dy_i = 1.$$

Здесь произведение интегралов равно произведению единиц. Следовательно, $\mathbf{M}y_i^2$ есть математическое ожидание квадрата нормально распределенной случайной величины с математическим ожиданием, равным нулю, и дисперсией, равной единице.

Если $i \neq j$, то

$$\mathbf{M}y_iy_j = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_i \exp\left(-\frac{1}{2}\, y_i^2\right) dy_i \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} y_j \exp\left(-\frac{1}{2}\, y_j^2\right) dy_j \cdot$$

$$\cdot \prod_{\substack{m=1\\ m \neq i,j}}^{n} \left\{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2}\, y_m^2\right) dy_m\right\} = 0;\ (i \neq j).$$

Так как $y_i e^{-\frac{1}{2}y_i^2}$ является нечетной функцией, то нечетный момент равен нулю, поэтому и произведение интегралов тоже равно нулю.

Из характера элементов матрицы ($\mathbf{Myy'}$) можем заключить, что ее диагональные элементы являются единицами, а недиагональные элементы являются нулевыми, значит, $\mathbf{Myy'} = \mathbf{E}$. Основываясь на этом, можно переписать выражение (1.12) так:

$$\mathbf{M}(\mathbf{x} - \mu)\ (\mathbf{x} - \mu)' = \mathbf{CEC'} = \mathbf{CC'}. \tag{1.13}$$

Обратимся к выражению (1.4) и умножим его части слева на $(\mathbf{C'})^{-1}$, а справа на $\mathbf{C}^{-1}$, получим:

$$(\mathbf{C'})^{-1}\, \mathbf{C'ACC}^{-1} = (\mathbf{C'})^{-1}\mathbf{EC}^{-1};$$

$$\mathbf{A} = (\mathbf{C'})^{-1}\mathbf{C}^{-1}.$$

Перейдем к обратным матрицам и получим:

$$\mathbf{CC'} = \mathbf{A}^{-1}. \tag{1.14}$$

Итак, ковариационная матрица (1.13) случайного вектора $\mathbf{x}$ есть обратная матрица $\mathbf{A}^{-1}$:

$$\Sigma = \mathbf{M}\ (\mathbf{x} - \mu)(\mathbf{x} - \mu)' = \mathbf{A}^{-1}.$$

Из выражения (1.14) заключаем, что ковариационная матрица Σ положительно определена. Доказанное при выяснении роли $\mathbf{b}$ и $\mathbf{A}$ обычно формулируется в виде теорем [6].

Теорема 1.2. Если плотность распределения вероятностей n-мерного случайного вектора $\mathbf{x}$ есть

$$\sqrt{|\mathbf{A}|}\ \ (2\pi)^{-\frac{1}{2}n} \exp\left(-\frac{1}{2}(\mathbf{x} - \mathbf{b})'\mathbf{A}(\mathbf{x} - \mathbf{b})\right),$$

то среднее значение $\mathbf{x}$ есть $\mathbf{b}$ и ковариационная матрица есть $\mathbf{A}^{-1}$.

Теорема 1.3. Обратная теорема. Если даны вектор μ и положительно определенная матрица Σ, то существует такая многомерная нормальная плотность распределения вероятностей:

$$n(\mathbf{x} \mid \mu, \Sigma) = (2\pi)^{-\frac{1}{2}n}\ |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu)\right), \tag{1.15}$$

что математическое ожидание случайного вектора $\mathbf{x}$ с этой плотностью распределения есть μ и ковариационная матрица есть Σ.

Обычно плотность распределения вероятностей обозначают так, как записано слева в формуле (1.15), а многомерный нормальный закон распределения обозначают $N(\mu, \Sigma)$. В данном распределении нас должна заинтересовать структура ковариационной матрицы и ее связь с корреляционной матрицей. Это можно сделать в общем виде для случайного вектора n-го порядка. Однако с методической точки зрения удобнее обратиться к простейшему многомерному распределению — двумерному нормальному распределению.

1.3. Двумерное нормальное распределение

Рассмотрим коэффициент корреляции между случайными величинами X_i и X_j:

$$\rho_{ij} = \frac{\sigma_{ij}}{\sqrt{\sigma_{ii}}\sqrt{\sigma_{jj}}} = \frac{\sigma_{ij}}{\sigma_i\sigma_j},$$

где σ_{ii}—i-й диагональный элемент ковариационной матрицы;

σ_{ij}—ковариация случайных величин X_i и X_j (или корреляционный момент случайных величин X_i и X_j);

σ_i —среднее квадратическое отклонение X_i.

Ковариационная матрица является симметрической, т. е. $\rho_{ij} = \rho_{ji}$. Ковариационная матрица положительно определена:

$$\begin{pmatrix} \sigma_{ii} & \sigma_{ij} \\ \sigma_{ji} & \sigma_{jj} \end{pmatrix} = \begin{pmatrix} \sigma_i^2 & \sigma_i\sigma_j\rho_{ij} \\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2 \end{pmatrix}, \qquad (1.16)$$

следовательно, определитель ее положителен:

$$\begin{vmatrix} \sigma_i^2 & \sigma_i\sigma_j\rho_{ij} \\ \sigma_i\sigma_j\rho_{ij} & \sigma_j^2 \end{vmatrix} = \sigma_i^2\sigma_j^2(1-\rho_{ij}^2) > 0.$$

Отсюда для невырожденных распределений $-1 < \rho_{ij} < 1$.

Рассмотрим теперь ковариационную матрицу двумерного нормального распределения $X_i = X_1$; $X_j = X_2$. Определим среднее значение:

$$\mathbf{M}\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix},$$

т. е. $\mathbf{Mx} = \mu$.

Определим ковариационную матрицу в развернутом виде и приведем к виду ковариационной матрицы (1.16):

$$\Sigma = \mathbf{M}\begin{pmatrix} (x_1-\mu_1)^2 & (x_1-\mu_1)(x_2-\mu_2) \\ (x_2-\mu_2)(x_1-\mu_1) & (x_2-\mu_2)^2 \end{pmatrix} =$$

$$= \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 \end{pmatrix},$$

где ρ —коэффициент корреляции между X_1 и X_2;

σ_1—среднее квадратическое отклонение X_1;

σ_2—среднее квадратическое отклонение X_2.

В выражении (1.15) в показателе степени стоит обращенная ковариационная матрица

$$\Sigma^{-1} = \begin{pmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho \\ \sigma_2\sigma_1\rho & \sigma_2^2 \end{pmatrix}^{-1}.$$

Так как это матрица второго порядка, то для присоединенной матрицы легко получить алгебраические дополнения. Присоединенная матрица

$$\Sigma^* = \begin{pmatrix} \sigma_2^2 & -\sigma_1\sigma_2\rho \\ -\sigma_2\sigma_1\rho & \sigma_1^2 \end{pmatrix}.$$

Разделив присоединенную матрицу на определитель $|\Sigma|$, получим:

$$\Sigma^{-1} = \frac{\Sigma^*}{|\Sigma|} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{\sigma_2^2}{\sigma_1^2\sigma_2^2} & \frac{-\sigma_1\sigma_2\rho}{\sigma_1^2\sigma_2^2} \\ \frac{-\sigma_2\sigma_1\rho}{\sigma_1^2\sigma_2^2} & \frac{\sigma_1^2}{\sigma_1^2\sigma_2^2} \end{pmatrix} = \frac{1}{1-\rho^2}\begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_2\sigma_1} & \frac{1}{\sigma_2^2} \end{pmatrix}.$$

Определим плотность вероятности X_1 и X_2:

$$n(\mathbf{x} \mid \mu, \Sigma) = f(x_1, x_2) = (2\pi)^{-\frac{1}{2}n}|\Sigma|^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu)\right).$$

Преобразуем выражение перед экспонентой:

$$(2\pi)^{-\frac{1}{2}n}|\Sigma|^{-\frac{1}{2}} = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}}.$$

Преобразуем выражение, стоящее в показателе степени:

$$-\frac{1}{2}(\mathbf{x}-\mu)'\Sigma^{-1}(\mathbf{x}-\mu) =$$

$$= -\frac{1}{2(1-\rho^2)}((x_1-\mu_1)(x_2-\mu_2))\begin{pmatrix} \frac{1}{\sigma_1^2} & \frac{-\rho}{\sigma_1\sigma_2} \\ \frac{-\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} =$$

$$= \frac{-1}{2(1-\rho^2)}\left(\left(\frac{x_1-\mu_1}{\sigma_1^2}-\frac{\rho(x_2-\mu_2)}{\sigma_2\sigma_1}\right)\left(-\frac{\rho(x_1-\mu_1)}{\sigma_1\sigma_2}+\frac{x_2-\mu_2}{\sigma_2^2}\right)\right)\begin{pmatrix} x_1-\mu_1 \\ x_2-\mu_2 \end{pmatrix} =$$

$$= -\frac{1}{2(1-\rho^2)}\left(\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right).$$

Тогда плотность вероятности X_1 и X_2 будет такова:

$$n(\mathbf{x} \mid \mu, \Sigma) = f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot$$

$$\cdot \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2}-2\rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2}+\frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}. \quad (1.17)$$

Таким образом, при рассмотрении двумерного нормального распределения мы легко убеждаемся в том, что коэффициенты корреляции и дисперсии случайных величин являются основными числовыми характеристиками наряду с математическими ожиданиями. Если конечное число случайных величин превосходит $n = 2$, то роль выражения (1.16) выполняет ковариационная или корреляционная матрица более высокого порядка. Элементы этой матрицы получаются на практике из экспериментальных или статистических данных и являются статистическими величинами, требующими своей оценки.

Рассмотрим критерии значимости выборочного коэффициента корреляции.

1.4. Критерий значимости коэффициента корреляции

Значение коэффициента корреляции является случайной величиной, которая зависит от объема выборки N. С уменьшением числа наблюдений доверительная вероятность коэффициента корреляции снижается при данном доверительном интервале. При возрастании N распределение r_{jk} (не превышающее 0,9) стремится к нормальному со средним значением и дисперсией, равными соответственно $\bar{x} = 0$; σ^2.

Коэффициент корреляции, полученный из выборки, является оценкой соответствующего коэффициента корреляции генеральной совокупности. Тесноту корреляционной связи между двумя признаками X_j и X_k характеризует коэффициент корреляции генеральной совокупности; этому коэффициенту корреляции не представляется возможным дать одно единственное значение. Можно определить доверительный интервал, в котором с определенной доверительной вероятностью находится коэффициент корреляции генеральной совокупности ρ_{jk}.

Величина доверительного интервала

$$r_{jk} - t_\alpha S_r \leqslant \rho_{jk} \leqslant r_{jk} + t_\alpha S_r,$$

где S_r — среднеквадратическая ошибка выборочного коэффициента корреляции;

t_α — табличное значение, имеющее распределение Стьюдента с числом степеней свободы $N-2$;

α — величина риска, выраженная в процентах или долях единицы.

При нормальном распределении случайной величины вероятность ее отклонения от своего среднего значения на величины не более $1{,}96\sigma_r$; $2{,}58\sigma_r$ и $3\sigma_r$ соответственно равна 0,95; 0,99; 0,997, а вероятность отклонения на большую величину соответственно выразится:

$$P\{r_{jk} + 1{,}96\sigma_r \leqslant \rho_{jk} \leqslant r_{jk} - 1{,}96\sigma_r\} = 0{,}05 = \alpha_1;$$

$$P\{r_{jk} + 2{,}58\sigma_r \leqslant \rho_{jk} \leqslant r_{jk} - 2{,}58\sigma_r\} = 0{,}01 = \alpha_2;$$

$$P\{r_{jk} + 3\sigma_r \leqslant \rho_{jk} \leqslant r_{jk} - 3\sigma_r\} = 0{,}003 = \alpha_3.$$

Таким образом, каждый раз весь интервал вероятностей разбивается на два интервала. В одном интервале находятся все возможные значения ρ_{jk} с вероятностью $P = 1 - \alpha$, а в другом интервале — остальные значения ρ_{jk} с вероятностью α.

Значение α называется уровнем значимости (для случая двухстороннего интервала). Если исследуется односторонний доверительный интервал, то учитывается значение $\alpha/2$. При проверке значимости полученного коэффициента корреляции проверяется гипотеза о том, что коэффициент корреляции генеральной совокупности $\rho_{jk} = 0$. Если данная гипотеза подтвердится, то мы не имеем права считать с доверительной вероятностью $P = 1 - \alpha$, что корреляционная связь между изучаемыми признаками существует. Для облегчения расчетов Фишер и Иейтс вычислили и табулировали распределение r_{jk}, соответствующее выборкам, извлекаемым из некоррелированной двумерной нормальной совокупности (см. приложение 2) [12]. Вместо α в таблице дан символ Q, так как часто в литературе говорят о Q%-ной точке выборочного коэффициента корреляции. Порядок пользования данной таблицей проиллюстрируем на примере.

Пример 1.2

Коэффициент корреляции вычислен на основании выборки объемом $N = 102$ и равен 0,68.

Определите, является ли обнаруженная связь значимой при $P = 0{,}99$.

Решение

1. Число степеней свободы

$\nu = N - 2 = 102 - 2 = 100.$

2. Определим абсолютную величину границы для гипотезы об отсутствии корреляции двух признаков ($\rho_{jk} = 0$). Для этого по входу в таблицу $\nu = 100$ выбираем в строке значение, соответствующее минимальному: $\alpha = 1 - P = 0{,}01$. Оно равно 0,25. Следовательно, заданное значение $r_{jk} = 0{,}68$ является значимым, так как оно превосходит границу интервала, включающего $\rho_{jk} = 0$. Значит, связь между признаками X_j и X_k является существенной.

Если число наблюдений в выборке невелико и получено высокое значение коэффициента корреляции, то для проверки гипотезы о наличии корреляционной связи или для построения доверительного интервала используется преобразование Фишера [69].

Вводятся новые переменные:

$$Z' = \frac{1}{2}\ln\frac{1 + r_{jk}}{1 - r_{jk}};$$

$$Z = \frac{1}{2}\ln\frac{1 + \rho_{jk}}{1 - \rho_{jk}}.$$

Разность $Z' - Z$ распределена по закону, близкому к нормальному, со средним значением

$$m_{(Z' - Z)} = \frac{\rho}{2(N-1)} \approx 0;$$

$$\sigma^2_{(Z' - Z)} = \frac{1}{N-1} + \frac{4 - \rho^2}{2(N-1)} \approx \frac{1}{N-3}.$$

Значит, с доверительной вероятностью α

$$-t_\alpha\sqrt{\frac{1}{N-3}} \leqslant Z; \qquad -Z \leqslant t_\alpha\sqrt{\frac{1}{N-3}}.$$

Доверительный интервал для Z определится так:

$$Z' - t_\alpha\sqrt{\frac{1}{N-3}} \leqslant Z \leqslant Z' + t_\alpha\sqrt{\frac{1}{N-3}}.$$

Для перехода от Z к r_{jk} можно воспользоваться приложением 3. Порядок использования данной таблицы покажем на следующем примере.

Пример 1.3

При уровне критерия значимости $\alpha = 5\%$ найдите доверительный интервал для значения коэффициента корреляции, равного 0,68, определенного при $N = 28$.

Решение

1. По приложению 3 для подходящего значения $r_{jk} = 0{,}6805$ находим значение $Z' = 0{,}83$.

2. Вычисляем доверительные границы для Z при $\alpha = 0{,}05$, когда $t_\alpha = 1{,}96$:

$$0{,}83 - 1{,}96\sqrt{\frac{1}{28-3}} \leqslant Z \leqslant 0{,}83 + 1{,}96\sqrt{\frac{1}{28-3}};$$

$$0{,}44 \leqslant Z \leqslant 1{,}22.$$

3. Для перехода к доверительным границам коэффициента корреляции генеральной совокупности по приложению 3 находим значения r_{jk}, соответствующие значениям Z, равным 0,44 и 1,22:

$$0{,}41 \leqslant \rho \leqslant 0{,}84.$$

В случае нелинейной зависимости между изучаемыми признаками коэффициенты корреляции теряют всякий физический смысл.

Для измерения тесноты криволинейной корреляционной связи между переменными X_j и X_k служит корреляционное отношение η_{jk}. Это отношение имеет тот же смысл, что и коэффициент корреляции. Во всех случаях, когда отсутствует линейная зависимость, $\eta_{jk} \neq \eta_{kj}$. Выполнение равенства $\eta_{jk} = |r_{jk}|$ является необходимым и достаточным условием того, чтобы регрессия X_j на X_k была точно линейной. Выполнение равенства $\eta_{kj} = |r_{jk}|$ является необходимым и достаточным условием того, чтобы регрессия X_k на X_j была точно линейной.

Корреляционные отношения неотрицательны и не превосходят единицы:

$$0 \leqslant \eta_{jk} \leqslant 1;\ 0 \leqslant \eta_{kj} \leqslant 1.$$

Если $\eta_{jk} = 0$, то связь между признаками X_j и X_k отсутствует. Если $\eta_{jk} = 1$, то между признаками существует однозначная функциональная связь. Для одной и той же корреляционной таблицы, характеризующей связь переменных X_j и X_k, оба корреляционных отношения не меньше абсолютной величины коэффициента корреляции:

$$\eta_{jk} \geqslant |r_{jk}|;\ \eta_{kj} \geqslant |r_{jk}|.$$

Надо учитывать, что корреляционное отношение может быть использовано для определения тесноты связи переменных (признаков) для функций криволинейных, но линейных относительно параметров. Для функций, нелинейных относительно параметров, корреляционное отношение не может служить точным критерием тесноты связи.

Известно, что корреляционным отношением X_j и X_k называется отношение межгруппового среднего квадратического отклонения δ_j переменной X_j к общему среднему квадратическому отклонению этой переменной σ_j:

$$\eta_{jk} = \delta_j/\sigma_j.$$

Корреляционным отношением X_k и X_j называется отношение межгруппового среднего квадратического отклонения δ_k переменной X_k к общему среднему квадратическому отклонению этой переменной σ_k:

$$\eta_{kj} = \delta_k/\sigma_k.$$

Таким образом, корреляционное отношение может быть одним из критериев для подтверждения гипотезы о линейной зависимости.

Для проведения исследования методом главных компонент необходимо вычислить коэффициенты корреляции (парные коэффициенты корреляции) между парами признаков (можно вычислять и коэффициенты ковариации). Так как мы пока не ставим ограничений на количество признаков, то надо полагать, что их может оказаться достаточно большое, но конечное число, поэтому преобразования проводятся в матричной форме.

Для оценки результатов можно воспользоваться готовыми таблицами нормального закона распределения.

Глава 2
МЕТОД ГЛАВНЫХ КОМПОНЕНТ

2.1. Метод главных компонент в ряду других методов многомерного статистического анализа

Длительное время метод главных компонент (компонентный анализ) рассматривался многими авторами [42], [59], [106] как одна из разновидностей методов факторного анализа. При этом под факторным анализом понимался способ приведения множества непосредственно наблюдаемых признаков к меньшему числу неявных, но объективно существующих факторов. В настоящее время метод главных компонент часто отделяют от факторного анализа. Впервые он был предложен в 1901 г. К. Пирсоном, затем развит, доработан, описан и обоснован в работах Г. Хотеллинга, Г. Хармана, С. Рао, П. Андруковича, С. А. Айвазяна, В. С. Мхитаряна [5], [7], [59], [83], [106].

В России метод главных компонент стал распространяться с появлением ЭВМ вследствие удобства математических процедур и наличия стандартных программ, которые оказалось возможным применить для организации математического обеспечения алгоритма расчета.

Метод главных компонент обладает определенными преимуществами перед другими методами факторного анализа. Он не требует, например, никаких гипотез о переменных, является линейным и аддитивным.

Несмотря на то, что при методе главных компонент для точного воспроизведения коэффициентов корреляции между переменными надо найти все n компонент, большая доля изменчивости признаков (дисперсии) объясняется небольшим числом (m) компонент. Кроме того, при методе главных компонент можно по признакам описать компоненты, а по компонентам — признаки. В книге Д. Лоули и А. Максвелла [59] сказано, что факторный анализ изучает ковариации, а метод главных компонент — дисперсии. Однако С. Рао показал, что метод главных компонент одинаково хорошо приближает дисперсии и ковариации.

В зависимости от решаемых задач каждый из методов факторного анализа, как и метод главных компонент, имеет свои достоинства и недостатки. Несмотря на то что метод главных компонент считается статистическим, есть подход, не являющийся статистическим. Этот подход связан с получением наилучшей проекции совокупности точек наблюдения в пространство меньшей размерности. В этом случае надо знать матрицу вторых моментов.

При статистическом подходе, которому и посвящена данная работа, задача заключается в выделении линейных комбинаций случайных величин, имеющих максимально возможную дисперсию. Общим между статистическим и нестатистическим подходами является использование матрицы вторых моментов как исходной для начала анализа.

2.2. Обратные и ортогональные преобразования в методе главных компонент

При получении главных компонент и построении по ним уравнений регрессии используются ортогональные и обратные преобразования. Для проведения данных преобразований необходимо остановиться на способах обращения квадратных матриц и свойствах обратных и ортогональных матриц. Начнем с определений.

Квадратная матрица $\mathbf{A}^{-1}$ *называется обратной к матрице* $\mathbf{A}$, *если она, будучи умноженной как слева, так и справа на матрицу* $\mathbf{A}$, *дает единичную матрицу:*

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{E}. \tag{2.1}$$

Для получения матрицы, обратной матрице $\mathbf{A}$, можно построить союзную (присоединенную) матрицу $\mathbf{A}^*$ и разделить ее на определитель матрицы $\mathbf{A}$.

Союзная матрица — это матрица, элементами которой являются алгебраические дополнения каждого элемента определителя матрицы $\mathbf{A}$, *причем алгебраические дополнения элементов строк определителя матрицы A помещены в соответствующих столбцах.*

Следовательно, алгоритм построения обратной матрицы может быть представлен так:

1) вычислить алгебраические дополнения A_{ij} каждого элемента a_{ij} определителя матрицы **A**;

2) вычислить определитель матрицы **A**;

3) составить транспонированную союзную матрицу:

$$\tilde{\mathbf{A}} = \begin{pmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \dots & \dots & \dots & \dots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{pmatrix};$$

4) произвести операцию транспонирования матрицы **A** и получить присоединенную матрицу:

$$\mathbf{A}^* = \tilde{\mathbf{A}}' = \begin{pmatrix} A_{11} & A_{21} & \dots & A_{n1} \\ A_{12} & A_{22} & \dots & A_{n2} \\ \dots & \dots & \dots & \dots \\ A_{1n} & A_{2n} & \dots & A_{nn} \end{pmatrix};$$

5) составить обратную матрицу:

$$\mathbf{A}^{-1} = \frac{\tilde{\mathbf{A}}'}{|\mathbf{A}|} = \begin{pmatrix} \frac{A_{11}}{|\mathbf{A}|} & \frac{A_{21}}{|\mathbf{A}|} & \dots & \frac{A_{n1}}{|\mathbf{A}|} \\ \frac{A_{12}}{|\mathbf{A}|} & \frac{A_{22}}{|\mathbf{A}|} & \dots & \frac{A_{n2}}{|\mathbf{A}|} \\ \dots & \dots & \dots & \dots \\ \frac{A_{1n}}{|\mathbf{A}|} & \frac{A_{2n}}{|\mathbf{A}|} & \dots & \frac{A_{nn}}{|\mathbf{A}|} \end{pmatrix}.$$

Эту запись можно привести в сокращенном виде так:

$$\mathbf{A}^{-1} = \frac{A_{ji}}{|\mathbf{A}|}. \tag{2.2}$$

Пример 2.1

Найдите матрицу, обратную следующей матрице:

$$\mathbf{A} = \begin{pmatrix} 1 & 4 & 1 \\ 2 & 5 & 4 \\ 3 & 0 & 5 \end{pmatrix}.$$

Решение

1. Вычислим алгебраические дополнения:

$$A_{11} = (-1)^2 \begin{vmatrix} 5 & 4 \\ 0 & 5 \end{vmatrix} = 25; \; A_{32} = (-1)^5 \begin{vmatrix} 1 & 1 \\ 3 & 4 \end{vmatrix} = 2;$$

$$A_{21} = (-1)^3 \begin{vmatrix} 4 & 1 \\ 0 & 5 \end{vmatrix} = -20; \; A_{13} = (-1)^4 \begin{vmatrix} 2 & 5 \\ 3 & 0 \end{vmatrix} = -15;$$

$$A_{31} = (-1)^4 \begin{vmatrix} 4 & 1 \\ 5 & 4 \end{vmatrix} = 11; \; A_{23} = (-1)^5 \begin{vmatrix} 1 & 4 \\ 3 & 0 \end{vmatrix} = 12;$$

$$A_{12} = (-1)^3 \begin{vmatrix} 2 & 4 \\ 3 & 5 \end{vmatrix} = 2; \; A_{33} = (-1)^6 \begin{vmatrix} 1 & 4 \\ 2 & 5 \end{vmatrix} = -3;$$

$$A_{22} = (-1)^4 \begin{vmatrix} 1 & 1 \\ 3 & 5 \end{vmatrix} = 2.$$

2. Вычислим определитель матрицы **A**:

$$|\mathbf{A}| = a_{12}A_{12} + a_{22}A_{22} + a_{32}A_{32} = 4 \cdot 2 + 5 \cdot 2 + 0 \cdot (-2) = 18.$$

3. Составим транспонированную союзную матрицу:

$$\tilde{\mathbf{A}} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix} = \begin{pmatrix} 25 & 2 & -15 \\ -20 & 2 & 12 \\ 11 & -2 & -3 \end{pmatrix}.$$

4. Составим союзную матрицу:

$$\mathbf{A}^* = \tilde{\mathbf{A}}' = \begin{pmatrix} 25 & -20 & 11 \\ 2 & 2 & -2 \\ -15 & 12 & -3 \end{pmatrix}.$$

5. Составим обратную матрицу:

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{25}{18} & \frac{-10}{9} & \frac{11}{18} \\ \frac{1}{9} & \frac{1}{9} & \frac{-1}{9} \\ \frac{-5}{6} & \frac{2}{3} & \frac{-1}{6} \end{pmatrix}.$$

Обратные матрицы обладают следующими свойствами:

- обратная матрица произведения квадратных матриц равна произведению квадратных матриц сомножителей, взятых в обратном порядке:
$$[\mathbf{AB}]^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}, \text{ или } [\mathbf{ABC}]^{-1} = \mathbf{C}^{-1}\mathbf{B}^{-1}\mathbf{A}^{-1};$$
- транспонированная обратная матрица равна обратной транспонированной данной матрице:
$$[\mathbf{A}^{-1}]' = [\mathbf{A}']^{-1};$$
- определитель обратной матрицы равен величине, обратной определителю заданной матрицы:
$$|\mathbf{A}^{-1}| = 1/|\mathbf{A}|;$$
- обратное линейное преобразование переменных $\mathbf{y}(y_1, y_2, ..., y_n)$ в переменные $\mathbf{f}(f_1, f_2, f_3, ..., f_n)$ может быть представлено при помощи обратной матрицы:

$$\mathbf{f} = \mathbf{A}^{-1}\mathbf{y},$$

если матрица **A** квадратная порядка n и при этом векторы **y** и **f** того же порядка.

Для обратного преобразования можно получить обратную матрицу более простым способом — при помощи метода Жордана-Гаусса. Суть его в следующем:

1) к матрице **A** приписываем единичную матрицу того же порядка через разделительную черту $(\mathbf{A} \mid \mathbf{E})$;

2) обе части полученной матрицы умножаем на $\mathbf{A}^{-1}$ и получаем $(\mathbf{A}^{-1}\mathbf{A} \mid \mathbf{A}^{-1}\mathbf{E}) = (\mathbf{E} \mid \mathbf{A}^{-1})$.

Вывод. Матрицу $(\mathbf{A} \mid \mathbf{E})$ надо преобразовать таким образом, чтобы слева получилась единичная матрица, тогда справа от черты получится матрица, обратная заданной **A**. При этом можно производить следующие элементарные преобразования матрицы $(\mathbf{A} \mid \mathbf{E})$:

- перемену местами любых строк;
- умножение некоторой строки на произвольное действительное число, отличное от нуля;
- прибавление к некоторой строке другой строки, умноженной на произвольное действительное число.

В результате этих преобразований в каждом столбце подматрицы **A** последовательно исключаются все элементы, кроме одного элемента, расположенного на ее главной диагонали. Полученная матрица $(\mathbf{E} \mid \mathbf{A}^{-1})$ должна иметь следующий вид:

$$(\mathbf{E} \mid \mathbf{A}^{-1}) = \left(\begin{array}{c|cccc} 1\,0\,0\,...\,0 & \frac{A_{11}}{|\mathbf{A}|} & \frac{A_{21}}{|\mathbf{A}|} & ... & \frac{A_{n1}}{|\mathbf{A}|} \\ 0\,1\,0\,...\,0 & \frac{A_{12}}{|\mathbf{A}|} & \frac{A_{22}}{|\mathbf{A}|} & ... & \frac{A_{n2}}{|\mathbf{A}|} \\ .... & ... & ... & ... & ... \\ 0\,0\,0\,...\,1 & \frac{A_{1n}}{|\mathbf{A}|} & \frac{A_{2n}}{|\mathbf{A}|} & ... & \frac{A_{nn}}{|\mathbf{A}|} \end{array}\right).$$

Пример 2.2

Методом Жордана-Гаусса получите матрицу, обратную матрице, представленной в примере 2.1.

Решение

1.

$$(\mathbf{A} \mid \mathbf{E}) = \left(\begin{array}{ccc|ccc} 1 & 4 & 1 & 1 & 0 & 0 \\ 2 & 5 & 4 & 0 & 1 & 0 \\ 3 & 0 & 5 & 0 & 0 & 1 \end{array}\right).$$

2. Получим на месте второго и третьего элементов первого столбца нули:

а) умножим для этого все элементы первой строки на (−2) и сложим с элементами второй строки:

$$\left(\begin{array}{ccc|ccc} 1 & 4 & 1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 3 & 0 & 5 & 0 & 0 & 1 \end{array}\right);$$

б) умножим все элементы первой строки на (−3) и сложим с элементами третьей строки:

$$\left(\begin{array}{ccc|ccc} 1 & 4 & 1 & 1 & 0 & 0 \\ 0 & -3 & 2 & -2 & 1 & 0 \\ 0 & -12 & 2 & -3 & 0 & 1 \end{array}\right).$$

3. Получим на месте второго элемента второго столбца преобразованной подматрицы **A** единицу, а вместо первого и третьего элементов этого столбца нули:

а) разделим все элементы второй строки на (-3):

$$\left(\begin{array}{ccc|ccc} 1 & 4 & 1 & 1 & 0 & 0 \\ 0 & 1 & \frac{-2}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & -12 & 2 & -3 & 0 & 1 \end{array}\right);$$

б) умножим все элементы второй строки на (-4) и сложим с элементами первой строки:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{11}{3} & \frac{-5}{3} & \frac{4}{3} & 0 \\ 0 & 1 & \frac{-2}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & -12 & 2 & -3 & 0 & 1 \end{array}\right);$$

в) умножим все элементы второй строки на 12 и сложим с элементами третьей строки:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{11}{3} & \frac{-5}{3} & \frac{4}{3} & 0 \\ 0 & 1 & \frac{-2}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & 0 & \frac{-18}{3} & \frac{15}{3} & \frac{-12}{3} & 1 \end{array}\right).$$

4. Получим на месте третьего элемента третьего столбца преобразованной подматрицы **A** единицу, а вместо остальных элементов этого столбца нули:

а) разделим на (-6) все элементы третьей строки:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{11}{3} & \frac{-5}{3} & \frac{4}{3} & 0 \\ 1 & 0 & \frac{-2}{3} & \frac{2}{3} & \frac{-1}{3} & 0 \\ 0 & 0 & 1 & \frac{-5}{6} & \frac{2}{3} & \frac{-1}{6} \end{array}\right);$$

б) умножим все элементы третьей строки матрицы на 2/3 и сложим с элементами второй строки:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & \frac{11}{3} & \frac{-5}{3} & \frac{4}{3} & 0 \\ 0 & 1 & 0 & \frac{1}{9} & \frac{1}{9} & \frac{-1}{9} \\ 0 & 0 & 1 & \frac{-5}{6} & \frac{2}{3} & \frac{-1}{6} \end{array}\right);$$

в) при помощи третьей строки получим второй нуль в третьем столбце преобразованной подматрицы **A**, чтобы она стала единичной подматрицей в последней матрице:

$$\left(\begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{25}{18} & \frac{-10}{9} & \frac{11}{18} \\ 0 & 1 & 0 & \frac{1}{9} & \frac{1}{9} & \frac{-1}{9} \\ 0 & 0 & 1 & \frac{-5}{6} & \frac{2}{3} & \frac{-1}{6} \end{array}\right).$$

Следовательно,

$$\mathbf{A}^{-1} = \begin{pmatrix} \frac{25}{18} & \frac{-10}{9} & \frac{11}{18} \\ \frac{1}{9} & \frac{1}{9} & \frac{-1}{9} \\ \frac{-5}{6} & \frac{2}{3} & \frac{-1}{6} \end{pmatrix}.$$

Ортогональная матрица — это матрица, у которой обратная и транспонированная матрицы совпадают:

$$\mathbf{A}^{-1} = \mathbf{A}'; \ \mathbf{AA}' = \mathbf{A}'\mathbf{A} = \mathbf{E}.$$

Квадратная матрица может быть названа ортогональной, если сумма квадратов элементов каждого столбца равна единице и сумма произведений соответствующих элементов из двух разных столбцов равна нулю. Определитель ортогональной матрицы равен ± 1.

Если матрица состоит из одного столбца порядка n, который обозначим **b**, то скалярное произведение

$$\mathbf{b}'\mathbf{b} = \sum_{j=1}^{n} b_j^2 = |\mathbf{b}|^2 \qquad (2.3)$$

представляет собой сумму квадратов элементов вектора **b**. Величина $|\mathbf{b}|$ называется нормой вектора **b**. Она равна корню квадратному из суммы квадратов элементов вектора.

Вектор **b** *называется нормализованным, если* $|\mathbf{b}| = 1$.

Пример 2.3

Проверьте, является ли вектор **a** нормализованным:

$$\mathbf{a} = (0{,}2\ 0{,}6\ 0{,}3\ 0{,}7\ 0{,}1\ 0{,}1).$$

Решение

Получим произведение вектора-строки **a**′ на вектор-столбец **a**:

$$\mathbf{a}'\mathbf{a} = (0{,}2\ 0{,}6\ 0{,}3\ 0{,}7\ 0{,}1\ 0{,}1)\begin{pmatrix} 0{,}2 \\ 0{,}6 \\ 0{,}3 \\ 0{,}7 \\ 0{,}1 \\ 0{,}1 \end{pmatrix} =$$

$$= 0{,}04 + 0{,}36 + 0{,}09 + 0{,}49 + 0{,}01 + 0{,}01 = 1.$$

Итак, вектор **a** является нормализованным. Его норма $|\mathbf{a}| = 1$.

Отсюда можно записать, что *квадратная матрица* **A** *называется ортогональной, если каждый вектор-столбец в* **A** *нормализован и ортогонален к любому другому ее вектору-столбцу, так что* $\mathbf{A}'\mathbf{A} = \mathbf{E}$.

Пример 2.4

Покажите, что матрица **A**′ ортогональна:

$$\mathbf{A} = \begin{pmatrix} \dfrac{0{,}3}{\sqrt{0{,}1}} & \dfrac{0{,}1}{\sqrt{0{,}1}} \\ \dfrac{0{,}2}{\sqrt{0{,}4}} & \dfrac{-0{,}6}{\sqrt{0{,}4}} \end{pmatrix}.$$

Решение

1. Определим транспонированную матрицу

$$\mathbf{A}' = \begin{pmatrix} \dfrac{0{,}3}{\sqrt{0{,}1}} & \dfrac{0{,}2}{\sqrt{0{,}4}} \\ \dfrac{0{,}1}{\sqrt{0{,}1}} & \dfrac{-0{,}6}{\sqrt{0{,}4}} \end{pmatrix}.$$

2. Выясним, является ли произведение двух матриц **AA**′ единичной матрицей

$$\begin{pmatrix} \dfrac{0{,}3}{\sqrt{0{,}1}} & \dfrac{0{,}1}{\sqrt{0{,}1}} \\ \dfrac{0{,}2}{\sqrt{0{,}4}} & \dfrac{-0{,}6}{\sqrt{0{,}4}} \end{pmatrix}\begin{pmatrix} \dfrac{0{,}3}{\sqrt{0{,}1}} & \dfrac{0{,}2}{\sqrt{0{,}4}} \\ \dfrac{0{,}1}{\sqrt{0{,}1}} & \dfrac{-0{,}6}{\sqrt{0{,}4}} \end{pmatrix} =$$

$$= \begin{pmatrix} \dfrac{0{,}09}{0{,}1} + \dfrac{0{,}01}{0{,}1} & \dfrac{0{,}06}{0{,}2} + \left(\dfrac{-0{,}06}{0{,}2}\right) \\ \dfrac{0{,}06}{0{,}2} + \left(\dfrac{-0{,}06}{0{,}2}\right) & \dfrac{0{,}04}{0{,}4} + \dfrac{0{,}036}{0{,}4} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \mathbf{E}.$$

Любая матрица перестановок является ортогональной. Отсюда следует, что единичная матрица также является ортогональной матрицей. Строки (столбцы) единичной матрицы являются единичными векторами:

$$\mathbf{e}_1 = (1, 0, 0, ..., 0),\ \mathbf{e}_2 = (0, 1, 0, ..., 0),\ ...,\ \mathbf{e}_n = (0, 0, 0, ..., 1).$$

Обратная и транспонированная матрицы ортогональной матрицы также являются ортогональными матрицами.

Если f_i и y_i представляют прямоугольные, или ортогональные, координаты, то преобразование (2.2) и матрица, определяемая им, называются ортогональными.

Из математики известно, что квадрат расстояния от точки до начала координат остается неизменным при замене координатных осей (ортогональном преобразовании). В матричной форме это заключение получается из свойств ортогонального преобразования

$$\left.\begin{aligned} \mathbf{y} &= \mathbf{Af} \\ \mathbf{f} &= \mathbf{A}'\mathbf{y} \end{aligned}\right\}.$$

Следовательно,

$$\mathbf{y}'\mathbf{y} = (\mathbf{f}'\mathbf{A}')(\mathbf{Af}) = \mathbf{f}'(\mathbf{A}'\mathbf{A})\mathbf{f} = \mathbf{f}'\mathbf{f}.$$

Любая ортогональная матрица размером 2×2 с определителем, равным +1, может быть записана в форме

$$\begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}.$$

Это преобразование координат от (f_1, f_2) к (y_1, y_2) представляет собой поворот на угол θ. Для проведения линейных преобразований исходной корреляционной матрицы в методе главных компонент необходимо воспользоваться собственными векторами и собственными значениями.

2.3. Собственные векторы и собственные значения матрицы

Собственным вектором матрицы **A**, соответствующим собственному значению (числу) λ, называется отличный от нуля вектор

$$\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{pmatrix},$$

удовлетворяющий уравнению

$$\mathbf{Ax} = \lambda\mathbf{x}. \qquad (2.4)$$

Матрицу **A** можно рассматривать как матрицу линейного преобразования, преобразующего собственный вектор **x** в вектор $\lambda\mathbf{x}$, который отличается от вектора **x** скалярным множителем λ. Можно уравнение (2.4) записать так:

$$(\mathbf{A} - \lambda\mathbf{E})\mathbf{x} = 0. \qquad (2.5)$$

Уравнение (2.5) является однородным и имеет ненулевое решение только тогда, когда характеристический многочлен равен нулю: $|\mathbf{A} - \lambda\mathbf{E}| = 0$, или в развернутом виде:

$$|\mathbf{A} - \lambda\mathbf{E}| = \begin{bmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22}-\lambda & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mm}-\lambda \end{bmatrix}. \qquad (2.6)$$

Данное выражение также называется характеристическим уравнением. Если корни этого уравнения λ_r все различны, то каждому из них соответствует характеристический вектор $\boldsymbol{x}_r$, определяемый с точностью до произвольного множителя, с системой линейных уравнений, эквивалентной формуле (2.6):

$$\left.\begin{array}{r} a_{11} - \lambda x_1 + a_{12}x_2 + \dots + a_{1m}x_m = 0; \\ a_{21}x_1 + (a_{22} - \lambda)x_2 + \dots + a_{2m}x_m = 0; \\ \dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + (a_{mm} - \lambda)x_m = 0. \end{array}\right\}$$

Для того чтобы эта система линейных однородных уравнений имела ненулевое решение, необходимо и достаточно, чтобы определитель ее (формула 2.6) был равен нулю. Покажем в развернутом виде квадратную матрицу **A** и матрицу $|\mathbf{A} - \lambda\mathbf{E}|$ порядка $m \times m$:

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mm} \end{pmatrix};$$

$$\mathbf{A} - \lambda\mathbf{E} = \begin{pmatrix} a_{11}-\lambda & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22}-\lambda & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mm}-\lambda \end{pmatrix},$$

где λ — независимая переменная;
E — единичная матрица;
$|\mathbf{A} - \lambda\mathbf{E}|$ — характеристическая матрица матрицы **A**.

Определитель матрицы $|\mathbf{A} - \lambda\mathbf{E}|$ является многочленом степени m относительно λ:

$$|\mathbf{A} - \lambda\mathbf{E}| = (-1)^m\lambda^m + (-1)^{m-1}p_1\lambda^{m-1} + \dots + p_m, \qquad (2.7)$$

где $p_1 = \mathrm{tr}\,\mathbf{A} = (a_{11} + a_{22} + \dots + a_{mm})$, $p_m = |\mathbf{A}|$.

Этот многочлен называется характеристическим многочленом матрицы **A**. Таким образом, первый, второй и последний члены определяются просто. Значительно труднее определять третий, четвертый и до $(m-1)$-го члена. Д.К. Фаддеев предложил следующие формулы [103]:

$$\mathbf{A}_1 = \mathbf{A};\ p_1 = \mathrm{tr}(\mathbf{A}_1);\ \mathbf{B}_1 = \mathbf{A}_1 - p_1\mathbf{E};$$

$$\mathbf{A}_2 = \mathbf{A}\mathbf{B}_1;\ p_2 = \frac{1}{2}\mathrm{tr}(\mathbf{A}_2);\ \mathbf{B}_2 = \mathbf{A}_2 - p_2\mathbf{E};$$

. . . .

$$\mathbf{A}_{m-1} = \mathbf{A}\mathbf{B}_{m-2};\ p_{m-1} = \frac{1}{m-1}\mathrm{tr}(\mathbf{A}_{m-1});\ \mathbf{B}_{m-1} = \mathbf{A}_{m-1} - p_{m-1}\mathbf{E};$$

$$\mathbf{A}_m = \mathbf{A}\mathbf{B}_{m-1};\ p_m = \frac{1}{m}\mathrm{tr}(\mathbf{A}_m);\ \mathbf{B}_m = \mathbf{A}_m - p_m\mathbf{E}.$$

Корни характеристического многочлена $|\mathbf{A} - \lambda\mathbf{E})$ называются характеристическими числами или собственными значениями матрицы **A**. Сумма собственных чисел матрицы **A** равна ее следу $\lambda_1 + \lambda_2 + \ldots + \lambda_m = \mathrm{tr}(\mathbf{A})$. Произведение собственных значений матрицы **A** $\lambda_1\lambda_2 \ldots \lambda_m = (-1)^m|\mathbf{A}|$. Каждому собственному значению соответствует свой собственный вектор матрицы **A**. Собственные значения вещественной симметричной матрицы вещественны. Собственные векторы, соответствующие различным собственным значениям вещественной симметричной матрицы, ортогональны. Вещественная симметричная матрица **A** может быть приведена к диагональной форме с помощью некоторого ортогонального преобразования.

Пример 2.5

Вычислите собственные значения и собственные векторы матрицы

$$\mathbf{A} = \begin{pmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{pmatrix},$$

приведенной в работе [40], и составьте матрицу собственных значений, пользуясь методом Фадеева [103].

Решение

1. Составим характеристическое уравнение

$$|\mathbf{A} - \lambda\mathbf{E}| = \begin{vmatrix} 7-\lambda & -2 & 0 \\ -2 & 6-\lambda & -2 \\ 0 & -2 & 5-\lambda \end{vmatrix} = 0.$$

2. Определим члены характеристического многочлена степени $m = 3$ относительно λ по формуле (2.7):

а) первый член $(-1)^3\lambda^3 = -\lambda^3$;

б) второй член $(-1)^{3-1}p_1\lambda^{3-1}$;

$$p_1 = \mathrm{tr}(\mathbf{A}) = 7 + 6 + 5 = 18;$$

$$(-1)^2 18\lambda^2 = 18\lambda^2;$$

в) третий член $(-1)^{3-2}p_2\lambda^{3-2}$;

$$p_2 = \frac{1}{2}\mathrm{tr}(\mathbf{A}_2);$$

$$\mathbf{A}_2 = \mathbf{A}\mathbf{B}_1;$$

$$\mathbf{B}_1 = \mathbf{A} - p_1\mathbf{E};$$

$$\mathbf{B}_1 = \begin{bmatrix} (7-18) & -2 & 0 \\ -2 & (6-18) & -2 \\ 0 & -2 & (5-18) \end{bmatrix} = \begin{bmatrix} -11 & -2 & 0 \\ -2 & -12 & -2 \\ 0 & -2 & -13 \end{bmatrix};$$

$$\mathbf{A}_2 = \mathbf{A}\mathbf{B}_1 = \begin{bmatrix} 7 & -2 & 0 \\ -2 & 6 & -2 \\ 0 & -2 & 5 \end{bmatrix} \cdot \begin{bmatrix} -11 & -2 & 0 \\ -2 & -12 & -2 \\ 0 & -2 & -13 \end{bmatrix} =$$

$$= \begin{bmatrix} -77+4+0 & -14+24+0 & 0+4+0 \\ 22-12+0 & 4-72+4 & 0-12+26 \\ 0+4+0 & 0+24-10 & 0+4-65 \end{bmatrix} = \begin{bmatrix} -73 & 10 & 4 \\ 10 & -64 & 14 \\ 4 & 14 & -61 \end{bmatrix};$$

$$p_2 = -\frac{1}{2}(73 + 64 + 61) = -99;$$

г) четвертый член $p_3 = p_m = |\mathbf{A}|$,

$$|\mathbf{A}| = a_{31}A_{31} + A_{32}a_{32} + a_{33}A_{33} =$$

$$= 0 + (-1)(-2)\begin{vmatrix} 7 & 0 \\ -2 & -2 \end{vmatrix} + 1{,}5\begin{vmatrix} 7 & -2 \\ -2 & 6 \end{vmatrix} = -28 + 190 = 162.$$

Итак, получены все члены характеристического многочлена:

$$\lambda^3 - 18\lambda^2 + 99\lambda - 162 = 0.$$

3. Определим корни характеристического уравнения, т. е. получим собственные значения матрицы **A**:

$$\lambda_1 = 3,\ \lambda_2 = 6 \text{ и } \lambda_3 = 9.$$

4. Определим собственный вектор, соответствующий $\lambda_3 = 9$. Составим систему уравнений:

$$\left.\begin{aligned} (7-9)x_1 - 2x_2 + 0x_3 &= 0; \\ -2x_1 + (6-9)x_2 - 2x_3 &= 0; \\ 0x_1 - 2x_2 + (5-9)x_3 &= 0. \end{aligned}\right\}$$

Решение этой системы позволяет определить следующие соотношения: $x_1 = -2$, $x_2 = 2$, $x_3 = -1$. Собственный вектор:

$$\mathbf{x}_3 = c\begin{pmatrix} -2 \\ 2 \\ -1 \end{pmatrix}.$$

5. Определим собственный вектор, соответствующий собственному значению $\lambda_2 = 6$. Составим систему уравнений:

$$\left.\begin{aligned} (7-6)x_1 - 2x_2 + 0x_3 &= 0; \\ -2x_1 + (6-6)x_2 - 2x_3 &= 0; \\ 0x_1 - 2x_2 + (5-6)x_3 &= 0. \end{aligned}\right\}$$

Получаем $x_1 = 2$, $x_2 = 1$, $x_3 = -2$. Собственный вектор

$$\mathbf{x}_2 = c\begin{pmatrix} 2 \\ 1 \\ -2 \end{pmatrix}.$$

6. Определим собственный вектор, соответствующий собственному значению $\lambda_1 = 3$. Составим уравнения:

$$\left.\begin{aligned} (7-3)x_1 - 2x_2 + 0x_3 &= 0; \\ -2x_1 + (6-3)x_2 - 2x_3 &= 0; \\ 0x_1 - 2x_2 + (5-3)x_3 &= 0. \end{aligned}\right\}$$

Получим $x_1 = 1$, $x_2 = 2$, $x_3 = 2$. Собственный вектор

$$\mathbf{x}_1 = c\begin{pmatrix} 1 \\ 2 \\ 2 \end{pmatrix}.$$

7. Матрица собственных значений имеет вид:

$$\Lambda = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

2.4. Квадратичные формы и главные компоненты

Для того чтобы представить в геометрическом плане главные компоненты, рассмотрим простейшие случаи: на плоскости и в трехмерном пространстве. Пусть дано уравнение линии второго порядка

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 = p. \qquad (2.8)$$

Левая часть уравнения не меняется, если x, y заменить на $-x$, $-y$. Значит, точки линии расположены парами симметрично относительно начала координат. *Линия второго порядка, заданная уравнением (2.8), во-первых, обладает центром (симметрии), и начало координат, во-вторых, находится в этом центре.*

Левая часть уравнения (2.8) представляет собой однородный многочлен второй степени. Многочлен называется *квадратичной формой от двух переменных* x и y. Приведем эту квадратичную форму к каноническому виду. Для этого надо повернуть координатные оси x и y так, чтобы в новых координатах исчез член с произведением новых текущих координат. Переход к новым координатам производится по известным формулам:

$$\left.\begin{aligned} x' &= l_1x + m_1y; \\ y' &= l_2x + m_2y. \end{aligned}\right\}.$$

Обратные преобразования связаны с обратной матрицей, но в ортогональном преобразовании транспонированная матрица совпадает с обратной матрицей, поэтому связь старых координат с новыми выражается так:

$$\left.\begin{aligned} x &= l_1x' + l_2y'; \\ y &= m_1x' + m_2y'. \end{aligned}\right\} \tag{2.9}$$

На рис. 2.1 показаны коэффициенты при старых координатах. На новой оси абсцисс отложен отрезок $\mathbf{x}_1'$ единичной длины. Его проекции на старые координатные оси составят $l_1 = \cos\alpha$; $m_1 = \sin\alpha$, где α — угол поворота осей x и y. Значит, вектор с компонентами l_1 и m_1 является единичным вектором, определяющим направление новой оси абсцисс x'; $\{l_1; m_1\} = \{\cos\alpha; \sin\alpha\}$. Единичный вектор, определяющий направление новой оси ординат, имеет вид $\{l_2; m_2\} = \{-\sin\alpha; \cos\alpha\}$.

Рассматриваемые коэффициенты обладают следующими свойствами:

$$l_1^2 + m_1^2 = 1; \tag{2.10}$$

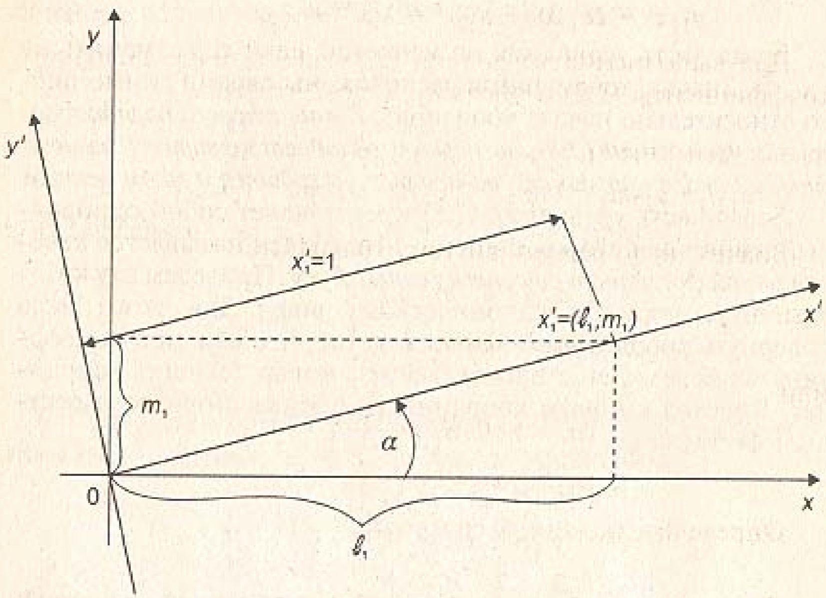


Рис. 2.1. Поворот осей координат

$$l_2^2 + m_2^2 = 1; \tag{2.11}$$

$$l_2l_2 + m_1m_2 = 0; \tag{2.12}$$

$$\Delta = \begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix} = 1. \tag{2.13}$$

Если ось x повернуть на угол α, а ось y —на угол $\alpha + \pi$, то нарушается ориентация осей, но при этом формулы (2.10)—(2.12) справедливы, а вместо формулы (2.13) получается определитель

$$\Delta = \begin{vmatrix} l_1 & m_1 \\ l_2 & m_2 \end{vmatrix} = -1.$$

Таким образом может быть совершен поворот осей прямоугольных координат с неизменным масштабом. Квадратичную форму можно привести к каноническому виду (средний коэффициент равен нулю), если по формуле перейти к новым координатам:

$$a_{11}x^2 + 2a_{12}xy + a_{22}y^2 = \lambda_1x'^2 + \lambda_2y'^2. \tag{2.14}$$

Для выполнения уравнения (2.14) достаточно подобрать коэффициенты (2.2) и числа λ_1, λ_2 так, чтобы

$$\left.\begin{aligned} a_{11}l_1 + a_{12}m_1 &= \lambda_1l_1; \quad & a_{11}l_2 + a_{12}m_2 &= \lambda_2l_2; \\ a_{12}l_1 + a_{22}m_1 &= \lambda_1m_1; \quad & a_{12}l_2 + a_{22}m_2 &= \lambda_2m_2. \end{aligned}\right\}$$

Значит, надо решить систему уравнений

$$\left.\begin{aligned} a_{11}l + a_{12}m &= \lambda l; \\ a_{12}l + a_{22}m &= \lambda m, \end{aligned}\right\}$$

или

$$\left.\begin{aligned} (a_{11} - \lambda)l + a_{12}m &= 0; \\ a_{12}l + (a_{22} - \lambda)m &= 0. \end{aligned}\right\} \tag{2.15}$$

Определитель данной системы

$$\begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{12} & a_{22} - \lambda \end{vmatrix} = 0. \tag{2.16}$$

Определим $\lambda_{1,2}$ из уравнения

$$\lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}^2) = 0;$$

$$\lambda_{1,2} = \frac{1}{2}[(a_{11} + a_{22}) \pm \sqrt{(a_{11}+a_{22})^2 - 4(a_{11}a_{22} - a_{12}^2)}]. \qquad (2.17)$$

Уравнение (2.16) является характеристическим уравнением квадратичной формы, а корни этого уравнения λ_1 и λ_2 являются характеристическими числами этой формы. После приведения формы к каноническому виду λ_1 и λ_2 являются коэффициентами при неизвестных.

Так как выражение под радикалом неотрицательно, т. е. $(a_{11} - a_{22})^2 + 4a_{12}^2 \geqslant 0$, то уравнение (2.16) имеет только действительные корни.

Отдельно рассмотрим случай, когда $(a_{11} - a_{22})^2 + 4a_{12}^2 > 0$. При этом условии $\lambda_1 \neq \lambda_2$. Подставим в уравнения (2.15) $\lambda = \lambda_1$, система будет иметь ненулевое решение l и m. Полученный вектор будет направлен по главному направлению квадратичной формы, которое соответствует характеристическому числу λ_1. С этим же главным направлением совпадает и вектор $\{l_1; m_1\}$, $l_1 = \mu l$, $m_1 = \mu m$ $(\mu \neq 0)$. Если примем, что $\mu\sqrt{l^2 + m^2} = \pm 1$, то $l_1^2 + m_1^2 = 1$. Вектор $\{l_1; m_1\}$ является единичным вектором главного направления. Естественно, что вектор $\{l_2; m_2\}$ определяет другое главное направление квадратичной формы. Согласно формуле (2.12), если $\lambda_1 \neq \lambda_2$, векторы главных направлений взаимно перпендикулярны.

Другой случай соответствует $(a_{11} - a_{22})^2 + 4a_{12}^2 = 0$. При этом $\lambda_1 = \lambda_2 = \lambda$; $a_{11} = a_{22}$; $a_{12} = 0$. По уравнению (2.17) $\lambda_1 = a_{11} = a_{22}$. Подставим в систему (2.15) полученное значение λ и убедимся в том, что все коэффициенты системы обращаются в нуль. Таким образом, система (2.15) будет состоять из тождеств. Ей подходят любые числа l и m.

В результате можно заключить, что если $\lambda_1 = \lambda_2$, то для квадратичной формы любое направление является главным. При повороте осей на любой угол форма сохранит свой канонический вид $a_{11}x^2 + a_{11}y^2$.

При любом преобразовании квадратичной формы к любым прямоугольным координатам ее инварианты не меняются:

$$\left.\begin{aligned} a_{11} + a_{22} &= a'_{11} + a'_{22}; \\ a_{11}a_{22} - a_{12}^2 &= a'_{11}a'_{22} - a'^2_{12}. \end{aligned}\right\}$$

Согласно теореме Виета, $a_{11}a_{22} - a_{12}^2 = \lambda_1\lambda_2$.

1. Если $\lambda_1 \neq 0$; $\lambda_2 \neq 0$ и имеют одинаковые знаки, то квадратичная форма называется **эллиптической** $a_{11}a_{22} - a_{12}^2 > 0$.

2. Если $\lambda_1 \neq 0$; $\lambda_2 \neq 0$, но знаки у них разные, то форма называется **гиперболической** $a_{11}a_{22} - a_{12}^2 < 0$.

3. Если одно из чисел λ_1, λ_2 равно нулю, т. е. $a_{11}a_{22} - a_{12}^2 = 0$, то форма называется **параболической**.

В методе главных компонент характеристические числа по своему физическому смыслу не могут равняться нулю и не могут быть отрицательными. Значит, $\lambda_1 > 0$ и $\lambda_2 > 0$. Квадратичная форма будет называться в этом смысле положительно определенной эллиптической формой.

На рис. 2.2 показан переход от произвольной системы координат с началом координат в центре эллипса и поворот осей, осуществленный для приведения квадратичной формы к каноническому виду.

После приведения к каноническому виду ось абсцисс, соответствующая λ_1, направлена по одной главной оси эллипса (главному направлению), а ось ординат, соответствующая λ_2, — перпендикулярно к ней, по другой главной оси эллипса (другому главному направлению).

В качестве примера обратимся к геометрической интерпретации плотности вероятностей для двумерного нормального распределения. Плотность вероятностей n-мерного случайного вектора (2.15) в n-мерном евклидовом пространстве постоянна на эллипсоидах, определяемых положительно определенными квадратичными формами для каждого положительного значения p, $(\mathbf{x} - \mu)'\Sigma^{-1}(\mathbf{x} - \mu) = p$.

Центром каждого эллипсоида является точка μ. Форма и положение эллипсоида определяются значениями ковариа-

4—1447

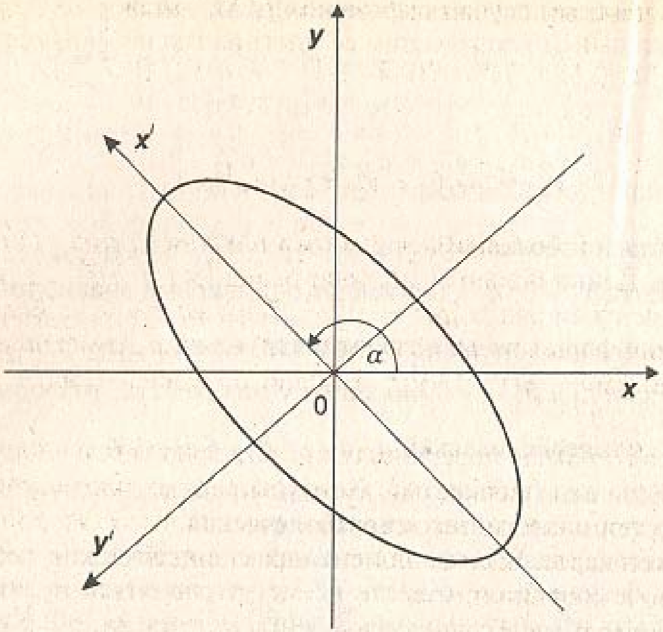


Рис. 2.2. Геометрическая интерпретация приведения квадратичной формы к каноническому виду, если начало координат в центре эллипса

ционной матрицы. Если ковариационная матрица получена на основании обработки конкретных статистических данных, то размеры эллипсоида определятся конкретным значением величины p.

Для удобства в двумерном случае введем нормированные отклонения от математического ожидания

$$y_i = (x_i - \mu_i)/\sigma_i.$$

В результате центры линий, на которых плотность распределения постоянна, оказались в начале координат. Если воспользоваться формулой (1.17), то в новых обозначениях можно записать:

$$\frac{1}{1-\rho^2}(y_1^2 - 2\rho y_1 y_2 + y_2^2) = p.$$

Это частный случай выражения (2.8), когда

$$a_{11} = a_{22} = 1/(1 - \rho^2);\ a_{12} = -\rho/(1 - \rho^2);\ x = y_1;\ y = y_2,$$

или

$$y_1^2 - 2\rho y_1 y_2 + y_2^2 = (1 - \rho^2)p.$$

При $\rho > 0$ большая ось эллипса наклонена под углом 45° к оси x. Длина большой полуоси равна $[p(1 + \rho)]^{1/2}$, а длина малой полуоси равна $[p(1 - \rho)]^{1/2}$. Если же $\rho < 0$, то большая ось эллипса наклонена под углом 135° к оси x. Длина большой полуоси равна $[p(1 - \rho)]^{1/2}$, а малой полуоси — $[p(1 + \rho)]^{1/2}$ (см. рис. 2.2).

Вероятностная поверхность представляет собой «нормальный холм» над плоскостью. Контуры равных плотностей вероятностей аналогичны контурам равных высот на топографической карте. Если от эллипсоида в координатах y_i перейти обратно к координатам $x_i = \sigma_i y_i + \mu_i$, то это означает, что каждый контур растягивается в σ_i раз в направлении i-й оси, а центр переносится в точку (μ_1, μ_2).

Для двумерных случайных величин имеются таблицы, по которым определяются плотности вероятностей.

Рассмотрим некоторые примеры квадратичных форм и их представление в матричной форме.

Пример 2.6

Дана функция

$$f(\mathbf{x}) = x_1^2 - 4x_1x_2 + x_1x_3 - 2x_2x_3 - x_3^2.$$

Определите матрицу **A** и представьте функцию в матричном виде.

Решение

1.

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 0{,}5 \\ -2 & 0 & -1 \\ 0{,}5 & -1 & -1 \end{pmatrix}.$$

2.

$$f(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x};$$

$$f(\mathbf{x}) = (x_1, x_2, x_3)\begin{pmatrix} 1 & -2 & 0{,}5 \\ -2 & 0 & -1 \\ 0{,}5 & -1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Пример 2.7

Представьте функционал, имеющий линейную часть, в матричном виде и проверьте его соответствие исходному:

$$f(\mathbf{x}) = 4x_1 + 5x_2 + 2x_1x_2 - 2x_1^2 - 2x_2^2.$$

Решение

1.

$$f(\mathbf{x}) = \mathbf{a}'\mathbf{x} + \mathbf{x}'\mathbf{A}\mathbf{x};\ \mathbf{a}' = (4{,}5);$$

$$f(\mathbf{x}) = (4{,}5)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1, x_2)\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

2.

$$f(\mathbf{x}) = (4{,}5)\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + (x_1, x_2)\begin{pmatrix} -2 & 1 \\ 1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$= 4x_1 + 5x_2 + ((-2x_1 + x_2), (x_1 - 2x_2))\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} =$$

$$= 4x_1 + 5x_2 + (-2x_1^2 + x_1x_2 + x_1x_2 - 2x_2^2) =$$

$$= 4x_1 + 5x_2 + 2x_1x_2 - 2x_1^2 - 2x_2^2.$$

Таким образом, в случае многомерного нормального распределения о характере и тесноте стохастической связи между признаками можно судить по ковариационной матрице, элементы которой могут быть оценены.

Всестороннее изучение ковариационной или корреляционной матрицы и ее преобразование к новой системе координат позволяет вскрыть неявные, но объективно существующие закономерности.

2.5. Главные компоненты в трехмерном и конечномерном пространстве

Введем некоторые известные определения.

Множество всех n-мерных векторов, для которого определены операции сложения и умножения вектора на число, называется n-мерным векторным пространством.

Трехмерное пространство можно легко себе представить. Четырехмерное, ..., n-мерное пространство — это уже не пространство в обычном понимании этого слова. Это множество n-мерных векторов. Пространством называют его по аналогии с обычным трехмерным пространством. Определения и свойства векторов в трехмерном и n-мерном пространствах аналогичны.

В n-мерном пространстве любая система, содержащая более чем n векторов, линейно зависима. В n-мерном пространстве любая линейно-независимая система может содержать не более n векторов.

Наибольшее число линейно-независимых векторов системы называется ее рангом. Ранг n-мерного пространства совпадает с его размерностью.

Базисом множества векторов называется любой набор линейно-независимых векторов, число которых равно рангу множества.

В n-мерном пространстве может быть бесчисленное множество базисов. Один из базисов будет единичным (состоящим из единичных векторов).

Любой вектор данного n-мерного пространства может быть представлен в виде линейной комбинации базисных векторов.

Представление вектора в виде линейной комбинации базисных векторов называется разложением вектора по базису. Коэффициенты линейной комбинации называются координатами вектора в данном базисе. *Разложение вектора по данному базису единственно.*

Рассмотрим разложение вектора, чаще всего используемое в методе главных компонент (рис. 2.3).

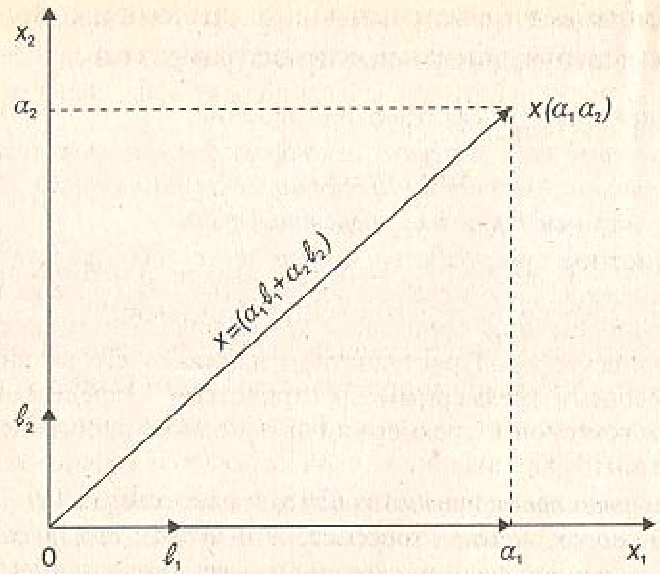


Рис. 2.3. Разложение вектора **x** по векторам ортогонального двумерного базиса

Вектор **x** разложим по векторам ортогонального двумерного базиса

$$\mathbf{x} = \alpha_1 l_1 + \alpha_2 l_2,$$

где α_1 и α_2 — коэффициенты разложения или координаты вектора **x** в базисе $\mathbf{e}_1$ и $\mathbf{e}_2$;

l_1 и l_2 — единичные векторы.

Ознакомившись с основными определениями конечномерного пространства, остановимся на линейных преобразованиях.

Преобразованием линейного пространства S называется закон, ставящий каждому вектору из S в соответствие некоторый вполне определенный вектор из S. Если при данном преобразовании произведение числа на вектор переводится в произведение того же числа на соответственный вектор, а сумма векторов переводится в сумму соответственных векторов, то данное преобразование называется линейным.

В принятых символах можно записать, что преобразование **A** называется линейным, если для любых векторов $\mathbf{x}_1$ и $\mathbf{x}_2$ пространства и любого числа α из поля коэффициентов имеют место равенства:

$$\mathbf{A}(\alpha\mathbf{x}) = \alpha(\mathbf{Ax}); \quad (2.18)$$

$$\mathbf{A}(\mathbf{x} + \mathbf{y}) = \mathbf{Ax} + \mathbf{Ay}. \quad (2.19)$$

Если $\alpha = 0$, то $\mathbf{A} \cdot 0 \cdot \mathbf{x} = 0$, т. е. всякое линейное преобразование переводит нулевой вектор в нулевой.

Возвратимся к интересующей нас задаче преобразования координат в трехмерном пространстве. Задан вектор **x**, который надо перевести в вектор **x**′. Вектор **x**′ называется образом вектора **x**. Преобразование $\mathbf{x}' = \mathbf{Ax}$ называется линейным, если соблюдены условия равенства (2.18) и равенства (2.19). Запишем общее уравнение второй степени относительно x, y, z, применяя для коэффициентов при неизвестных одну букву с двумя числовыми индексами:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz +$$
$$+ 2a_{23}yz + 2a_{14}x + 2a_{24}y + 2a_{34}z + a_{44} = 0.$$

Из этого общего уравнения рассмотрим неполное уравнение второй степени — квадратичную форму от трех переменных:

$$a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz + 2a_{23}yz = p. \quad (2.20)$$

В нем нет членов первой степени и левая часть не меняется при замене x, y, z на $-x$, $-y$, $-z$. Точки поверхности расположены парами симметрично относительно начала координат.

Таким образом, поверхность второго порядка, описываемая равенством (2.20), имеет в центре начало координат и обладает центром симметрии. Приведем форму к каноническому виду. Следует произвести такое ортогональное преобразование координат, при котором исчезли бы все члены с произведением новых текущих координат.

Как известно, надо найти

$$
\left.
\begin{aligned}
x &= l_1x' + l_2y' + l_3z'; \\
y &= m_1x' + m_2y' + m_3z'; \\
z &= n_1x' + n_2y' + n_3z'.
\end{aligned}
\right\}.
$$

Из данных формул следует тождество

$$
a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{13}xz +
$$
$$
+ 2a_{23}yz = \lambda_1 x'^2 + \lambda_2 y'^2 + \lambda_3 z'^2.
$$

Решение по схеме для случая двух переменных приводит к характеристическому определителю

$$
\begin{vmatrix} a_{11}-\lambda & a_{12} & a_{13} \\ a_{21} & a_{22}-\lambda & a_{23} \\ a_{31} & a_{32} & a_{33}-\lambda \end{vmatrix} = 0.
$$

Отметим, что характеристический многочлен матрицы линейного преобразования **A** не зависит от выбора базиса. После развертывания определителя получим характеристический многочлен, где наивысшая степень λ будет равна трем. Решив его, найдем характеристические корни данного уравнения.

Нас интересует случай, когда $\lambda_1 > 0$; $\lambda_2 > 0$; $\lambda_3 > 0$. Значит, квадратичная форма данного уравнения будет эллиптической. При этом $p > 0$ и в пространстве получим поверхность второго порядка — эллипсоид. У данного эллипсоида три главных направления, соответствующие трем значениям собственных чисел λ_1, λ_2 и λ_3, совпадают с его главными осями.

Если будем осуществлять линейные преобразования в n-мерном пространстве, то получим некоторые гиперповерхности, главные оси которых будут совпадать с главными направлениями после приведения их к каноническому виду. Все главные направления взаимно перпендикулярны.

Всякому собственному числу λ_j соответствует свой собственный вектор, совпадающий с j-м главным направлением.

Таким образом, при решении задачи методом главных компонент мы переходим к новой системе взаимно перпендикулярных осей, при которой взаимно перпендикулярные главные направления совпадают соответственно с этой новой системой координат.

2.6. Главные компоненты и ортогональная регрессия

При построении прямой линии регрессии для двух признаков, например X_2 относительно X_1, расстояния от наблюдаемых точек до прямой регрессии измеряются по линиям, параллельным осям координат x_1 и x_2. В отличие от простой регрессии будем измерять расстояние от точки до выравнивающей прямой по перпендикуляру к этой прямой. Рассмотрим, что собой представляют главные компоненты, используя схему [61].

Пусть уравнение прямой

$$
a_1x_1 + a_2x_2 = b,
$$

где a_1 и a_2 — соответственно косинус и синус угла, который образуется между осью абсцисс и нормалью к прямой;
b — расстояние от начала координат до выравнивающей прямой.

Тогда можно будет записать:

$$
a_1^2 + a_2^2 = 1. \tag{2.21}
$$

Запишем сумму квадратов расстояний точек от выравнивающей прямой:

$$
u = \sum_{k=1}^{N} (a_1x_{1k} + a_2x_{2k} - b)^2. \tag{2.22}
$$

Определим коэффициенты выравнивающей прямой a'_1, b', a'_2. Для этого надо из выражения (2.22) вычесть произведение

$$
\lambda(a_1^2 + a_2^2 - 1),
$$

где λ — множитель Лагранжа.

Далее надо взять производные по a_1, a_2 и b от полученного выражения

$$
\sum_{k=1}^{N} (a_1x_{1k} + a_2x_{2k} - b)^2 - \lambda(a_1 + a_2^2 - 1).
$$

Полученные производные приравнивают к нулю. В результате получим при дифференцировании по b:

$$b' = a'_1\bar{x}_1 + a'_2\bar{x}_2,$$

где

$$\bar{x}_1 = \frac{1}{N}\sum_{k=1}^{N} x_{1k}; \qquad \bar{x}_2 = \frac{1}{N}\sum_{k=1}^{N} x_{2k}.$$

Итак, выравнивающая прямая проходит через среднюю точку $(\bar{x}_1, \bar{x}_2)$:

$$a'_1(x - \bar{x}_1) + a'_2(x_2 - \bar{x}_2) = 0.$$

После дифференцирования по a_1 и a_2 и исключения b' получим:

$$a'_1[\sum_{k=1}^{N}(x_{1k} - \bar{x}_1)^2 - \lambda] + a'_2 \sum_{k=1}^{N}(x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) = 0; \quad (2.23)$$

$$a'_1 \sum_{k=1}^{N}(x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2) + a'_2[\sum_{k=1}^{N}(x_{2k} - \bar{x}_2)^2 - \lambda] = 0. \quad (2.24)$$

Перепишем эти выражения с введением обозначений центральных моментов второго порядка m_{11} и m_{22} и центрального смешанного момента второго порядка m_{12} для x_{1k} и x_{2k}:

$$\left.\begin{aligned} m_{11} &= \frac{1}{N}\sum_{k=1}^{N}(x_{1k} - \bar{x}_1)^2; \\ m_{22} &= \frac{1}{N}\sum_{k=1}^{N}(x_{2k} - \bar{x}_2)^2; \\ m_{12} &= \frac{1}{N}\sum_{k=1}^{N}(x_{1k} - \bar{x}_1)(x_{2k} - \bar{x}_2). \end{aligned}\right\}$$

Тогда можно переписать уравнения (2.23) и (2.24):

$$\left.\begin{aligned} a'_1\,(m_{11} - (\lambda/N)) + a'_2\, m_{12} &= 0; \\ a'_1\, m_{12} + a'_2\,(m_{22} - (\lambda/N)) &= 0. \end{aligned}\right\}$$

Эти выражения представляют собой линейную однородную систему алгебраических уравнений. Она не имеет нулевого решения только тогда, когда ее определитель равен нулю:

$$\begin{vmatrix} m_{11} - (\lambda/N) & m_{12} \\ m_{12} & m_{22} - (\lambda/N) \end{vmatrix} = 0. \quad (2.25)$$

При условии (2.21) придаем λ значение наименьшего корня уравнения (2.25). При этом минимизируемое значение u формулы (2.22) определяется как $u = \lambda$. Значит, наименьший корень уравнения (2.25) определяет минимальную суммарную дисперсию относительно выравнивающей линии.

Таким образом, ортогональная регрессия соответствует главной компоненте облака точек наблюдения двух величин X_1 и X_2. Главная компонента множества наблюдений двух величин является линейной формой:

$$v = f_1(x_1 - \bar{x}_1) + f_2(x_2 - \bar{x}_2).$$

Она удовлетворяет условию $f_1^2 + f_2^2 = 1$. У этой линейной формы среднее квадратическое по множеству наблюдений является максимальным.

Таким образом, *прямая, минимизирующая сумму квадратов расстояний* (u_k), *совпадает с прямой, максимизирующей дисперсию по множеству наблюдений. Эта прямая является прямой ортогональной регрессии — главной компонентой множества наблюдений двух величин* X_1 и X_2 [61].

Следовательно,

$$f_1 = a'_2; \; f_2 = -a'_1; \; (a'_2)^2 + (-a'_1)^2 = 1.$$

Сумма λ_1 и λ_2, найденных из уравнения (2.25), равна:

$$\lambda_1 + \lambda_2 = \sum_{k=1}^{N}(x_{1k} - \bar{x}_1)^2 + \sum_{k=1}^{N}(x_{2k} - \bar{x}_2)^2 = u + v.$$

Рассмотрим на рис. 2.4 одну k-ю точку из множества точек, составляющих облако наблюдений двух величин.

Средняя точка $(\bar{x}_1, \bar{x}_2)$ помещена таким образом, что $x_1 = 0$, а $x_2 = b$. Сумма

$$u_k^2 + v_k^2 = (x_{1k} - \bar{x}_1)^2 + (x_{2k} - \bar{x}_2)^2$$

не зависит от положения выравнивающей прямой.

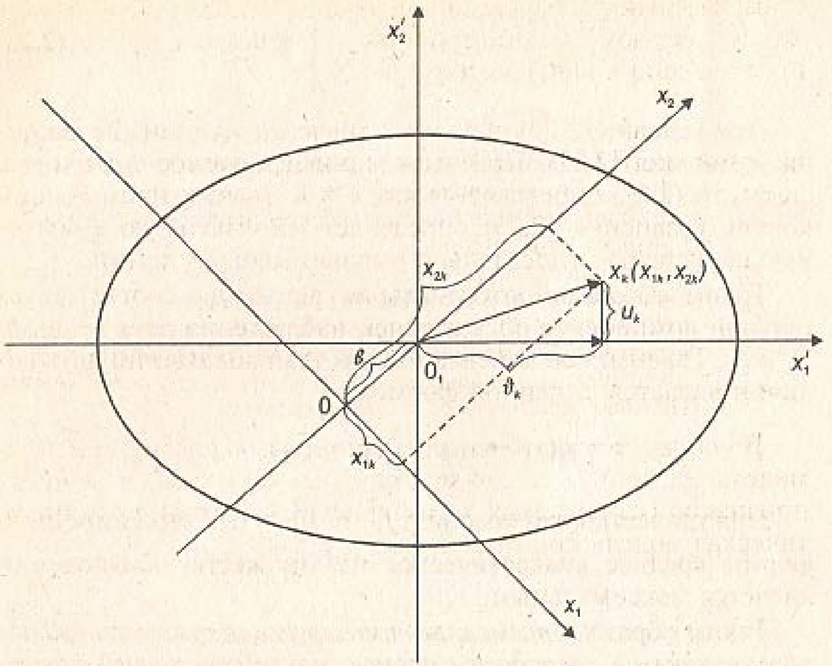


Рис. 2.4. Эллипс, построенный на облаке точек наблюдения

Таким образом, λ_1 и λ_2 являются корнями характеристического определителя. Им соответствуют ортогональные главные компоненты, совпадающие с главными осями эллипса, построенного на облаке точек наблюдения.

2.7. Математическая модель метода главных компонент

Истинная величина изучаемого явления содержит, по крайней мере, два компонента:

- истинную характеристику оцениваемого явления;
- ошибку измерения, которая зависит от большого числа причин.

Если измерения проводятся в психологии, биологии или экономике, то добавляется третья составляющая, зависящая

от вариабельности изучаемого признака, индивида или объекта. Таким образом, зарегистрированное значение может быть представлено в виде суммы

$$x_{ji} = x_{mi} + x_{si} + x_{ei},$$

где x_{ji} — зарегистрированное значение измеряемого признака (параметра) у i-го индивида (объекта исследования);
x_{mi} — истинное значение (математическое ожидание) измеряемого признака у i-го индивида;
x_{si} — вариативное значение измеряемого признака у i-го индивида;
x_{ei} — ошибка измерения при определении j-го признака у j-го объекта исследования. Так как ошибка измерения значительно меньше вариативной части признака, то ее часто объединяют с частью, зависящей от индивидуальных, специфических особенностей. При необходимости ее можно выделить.

В основу метода главных компонент положена линейная модель. Если N — число исследуемых объектов, n — число признаков (измеряемых характеристик объекта), то математическая модель принимает вид:

$$y'_j = \sum_{r=1}^{n} a_{jr} f_r, \tag{2.26}$$

где y'_j — нормированное значение j-го признака, полученное из модели (y_j — нормированное значение j-го признака, полученное из эксперимента, на основе наблюдений);
a_{jr} — вес r-й компоненты в j-й переменной;
f_r — r-я главная компонента;
$r, j = 1, 2, ..., n$.

В матричной форме

$$\mathbf{y} = \mathbf{Af}. \tag{2.27}$$

Для исследования начальными основными данными являются ковариации или коэффициенты корреляции. В дальнейшем будем пользоваться коэффициентами корреляции.

Для установления связи между главными компонентами и коэффициентами корреляции перепишем формулу (2.26) для любого изучаемого объекта i:

$$y_{ji} = a_{j1} f_{1i} + a_{j2} f_{2i} + \cdots + a_{jn} f_{ni}, \qquad i = 1, 2, ..., N,$$

где y_{ji} — нормированное значение j-го признака для i-го объекта;
f_{1i} — значение первой главной компоненты для i-го индивида;
N — число исследуемых объектов (индивидов);
f_{ni} — значение n главной компоненты для i-го индивида;

a_{j1} — вес первой компоненты в j-й переменной;
a_{jn} — вес n-й компоненты в j-й переменной.

Вариабельность, зависящая от особенностей объектов, является причиной разброса показаний признаков от объекта к объекту относительно математического ожидания. Полная дисперсия статистического признака выражается через дисперсию главных компонент [21]:

$$\sigma_j^2 = \frac{1}{N}\sum_{i=1}^{N} y_{ji}^2 = \frac{1}{N}\left[a_{j1}^2\sum_{i=1}^{N} f_{1i}^2 + a_{j2}^2\sum_{i=1}^{N} f_{2i}^2 + \ldots + a_{jn}\sum_{i=1}^{N} f_{ni} + \right.$$

$$\left. + 2\left(a_{j1}a_{j2}\sum_{i=1}^{N} f_{1i}f_{2i} + a_{j1}a_{j3}\sum_{i=1}^{N} f_{1i}f_{3i} + \ldots + a_{j(n-1)}a_{jn}\sum_{i=1}^{N} f_{(n-1)i}f_{ni}\right)\right].$$

Так как дисперсии нормированных величин равны единице, то можно записать:

$$\sigma_j^2 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jn}^2 + 2(a_{j1}a_{j2}r_{f_1f_2} +$$

$$+ a_{j1}a_{j3}r_{f_1f_3} + \ldots + a_{j(n-1)}a_{jn}r_{f_{(n-1)}f_n}) = 1.$$

Поскольку главные компоненты ортогональны, то выражение упрощается:

$$\sigma_j^2 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jn}^2 = 1.$$

Слева записана полная дисперсия, а справа — доли полной дисперсии, относящиеся к соответствующим главным компонентам. Дисперсия является характеристикой изменчивости случайной величины, ее отклонений от среднего значения. Полный вклад r-го фактора в дисперсию всех n признаков определяет ту долю общей дисперсии, которую данная главная компонента объясняет. Этот вклад вычисляется по формуле

$$V_r = \sum_{j=1}^{n} a_{jr}^2, \tag{2.28}$$

где j — индекс признака;
r — индекс главной компоненты.

Различают генеральные главные компоненты и общие. Генеральные главные компоненты связаны со всеми признака-

Таблица 2.1

Матрица — модель коэффициентов веса в методе главных компонент

Признак	Коэффициенты веса			
	a_1	a_2	a_3	a_4
y_1	0,26	0,00	0,00	−0,70
y_2	0,30	0,00	0,00	−0,44
y_3	0,43	−0,65	0,00	0,00
y_4	0,28	−0,64	0,00	0,00
y_5	0,34	0,00	0,86	0,00
y_6	0,24	−0,67	0,00	0,24
y_7	0,40	0,51	0,00	0,00
y_8	−0,60	−0,30	0,00	0,00
y_9	0,59	0,00	−0,46	0,00
y_{10}	0,57	0,00	0,00	0,45
y_{11}	0,63	0,00	−0,48	0,00

ми задачи, общие главные компоненты связаны более чем с одним признаком.

В табл. 2.1 приведены четыре главные компоненты, полученные при исследовании влияния подготовки специалистов на их работу на производстве.

Как видно из табл. 2.1, первая главная компонента имеет весовые коэффициенты, связывающие ее с каждым из признаков. Она связана со всеми переменными, поэтому назовем ее генеральной компонентой. Остальные три компоненты связаны только с некоторыми переменными, поэтому их надо отнести к групповым.

Представим результаты решения графически. На рис. 2.5*а* обозначены главные компоненты $\mathbf{f}_r$, а весовые коэффициенты проставлены в клеточках внутри компоненты и указывают на тесноту связи между данным признаком и данной главной компонентой.

Итак, все признаки, за исключением одного, оказались связанными с двумя главными компонентами, а признак y_6 —

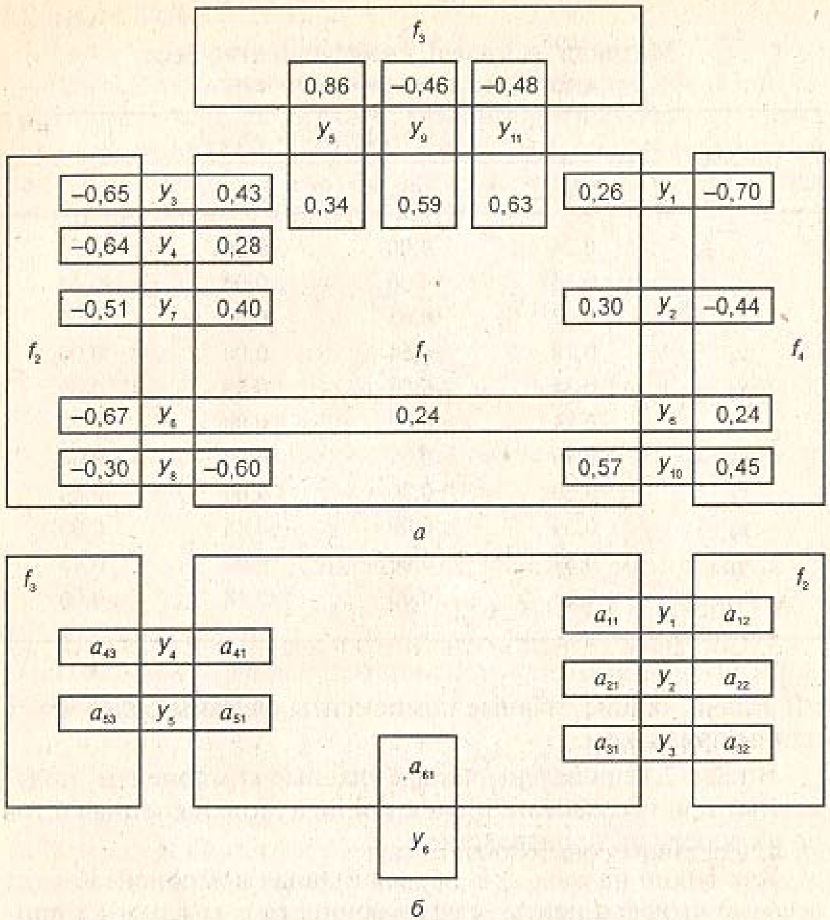


Рис. 2.5. Графическая схема результатов, полученных методом главных компонент:

а — значение главных компонент;

б — общее отображение связей между главными компонентами.

с тремя главными компонентами. В другом случае может оказаться, что некоторый признак связан только с одной главной компонентой, как, например, y_6 на рис. 2.5*б*.

Несмотря на то что вместо *n* признаков получено такое же количество главных компонент, вклад большей части главных компонент в объясняемую дисперсию оказывается небольшим. Исключаем из рассмотрения те главные компоненты, вклад которых мал. Оказывается, что при помощи *m* первых (наиболее весомых) главных компонент можно объяснить основную часть суммарной дисперсии.

Опыт исследования в различных областях показал, что число наиболее весомых компонент часто составляет 10—25% числа признаков. В зависимости от конкретных задач решается вопрос о том, сколько и каких компонент следует оставить для дальнейшего исследования: одну, две, 25% общего числа извлеченных компонент или более. В связи с этим часто используется понятие объясняемой групповой дисперсии, при помощи которой проводится содержательная интерпретация полученных результатов:

$$h_j^2 = a_{j1}^2 + a_{j2}^2 + \ldots + a_{jm}^2,$$

где *m* —число используемых главных компонент ($m < n$).

Таким образом, можно сделать следующие **выводы**:

- метод позволяет описать большие наборы *n* признаков небольшим числом *m* главных компонент ($m < n$);
- различия между объектами (индивидами) зависят от доли изменчивости, связанной с данной главной компонентой;
- каждому признаку свойственна факторная структура;
- связи между признаками и факторами (главными компонентами) линейные;
- для данного признака эффект воздействия факторов суммируется.

Если изучаем зависимости между *n* переменными (признаками), то на основании формулы (2.26) можно написать *n* линейных уравнений, которые свяжут признаки с главными компонентами. Эти *n* линейных уравнений составят математическую модель при решении задачи методом главных компонент:

$$\left.\begin{array}{l} y_1 = a_{11}f_1 + a_{12}f_2 + \ldots + a_{1n}f_n \\ y_2 = a_{21}f_1 + a_{22}f_2 + \ldots + a_{2n}f_n \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ y_n = a_{n1}f_1 + a_{n2}f_2 + \ldots + a_{nn}f_n \end{array}\right\}. \qquad (2.29)$$

5—1447

Итак, при проведении эксперимента мы получаем результаты в виде матрицы наблюдаемых величин:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ x_{n1} & x_{n2} & \cdots & x_{nN} \end{pmatrix},$$

где N — число наблюдаемых объектов;
n — число измеряемых признаков.

Элементы данной матрицы центрируются и нормируются, и мы получаем матрицу $\mathbf{Y}$ порядка $n \times N$. Теперь в развернутом матричном виде можно представить модель (2.29):

$$\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ y_{n1} & y_{n2} & \cdots & y_{nN} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} f_{11} & f_{12} & \cdots & f_{1N} \\ f_{21} & f_{22} & \cdots & f_{2N} \\ \ldots & \ldots & \ldots & \ldots \\ f_{n1} & f_{n2} & \cdots & f_{nN} \end{pmatrix},$$

или

$$\begin{pmatrix} y_1 \\ y_2 \\ \ldots \\ y_n \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \ldots \\ f_n \end{pmatrix},$$

где матрица $\mathbf{Y}$ представляет собой совокупность всех N наблюдаемых значений всех n параметров;
вектор y — совокупность n рассматриваемых признаков.

Точно так же представлена матрица $\mathbf{F}$, включающая совокупность всех N полученных значений всех n главных компонент, а вектор $\mathbf{f}$ представляет собой совокупность всех n главных компонент.

Коротко это запишется в матричной форме как выражение (2.27).

Под математической моделью процесса понимают оператор, которым он описывается. Значит, математическое описание изучаемого процесса заключается в определении оператора данного процесса. Этот оператор должен поставить в соответствие зависимую переменную независимой. Оператор, в данном случае матрица $\mathbf{A}$, указывает на совокупность математических действий, которые необходимо осуществить, чтобы по главным компонентам определить признаки, которые мы считаем линейно-зависимыми от этих компонент. Элементами матрицы являются весовые коэффициенты.

Выясним, что собой представляют весовые коэффициенты между признаками и главными компонентами. Для этого умножим y_j на первую главную компоненту и получим:

$$y_j f_1 = a_{j1} f_1 f_1 + a_{j2} f_2 f_1 + \cdots + a_{jr} f_r f_1 + \cdots + a_{jn} f_n f_1.$$

Чтобы получить коэффициент корреляции между j-м признаком и первой главной компонентой, просуммируем левую часть по всем N наблюдениям и разделим сумму на число наблюдений N, тогда правая часть примет вид:

$$\frac{1}{N}\sum_{i=1}^{N} y_{ji} f_{1i} = \frac{1}{N}\left[a_{j1}\sum_{i=1}^{N} f_{1i} f_{1i} + a_{j2}\sum_{i=1}^{N} f_{2i} f_{1i} + \ldots + a_{jn}\sum_{i=1}^{N} f_{ni} f_{1i}\right].$$

Учитывая, что $r_{f_r f_r} = 1$, перепишем выражение:

$$r_{y_j f_1} = a_{j1} + a_{j2}\, r_{f_2 f_1} + \cdots + a_{jr}\, r_{f_r f_1} + \cdots + a_{jn} r_{f_n f_1},$$

где $r_{y_j f_1}$ — коэффициент корреляции между j-м признаком и первой главной компонентой ($r_{y_j f_r}$ называют элементом структуры);
$r_{f_n f_1}$ — коэффициент корреляции между n-й и первой главной компонентой;
a_{jr} — весовые коэффициенты, которые в факторном анализе называются коэффициентами отображения.

В общем случае при проведении факторного анализа элементы структуры не совпадают с элементами отображения. Поскольку в методе главных компонент компоненты не коррелированы между собой, можно записать $r_{f_r f_k} = 0$ $(r \neq k)$, поэтому $r_{y_j f_1} = a_{j1}$. В общем случае в методе главных компонент $r_{y_j f_r} = a_{jr}$.

Матрицу коэффициентов корреляции между главными компонентами и признаками называют факторной структурой и обозначают так:

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1n} \\ s_{21} & s_{22} & \dots & s_{2n} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nn} \end{pmatrix}.$$

Такая полная матрица является квадратной. Она может иметь обратную матрицу $\mathbf{S}^{-1}$. На практике чаще используется прямоугольная матрица. Если при анализе берутся не все n компонент, а только их часть m, то

$$\mathbf{S} = \begin{pmatrix} s_{11} & s_{12} & \dots & s_{1m} \\ s_{21} & s_{22} & \dots & s_{2m} \\ \dots & \dots & \dots & \dots \\ s_{n1} & s_{n2} & \dots & s_{nm} \end{pmatrix}.$$

Эта матрица, естественно, не имеет обратной матрицы, но в рассматриваемом методе главных компонент она соответствует матрице-оператору $\mathbf{A}$ в формуле (2.27): $\mathbf{S} = \mathbf{A}$.

Матрица наблюдаемых коэффициентов корреляции на основе [106] может быть представлена так:

$$\mathbf{R} = \mathbf{YY}'\frac{1}{N}, \tag{2.30}$$

где $\mathbf{Y}$ — матрица нормированных значений признаков;
$\mathbf{Y}'$ — транспонированная матрица.

Рассмотрим элементы R:

$$r_{jj} = r_{kk} = 1; \; r_{jk} = r_{kj}.$$

Известно, что в математической статистике связь между случайными величинами часто характеризуется коэффициентом корреляции, корреляционным отношением, показателем корреляции рангов, коэффициентом взаимной сопряженности, показателем сходства, множественными и частными коэффициентами корреляции. Рассмотрим коэффициент корреляции, который необходим для проведения анализа методом главных компонент.

Коэффициент корреляции r_{jk} характеризует связь между двумя случайными величинами X_j и X_k в случае линейной корреляции между ними. Иногда его называют коэффициентом парной корреляции. Коэффициент корреляции представляет эмпирический первый основной смешанный момент. Для любых признаков и случайных величин

$$r_{jk} = \frac{1}{N}\sum_{i=1}^{N} y_{ji}y_{ki}, \tag{2.31}$$

где y_{ji} — нормированное и центрированное значение X_j-й случайной величины X_j для i-го индивида (i-го измерения);
y_{ki} — нормированное и центрированное значение другой случайной величины X_k для i-го измерения (индивида);
N — количество индивидов (измерений случайной величины).

$$y_{ji} = (x_{ji} - \bar{x}_j)/\sigma_j,$$

где x_{ji} — значение случайной величины X_j при i-м измерении;
$\bar{x}_j$ — среднее значение случайной величины X_j по результатам N измерений;
σ_j — среднее квадратическое отклонение X_j.

Известно, что при $r_{jk} = \pm 1$ между случайными величинами существует прямая или обратная линейная функциональная связь. При $r_{jk} = 0$ между ними не существует линейной корреляционной связи. Следовательно, диагональные элементы матрицы $\mathbf{R}$ представляют собой общности $h^2 = r_{jj} = r_{kk} = 1$, т. е. состоят из единиц.

Если $0 < r_{jk} < 1$, или $-1 < r_{jk} < 0$, то между величинами X_j и X_k существует прямая или обратная корреляционная связь. Равенство $r_{jk} = r_{kj}$ означает, что элементы, симметричные относительно главной диагонали, равны между собой. Значит, коэффициент корреляции служит показателем тесноты связи между случайными величинами X_j и X_k.

Отметим, что если $r_{jk} = 1$, то все точки выборочного наблюдения лежат на прямой линии. Угловой коэффициент этой линии положителен. Если $r_{jk} = -1$, то все точки выборочного наблюдения лежат на прямой с отрицательным угловым коэффициентом. Значение r_{jk} характеризует только степень разброса точек по отношению к линии регрессии, но не дает никакой информации о величине углового коэффициента, который характеризует линию регрессии.

Если коэффициент корреляции равен единице по абсолютному значению, то говорят, что между выборочными значениями наблюдается полная корреляция.

Если предполагается двумерное нормальное распределение, то оно характеризуется пятью параметрами: $\bar{x}_j$, $\bar{x}_k$, σ_j^2, σ_k^2, r_{jk}, т. е. средними значениями и дисперсиями частных распределений и коэффициентом корреляции.

Условные распределения X_j при заданном X_k и условные распределения X_k при заданном X_j также нормальны и определяются их средними значениями и дисперсиями:

$$\overline{x_{jk}} = \bar{x}_j + r_{jk}\frac{\sigma_j}{\sigma_k}(x_{ki} - \bar{x}_k); \tag{2.32}$$

$$\overline{x_{kj}} = \bar{x}_k + r_{jk}\frac{\sigma_k}{\sigma_j}(x_{ji} - \bar{x}_j); \tag{2.33}$$

$$\sigma_{jk}^2 = \sigma_j^2(1 - r_{jk}^2); \tag{2.34}$$

$$\sigma_{kj}^2 = \sigma_k^2(1 - r_{jk}^2), \tag{2.35}$$

где $\overline{x_{jk}}$ — среднее значение условного распределения X_j при заданном X_k;

$\overline{x_{kj}}$ — среднее значение условного распределения X_k при заданном X_j;

σ_{jk}^2 — дисперсия условного распределения X_j при заданном X_k (она не зависит от X_k);

σ_{kj}^2 — дисперсия условного распределения X_k при заданном X_j (она не зависит от X_j).

Среднее значение случайной величины X_j определяется по формуле:

$$\bar{x}_j = \frac{1}{N}\sum_{i=1}^{N} x_{ji}, \tag{2.36}$$

а среднее квадратическое отклонение

$$\sigma_j = \sqrt{\frac{\sum_{i=1}^{N}(x_{ji} - \bar{x}_j)^2}{N-1}}. \tag{2.37}$$

Из формул (2.32)—(2.35) видно, что, когда случайные величины принимают различные значения, средние значения условных распределений описываются прямыми, которые называются прямыми регрессии, а дисперсии остаются неизменными. Прямая регрессии X_j по X_k запишется так:

$$x_{ji} - \bar{x}_j = \beta_{jk}(x_{ki} - \bar{x}_k),$$

где $\beta_{jk} = r_{jk}\dfrac{\sigma_j}{\sigma_k}$.

Прямая регрессии X_k по X_j:

$$x_{ki} - \bar{x}_k = \beta_{kj}(x_{ji} - \bar{x}_j),$$

где $\beta_{kj} = r_{jk}\dfrac{\sigma_k}{\sigma_j}$;

β_{kj} и β_{jk} — коэффициенты регрессии.

В отличие от коэффициентов корреляции здесь порядок индексов играет существенную роль.

Кроме того, известно, что

$$r_{jk}^2 = \beta_{jk}\beta_{kj},$$

где r_{jk}^2 — коэффициент детерминации.

Он показывает долю дисперсии функции, объясняемую аргументом при данном значении коэффициента регрессии.

При интерпретации коэффициентов корреляции не интересуются причинностью, так как $r_{jk} = r_{kj}$. При интерпретации коэффициентов регрессии надо учитывать причинно-следственную связь. Например, можно изучать влияние подготовленности студентов в вузе на их практическую работу на производстве заданного профиля, но нельзя рассуждать о влиянии практической работы специалиста на предшествующую ей учебу в вузе. Если $r_{jk} = +1$, то прямые регрессии X_j на X_k и X_k на X_j совпадают.

Таким образом, корреляционная матрица наблюденных коэффициентов корреляции является симметрической вещественной, у которой

$$r_{11} = 1;$$

$$r_{12} = \frac{1}{N}\,[y_{11}y_{21} + y_{12}y_{22} + \cdots + y_{1N}y_{2N}];$$

....

$$r_{jk} = \frac{1}{N}\,[y_{j1}y_{k1} + y_{j2}y_{k2} + \cdots + y_{jN}y_{kN}];$$

....

$$r_{nn} = 1.$$

Данная матрица наблюденных коэффициентов **R** является исходной для проведения анализа методом главных компонент. Для этого надо провести ортогональные преобразования.

2.8. Преобразование корреляционной матрицы

Остановимся вначале на некоторых теоремах, которые мы должны применить в методе главных компонент [6]. Первая из них — это известная нам теорема 1.1, которую приведем в нужных индексах и воспользуемся двумя ранее не рассмотренными следствиями.

Теорема 2.1. Для любой симметрической матрицы **R** существует ортогональная матрица **U**, такая, что

$$\mathbf{U}'\mathbf{R}\mathbf{U} = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

Если при этом **R** — положительно определенная матрица, то все $\lambda_r > 0$.

Следствие 1. Если матрица **R** положительно определена, то $|\mathbf{R}| > 0$.

Следствие 2. Если матрица **R** положительно определена, то все главные ее миноры положительны.

Остается только пояснить, что λ_r — собственные значения матрицы **R**, а **U** является такой ортогональной матрицей, в которой r-й столбец является r-м собственным вектором, соответствующим r-му собственному числу λ_r.

Элементы в матрице Λ расположены в порядке убывания:

$$\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n \geqslant 0. \tag{2.38}$$

Вернемся к основной задаче:

$$\mathbf{y} = \mathbf{A}\mathbf{f}. \tag{2.39}$$

Для определения дальнейших преобразований, при которых мы сможем воспользоваться теоремой [6], [59], произведем преобразование к новым переменным $z_1, z_2, \ldots, z_n$ и введем ортогональную матрицу **U**. Пусть при этом удовлетворяются равенства

$$\mathbf{z} = \mathbf{U}'\mathbf{y};\ \mathbf{y} = \mathbf{U}\mathbf{z},$$

где $\mathbf{z} = \{Z_1, Z_2, \ldots, Z_n\}$ $(j = 1, 2, \ldots, n)$.

Будем считать, что $\mathbf{u}_r$ — r-й столбец матрицы **U**, тогда $\mathbf{u}'_r$ — транспонированный вектор $\mathbf{u}_r$ или r-я строка матрицы $\mathbf{U}'$. Первый столбец матрицы U выбирается так, чтобы дисперсия Z_1 была максимальной. Остальные векторы выбираются таким образом, чтобы дисперсии Z_r были максимальными. При этом переменная Z_r не должна коррелировать ни с одной из ранее выбранных переменных: $Z_1, Z_2, \ldots, Z_{r-1}$.

Пусть λ_r — дисперсия Z_r, но $Z_r = \mathbf{u}'_r\mathbf{y}$, тогда λ_r по определению дисперсии равна математическому ожиданию квадрата центрированной величины Z_r:

$$\lambda_r = M(\mathbf{u}'_r\,\mathbf{y})^2 = M(\mathbf{u}'_r\mathbf{y}\mathbf{u}'_r\,\mathbf{y}).$$

Так как скалярные произведения равны $\mathbf{u}'_r\,\mathbf{y} = \mathbf{y}'\mathbf{u}_r$, то $\lambda_r = \mathbf{u}'_r\,\mathbf{R}\mathbf{u}_r$, где **R** —выборочная ковариационная матрица.

Так как Z_r не коррелированы по условию, то $\mathbf{u}'_r\,\mathbf{R}\mathbf{u}_s = 0$, если $r \neq s$. Следовательно, по теореме 2.1 полученный результат в матричном виде означает, что

$$\Lambda = \mathbf{U}'\mathbf{R}\mathbf{U} \tag{2.40}$$

является диагональной матрицей с элементами, расположенными в порядке убывания, где λ_r — r-е по величине собственное значение матрицы **R**.

Умножив матрицу (2.40) слева на **U**, получим $\mathbf{R}\mathbf{U} = \mathbf{U}\Lambda$, или в терминах векторов-столбцов $\mathbf{R}\mathbf{u}_r = \lambda_r\mathbf{u}_r$, т. е. по формуле (2.4) $\mathbf{u}_r$ — собственный вектор матрицы **R**, а λ_r — соответствующее ему собственное значение.

Для того чтобы дисперсии Z_r для всех $r = 1, 2, ..., n$ равнялись единицам, надо полагать, что

$$\mathbf{f}_r = \lambda_r^{-1/2}; \quad Z_r = \lambda_r^{-1/2}\, \mathbf{u}_r' \mathbf{y}.$$

Это условие нормирования Z_r для получения r-й главной компоненты. Его можно переписать в виде

$$\mathbf{f} = \Lambda^{-1/2}\mathbf{z} = \Lambda^{-1/2}\mathbf{U}'\mathbf{y}.$$

Отсюда надо определить $\mathbf{y}$: $\mathbf{y} = \mathbf{U}\Lambda^{1/2}\mathbf{f}$. Учитывая формулу (2.39), можно заключить, что искомая матрица $\mathbf{A} = \mathbf{U}\Lambda^{1/2}$. Столбец матрицы $\mathbf{A}$ может быть легко определен, если известны собственные векторы $\mathbf{u}_r$:

$$\mathbf{a}_r = \lambda_r^{1/2}\mathbf{u}_r. \tag{2.41}$$

Норма вектора $\mathbf{a}_r$ по формуле (2.3) определяется при помощи скалярного произведения. Оно равно собственному значению λ_r:

$$\mathbf{a}_r' \mathbf{a}_r = \lambda_r, \tag{2.42}$$

а норма вектора $\mathbf{a}_r$ равна корню квадратному из этого произведения.

Теперь важно выяснить, как соотносятся дисперсии векторов и обобщенная дисперсия модели до и после ортогональных преобразований. Обратимся к теореме [6].

Теорема 2.2. Ортогональное преобразование $\mathbf{y} = \mathbf{Af}$ случайного вектора $\mathbf{f}$ оставляет инвариантной обобщенную дисперсию и сумму дисперсий компонент.

Из теоремы 2.2 вытекает следствие, которым необходимо воспользоваться.

Следствие. Обобщенная дисперсия вектора главных компонент равна обобщенной дисперсии исходного вектора, а сумма дисперсий — сумме дисперсий исходных величин.

На основании данного следствия и выражения (2.42) ясно, что $\mathrm{tr}\,(\mathbf{R}) = \mathrm{tr}\,(\Lambda)$, где $\mathrm{tr}\,(\mathbf{R})$ — след корреляционной матрицы; $\mathrm{tr}\,(\Lambda)$ — след диагональной матрицы; $\mathrm{tr}\,(\mathbf{R}) = n$, так как на главной диагонали корреляционной матрицы всегда стоят единицы ($r_{jj} = 1$).

Вклад данного вектора $\mathbf{a}_r$ в общую дисперсию по формулам (2.42) и (2.28) определится по формуле

$$V_r = \mathbf{a}_r' \mathbf{a}_r = \lambda_r = \sum_{j=1}^{n} a_{jr}^2.$$

Если мы хотим воспроизвести по матрице-модели $\mathbf{A}$ исходную матрицу наблюденных коэффициентов корреляции, то

$$\left.\begin{array}{l} \mathbf{R} = \mathbf{AA}'; \\ \mathbf{A}'\mathbf{A} = \Lambda. \end{array}\right\} \tag{2.43}$$

Таким образом, мы рассмотрели преобразование матрицы наблюденных коэффициентов корреляции и определение матрицы-модели $\mathbf{A}$ весовых коэффициентов, учитывающих тесноту связи между признаками и главными компонентами. Наконец, показали, как, зная главные компоненты, можно однозначно описать признаки матриц (2.43).

Разобрав кардинальные вопросы метода, остановимся на алгоритме решения задачи методом главных компонент.

2.9. Блок-схема алгоритма

Рассмотрим последовательность решения задачи методом главных компонент на основании результатов измерения признаков. Признаков по каждому объекту измерено n, общее число объектов равно N, полученное число наиболее весомых главных компонент, объясняющих удовлетворяющую исследователя долю общей дисперсии, ограничено m. Запишем в операторной форме блок-схему алгоритма (рис. 2.6):

$$\left.\begin{array}{l} A_1 A_2 A_3 A_4^{7} A_5 R_6^{\;4} A_7^{14} A_8 A_9 A_{10} A_{11}^{14} A_{12} \\ R_{13}^{\;7,11} A_{14} F_{15} A_{16}^{19,20} A_{17} R_{18}^{\;23,16} A_{19}^{26}\;{}^{23,16} A_{20}^{23,26} A_{21} \\ R_{22}^{20} P_{23\downarrow 19,20} A_{24} P_{25\downarrow 30}\;{}^{19,20} A_{26} F_{27} A_{28} R_{29}^{\;25} R_{30} \end{array}\right\}.$$

В операторной форме блок-схемы алгоритма представлены следующие обозначения:

A_i — операторы действия;
F_i — операторы формирования;
P_i — операторы сравнения;

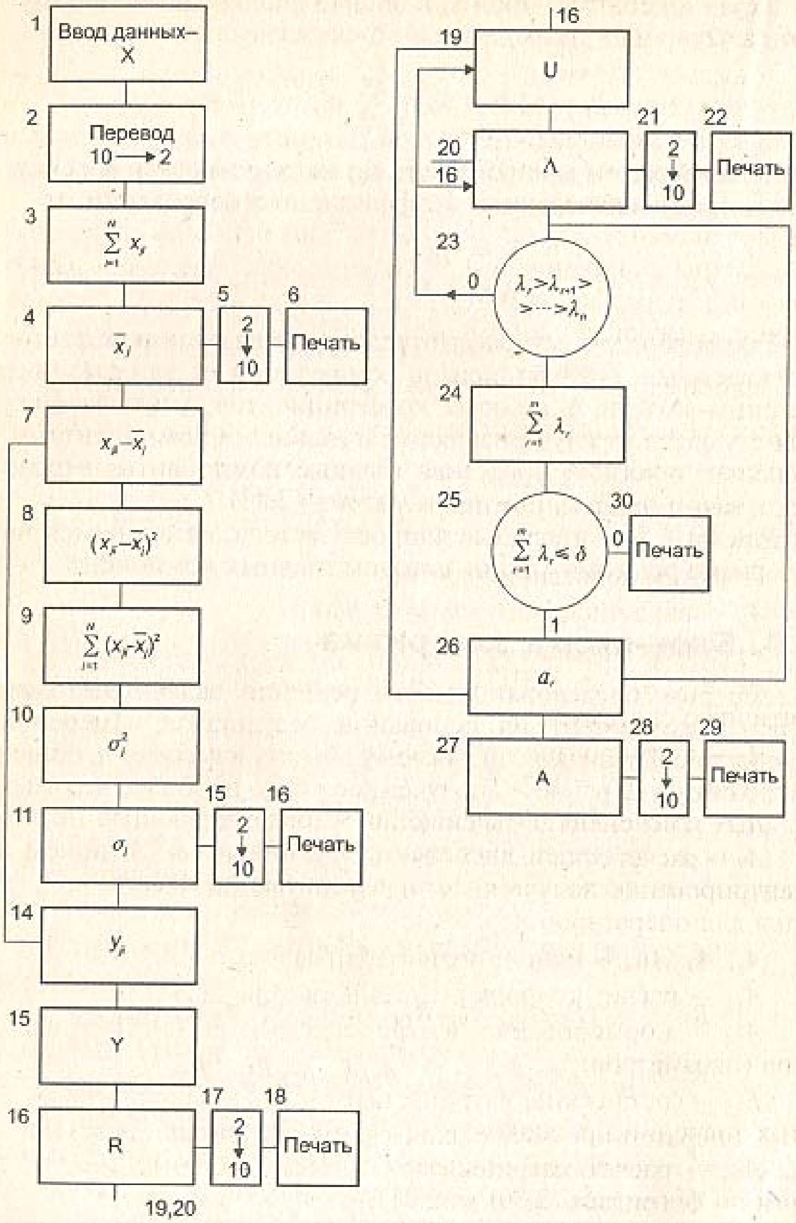


Рис. 2.6. Блок-схема алгоритма метода главных компонент

R_i — операторы окончания самостоятельной операции или алгоритма в целом.

Каждый оператор, кроме R_i, передает команду справа стоящему оператору. Внизу справа стоит номер оператора (i). Слева наверху стоят номера операторов, передающих свои команды другому оператору, если они не являются соседями слева. Справа наверху стоят номера операторов, которым передает команду другой оператор. Исключение составляют операторы сравнения, у которых номера операторов, которым они передают команду, стоят справа внизу после стрелочки, следующей за номером оператора.

Например, $^{23,16}A^{19}_{26}$ получает команды от операторов 23 и 16, а передает команды оператору 26 и соседнему 20. Оператор сравнения $^{20}P_{23\downarrow 19,20}$ получает команду от 20-го оператора, а передает команды операторам 19 и 20 и соседнему оператору 24.

Поясним действия операторов:

A_1 — ввод числового массива (данных);

A_2 — перевод данных из десятичной в двоичную систему;

A_5, A_{12}, A_{17}, A_{21}, A_{28} — перевод данных из двоичной в десятичную систему;

A_3 — суммирование N результатов измерений j-го признака ($j = 1, 2, ..., n$; $i = 1, 2, ..., N$);

A_4 — расчет по формуле (2.36);

A_7 — расчет отклонений результатов от среднего значения — центрирование полученных значений параметров (используется для операторов A_8 и A_{14});

A_8, A_9, A_{10} — используются для расчета σ_j (рис. 2.6);

A_{11} — расчет по формуле (2.37) σ_j;

A_{14} — нормирование центрированных значений признаков (параметров);

F_{15} — составление матрицы центрированных нормированных значений признаков;

A_{16} — расчет матрицы парных коэффициентов корреляции по формулам (2.30) и (2.31);

A_{19}, A_{20} — получение по стандартным программам матриц собственных значений Λ и собственных векторов **U**;

P_{23} — сравнение порядка следования собственных значений (проверяется требование $\lambda_r \geqslant \lambda_{r+1}$; при его нарушении перестраиваются матрицы Λ и соответственно $\mathbf{U}$);

A_{24} — расчет $\sum\limits_{r=1}^{m} \lambda_r$ для проверки в следующем операторе; эта сумма не должна превосходить некоторой заданной величины δ.

Величина δ представляет собой заданное значение объясняемой общей дисперсии матрицы $\mathbf{R}$.

$$\delta = \gamma \operatorname{tr}(\mathbf{R}),$$

где

$$\gamma = \frac{1}{\operatorname{tr}(\mathbf{R})} \sum_{r=1}^{m} V_r. \tag{2.44}$$

Предел этой суммы определяется в зависимости от требований поставленной задачи. Обычно задается $\gamma = (0{,}80 \div 0{,}95)$ и по этой величине определяется $\delta = \sum\limits_{r=1}^{m} V_r$. Однако есть целый ряд задач, когда удается сделать очень важные и интересные выводы по небольшому числу главных компонент при $\gamma < 0{,}80$, где γ — доля дисперсии, объясняемой извлеченными главными компонентами. Иногда вместо ограничений по величине суммарной дисперсии извлеченных главных компонент используется грубое правило: вклад очередной извлекаемой главной компоненты V_r должен превосходить единицу ($V_r > 1$).

Однако после окончания описанного анализа можно построить регрессию по главным компонентам. Поэтому иногда целесообразно получить все n главных компонент и построить полную матрицу $\mathbf{A}$. В этом случае оператор сравнения P_{25} исключается.

P_{25} — сравнение извлеченной суммы собственных значений и δ;

A_{26} — расчет векторов $\mathbf{a}_r$ матрицы $\mathbf{A}$, т. е. весовых коэффициентов связи признаков y_j с главными компонентами $\mathbf{f}_r$ по формуле (2.41); для расчета поступают данные: собственное

значение λ_r из матрицы Λ (оператор A_{20}) и собственный вектор $\mathbf{u}_r$ ортогональной матрицы $\mathbf{U}$ (оператор A_{19});

F_{27} — формирование из векторов $\mathbf{a}_r$ с целью удовлетворения некоторого поставленного требования по формированию матрицы весовых коэффициентов.

По достижении заданной величины γ процедуры расчетов заканчиваются оператором R_{30}. На этом алгоритм расчета матрицы весовых коэффициентов, т. е. оператора системы (2.27), заканчивается.

Отметим, что в стандартных программах математического обеспечения ЭВМ типа ЕС оператор P_{23} отсутствует, так как в них предусматривается следование $\lambda_r > \lambda_{r+1}$ при извлечении главных компонент. При использовании этих стандартных программ для ввода и вывода данных в статистических расчетах не понадобятся операторы перевода из двоичной системы в десятичную и обратно.

Глава 3
ИССЛЕДОВАНИЕ ПРАКТИЧЕСКИХ ЗАДАЧ МЕТОДОМ ГЛАВНЫХ КОМПОНЕНТ

3.1. Исследование путей повышения экономической эффективности СУП химического производства

При изготовлении высококачественных лаков и красок некоторые компоненты исходного сырья и промежуточные вещества, получаемые по технологии производства, обладают токсичностью. Поэтому производство продукции автоматизируется; в частности, для управления технологическими процессами внедряется типовая СУП. Специалистов для работы в СУП готовят в вузах и техникумах. Однако, несмотря на принимаемые меры совершенствования системы управления производством, на предприятиях случаются аварии, которые наносят большой ущерб государству. Стоимость одной аварии намного превосходит стоимость СУП и подготовки 100 специалистов для работы в этой системе.

Анализ причин возникновения аварий показал, что 95% из них происходит из-за ошибок и неправильных действий обслуживающего персонала (операторов СУП), а 5% — из-за отказов технических устройств СУП вследствие конструктивных и производственных недоработок. Таким образом, если в качестве сложной системы рассматривать все заводы данного профиля, вузы и техникумы, готовящие специалистов для СУП, то экономический анализ позволит минимизировать стоимость функционирования большой системы.

Запишем следующее соотношение:

$$C_у \gg C_{пр} > C_в > C_т. \tag{3.1}$$

В соотношении (3.1):

$C_у$ — средняя стоимость ущерба от одной аварии;

$C_в$ — средняя стоимость подготовки в вузах 100 специалистов;

$C_т$ — средняя стоимость подготовки 100 специалистов в техникуме;

$C_{пр}$ — стоимость проектирования и производства технической подсистемы СУП.

При этом $\sum_{i=1}^{4} C_i = W$ — критерий эффективности функционирования большой системы по стоимости. Значит, для повышения эффективности надо минимизировать W.

Учитывая соотношение (3.1), запишем, что минимизации следует подвергнуть $W' = C_у$.

Из анализа причин аварий известно, что

$$C_у = C_{уч} + C_{ут},$$

где $C_{уч}$ — часть средней стоимости ущерба по вине обслуживающего персонала;

$C_{ут}$ — часть средней стоимости ущерба из-за отказов технических устройств СУП по конструктивным и производственным причинам.

При этом

$$\left.\begin{aligned} C_{уч} &= \alpha C_у; \\ C_{ут} &= (1-\alpha) C_у, \end{aligned}\right\} \tag{3.2}$$

где $\alpha = 0{,}95$ — доля среднего ущерба от аварий по вине обслуживающего персонала.

Таким образом, экономический анализ показал, что в первую очередь надо решать две основные задачи снижения вероятности появления аварий:

- авария по вине операторов СУП;
- авария из-за отказов технических устройств по конструктивным и производственным причинам.

Остановимся на первой задаче; она намного сложнее второй, но эффект от ее решения значительно выше (см. выражение (3.2)).

Уточним цели исследования и составим блок-схему задачи (рис. 3.1). Пусть большая система включает подсистему производства продукции и подсистему подготовки специалистов для СУП.

В подсистеме производства специалист (1), окончивший вуз или техникум, сам выполняет работу (3) при помощи СУП или руководит подчиненными специалистами (2), которые выполняют работу (3) при помощи СУП. Однако никто специально не оценивает работу специалистов-выпускников для передачи информации в вузы.

Подсистема подготовки специалистов для работы в СУП включает вузы, техникумы и другие учебные заведения по подготовке и переподготовке специалистов. В учебных заведениях применяются методы обучения с обратной связью и корректировкой процесса внутри учебного заведения, с обратной связью, но без корректировки, открытые методы и т. д. Между методами существуют взаимные связи, а также связи между обучаемыми. Эффект подготовки зависит от качества учебного процесса, и в частности от педагогического состава (0). Однако в течение всего периода обучения буду-

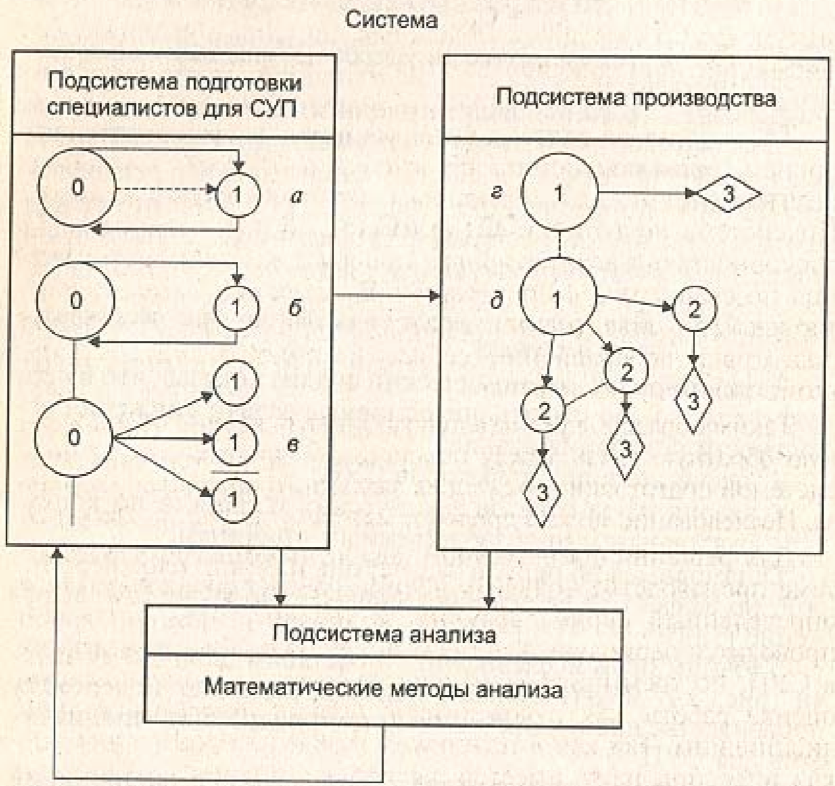


Рис. 3.1. Схема функционирования исследуемой системы:
а — методы обучения с обратной связью и корректировкой процесса внутри учебного заведения;
б — методы обучения с обратной связью без корректировки процесса внутри учебного заведения;
в — открытые методы;
г — самостоятельная деятельность специалиста при помощи СУП;
д — специалист руководит подчиненными специалистами, которые выполняют работу при помощи СУП.

щие специалисты оцениваются в зависимости от их успехов по тем или иным дисциплинам.

Большое количество причин влияет на работу специалистов на производстве. Сюда входят условия труда и дисциплина производства на данном предприятии, личные качества работника и т. д. Существует определенное влияние и уровня подготовки специалиста.

Интересно было бы выявить скрытые связи, существующие между этими двумя подсистемами. Для обработки информации, поступающей из этих двух подсистем, целесообразно создать третью подсистему (подсистему анализа).

Тогда большая система будет состоять из трех подсистем, причем третья подсистема призвана вырабатывать управляющее воздействие для подсистемы подготовки специалистов. Подсистема подготовки специалистов обладает наибольшей способностью к корректировке процесса ее функционирования по сравнению с процессом производства. Сама корректировка (по объективным критериям) программ и методов подготовки по сравнению со всеми затратами на обучение стоит значительно дешевле.

Таким образом, предлагается установить научно обоснованную обратную связь между подсистемой производства и подсистемой подготовки через вновь введенную подсистему анализа. Исследование можно провести методом главных компонент.

Для решения поставленной задачи необходимо в подсистеме производства получить оценки работы специалистов за определенный период времени. В течение года, например, проводится периодическая оценка специалистов, работающих в СУП, по трем признакам: оценка работы как специалиста, оценка работы как руководителя, оценка производственной дисциплины. Так как в техникуме, вузе и на производстве по каждому признаку имеется несколько оценок, в качестве оценки признака берется средняя из четырехбалльной оценки. Кроме того, по четырехбалльной системе оценивается специальность обучаемого до поступления в техникум.

Известно, что после окончания учебного заведения молодому специалисту выделяется период для адаптации к практическим условиям работы в СУП. После этого специалист сдает зачет на право самостоятельной работы. Длительность этого периода является характеристикой выпускника (учитывается число дней от момента прибытия в СУП до даты получения права на самостоятельную работу).

Для удобства расчетов пронумеруем характеристики признаков:

1) оценка специальности до поступления в техникум;
2) «Электротехника»;
3) «Основы ремонта блоков СУП»;
4) «Конструкция блоков СУП»;
5) «Техника безопасности»;
6) «Электрооборудование СУП»;
7) «Эксплуатация оборудования СУП»;
8) «Организация работы в СУП»;
9) оценка продолжительности дополнительной подготовки на производстве до допуска к самостоятельной работе;
10) оценка специалиста при текущем контроле на производстве;
11) оценка способностей к руководству подчиненными;
12) оценка производственной дисциплины.

Признаки 2—8 выражаются оценками по предметам, изученным в техникуме.

По всем перечисленным признакам были получены оценки для 77 человек, обучавшихся в техникуме в течение трех лет и проработавших после этого год в СУП.

После обработки полученных результатов по алгоритму в соответствии с операторами A_1—A_{18} (рис. 3.2) была получена корреляционная матрица (табл. 3.1). Необходимо напомнить, что данная матрица является симметричной матрицей и ее элементы $r_{jk} = r_{kj}$, поэтому нижняя часть этой матрицы не заполнена.

При решении по стандартной программе задачи по получению матрицы-модели и диагональной матрицы собственных значений главные компоненты извлекали до величины $\gamma = 0{,}80$, поэтому была получена следующая диагональная матрица Λ.

$$\Lambda = \begin{bmatrix} 3{,}021 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1{,}943 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1{,}336 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1{,}173 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0{,}900 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{,}882 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0{,}708 \end{bmatrix}.$$

Таблица 3.1

Матрица парных коэффициентов корреляции

y_j	1	2	3	4	5	6	7	8	9	10	11	12
1	1	0,037	0,129	0,149	0,121	0,158	−0,013	0,096	−0,249	0,070	−0,023	0,199
2		1	0,019	0,054	0,181	−0,093	0,086	0,032	0,166	−0,045	−0,116	−0,141
3			1	0,188	0,048	0,098	0,044	0,111	−0,205	0,152	0,059	0,227
4				1	0,568	0,323	0,561	0,430	−0,059	0,264	−0,053	0,232
5					1	0,042	0,399	0,373	−0,053	0,206	0,020	0,131
6						1	0,113	0,081	−0,361	−0,079	0,387	−0,128
7							1	0,462	0,069	0,123	0,106	0,104
8								1	0,034	0,347	0,167	0,277
9									1	−0,354	−0,399	−0,381
10										1	0,388	0,425
11											1	0,394
12												1

В данной матрице собственные значения определяют вклад соответствующей главной компоненты в общую дисперсию. Суммарная дисперсия

$$V = \sum_{r=1}^{n} V_r = \mathrm{tr}\,(\mathbf{R}) = n = 12.$$

Приведем вклады в общую дисперсию каждой главной компоненты (V_r и $V_{r/n}$) и суммарный вклад первых компонент $\gamma_r = \frac{1}{n}\sum_{\pm j=1}^{r} V_r$ (табл. 3.2).

Таблица 3.2

Вклады в общую дисперсию

Критерий	f_1	f_2	f_3	f_4	f_5	f_6	f_7
V_r	3,02	1,94	1,34	1,17	0,90	0,88	0,71
$V_{r/n}$, %	25,17	16,17	11,17	9,75	7,50	7,33	5,92
γ_r, %	25,17	41,34	52,51	62,25	69,76	77,09	83,01

Доля дисперсии, объясняемой извлеченными семью главными компонентами, определится по формуле (2.44):

$$\gamma = \frac{\sum_{r=1}^{7} V_r}{V} = \frac{3{,}02 + 1{,}94 + 1{,}34 + 1{,}17 + 0{,}90 + 0{,}88 + 0{,}71}{12} = 0{,}83;$$

$$\gamma = \frac{9{,}96 \cdot 100\%}{12} = 83\%.$$

Таким образом, общий вклад остальных пяти главных компонент составляет около 17% (начиная с пятого компонента, $V_r < 1$). Если воспользоваться этим критерием, то надо остановиться на первых четырех главных компонентах, которые объясняют 62,3% общей дисперсии изучаемого процесса. Приведем весовые коэффициенты матрицы-модели.

Характер вклада главных компонент в общую дисперсию процесса представлен на рис. 3.2. Наиболее существенный вклад дает первая главная компонента. Она характеризует способности к обучению и практической работе.

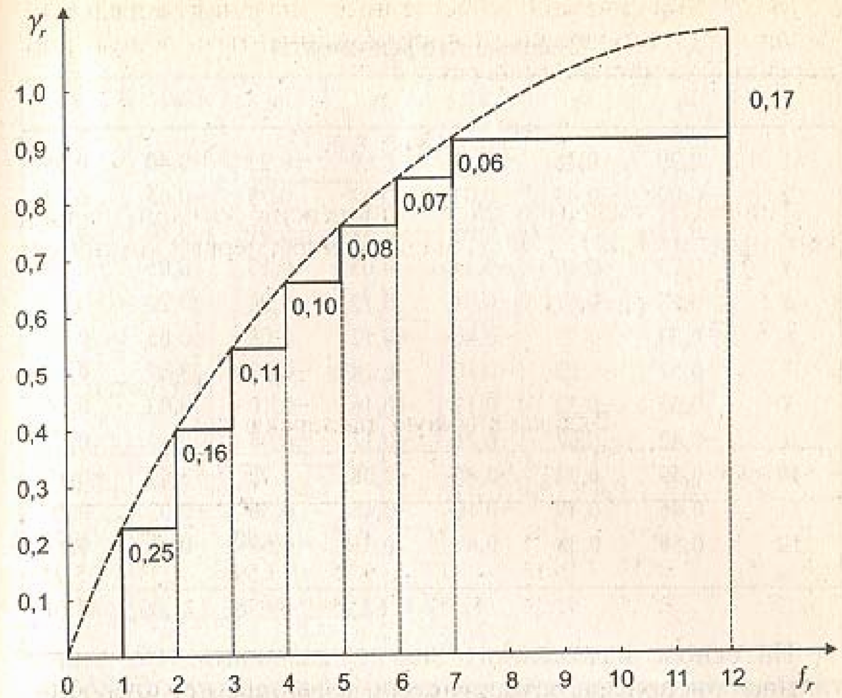


Рис. 3.2. Изменение суммарного вклада главных компонент в общую дисперсию исследуемого процесса

На этой интерпретации остановимся более подробно. Первая главная компонента, как видно из табл. 3.3, связана со всеми признаками, определяющими успехи специалиста при обучении в техникуме (кроме второго признака), и одновременно со всеми признаками, характеризующими работу специалиста на производстве. Наиболее высокие значения коэффициентов веса связывают первую главную компоненту с признаками 4, 8, 10, 12, 5, 7, 11, 9.

Все перечисленные признаки, кроме последнего, связаны с первой главной компонентой положительными коэффициентами веса, а последняя компонента — с отрицательным коэффициентом веса. Действительно, при более высоких успехах во время учебы уменьшается период подготовки до получения права самостоятельной работы на производстве.

Таблица 3.3

Весовые коэффициенты

y_j	a_1	a_2	a_3	a_4	a_5	a_6	a_7
1	0,29	0,16	−0,17	0,69	−0,23	−0,40	0,35
2	−0,02	−0,43	0,03	0,37	0,79	−0,05	0,12
3	0,33	0,19	0,03	0,44	0,01	0,80	0,01
4	0,73	−0,40	−0,21	0,03	−0,15	0,05	−0,19
5	0,58	−0,47	−0,01	0,12	0,04	−0,20	−0,39
6	0,33	0,23	−0,86	−0,12	0,08	0,05	0,05
7	0,57	−0,52	−0,10	−0,24	−0,09	0,09	0,11
8	0,65	−0,32	0,17	−0,16	−0,10	0,03	0,39
9	−0,42	−0,67	0,16	−0,17	−0,08	0,13	0,32
10	0,59	0,24	0,46	−0,08	0,15	−0,09	−0,19
11	0,46	0,49	−0,10	−0,45	0,36	−0,03	0,24
12	0,58	0,38	0,47	0,05	−0,08	−0,04	0,13

На основе изложенного можно заключить, что первая главная компонента действительно характеризует способности специалиста, проявленные в работе и учебе.

Остановимся на второй главной компоненте. Наиболее высокий положительный коэффициент веса соответствует признаку 11 — оценка способностей к руководству подчиненными. Меньшими, но положительными значениями коэффициента веса эта компонента связана с признаками 10 и 12. Отрицательные коэффициенты веса у признаков, характеризующих успехи в обучении, показывают, что способности к руководству воздействуют на вторую главную компоненту в сторону, противоположную успехам в обучении (признакам 2, 4, 5, 7, 8). Значит, изучение дисциплин в техникуме поставлено так, что они не формируют способности к руководству.

Можно полагать, что способности к руководству подчиненными связаны во второй главной компоненте с признаком 6 — оценка по предмету «Электрооборудование СУП». С учетом того, что с этой главной компонентой больше связаны признаки, определяющие производство (10, 11, 12), она и была названа «способности к руководству производством». По-

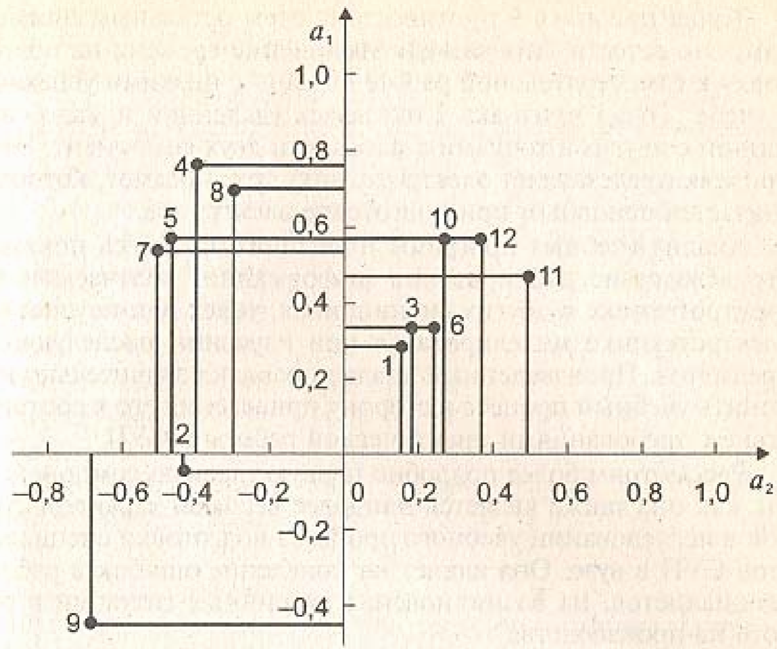


Рис. 3.3. Результаты анализа в координатах двух главных компонент

скольку способности к руководству производством в техникуме не формируются, они формируются на самом производстве.

Следует заметить, что основной анализ в рассматриваемом случае был направлен на первые две компоненты, где все три характеристики производственной деятельности имеют один знак.

Первая компонента могла бы быть генеральной, если бы не второй признак, для которого $a_{21} = -0,02$ (можно считать, что $a_{21} = 0$). В связи с этим ее надо отнести, как и вторую, к общим главным компонентам. Среди всех оставшихся весовых коэффициентов только девятый имеет знак минус.

Полученные результаты рассмотрим в координатах двух главных компонент, где весовые коэффициенты связи признаков с данными компонентами определяются соответствующими точками признаков на плоскости. На рис. 3.3 явно выделяются две грозди точек.

Точка признака 9 противостоит всем остальным признакам. Это естественно, так как увеличение времени на подготовку к самостоятельной работе связано с низкими успехами в учебе. Точка признака 2 оказалась удаленной и мало связанной с другими точками в плоскости двух компонент. Этот признак представляет электротехнику, т.е. предмет, который считается основным при подготовке электриков.

Анализ учебных программ и учебного процесса показал, что вследствие дублирования информации, получаемой по электротехнике в других дисциплинах, успех или неуспех по электротехнике нивелировался при изучении последующих предметов. Произведенный анализ позволил значительно изменить учебный процесс в сторону приведения его в соответствие с требованиями практической работы в СУП.

Рассмотрим более подробно первую главную компоненту, так как она также является наиболее весомой характеристикой в исследовании учебного процесса подготовки специалистов СУП в вузе. Она влияет на появление ошибок в работе специалистов, на возникновение аварийных ситуаций в работе на производстве.

3.2. Исследование первой главной компоненты при анализе работы специалистов СУП после окончания вуза

Для определения возможных путей снижения вероятности появления аварий особый интерес представляет анализ влияния первой главной компоненты, характеризующей способность к обучению и работе в СУП на производстве. Целесообразно проверить, является ли эта главная компонента свойством данной выборки из семидесяти семи человек или устойчивым общим свойством. Для этого был проведен анализ обучения в вузе и работы специалистов в СУП.

Условия эксперимента соответствовали условиям, описанным в п. 3.1, но срок обучения составлял не три, а пять лет. Под наблюдение были взяты три группы студентов: 106, 60 и 56 человек одной специализации. Не учитывались специальность до поступления в вуз и дополнительное время на подготовку специалистов. Для анализа было взято 17 признаков — предметов, изучаемых в вузе, признак 18 — оценка работы

Таблица 3.4

Первые главные компоненты

y_j	Число студентов			y_j	Число студентов		
	106	56	60		106	56	60
1	0,76	0,81	0,66	11	0,63	0,68	0,62
2	0,64	0,73	0,84	12	0,57	0,52	0,54
3	0,78	0,75	0,73	13	0,46	0,65	0,73
4	0,75	0,49	0,77	14	0,09	0,57	0,43
5	0,72	0,60	0,79	15	−0,09	0,52	0,51
6	0,72	0,77	0,63	16	0,15	0,21	0,31
7	0,63	0,59	0,76	17	0,21	−0,19	0,30
8	0,79	0,70	0,64	18	0,57	0,80	0,67
9	0,72	0,50	0,46	19	0,34	0,67	0,73
10	0,62	0,64	0,61	20	0,21	0,43	0,49

выпускника в течение года при текущем контроле на производстве, признак 19 — оценка способностей к руководству подчиненными, признак 20 — характеристика производственной дисциплины. Характеристика первой главной компоненты, их вклады в общую дисперсию процесса представлены в табл. 3.4.

Для того чтобы оценить роль первой главной компоненты в изучаемом процессе, для каждого эксперимента приведем

Таблица 3.5

Характеристика первой главной компоненты

Критерии	Число студентов		
	106	56	60
V_1	6,62	7,50	7,90
V_2	1,70	1,60	1,90
k	3,89	4,69	4,16
$V_{1/n}$, %	33,10	37,50	39,50

полученные характеристики вклада в общую дисперсию V_1 и $V_{1/n}$ и отношение вклада первой ко второй главной компоненте $k = V_1/V_2$ (см. табл. 3.5).

Из анализа табл. 3.4 и 3.5 видно, что первая главная компонента характеризуется весьма высоким вкладом в общую дисперсию процесса при подготовке специалистов СУП как в техникуме, так и в вузе. При этом в вузе она является более весомой по вкладу в общую дисперсию (от 33 до 40%), чем в техникуме, и в 4 раза превосходит вклад второй главной компоненты.

На основе выявления первой главной компоненты как характеристики способностей к обучению и работе в СУП возникла идея проведения компонентного анализа для выявления индивидуальных значений первой главной компоненты и ранжирования специалистов по весу этих значений. Это позволяет на наиболее ответственные должности выбирать специалистов, обладающих наиболее высокой общей характеристикой по данному новому критерию.

3.3. Экономический анализ производственно-хозяйственной деятельности предприятий методом главных компонент

Целесообразно выяснить, какую дополнительную информацию можно получить после решения некоторой практической экономической задачи методами корреляционного и регрессионного анализа на основе полученной матрицы парных коэффициентов корреляции. Эта информация может быть извлечена за счет преобразования матрицы парных коэффициентов корреляции методом главных компонент.

По результатам работы 102 строительных организаций собраны статистические данные по десяти признакам, влияющим на прибыль:

1) объем выполненных работ первого основного производства;
2) объем выполненных работ по капитальному ремонту;
3) количество работающих в подсобных цехах и мастерских;
4) объем продукции подсобных цехов и мастерских;
5) строительство цехов первого основного производства;

6) строительство цехов второго основного производства;
7) монтаж оборудования первого основного производства;
8) строительство по договорам строительного управления;
9) монтаж оборудования второго основного производства;
10) строительство жилых домов.

Все характеристики выражены в рублях. На основе этих данных получена матрица парных коэффициентов корреляции и проведен регрессионный анализ по определению влияния рассмотренных признаков на прибыль.

Проведем дальнейшее исследование признаков, влияющих, как показал регрессионный анализ, определенным образом на прибыль. Для этого получим из ранее использованной корреляционной матрицы матрицу собственных значений и матрицу собственных векторов.

Матрица собственных значений имеет следующий вид:

$$\Lambda = \begin{pmatrix} 4{,}20 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1{,}36 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0{,}98 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0{,}94 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0{,}79 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0{,}70 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0{,}37 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}32 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}20 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0{,}12 \end{pmatrix}.$$

На основании этой матрицы следует извлекать первые три главные компоненты, если будет использоваться критерий $V_r \geqslant 1$. Можно воспользоваться рис. 3.4*а* для того, чтобы решить, какую долю общей дисперсии необходимо охватить извлекаемыми компонентами.

Приведем первые восемь главных компонент (табл. 3.6) и проведем их интерпретацию. Как видно из табл. 3.6, первая главная компонента может быть названа «мощность предприятия». Эта новая характеристика наиболее тесно связана с признаками, определяющими размер предприятия.

Так, с объемом выполненных работ первого производства связь характеризуется коэффициентом $a_{11} = 0{,}88$, с объемом выполненных работ по капитальному ремонту — $a_{21} = 0{,}85$, с

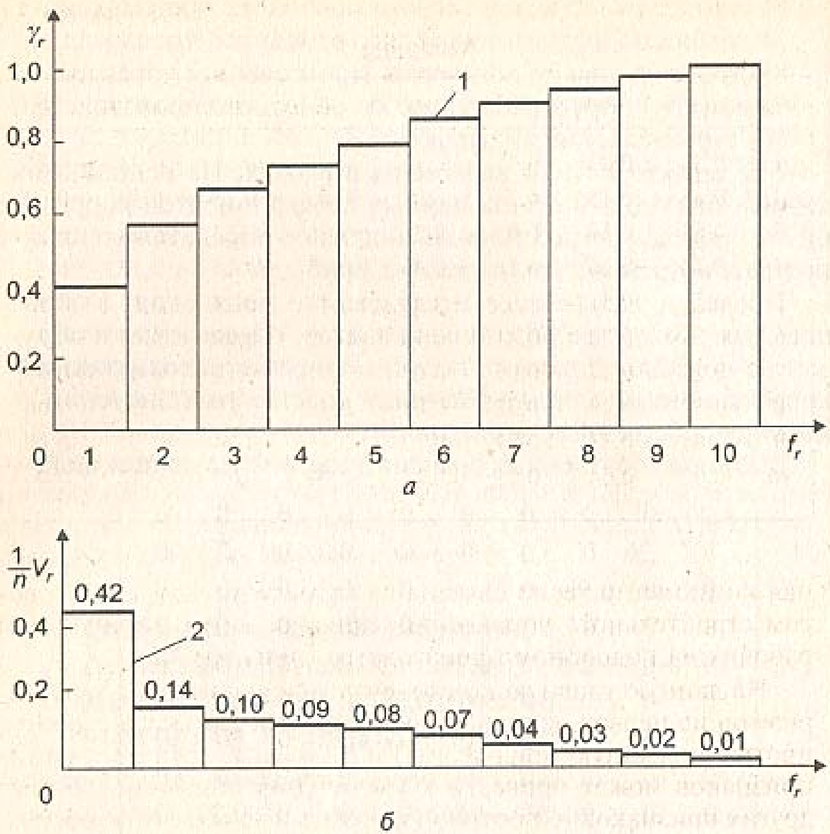


Рис. 3.4. Изменение суммарного вклада главных компонент в общую дисперсию в решаемой экономической задаче:
а — доля общей дисперсии, которую необходимо охватить извлекаемыми компонентами;
б — вклад компонент в общую дисперсию

количеством рабочих — $a_{31} = 0,68$, с объемом продукции подсобных цехов — $a_{41} = 0,77$ и т. д. Значит, можно отметить, что изменения в получаемой прибыли на 42% зависят от мощности предприятия. С увеличением мощности предприятия следует ожидать роста прибыли при существующих условиях работы.

Вторую главную компоненту можно назвать «характеристика основного и нецентрализованного строительства». Дан-

Таблица 3.6

Матрица А

y_j	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8
1	0,88	−0,05	−0,16	0,13	−0,17	−0,17	−0,28	−0,26
2	0,85	0,15	−0,21	0,01	−0,02	−0,23	0,50	−0,07
3	0,68	0,56	0,04	0,15	−0,38	−0,23	−0,14	0,27
4	0,77	−0,26	−0,16	−0,46	0,17	−0,10	−0,12	−0,13
5	0,54	−0,51	0,03	−0,40	−0,41	0,29	0,05	0,16
6	0,34	0,30	0,81	0,10	−0,02	0,09	0,05	−0,23
7	0,28	−0,47	−0,05	0,69	−0,07	0,12	0,04	−0,09
8	0,57	0,49	−0,32	−0,01	0,42	0,39	−0,06	0,04
9	0,68	−0,38	0,37	0,13	0,48	−0,24	−0,04	0,25
10	0,60	0,04	0,11	0,18	−0,04	0,50	0,02	0,04

ная компонента тесно связана со строительством по договорам строительного управления (признак 8) и с количеством рабочих на подсобном производстве (признак 3).

На данную главную компоненту признаки 5 и 7, характеризующие первое основное производство, влияют в сторону, противоположную признакам 3 и 8. Значит, если рост одних признаков может привести к увеличению прибыли, то рост других признаков (с противоположными знаками), наоборот, к ее уменьшению. Данные о вкладе компонент в общую дисперсию представлены на рис. 3.4*б*, а результаты анализа по двум компонентам — на рис. 3.5.

Третью главную компоненту можно назвать «характеристика второго основного производства». Она наиболее тесно связана со строительством цехов второго основного производства ($a_{63} = 0,81$) и с монтажом оборудования второго основного производства. В отличие от второй компоненты вес третьей компоненты составляет только 10%.

Четвертую компоненту можно назвать «характеристика первого основного производства». Ее вес невелик. Он составляет около 9%. При этом следует заметить, что если монтаж оборудования на первом основном производстве ($a_{74} = 0,69$) приводит к росту прибыли, то при тех же условиях работы и

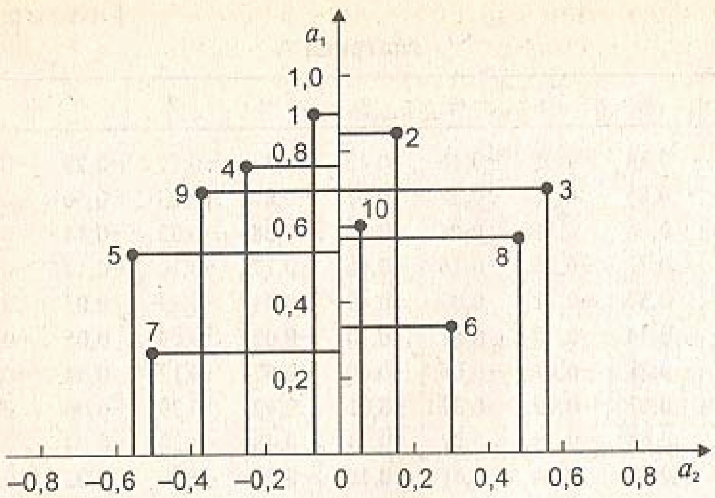


Рис. 3.5. Результаты анализа в координатах двух главных компонент

существующих нормах строительные работы в цехах первого основного производства могут привести к противоположным результатам. Это свидетельствует о том, что некоторые нормативы или организация работы на данном участке строительства требуют пересмотра. Однако, как указано выше, удельный вес влияния этих причин невелик. Остальные компоненты имеют еще меньший вес.

Таким образом, данный анализ позволил установить:

- колебания прибыли в исследованных организациях в первую очередь зависят от мощности предприятия, нецентрализованного строительства, хода строительства цехов второго и первого основных производств;
- монтаж и строительство цехов первого основного производства по-разному влияют на главную компоненту.

3.4. Обратная факторная задача

Каждая главная компонента дает некоторую общую новую характеристику всем изучаемым объектам. При этом каждая компонента является функцией особенностей каждого из изучаемых объектов.

Представим случай, когда нас интересуют качества объектов, связанные с одной или некоторыми главными компонентами. Например, в п. 3.1 нас заинтересовала первая общая главная компонента, характеризующая способности специалистов к обучению и практической работе в СУП. Если бы нам удалось получить значение данной компоненты для каждого из обучаемых, то их можно было бы ранжировать или классифицировать по этой очень важной способности. В данном случае мерой было бы индивидуальное значение главной компоненты для каждого специалиста.

Обратимся к методу, изложенному в [59]. Для этого развернем равенство (2.26) для j-го признака:

$$y_j = a_{j1}f_1 + a_{j2}f_2 + \cdots + a_{jn}f_n.$$

Выразим значения главных компонент через значения признаков. Для r-й главной компоненты

$$f_r = \frac{1}{\lambda_r}(a_{1r}y_1 + a_{2r}y_2 + \cdots + a_{nr}y_n).$$

Прежде чем перейти к примерам, отметим, что предложенный метод не является единственным; он достаточно прост для программирования на ЭВМ.

Пример 3.1

Выразите первый признак через главные компоненты и первую компоненту через признаки (см. п. 3.1 табл. 3.3).

Решение

1. Значение первого признака:

$$y_1 = 0{,}29f_1 + 0{,}16f_2 - 0{,}17f_3 + 0{,}69f_4 - 0{,}23f_5 - 0{,}40f_6 + 0{,}35f_7.$$

2. Значение первой главной компоненты:

$$f_1 = \frac{1}{3{,}02}(0{,}29y_1 + 0{,}33y_3 + 0{,}73y_4 + 0{,}58y_5 + 0{,}33y_6 + 0{,}57y_7 + 0{,}65y_8 - 0{,}42y_9 + 0{,}59y_{10} + 0{,}46y_{11} + 0{,}58y_{12}).$$

В данном выражении пропущен член с y_2, так как его весовой коэффициент мал.

Пример 3.2

Выразите главную компоненту, характеризующую мощность предприятия, воспользовавшись данными п. 3.3 (см. табл. 3.6).

Решение

$$f_1 = \frac{1}{4{,}2}(0{,}88y_1 + 0{,}85y_2 + 0{,}68y_3 + 0{,}77y_4 + 0{,}54y_5 + 0{,}34y_6 + 0{,}28y_7 + 0{,}57y_8 + 0{,}68y_9 + 0{,}60y_{10}).$$

Пример 3.3

Выразите второй признак через главные компоненты в примере по изучению прибыли предприятия (табл. 3.6). Коэффициенты веса, меньшие 0,10, в данном примере разрешается не учитывать.

Решение

$$y_2 = 0{,}85f_1 + 0{,}15f_2 - 0{,}21f_3 - 0{,}23f_6 + 0{,}50f_7.$$

Пример 3.4

Вернемся к п. 3.2, в котором было проанализировано 106 специалистов. Для каждого специалиста желательно получить индивидуальные значения первой главной компоненты f_{1i} и по ним разбить специалистов на четыре класса.

Решение

Разбиение на классы рекомендуется производить следующим образом: разбить всех специалистов на две группы (с положительными и отрицательными значениями f_{1i}, затем каждую группу разделить пополам. В результате у первой группы будут I и II классы и у второй — III и IV классы. В класс I войдут специалисты с наибольшими положительными значениями, а в класс IV — с наибольшими по абсолютной величине отрицательными значениями первой главной компоненты для индивидов. Первая главная компонента характеризует способности специалистов к обучению и работе в СУП.

Таким образом, по нашему мнению, в класс I вошли специалисты, наиболее способные к обучению в вузе и практической работе в СУП на производстве.

Жизненность рекомендаций в любых теоретических исследованиях может подтвердить только практика. Рассматриваемые специалисты после окончания вуза проработали год на производстве. За это время они получили оценки по трем признакам и, кроме того, были представлены все замечания по их работе за год. При этом объяснялось, за какие упущения оценки были снижены, какие замечания они получили, у кого во время работы сложились аварийные ситуации и т. д.

Количество специалистов, которые в течение года работы на производстве получили замечания, было принято за 100%. Их деятельность была проанализирована в соответствующих классах. Оказалось, что 70% специалистов, получивших замечания, относились к классам III и IV, 30% — к классу II. В результате в классе I не оказалось ни одного специалиста, который получил хотя бы одно замечание за практическую работу в течение года. Так была подтверждена объективность классификации, полученной в результате решения обратной факторной задачи.

После проведения описанного эксперимента специалисты класса I (после первого года работы) были выдвинуты на наиболее ответственные инженерные и руководящие должности в СУП. Наблюдение за специалистами в течение последующих двух лет показало, что они успешно справляются с возложенными на них ответственными обязанностями.

В заключение отметим, что если при проведении преобразований по методу главных компонент достаточно знания вектора средних и матрицы ковариаций, то для обеспечения устойчивости выборочных значений весьма существенно предположение о нормальности распределения исходных признаков. Необходимо рассмотреть вопрос обеспечения более точных значений числовых характеристик законов распределения исходных параметров модели.

3.5. Анализ моделей функционирования объединений и отраслей народного хозяйства

Анализ производительности труда на предприятиях текстильной промышленности

Методом главных компонент было проведено исследование производительности труда ткачей двух предприятий текстильной промышленности [66]. Под наблюдение было взято 430 ткачих. Исследование ставило две задачи:

- получить интегральный показатель производительности труда, более объективный, чем каждый из четырех ранее используемых показателей в отдельности;
- выявить зависимость уровня производительности труда от производственных условий и индивидуальных особенностей работника.

Для решения первой задачи были использованы данные по четырем показателям качества работы ткачих: выработка в метроуточинах (учитывается количество выработанных метров, плотность ткани по утку и ширина суровья); процент выполнения нормы выработки; сортность продукции; среднемесячная заработная плата.

Каждый из перечисленных показателей не дает полной характеристики производительности труда ткачих. Был получен интегральный критерий, в котором все перечисленные показатели были приведены к соизмеримому основанию. Интегральный показатель позволил учесть влияние внешних факторов на работу ткачих.

Первые две главные компоненты объясняли на одном комбинате 45%, а на втором комбинате 55% общей дисперсии. Вторая компонента объясняла соответственно 24 и 21% общей дисперсии.

В результате проведенного анализа по главным компонентам удалось получить распределение работниц по индивидуальным значениям первой компоненты на обоих предприятиях и выделить по уровню производительности труда передовую группу и группу отстающих работников, более точно оценить трудовой вклад отдельных групп работников. Это имело большое значение для развития и выявления резервов роста производительности труда.

В этом же исследовании приведены весьма интересная группировка ткачих по условиям труда и количественная оценка условий труда. Рассматривались следующие семь показателей: номер уточной пряжи; ширина суровой ткани; скорость станка; обрывность основы; количество артикулов, вырабатываемых на одном рабочем месте; простои по организационно-техническим причинам; количество месяцев работы станка после капитального ремонта.

На первую главную компоненту пришлось около 47% объяснимой дисперсии, а на вторую — 18%.

Были получены индивидуальные значения первых двух главных компонент для каждого работника. Распределение ткачих по первым двум компонентам позволило разбить исследуемых работников на четыре однородные группы по виду выпускаемой продукции, типу оборудования и уровню специализации рабочего места. Внутри однородных групп оказалось возможным использовать существующие показатели выработки в метроуточинах и процент выполнения норм выработки для сравнения уровня производительности труда, потому что в группах показатели выработки рассматриваются

в однородных условиях приложения труда. Следовательно, они отражают различия в производительности труда, связанные главным образом с особенностями работы самих ткачих.

Анализ производственно-хозяйственной деятельности в угольной промышленности

В работе [66] подчеркиваются трудности, возникающие на пути интерпретации извлеченных главных компонент. Трудности возрастают, если для анализа используется регрессионная модель по главным компонентам. В этом случае не рекомендуется сразу отбрасывать главные компоненты, имеющие малый вес, т. е. объясняющие малую долю общей дисперсии исходных признаков (параметров). Однако, несмотря на эти трудности, представляется возможным получить новую дополнительную содержательную информацию, которую регрессионным анализом по исходным признакам получить не удается. Более того, если входные параметры коррелированы, то коэффициенты этой модели не отражают реального влияния входных признаков. На практике реальные технико-экономические процессы часто имеют коррелированные входные признаки. Моделировать такие процессы рекомендуется методом главных компонент.

В качестве исходных параметров было взято 26 показателей, характеризующих работу 91 шахты Подмосковного бассейна. Из 26 признаков 17 имели количественные оценки, а 9 — только качественные характеристики; 9 признаков с качественной характеристикой по определенной методике были преобразованы в количественные оценки.

В качестве оцениваемых параметров были взяты следующие:

1) среднесуточная добыча угля на шахте (нагрузка на шахту);
2) характеристика обводненности шахты;
3) характеристика отжима угля в очистных забоях;
4) среднесуточная нагрузка на лаву;
5) уровень добычи угля из комплексно-механизированных лав в процентах;
6) характеристика гипсометрии;
7) характеристика типа кровли;
8) среднедействующее число очистных забоев;
9) характеристика раскарстованности месторождений;
10) среднемесячное подвигание очистной линии забоя.

Причем для анализа специально не собирались статические данные. В качестве показателей для исследования были использованы данные, формирующие обычный набор признаков для моделирования таких показателей предприятий, как производительность труда, себестоимость и др. В результате проведенного анализа 26 признаков методом главных компонент было получено двадцать шесть компонент.

Автор сумел провести инженерно-экономическую интерпретацию одиннадцати полученных главных компонент, ему не удалось надежно интерпретировать пятнадцать главных компонент. Несмотря на это обстоятельство, было установлено, что первая главная компонента по набору существенно определяющих ее признаков дает обобщенную характеристику горно-геологических условий производства на шахтах.

Остальные десять главных компонент получили следующую интерпретацию:

f_2 — уровень технической оснащенности шахты;

f_4 —характеристика концентрации очистных работ;

f_8 —характеристика процесса подготовительных работ;

f_{10} —условия поддержания горных выработок;

f_{14} —показатели, характеризующие процесс ремонта горных выработок;

f_{17} — характеристика внутришахтного транспорта;

f_{19} — показатели, характеризующие концентрацию горных работ на шахте;

f_{21} — показатели, характеризующие производственные процессы на поверхности шахты;

f_{23} — показатели, характеризующие производственные процессы на прочих подземных работах;

f_{25} — характеристика протяженности горных выработок и разбросанности горных работ на шахте.

Полученные главные компоненты группировались в три класса. Первый класс включал одну первую главную компоненту. Второй класс включал f_4, f_8, f_{19}, f_{21} и f_{23}. Этот класс характеризовал основные производственные процессы на шахтах. Третий класс включал f_2, f_{10}, f_{19} и f_{25}. Этот класс был назван обобщенной характеристикой производственно-технического облика шахты и масштаба производственных мощностей.

Оставшиеся 15 компонент автор работы отнес в четвертый класс. Следует отметить, что такой подход с последующей

классификацией извлеченных главных компонент редко встречается в литературе.

В дальнейшем методом главных компонент были получены экономико-математические модели производительности труда и себестоимости угля на шахтах Подмосковного бассейна.

Анализ производственно-хозяйственной деятельности в цементной промышленности

Было проведено исследование технико-экономических показателей предприятий с помощью метода главных компонент [66]. Под наблюдение было взято 71 предприятие цементной промышленности, и на них изучены следующие показатели:

1) средняя часовая производительность вращающихся печей;
2) часовая производительность цементных мельниц;
3) часовая производительность сырьевых мельниц;
4) среднегодовая стоимость основных промышленно-производственных фондов;
5) удельный вес активной части основных фондов;
6) средняя марка цемента по всем видам;
7) ввод добавок при помоле клинкера;
8) использование календарного времени вращающихся печей (отношение отработанного времени к календарному);
9) использование календарного времени цементных мельниц;
10) использование календарного времени сырьевых мельниц;
11) электровооруженность;
12) численность промышленно-производственного персонала;
13) оборачиваемость оборотных фондов;
14) удельный расход условного топлива;
15) выпуск цемента;
16) расход электроэнергии на 1 т цемента;
17) текучесть кадров;
18) машиновооруженность;
19) ввод в действие новых основных фондов.

Эти признаки охватили технологию производства, качество выпускаемой продукции, степень использования оборудования, показатели, характеризующие экономическую эффективность производства, и др. Для расчетов были взяты следующие четыре производственных показателя:

- выработка натурального цемента на одного работающего;
- фондоотдача;
- среднезаводская себестоимость 1 т цемента;
- рентабельность.

Авторы поставили перед собой задачу исследовать влияние признаков, которые определяют разные стороны производства (19 параметров), на производственно-технические показатели. Для анализа были выделены первые четыре главные компоненты. Они объясняли 70% общей дисперсии исходных признаков. При этом первая компонента объясняла 29,5% суммарной дисперсии процесса, вторая, третья и четвертая компоненты отражали соответственно 17,2; 12,5 и 9,1%.

Остановимся на одном из важных моментов метода — на интерпретации главных компонент. Авторы проанализировали весовые коэффициенты между признаками и главными компонентами. С первой главной компонентой наиболее высокие положительные значения коэффициентов веса были у следующих девяти признаков из девятнадцати: 1, 2, 3, 4, 5, 11, 15, 19, 18. Первые три признака характеризуют производительность оборудования. Размер основных фондов и их структуру определяют признаки 4 и 5. Уровень автоматизации и механизации производства определяют признаки 11 и 18. Признак 15 входит в характеристику мощности предприятий, а признак 19 определяет степень обновления в отрасли.

Все перечисленные девять признаков характеризуют технический уровень в отрасли. Авторы и назвали первую главную компоненту «Технический уровень производства предприятий цементной промышленности».

Если при интерпретации первой компоненты учитывались признаки с наиболее высокими положительными коэффициентами связи, то при интерпретации второй главной компоненты брались во внимание коэффициенты как с положительными, так и с отрицательными значениями. Всего признаков выделено семь. Так, признаки 8, 9 и 10 входят в модель с положительными значениями. Они определяют эксплуатацию основного оборудования в отрасли. Признаки 13, 14, 16, 17 входят в модель с наибольшими коэффициентами веса, имеющими знак «минус». Итак, воздействие этих признаков противоположно результатам действия первых трех признаков. Так, при хорошей организации производства уменьшается период оборачиваемости оборотных фондов. При улучшении организации производства снижается расход топлива (признак 14) и электроэнергии (признак 16). Точно так же можно полагать, что с повышением уровня организации производства уменьшается текучесть кадров. Таким образом, все перечисленные семь признаков влияют на уровень организации производства.

Учитывая, что признаки, имеющие наиболее высокие коэффициенты веса со второй главной компонентой, отражают элементы планирования производства, учета и контроля материально-технических ресурсов, организацию труда, авторы интерпретировали эту компоненту как «Уровень организации производства на предприятиях цементной промышленности».

Третья компонента определяет качество продукции (признаки 6 и 7). Четвертая главная компонента — размер предприятия.

Таким образом, приходим к выводу, что в основе изменчивости девятнадцати взятых для анализа признаков лежат четыре главные компоненты. Если исходные признаки были непосредственно получены из статистики (измерены), то технический уровень производства, уровень организации производства, уровень качества продукции, размер предприятия непосредственно не измеряются. Они получены на основе математического анализа составленной первоначальной модели из девятнадцати признаков методом главных компонент.

Если мы ранее каждое предприятие характеризовали девятнадцатью признаками, то теперь можно судить о предприятии по четырем новым, но обобщенным характеристикам — главным компонентам. Извлеченные четыре главные компоненты были использованы в качестве обобщенных факторов для аппроксимации экономических показателей при помощи линейной регрессионной модели.

Значит, при помощи регрессионного анализа было определено влияние главных компонент на экономические показатели.

Полученные результаты следует представить в таблице, чтобы убедиться в том, какую роль может сыграть подобный анализ в руках руководителя как средство научного подхода к принятию управляющего решения (см. табл. 3.7) [66].

Таблица 3.7

Регрессионный анализ на главных компонентах

Показатель	Обозначение	Главные компоненты				F-отношение	R
		1	2	3	4		
Выработка натурального цемента на одного работающего	b_r	96,783	132,436	−3,789	−66,180	3,232	0,691
	t_{b_r}	8,502	8,820	0,213	3,184		
	π_r, %	46,000	48,000	—	6,000		
Фондоотдача на 1 руб. основных промышленно-производственных фондов	b_r	−1,784	7,304	−0,235	−1,014	2,233	0,552
	t_{b_r}	2,901	9,006	0,245	0,903		
	π_r, %	10,000	89,000	—	1,000		
Среднезаводская себестоимость 1 т цемента	b_r	−0,182	−1,618	0,020	−0,251	2,886	0,654
	t_{b_r}	1,701	11,475	0,120	1,287		
	π_r, %	2,000	96,000	—	2,000		
Рентабельность	b_r	−0,908	1,887	0,594	−0,626	1,947	0,593
	t_{b_r}	2,929	4,450	1,186	1,068		
	π_r, %	26,000	65,000	5,000	4,000		

В табл. 3.7 приняты следующие обозначения:

b_r — коэффициент в уравнении регрессии при r-й главной компоненте;

t_{b_r} — значение t-критерия для проверки значимости отличия от нуля коэффициентов модели;

π_r — доля дисперсии (в процентах);

F — значение F-критерия для проверки значимости модели;

R — коэффициент множественной корреляции.

На основании данных табл. 3.7 можно, например, для рентабельности (y_i) записать:

$$y_i = -0{,}908f_{1i} + 1{,}887f_{2i} + 0{,}594f_{3i} - 0{,}626f_{4i}.$$

Проведенная оценка регрессионной модели и ее составляющих иллюстрирует точность представленных исследований.

Исследование производительности труда на предприятиях цементной промышленности

Анализ динамики отрасли и прогнозирование производительности труда на основе модели главных компонент провели авторы в работе [66]. В качестве исследуемых признаков они выбрали следующие четыре из девятнадцати: 1, 7, 8 и 11. При помощи двух главных компонент им удалось объяснить более 97% общей дисперсии исходных признаков. При этом первая главная компонента явилась генеральной, т. е. связанной коэффициентами веса со всеми четырьмя признаками и признаком t — временем.

Представим матрицу, связывающую обе компоненты с признаками (табл. 3.8) [66].

Из табл. 3.8 видно, что наиболее высокие коэффициенты связи наблюдаются между первой главной компонентой, признаками 1, 8, 11 и временем. Авторы предложили рассматривать первую главную компоненту (f_1) как «Обобщенный фактор развития отрасли». Вторая компонента не связана со временем, значимым коэффициентом веса.

Таблица 3.8

Коэффициенты веса пяти признаков и двух главных компонент

№ признака	Признак	f_1	f_2
1	Средняя часовая производительность вращающихся печей	0,99	0,05
7	Ввод добавок при размоле клинкера	0,47	0,88
8	Использование календарного времени вращающихся печей	0,82	0,49
11	Электровооруженность	0,99	0,03
	Время	0,99	0,04

Регрессионная модель развития отрасли, построенная по полученным двум главным компонентам, имеет вид:

$$y = 167{,}83f_1 + 4{,}18f_2.$$

В этой модели только первая главная компонента оказывается значимой. Из известного предположения, что динамика развития производительности труда определяется перспективами развития производственных возможностей отрасли, отражаемых в данном случае признаками 1, 8, 11, следует, что первую главную компоненту можно использовать для прогнозирования производительности труда. Таким образом, показано, что регрессионный анализ экономических показателей на главных компонентах имеет более широкие возможности, чем «классический» регрессионный анализ.

В нашем исследовании полученные результаты хорошо согласуются с показателями конкретной экономики отрасли. Более того, при обратном преобразовании главной компоненты f_1 через исходные признаки можно получить модель для прогнозирования при помощи исходных признаков и сравнить ее с моделью, полученной обычным регрессионным анализом.

Анализ текучести рабочей силы на шахтах

Анализу причин текучести рабочей силы на шахтах пяти каменноугольных бассейнов была посвящена работа [7]. В ней рассматривались 35 показателей работы 207 шахт. Указанные 207 шахт распределялись по бассейнам в соответствии с данными табл. 3.9 [7].

При этом в состав 35 показателей вошли показатели, которые характеризовали квалификацию рабочих, размеры заработной платы, социально-бытовые условия, возрастной состав, техническую оснащенность шахты, горно-геологические условия и т. д. Были выделены главные компоненты. При помощи регрессионной модели по главным компонентам были изучены различия между шахтами рассматриваемых бассейнов.

Таблица 3.9

Распределение шахт по бассейнам

Бассейн	Количество шахт	Бассейн	Количество шахт
Донбасс	75	Карагандинский	21
Кузбасс	44	Печорский	23
Подмосковный	24	Прочие бассейны	20

При помощи первой главной компоненты было определено, что наибольшие различия существуют между шахтами Подмосковного и Печорского бассейнов. Эти различия были выявлены по второй компоненте.

Третья и последующие компоненты выражали различия, не связанные с принадлежностью к соответствующему бассейну. Так, третья компонента характеризовала шахты по их технической оснащенности, возрастной структуре и заработной плате.

При построении уравнения регрессии по главным компонентам оказалось, что значимыми являются четыре компоненты. При этом наибольшее влияние на текучесть кадров оказывала третья компонента. При первом рассмотрении была обнаружена зависимость между заработной платой и текучестью рабочей силы. С ростом заработной платы росла текучесть рабочей силы на шахтах.

Для выявления причин обнаруженного явления были получены проекции точек выборки на главные компоненты. Проведенные исследования показали, что шахты Печорского бассейна находятся в тяжелых климатических условиях и имеют отдаленное местоположение, однако работники шахт имеют высокую заработную плату. Несмотря на высокую заработную плату, на шахтах Печорского бассейна наблюдается высокая текучесть рабочей силы. На шахтах Подмосковного бассейна работники имеют более низкую заработную плату, но текучесть рабочей силы незначительная.

Таким образом, при построении регрессии по главным компонентам представляется возможным получить дополнительную информацию о связях между различными переменными. Для этого рекомендуется использовать проекции точек выборки на направления главных компонент.

Не следует думать, что методы многомерного статистического анализа стали применять в экономике только с широким внедрением ЭВМ. Они применялись и до 40-х гг. XX в., но не так массово, как в настоящее время.

Большой интерес представляет факторный анализ соответствий. Он позволяет применить метод главных компонент для анализа качественных данных [72]. Так, после проведения переписи населения на основании изучения отраслей, в которых работает та или иная группа населения в разных областях, можно выявить степень сходства областей по профессиональному и отраслевому распределению населения.

3.6. Направления приложения метода главных компонент

Направлений приложения метода главных компонент при проведении исследований в экономике существует множество. Остановимся на некоторых наиболее характерных направлениях. Все, что было рассмотрено выше, относится к так называемому направлению «*R*-техники». При помощи *R*-техники решения, как отмечается в работе [72], к настоящему моменту выполнено около 95% работ, посвященных многомерному статистическому анализу (факторному, методу главных компонент).

Суть заключается в том, что имеется N объектов, оцениваемых по n признакам. Определяется связь между n признаками и n главными компонентами. Для интерпретации, как правило, используется m наиболее весомых главных компонент. Менее распространена «*Q*-техника». При помощи *Q*-техники определяется степень взаимной близости N объектов на основе корреляций n признаков.

В первом случае определяются признаки, наиболее тесно связанные с данной главной компонентой. Во втором случае выделяются группы объектов, наиболее близких друг к другу по рассматриваемым признакам. Они объединяются в группу, характеризующуюся близкими величинами нагрузки (коэффициентов веса) той или иной главной компоненты для данных объектов. Составим матрицу первичных оценок признаков:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nN} \end{pmatrix}.$$

По данной матрице при применении *R*-техники изучается корреляция между признаками (между строками), а при *Q*-технике — корреляция между объектами (столбцами).

Определенный интерес могут представлять «*P*-техника» и «*O*-техника». Эти техники применяются не к N объектам, как это было в предыдущих двух случаях, а к одному объекту исследования.

Рассмотрим *P*-технику. Один объект исследования (одно предприятие) оценивается по некоторому небольшому числу признаков. Эти признаки оцениваются в разные интервалы времени. Можно производить оценки через сутки, неделю, месяц, квартал, год и т. д. Обычно изучается корреляция между каждой парой признаков. Исследования могут проводиться через разные интервалы времени, и тогда можно наблюдать колебания условий процесса. Все зависит от задачи и условий протекания процесса.

P-техника позволяет изучать индивидуальные различия объектов, в то время как при *R*-технике изучаются различные комбинации общих свойств. *O*-техника также используется при изучении одного объекта по большому числу признаков n. Оценка проводится по всем признакам в некоторые моменты времени, например один раз в квартал, ежедневно и т. д., и изучается корреляция уже не между парами признаков, а между парами дней (недель и т. д.).

Представим матрицу первичных замеров у одного объекта n признаков по $t_1, t_2, \dots, t_N$ моментам времени:

$$\mathbf{X} = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nN} \end{pmatrix},$$

где x_{nN} — измерение у одного объекта N-го признака в n-й момент времени ($N = 1, 2, \dots, k$; $n = 1, 2, \dots, l$).

Таким образом, корреляция между двумя столбцами изучается при помощи *O*-техники, а между строками — при помощи *P*-техники. Методами *O*-техники можно, например, изучить процесс реализации n видов изделий в течение N дней одним магазином.

Глава 4
ДАЛЬНЕЙШЕЕ СОВЕРШЕНСТВОВАНИЕ МЕТОДА ГЛАВНЫХ КОМПОНЕНТ И РАСШИРЕНИЕ ОБЛАСТИ ЕГО ПРИЛОЖЕНИЯ

4.1. Уточнение числовых характеристик закона распределения исходных параметров модели

Точность соответствия ковариационной матрицы истинным стохастическим связям будет зависеть от точности определения математического ожидания и дисперсии случайных величин, составляющих многомерное нормальное распределение. Исследования признаков, изучаемых при анализе производственно-хозяйственной деятельности предприятий, показали, что распределения бывают вытянутыми в ту или иную сторону или с утяжеленными хвостами.

Тьюки показал, что по мере удаления истинного распределения от нормального выборочное среднее значение быстро теряет свои свойства наилучшей оценки центра нормального распределения. Тьюки задался целью найти другие оценки центра распределения, которые были бы, если и несколько хуже выборочного среднего значения, но более устойчивыми к отклонениям от нормального распределения. Если рассматривается одномерный случай, то нормальная выборка считается засоренной нормальными выбросами с тем же средним, но с значительно большей дисперсией (табл. 4.1).

Таблица 4.1

Сравнение устойчивости оценок центра распределения

Оценка	Доля засоренности распределения						
	0	0,001	0,005	0,01	0,05	0,2	0,5
$\bar{x}$	1,0	1,011	1,051	1,103	1,554	4,07	20,10
m	1,0	1,010	1,037	1,065	1,258	2,05	5,93

Пусть ε — доля засоренности распределения, тогда эта выборка принадлежит генеральной совокупности с плотностью распределения [102].

$$f(x, \mu) = \frac{(1-\varepsilon)}{\sqrt{2\pi}}\exp\left(-\frac{(x-\mu)^2}{2}\right) + \frac{\varepsilon}{\sqrt{18\pi}}\exp\left(-\frac{(x-\mu)^2}{18}\right).$$

Если при этом наблюдаемая выборка $(x_1, x_2, ..., x_N)$, то $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ не является наилучшей оценкой для параметра μ, где μ — центр распределения генеральной совокупности.

Таким образом, если на основное распределение типа $N(0,1)$ (нормальное с математическим ожиданием, равным нулю, и единичной дисперсией) наложено засоряющее распределение со средним квадратическим отклонением, равным трем, то Тьюки предлагает следующую оценку параметра μ:

$$\bar{x}_\alpha = \frac{1}{N-2[\alpha N]}\sum_{i=[\alpha N]+1}^{N-[\alpha N]} x_i^*.$$

При этом $x_1^* \leqslant x_2^* \leqslant ... \leqslant x_N^*$ — вариационный ряд выборки $(x_1, x_2, ..., x_N)$; $\bar{x}_\alpha$ — α — урезанная оценка среднего значения; $[\alpha N]$ — наибольшее целое число в αN, $0 \leqslant \alpha \leqslant 0{,}5$. Отметим, что при $\alpha = 0$ $\bar{x}_\alpha$ совпадает со средним значением $\bar{x}$, а при $\alpha = 0{,}5$ урезанная оценка среднего значения совпадает с медианой.

Дисперсия $\bar{x}_\alpha$ имеет вид:

$$D_{\bar{x}_\alpha} = \frac{1}{(1-2\alpha)^2}\left\{\sum_{i=[\alpha N]+1}^{N-[\alpha N]} (x_i - \bar{x}_\alpha)^2 + \alpha(x_{[\alpha N]+1} - \bar{x}_\alpha)^2 + \alpha(x_{N-[\alpha N]} - \bar{x}_\alpha)^2\right\}.$$

Случайная величина $\frac{\bar{x}_\alpha - \mu}{\sigma \bar{x}_\alpha}$ имеет приближенно t-распределение с $N - 2\,[\alpha N] - 1$ степенями свободы. Про оценку $\bar{x}_\alpha$ известно, что при большом числе N она имеет приближенное нормальное распределение.

Отметим, что кроме модели Тьюки есть подход Хубера [122], теоретическое обоснование Л. Д. Мешалкина [102], ко-

8—1447

торый рассматривает случай, когда засоряющее распределение не обязательно симметрично относительно оцениваемого параметра, как у Тьюки.

В заключение интересно будет привести результаты расчетов, проведенных Б. П. Титаренко, из которых видно, что при засоренности распределения более 0,001 оценка Хубера m дает более устойчивые результаты [102].

Отметим, что оценка Хубера близка к оценке Тьюки. Изложенный подход дает возможность строить более адекватные модели метода главных компонент и регрессии на главных компонентах.

Если анализ методом главных компонент используется для выработки управляющих решений или корректирующих воздействий на управляемый процесс, то значение устойчивых оценок несомненно возрастает.

4.2. Динамическая модель

В п. 4.1 было показано, что использование робастных (устойчивых) оценок признаков может существенно повлиять на точность обработки исходных данных в методе главных компонент. На конечном этапе — интерпретации полученных результатов — появляются новые трудности. Они заключаются в том, что при изучении влияния некоторых признаков в течение ряда лет в некоторые годы они могут оказаться существенными, а в другие годы — несущественными. Остановимся на этом вопросе.

Интерпретацию полученных результатов компонентного анализа в области проведенного конкретного исследования специалисты обычно проводят совместно с математиками. В зависимости от требований и условий задачи все коэффициенты веса каждой главной компоненты делятся на группы. Первая группа — подмножество W_1 — включает «малые» коэффициенты веса. В данное подмножество обычно включают коэффициенты веса в пределах от нуля до 0,10÷0,15, назначая некоторый граничный (критический) весовой коэффициент $a_{\text{кр}}$. Считаем, что данное подмножество включает и нулевой коэффициент веса. Второе подмножество W_2 содержит все коэффициенты веса, превосходящие $a_{\text{кр}}$.

Некоторые исследователи при определении $a_{\text{кр}}$ используют Q%-ную точку для выборочного коэффициента корреляции. Известны попытки обоснования интервальной оценки для весовых коэффициентов со стороны как отечественных, так и зарубежных ученых, но пока еще нет единого мнения по данному вопросу. Значит, при применении таблиц Фишера—Йейтса [12] для определения $a_{\text{кр}}$ не следует переносить на весовые коэффициенты статистический смысл Q%-ной точки для выборочного коэффициента корреляции. Пока что данный подход можно рассматривать как искусственный прием для замены расплывчатой границы между подмножествами W_1 и W_2 на «четкую» границу.

Таким образом, при рассмотрении вектора весовых коэффициентов главной компоненты в любой статической задаче различают два подмножества:

$$W_1 = \{a_{jr} \leqslant a_{\text{кр}}\};\ a_{jr} = 0 \in W_1.$$

Тогда

$$W_2 = \{a_{jr} > a_{\text{кр}}\}.$$

При интерпретации полученных результатов основное внимание уделяется подмножеству W_2. Более того, название главной компоненты часто формируется по тем признакам, которые обладают наиболее высокими коэффициентами веса в данной главной компоненте. Граница между подмножеством признаков $(W_2 - W_3)$, участвующих в формировании названия, и подмножеством W_3, не участвующим в нем, не является четкой.

Таким образом,

$$W_3 \subset W_2;\ W_2 = W_3 + (W_2 - W_3).$$

Значит, у подмножества W_3 нижняя граница является искусственно четкой, а верхняя граница — нечеткой или обе границы у него нечеткие.

Итак, заключаем, что в статической задаче вертикальное множество коэффициентов веса вектора-столбца a_r состоит из трех вертикальных подмножеств W_1, W_3, $(W_2 - W_3)$.

Перейдем к динамической задаче и для простоты рассмотрим только один столбец весовых коэффициентов за ряд лет t_k $(k = 1, 2, ..., l)$.

Составим матрицу значений r-й главной компоненты за l лет. В этой матрице $n \times l$ имеется n горизонтальных подмножеств $W'_1, W'_2, ..., W'_n$, включающих значимые коэффициенты веса признаков за l лет. Одни коэффициенты веса меняются из года в год. Если они только значимые, то в некоторые годы t_k они могут отсутствовать, переходя в вертикальное подмножество W_1. Другие коэффициенты, наоборот, могут принимать такие числовые значения, что их желательно было бы перевести в подмножество $(W_2 - W_3)$.

Анализ производственно-хозяйственной деятельности предприятий, объединений и отраслей народного хозяйства за ряд лет методом главных компонент показал, что матрицы весовых коэффициентов, сохраняя (если не было коренных изменений) основную структуру, подвержены флуктуациям. Приступим к оценке этих колебаний.

Пусть задана совокупность $Y = \{y_j\}$ признаков y_j. Пусть $\mu w (y_j)$ показывает степень принадлежности y_j к W_j. Тогда можно записать совокупность упорядоченных пар $W_j = \{y_j, \mu w (y_j)\}$. Эта совокупность является расплывчатым множеством W_j и Y. Можно сказать, что $\mu w(y_j)$ — функция, отображающая Y в пространство М. Под М будем подразумевать интервал [0, 1]. Нуль в пространстве М означает низшую степень принадлежности, единица — высшую степень принадлежности. Нас интересует пересечение подмножеств W_i — вертикальных и W'_j — горизонтальных. Наибольшее расплывчатое множество, содержащееся как в W_i, так и в W'_j $(i = \overline{1,3}; j = \overline{1, n})$, обозначим $W_i \cap W'_j$. Функция принадлежности для $W_i \cap W'_j$ определится [43] из соотношения:

$$\mu_{w_i \cap w'_j}(y_j) = \min_{\mu} (\mu_{w_i}(y_j); \mu_{w'_j}(y_j)).$$

Отметим, что $W_1 \cap W'_j = 0$; $W_2 \cap W'_j \geqslant 0$. При определении функции принадлежности исследователь может подобрать такой ее вид, который будет наиболее объективно учитывать роль коэффициентов веса по годам.

Таким образом, можно с большей определенностью интерпретировать полученные результаты в динамических моделях, выявлять признаки, существенно влияющие на процесс в течение всего изучаемого периода (длительного или очень короткого). При этом явления, проявившие свое существенное влияние в течение последнего короткого промежутка времени, могут оказаться теми прогрессивно влияющими признаками, на которые, несмотря на непродолжительность действия, исследователю надо обратить особое внимание.

В заключение заметим, что применение теории расплывчатых множеств и устойчивых оценок функций распределения параметров позволяет более объективно в главной компоненте выделять группы признаков, наиболее существенно влияющих на нее. Такой подход облегчает интерпретацию и анализ многомерной статистической модели и делает эту модель более адекватной наблюдаемому на практике процессу. Следовательно, представляется возможным выработать более эффективное управляющее решение.

Пример 4.1

Определите степень принадлежности признаков к горизонтальным множествам W'_j на основании полученных результатов после проведения анализа методом главных компонент (в течение пяти лет) учебы в техникуме и последующей работы специалистов в СУП. Количество ежегодно испытуемых $N = 77$, причем каждый год это были новые (другие) студенты. Результаты обработки данных представлены в табл. 4.2.

Количество и наименование признаков соответствуют табл. 3.3. Данные за первый год (t_1) являются первым столбцом табл. 3.3. Пусть $a_{кр} = 0{,}19$.

Расчеты степени принадлежности проведем при помощи трех типов критериев.

1. В l-м году

$$\mu_{w'_j}(y_j) = \begin{cases} 1, \text{ если } a_{jl} > a_{кр}; \\ 0, \text{ если } a_{jl} \leq a_{кр}, \end{cases}$$

где a_{jl} — весовой коэффициент j-го признака на изучаемой главной компоненте в l-м году.

Подсчет за ряд лет:

$$\mu_{w'_j}(y_j) = q/p, \tag{4.1}$$

где q — число лет, соответствующих $a_{jl} > a_{кр}$;
p — общее количество рассматриваемых лет.

Таблица 4.2

Динамика изменения коэффициентов веса первой главной компоненты за пять лет

y_j	t_1	t_2	t_3	t_4	t_5
1	0,29	0,34	0,26	0,15	0,14
2	−0,02	0,04	0,21	0,34	0,42
3	0,33	0,35	0,30	0,42	0,36
4	0,73	0,69	0,75	0,70	0,69
5	0,58	0,56	0,60	0,62	0,70
6	0,33	0,40	0,42	0,38	0,36
7	0,57	0,53	0,55	0,52	0,60
8	0,65	0,60	0,62	0,66	0,62
9	−0,42	−0,50	−0,52	−0,55	−0,58
10	0,59	0,62	0,65	0,60	0,61
11	0,46	0,51	0,54	0,56	0,48
12	0,58	0,65	0,66	0,70	0,69

2. В l-м году

$$\mu_{w_j'}(y_j) = \begin{cases} l, \text{ если } a_{jl} > a_{\text{кр}}; \\ 0, \text{ если } a_{jl} \le a_{\text{кр}}. \end{cases}$$

Подсчет за ряд лет:

$$\mu_{w_j'}(y_j) = \sum_{l=1}^{q} l \Big/ \sum_{l=1}^{p} l. \qquad (4.2)$$

3. В l-м году

$$\mu_{w_j'}(y_j) = \begin{cases} l^2, \text{ если } a_{jl} > a_{\text{кр}}; \\ 0, \text{ если } a_{jl} \le a_{\text{кр}}. \end{cases}$$

Подсчет за ряд лет:

$$\mu_{w_j'}(y_j) = \sum_{l=1}^{q} l^2 \Big/ \sum_{l=1}^{p} l^2. \qquad (4.3)$$

Решение

1. Выделим горизонтальные множества. Из табл. 4.2 видно, что только признаки 1 и 2 имели весовые коэффициенты ниже $a_{\text{кр}}$. Выпишем только для них:

$$W'_1 = \{0{,}29;\ 0{,}34;\ 0{,}26\};$$
$$W'_2 = \{0{,}21;\ 0{,}34;\ 0{,}42\}.$$

2. Определим степень принадлежности признаков к горизонтальным множествам по каждому из трех указанных критериев.

По первому критерию:

$$\mu_{w_1'}(y_1) = \frac{3}{5} = 0{,}6;\ \mu_{w_2'}(y_2) = \frac{3}{5} = 0{,}6;$$
$$\mu_{w_j'}(y_j) = 1\ (j = 3, 4, \ldots, 12).$$

По второму критерию:

$$\mu_{w_1'}(y_1) = \frac{1+2+3}{1+2+3+4+5} = \frac{6}{15} = 0{,}4;$$
$$\mu_{w_2'}(y_2) = \frac{3+4+5}{15} = 0{,}8;$$
$$\mu_{w_j'}(y_j) = 1\ (j = 3, 4, \ldots, 12).$$

По третьему критерию:

$$\mu_{w_1'}(y_1) = \frac{1+4+9}{1+4+9+16+25} \approx 0{,}26;$$
$$\mu_{w_2'}(y_2) = \frac{9+16+25}{55} \approx 0{,}91;$$
$$\mu_{w_j'}(y_j) = 1\ (j = 3, 4, \ldots, 12).$$

Результаты, полученные по первому критерию, не вскрывают динамики изменения принадлежности коэффициентов к горизонтальным множествам. Динамика выявляется вторым и третьим критериями. В данной задаче достаточно было бы воспользоваться только вторым критерием. Содержательный причинный анализ показал, что изменения первого признака связаны с установлением определенных ограничений на прием в техникум лиц без специальности или лиц, работавших по специальности, не связанной с электротехникой. Изменение влияния второго признака было связано с пересмотром учебных программ и ликвидацией дублирования вопросов электротехники в других спецкурсах.

В заключение заметим, что метод главных компонент является развивающимся математическим аппаратом, более того, алгоритмы его решения находят применение и в факторном анализе, например в методе главных факторов.

4.3. Метод главных факторов

Метод главных факторов был предложен для модели компонентного анализа:

$$y_j = a_{j1}f_1 + a_{j2}f_2 + \dots + a_{jn}f_n \qquad (j = 1, 2, \dots, n).$$

Затем его применили к модели факторного анализа:

$$y_j = a_{j1}f_1 + a_{j2}f_2 + \dots + a_{jm}f_m + a_jv_j \qquad (4.4)$$
$$(j = \overline{1, n},\ r = \overline{1, m}),$$

где признак y_j зависит от m общих факторов f_r ($r = \overline{1, m}$) и одного характерного (специфического) фактора v_j;

a_{jr} — весовой коэффициент при общих факторах (a_{jr}^2 определяет вклад фактора f_r в общность параметра y_j);

a_j — весовой коэффициент при индивидуальном (характерном) факторе, присущем только j-му признаку.

Обычно под методом главных факторов понимают применение метода главных компонент к редуцированной корреляционной матрице. Редуцированной корреляционной матрицей называют матрицу парных коэффициентов корреляции, в которой на главной диагонали стоят вместо единиц значения оценок общностей $h^2 \neq 1$. Общность характеризует вклад данного признака в суммарную общность процесса.

Остановимся на некоторых методах определения общностей. Укажем, что одним из наиболее простых методов является метод подстановки вместо единицы максимального по модулю коэффициента корреляции данного параметра со всеми остальными параметрами. Это значит, что из строки или столбца, на пересечении которых находится заменяемая в корреляционной матрице единица, выбирается максимальный по модулю элемент. Такой метод оценки общностей весьма эффективен для матриц высокого порядка.

В качестве примера преобразуем корреляционную матрицу, полученную при исследовании путей повышения эффективности работы СУ химического производства. Для этого в матрице (табл. 3.1) заменим единицы, стоящие на диагонали, значениями наибольших коэффициентов корреляции в строке или в столбце (см. табл. 4.3).

Если число параметров мало, то этот способ менее эффективен. Есть и другие способы оценки общностей, например пользуются методом триад:

$$h_j^2 = \frac{r_{jk}r_{jl}}{r_{kl}}, \qquad (4.5)$$

где k и l — номера параметров, коэффициенты корреляции которых с y_j превышают его остальные коэффициенты корреляции.

В числителе оказываются два наибольших коэффициента корреляции в столбце или строке, стоящих на пересечении заменяемой (j-й) единицы. В знаменателе стоит коэффициент корреляции между k-м и l-м признаком. Иногда используют средний коэффициент корреляции данного признака y_j.

Пример 4.2

В матрице табл. 3.1 методом триад получите значение общности для: 1) $y_j = y_5$ и 2) $y_j = y_2$.

Решение

1. Определим два наибольших коэффициента корреляции в строке, соответствующей y_5 и r_{kl}:

$$r_{jk} = r_{54} = 0{,}568;\ r_{jl} = r_{57} = 0{,}399;$$
$$r_{kl} = r_{47} = 0{,}561.$$

Отсюда согласно триаде (4.5)

$$h_5^2 = \frac{r_{54}r_{57}}{r_{47}} = \frac{0{,}568 \cdot 0{,}399}{0{,}561} \approx 0{,}404.$$

2. Определим два наибольших коэффициента корреляции в строке, соответствующей y_2 и r_{kl}:

$$r_{25} = 0{,}181;\ r_{29} = 0{,}166;\ r_{59} = 0{,}059.$$

Отсюда согласно триаде (4.5)

$$h_2^2 = \frac{r_{25}r_{29}}{r_{59}} = \frac{0{,}181 \cdot 0{,}166}{0{,}059} \approx 0{,}567.$$

Таблица 4.3

Редуцированная матрица парных коэффициентов корреляции

y_j	1	2	3	4	5	6	7	8	9	10	11	12
1	0,249	0,037	0,129	0,149	0,121	0,158	−0,013	0,096	−0,249	0,070	−0,028	0,199
2		0,181	0,019	0,054	0,181	−0,093	0,086	0,032	0,166	−0,045	−0,116	−0,141
3			0,227	0,188	0,048	0,098	0,044	0,111	−0,205	0,152	0,059	0,227
4				0,568	0,568	0,323	0,561	0,430	−0,059	0,264	−0,053	0,232
5					0,568	0,042	0,399	0,373	−0,053	0,206	0,020	0,131
6						0,387	0,113	0,081	−0,361	−0,079	0,387	−0,128
7							0,561	0,462	0,069	0,123	0,106	0,104
8								0,462	0,034	0,347	0,167	0,277
9									0,381	−0,354	−0,399	−0,381
10										0,425	0,338	0,425
11											0,394	0,394
12												0,425

Существует много методов оценки общности, но ни один из них не гарантирует достижения минимального ранга редуцированной корреляционной матрицы. Однако опыт исследований показал, что конечные результаты факторного анализа мало чувствительны к значениям общностей, поставленных на главной диагонали корреляционной матрицы. Это может оказаться справедливым для матриц с числом признаков более десяти. Вот почему чаще общность определяют по модулю максимального элемента столбца или строки признака в матрице парных коэффициентов корреляции.

В методе главных факторов сначала находят первый фактор, затем — остаточную матрицу $\mathbf{R}_1$ и определяют ее собственные значения и собственные векторы. Это начало совпадает с методом главных компонент. В методе главных компонент коэффициенты при первом факторе определяются так, чтобы вклад первой главной компоненты в суммарную общность был максимальным. Так,

$$V_1 = \mathbf{a}_1' \mathbf{a}_1 = \sum_{j=1}^{n} a_{j1}^2. \tag{4.6}$$

Максимум V_1 обеспечивается при условии, что

$$r_{jk} = \sum_{r=1}^{m} a_{jr} a_{kr} \ (j, k = \overline{1, n}). \tag{4.7}$$

Если $j = k$, то r_{jj} — общность h_j^2 параметров y_j. Значит, в формуле (4.7) выборочные коэффициенты корреляции заменяются вычисленными, т. е. предполагаем, что остатки равны нулю. Необходимо, таким образом, максимизировать функцию V_1, зависящую от n переменных a_1 при имеющихся $0{,}5\,(n + 1)n$ условий формулы (4.7). Эти условия наложены на коэффициенты a_{jr}. При помощи множителей Лагранжа можно составить функцию [106]:

$$2T = V_1 - \sum_{j,k=1}^{n} \mu_{jk} r_{jk}.$$

Эта функция с учетом формулы (4.7) может быть представлена:

$$2T = V_1 - \sum_{j,k=1}^{n} \sum_{r=1}^{m} \mu_{jk} a_{jr} a_{kr},$$

где $\mu_{jk} = \mu_{kj}$ — множители Лагранжа.

Далее приравняем частные производные функции T по n переменным a_{j1} к нулю, примем, что $V_1 = \lambda_1$ и после преобразований получим однородную систему уравнений:

$$\sum_{k=1}^{n} r_{jk} a_{k1} - \lambda_1 a_{j1} = 0.$$

Перепишем систему уравнений в виде, хорошо знакомом из предыдущего материала. Эта система имеет следующий вид:

$$\left.\begin{array}{l} (h_1^2 - \lambda) a_{11} + r_{12} a_{21} + \ldots + r_{1n} a_{n1} = 0; \\ r_{21} a_{11} + (h_2^2 - \lambda) a_{21} + \ldots + r_{2n} a_{n1} = 0; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ r_{n1} a_{11} + r_{n2} a_{21} + \ldots + (h_n^2 - \lambda) a_{n1} = 0. \end{array}\right\}$$

Значит, максимизация функции V_1 при условиях формулы (4.7) привела к системе n уравнений с n неизвестными a_{j1}. Необходимым и достаточным условием существования ненулевого решения, как мы знаем, является равенство нулю характеристического уравнения:

$$\begin{vmatrix} (h_1^2 - \lambda) & r_{12} & \ldots & r_{1n} \\ r_{21} & (h_2^2 - \lambda) & \ldots & r_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ r_{n1} & r_{n2} & \ldots & (h_n^2 - \lambda) \end{vmatrix} = 0. \qquad (4.8)$$

Значение функции V_1 равно наибольшему значению из корней характеристического уравнения.

Остается найти коэффициенты a_{j1}, обеспечивающие максимально возможную долю суммарной общности h_1^2. Для этого достаточно подставить λ_1 в формулу (4.8) и получить одно из возможных значений $a'_{11}, a'_{21}, \ldots, a'_{n1}$. Полученное значение можно разделить на корень квадратный из суммы квадратов и умножить на квадратный корень из λ_1, тогда будет выполнено условие (4.6):

$$a_{j1} = \frac{a'_{j1} \sqrt{\lambda_1}}{\sqrt{(a'_{11})^2 + (a'_{21})^2 + \ldots + (a'_{n1})^2}},$$

где a_{j1} — коэффициент при первом факторе в факторном отображении формулы (4.4).

Если провести факторный анализ методом главных факторов, то можно получить не квадратную матрицу **A** порядка n, а прямоугольную матрицу **A** порядка $n \times m$. Эта матрица не имеет обратной матрицы. Остается в силе произведение матриц $\mathbf{A'A} = \Lambda$, где Λ — диагональная матрица порядка m:

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_m \end{pmatrix}.$$

Из свойств ортогональности известно, что $\sum_{j=1}^{n} a_{jr}^2 = \lambda_r$ и $\sum_{j=1}^{n} a_{jr} a_{js} = 0$, где $r, s = 1, 2, \ldots, m$, $r \neq s$. После извлечения первого фактора необходимо найти второй фактор, который обеспечит максимум остаточной общности h_2^2:

$$\mathbf{R}_1 = \mathbf{R} - \tilde{\mathbf{R}}_1;$$
$$\tilde{\mathbf{R}}_1 = \mathbf{a}_1\mathbf{a}_1',$$

где $\tilde{\mathbf{R}}_1$ — матрица рассеяния.

Данная матрица является симметрической порядка n.

При определении коэффициентов при втором факторе придется максимизировать функцию $V_2 = \sum_{j=1}^{n} a_{j2}^2$. Коэффициенты корреляции после исключения первого фактора можно записать так: ${}_l r_{jk}$. Например, для $l = 1$ (извлечен первый фактор) по формуле (4.7) получим:

$${}_l r_{jk} = r_{jk} - a_{j1}a_{k1} = a_{j2}a_{k2} + a_{j3}a_{k3} + \cdots + a_{jm}a_{km}.$$

Это выражение может играть ту же роль, что и условия формулы (4.7) при извлечении первого фактора. Однако повторять всю процедуру не придется, так как известно, что максимальное собственное значение матрицы $\mathbf{R}_1$ равно второму собственному значению матрицы $\mathbf{R}$.

Заметим, что кроме λ_1 собственные векторы и собственные значения матрицы $\mathbf{R}_1$ равны соответствующим собственным векторам и собственным значениям матрицы $\mathbf{R}$ [106]. Исключение составляет вектор $\mathbf{a}_1$. Его собственное значение в матрице $\mathbf{R}$ равно λ_1, а в матрице $\mathbf{R}_1$ — нулю. Собственное значение λ_2 матрицы $\mathbf{R}$ является наибольшим собственным значением матрицы $\mathbf{R}_1$.

Дальнейшее извлечение факторов производится таким же образом до получения m факторов, причем собственные векторы и собственные значения получаются по стандартным программам на основе матрицы $\mathbf{R}$, т. е. из матрицы редуцированной, у которой в отличие от матрицы наблюденных коэффициентов корреляции на главной диагонали стоят не единицы, а оценки общностей.

Г. Харман [106] предупреждает, что на практике при преобразовании редуцированной корреляционной матрицы, след которой меньше следа матрицы наблюденных коэффициентов корреляции, могут появиться как положительные, так и отрицательные собственные значения. Отрицательные собственные значения не имеют смысла в факторном анализе. Сумма положительных собственных значений превосходит сумму общностей. Процесс извлечения факторов рекомендуется прекращать в момент, когда сумма собственных значений становится равной сумме общностей.

Сумма общностей, как известно, представляет собой след редуцированной матрицы. Эта величина может быть определена одновременно с редуцированием матрицы. Следовательно, данная легко получаемая величина может стать критерием, определяющим окончание процедуры извлечения главных факторов в алгоритме решения задачи на ЭВМ.

Таким образом, мы с разных сторон рассмотрели широкие возможности приложения метода главных компонент. Вначале показали, что данный метод имеет и нестатистический аспект, а в заключение отметили, что многие процедуры метода главных компонент иногда могут быть использованы в факторном анализе, т. е. еще в одном из способов снижения размерности многомерной статистической информации.

ЧАСТЬ II

ОЦЕНКА ЭФФЕКТИВНОСТИ СИСТЕМ УПРАВЛЕНИЯ

Глава 5

ЭКОНОМИЧЕСКАЯ ЭФФЕКТИВНОСТЬ СИСТЕМ УПРАВЛЕНИЯ

5.1. Проблема экономической эффективности

Во второй части рассматриваются четыре группы показателей:

- использования труда;
- использования основных фондов, оборотных средств и капитальных вложений;
- использования материальных затрат;
- обобщающих показателей.

К *показателям экономической эффективности использования труда* для объединения и предприятия относятся темпы роста производительности труда, доля прироста чистой и товарной продукции в результате повышения производительности труда, экономия живого труда.

Показатели эффективности использования основных фондов, оборотных средств и капитальных вложений включают фондоотдачу, оборачиваемость оборотных средств, отношение прироста чистой продукции к приросту капитальных вложений, срок окупаемости капитальных вложений.

При определении *эффективности использования материальных затрат* учитывают следующие показатели: планируемый уровень материальных затрат на рубль товарной (валовой) продукции, уровень материальных затрат за прошлый период, нормативный уровень материальных затрат на рубль товарной продукции.

К *обобщающим показателям* относят темпы роста производства, производство чистой продукции на рубль затрат, относительную экономию использования прошлого и живого труда, общую рентабельность.

Таковы показатели, предусмотренные для предприятий и объединений. Разработана также система показателей для народного хозяйства всей страны, отраслей материального производства. Например, по народному хозяйству определяются национальный доход и связанные с ним показатели повышения экономической эффективности.

Технической базой СУ служит современная электронно-вычислительная техника, математической основой — широкое использование экономико-математических методов.

В дальнейшем достаточное внимание будет уделено эффективности использования СУ и ЭВМ. Общий подход к оценке эффективности должен учитывать, что современные СУ — это сложные человеко-машинные системы. Исследовать особенности этих систем надо с применением достижений современной математической науки в этой области.

Автор не ограничивается изложением только оценки экономической эффективности, являющейся важнейшей составляющей общей теории эффективности. Проблема рассмотрена шире, при этом учитываются ситуации, когда исследователь не в состоянии рассчитать экономический эффект, но может определить эффект выигрыша во времени, а также выигрыша в объеме обрабатываемой информации, оценить удобства эксплуатации и т. д.

С появлением СУ методы расчета экономической эффективности усложнились. Однако работы ученых советской и русской школ позволили довести до уровня расчетных алгоритмов и методов труднейшие задачи оценки эффективности сложных систем — объединений и отраслей народного хозяйства [63], [64].

Методика [63] являлась обязательной для всех отраслей народного хозяйства. Только на ее основе разрабатывались и утверждались министерствами и ведомствами отраслевые методические указания. Отраслевые методики, учитывающие особенности отрасли в расчетах экономической эффективности СУ, приобретают законную силу только после согласования их с управлением.

Особенностями методик [126], [127] является то, что они в соответствии с ГОСТ 198—75 более четко рассматривают автоматизированные системы управления как человеко-машинные системы.

Одним из важнейших критериев эффективности является экономическая эффективность. Для иллюстрации цикла работы по определению экономической эффективности ниже остановимся на расчетах по методике, изложенной в работе [127], для отраслевой автоматизированной системы управления (СУ).

5.2. Оценка экономической эффективности систем управления

Состав СУ представлен в типовом перечне. Он включает **подсистемы** [64]:

- перспективного развития отрасли;
- технико-экономического планирования;
- оперативного управления;
- управления сбытом продукции;
- управления финансовой деятельностью;
- планирования, учета и анализа труда и заработной платы;
- управления материально-техническим снабжением;
- планирования, учета и анализа кадров;
- управления капитальным строительством;
- бухгалтерского учета;
- научно-технической информации;
- управления научно-исследовательскими работами.

Этот состав может корректироваться в зависимости от особенностей отрасли. Данный перечень достаточно хорошо охватывает круг задач, которые решаются при оценке эффективности СУ.

Экономическая эффективность СУ определяется на основе факторов, имеющих количественные оценки. За базовое значение для сравнения с ожидаемой экономической эффективностью принимаются планируемые показатели производственно-хозяйственной деятельности отрасли, которые предполагается достигнуть в год планируемого ввода СУ, без учета результатов ее внедрения.

При проведении расчета экономии от внедрения конкретной отраслевой СУ принимают во внимание только те элементы экономии, на которые при помощи СУ можно оказать

реальное воздействие, с учетом проектируемого состава подсистем и комплекса задач. Экономические показатели определяются по действующим на момент расчета оптовым ценам, тарифам и ставкам заработной платы.

При определении экономической эффективности СУ необходимо обеспечить сопоставимость всех показателей во времени, по ценам, нормам и элементам затрат.

Экономическая эффективность СУ определяется тремя основными характеристиками [127]:

- годовой экономией от снижения себестоимости продукции;
- годовым экономическим эффектом;
- эффективностью затрат.

1. **Годовая экономия**, или приведенный годовой прирост прибыли от внедрения СУ, является результатом:

- снижения затрат на выпуск всей продукции;
- прибыли вследствие улучшения использования производственных мощностей, дополнительной реализации продукции в результате улучшения плана производства.

Введем следующие обозначения, млн руб.:

A_1 — годовой объем реализации продукции до внедрения СУ;

A_2 — годовой объем реализации продукции после внедрения СУ;

$П_1$ — прибыль от реализации продукции до внедрения СУ;

$\Delta A = A_2 - A_1$ — дополнительный прирост объема выпуска и реализации продукции;

$\Delta П^A$ — дополнительная прибыль, получаемая в результате ликвидации непроизводительных расходов в отрасли, не входящих в себестоимость выпускаемой продукции;

ΔC^A — изменение себестоимости реализуемой продукции вследствие функционирования СУ.

Годовой прирост прибыли может быть рассчитан по формуле

$$Э_{год} = \frac{A_2 - A_1}{A_1} П_1 + \Delta П^A + \Delta C^A, \tag{5.1}$$

где $Э_{год}$ — годовой прирост прибыли.

2. **Годовой экономический эффект** от внедрения СУ определяется по формуле

$$Э = \left(\frac{A_2 - A_1}{A_1}\right) П_1 + ΔП^А + ΔС^А - Е_Н К_Д^А = Э_{год} - Е_Н К_Д^А, \quad (5.2)$$

где $Е_Н$ — нормативный коэффициент экономической эффективности капитальных вложений в данной отрасли;

$К_Д^А$ — дополнительные затраты отрасли на создание СУ, т. е. затраты на его проектирование и внедрение, млн руб.

3. **Эффективность затрат** на создание СУ определяется так:

$$Е_р = Э_{год} / К_Д^А, \quad (5.3)$$

где $Е_р$ — эффективность затрат на создание СУ.

Эффективность затрат на создание СУ определяется также обратной величиной — сроком окупаемости (T):

$$T = К_Д^А / Э_{год}. \quad (5.4)$$

Величины $А_1$ и $Е_Н$, входящие в формулы (5.1) — (5.4), известны. Следовательно, для расчета трех основных характеристик, определяющих экономическую эффективность СУ, необходимо определить $ΔА$, $ΔП^А$, $ΔС^А$, $К_Д^А$.

Расчет дополнительного прироста объема выпуска и реализации продукции

Рассмотрим факторы, которые могут реально повлиять на прирост объема выпуска и реализации продукции в результате внедрения СУ.

1. Одним из основных факторов является оптимизация производственной программы, проводимая на ВЦ СУ. Соответствующий прирост объема выпуска и реализации продукции $ΔА_1^А$ (млн руб.) учитывается в плане. В подсистемах СУ решается одна или несколько следующих задач:

- определение основных показателей оптимального плана развития производства отрасли;
- расчет вариантов оптимальных перспективных планов развития и размещения производства отрасли;
- расчет оптимального текущего производственного плана на предприятий или подотраслей;
- выявление резервов дополнительных мощностей в результате анализа хода выполнения плана.

Перечисленные задачи можно решить, зная годовой объем выпуска и реализации продукции ($А_2^1$, млн руб.), получаемый в результате оптимизации производственной программы:

$$ΔА_1^А = А_2^1 - А_1. \quad (5.5)$$

2. В ходе выполнения плана на ВЦ СУ проводится анализ основных показателей производственно-хозяйственной деятельности отрасли. На основе этого анализа оперативно выявляются и включаются в производство резервы, принимаются меры по обеспечению ритмичности производства и ликвидации срывов в выпуске продукции. Благодаря оперативному получению информации и использованию ее для корректировки хода выполнения плана получают дополнительный прирост объема выпуска и реализации продукции:

$$ΔА_2^А = ΔА_Ф \cdot Г_{НП} \cdot 12, \quad (5.6)$$

где $ΔА_Ф$ — среднемесячное, фактически сложившееся невыполнение плана выпуска продукции до внедрения СУ, млн руб.;

$Г_{НП}$ — среднемесячный коэффициент возможного сокращения невыполнения плана выпуска продукции в результате своевременного поступления информации и принятия решения в условиях функционирования СУ.

3. В подсистемах СУ решается задача учета и контроля труда и заработной платы на предприятиях и в организациях отрасли.

Путем оперативного вмешательства с помощью СУ в кадровую политику, политику труда и заработной платы, приводящего к уменьшению текучести кадров основных рабочих, можно получить дополнительную продукцию. Ее объем ($ΔА_3^А$) определяется так:

$$ΔА_3^А = П_{ТР} \cdot Г_{ПТ} \cdot Ч_У \cdot Г_{СТ}, \quad (5.7)$$

где $Ч_У$ — среднегодовое число основных рабочих, уволенных по неуважительным причинам;

$Г_{СТ}$ — коэффициент снижения текучести основных рабочих в условиях функционирования СУ;

$П_{ТР}$ — фактическая среднегодовая выработка на одного основного рабочего, руб.;

$Г_{ПТ}$ — коэффициент снижения производительности труда основных рабочих при переходе на другую работу.

4. В подсистемах СУ решаются также задачи контроля за ходом выполнения плана капитальных вложений и контроля за ходом строительства пусковых и особо важных объектов. Использование результатов решения этих задач на СУ позволяет добиться досрочного ввода в действие производственных мощностей и дополнительного выпуска продукции. Объем дополнительного выпуска продукции вычисляют по формуле

$$\Delta A_4^A = K_Ф \frac{Ф_{отд}}{12} t_С, \qquad (5.8)$$

где $K_Ф$ — стоимость досрочно вводимых в эксплуатацию основных фондов в результате функционирования СУ, млн руб.;

$Ф_{отд}$ — годовая фондоотдача в отрасли от внедрения СУ, руб.;

$t_С$ — величина сокращения срока строительства в результате функционирования СУ, мес.

5. Кроме вышеперечисленных в методике четырех составляющих на прирост объема выпуска и реализации продукции могут оказывать действие и другие факторы. Обозначим общее число факторов через n, тогда

$$A_2 = A_1 + \sum_{i=1}^{n} \Delta A_i^A. \qquad (5.9)$$

Факторы с индексами $i = 1, 2, 3, 4$ описаны. Для других факторов $i = \overline{5, n}$.

С учетом формулы (5.9) можно записать:

$$A_2 = A_1 + A_2^1 - A_1 + \Delta A_Ф Г_{НП} \cdot 12 + П_{ТР} Г_{ПТ} Ч_У Г_{СТ} + K_Ф \frac{Ф_{отд}}{12} t_С =$$

$$= A_2^1 + \Delta A_Ф Г_{НП} \cdot 12 + П_{ТР} Г_{ПТ} Ч_У Г_{СТ} + K_Ф \frac{Ф_{отд}}{12} t_С. \qquad (5.10)$$

Приведем пример расчета годового прироста объема выпуска и реализации продукции после внедрения СУ.

Пример 5.1

Рассчитанное на ЭВМ значение годового объема выпуска и реализации продукции составляет 3 млрд руб., среднемесячное фактически сложившееся невыполнение плана выпуска

продукции до внедрения СУ равно 1 млн руб., среднемесячное значение коэффициента $Г_{НП} = 0{,}10$. Известны значения: $П_{ТР} = 6000$ руб.; $Г_{ПТ} = 0{,}15$; $Ч_У = 2000$ чел.; $Г_{СТ} = 0{,}40$. Сроки строительства в результате функционирования СУ не снизились, $A_1 = 2969{,}9$ млн руб.

Определите дополнительный прирост объема выпуска и реализации продукции в результате оптимизации производственной программы.

Решение

1. Запишем условие в принятых обозначениях с соблюдением размерности критериев:

$$A_2^1 = 3000 \text{ млн руб.}; \Delta A_Ф = 1 \text{ млн руб.};$$
$$Г_{НП}\ 0{,}10; П_{ТР} = 6000 \text{ руб.};$$
$$Г_{ПТ} = 0{,}15; Ч_У = 2000 \text{ чел.};$$
$$Г_{СТ} = 0{,}40; t_С = 0;$$
$$A_1 = 2969{,}9 \text{ млн руб.}$$

2. С учетом того, что $t_С = 0$, по формуле (5.10) подсчитаем:

$$A_2 = 3000 + 1 \cdot 0{,}10 \cdot 12 + 0{,}006 \cdot 0{,}15 \cdot 2000 \cdot 0{,}40 \approx 3001{,}92 \text{ млн руб.}$$

3. Дополнительный прирост объема выпуска и реализации продукции составит:

$$\Delta A = A_2 - A_1 = 3001{,}92 - 2969{,}9 = 32{,}02 \text{ млн руб.}$$

Расчет дополнительной прибыли в результате сокращения непроизводительных расходов, не входящих в себестоимость продукции

Благодаря оперативному контролю и регулированию финансовых отношений при помощи СУ сокращаются потери по уплате неустоек, штрафов и пени, что позволяет получить дополнительную прибыль. В СУ решаются следующие задачи:

- оптимизация плана распределения фондов по предприятиям и организациям отрасли;
- формирование плана поставок и оперативный контроль хода поставок;
- анализ финансового состояния предприятий и организаций отрасли;
- анализ факторов, влияющих на неудовлетворительное финансовое состояние предприятий.

Решение первых двух задач позволяет сократить потери материальных ресурсов, комплектующих изделий, готовой продукции при хранении на складах. Это дает рост прибыли, которую легко подсчитать по формуле

$$\Delta П_1^Д = У_i Г_{Пi}, \qquad (5.11)$$

где $У_i$ — годовые убытки в отрасли от i-х потерь, млн руб.;

$Г_{Пi}$ — коэффициент i-х потерь в условиях функционирования СУ.

Если известны $Н_{Пi}$ — величина непроизводительных расходов i-го вида в отрасли до внедрения СУ, млн руб., $Г_{непр\ i}$ — коэффициент снижения непроизводительных расходов i-го вида при функционировании СУ, то можно записать дополнительную прибыль, получаемую от использования результатов решения двух последних задач:

$$\Delta П_2^Д = Н_{Пi} Г_{непр\ i}.$$

Следовательно, вся величина дополнительной прибыли в результате сокращения непроизводительных расходов, не входящих в себестоимость продукции, определяется по формуле

$$\Delta П^Д = \sum_{i=1}^{m} \Delta П_i^Д \quad (i = 1, 2, ..., m).$$

Приведем пример расчета.

Пример 5.2

В год, предшествующий внедрению СУ, на хранении было, млн руб.: готовой продукции на складах — 1,5; комплектующих изделий — 2,5; запасных элементов — 1,2. Кроме того, из-за сверхнормативного простоя железнодорожного транспорта на станциях разгрузки предприятий выплачено штрафов 1,8 млн руб., за несвоевременную поставку продукции выплачено штрафов 500 тыс. руб.; за несвоевременную выплату по счетам пени составили 25 тыс. руб. После внедрения в СУ оптимального плана распределения фондов по предприятиям и внедрения оперативного контроля поставок коэффициент снижения потерь составил 0,2. Анализ финансового состояния предприятий и факторов, приводящих к нарушению установленных финансовых требований, позволил при помощи СУ получить коэффициент снижения непроизводительных расходов, равный 0,3.

Определите объем дополнительной прибыли, получаемой в результате сокращения непроизводительных расходов, не входящих в себестоимость продукции.

Решение

1. Запишем условие задачи в принятых обозначениях:

$$У_1 = 1,5 + 2,5 + 1,2 = 5,2 \text{ млн руб.}; \ Г_{П1} = 0,2;$$

$$Н_{П2} = 1,8 + 0,5 + 0,025 = 2,325 \text{ млн руб.}; \ Г_{непр\ 2} = 0,3.$$

2. Определим дополнительную прибыль, полученную в результате улучшения плана распределения фондов и лучшего формирования плана поставок:

$$\Delta П_1^Д = У_1 Г_{П1} = 5,2 \cdot 0,2 = 1,04 \text{ млн руб.}$$

3. Определим дополнительную прибыль, полученную в результате уменьшения потерь из-за штрафов, пени и неустоек:

$$\Delta П_2^Д = Н_{П2} \cdot Г_{непр\ 2} = 2,325 \cdot 0,3 = 0,6975 \text{ млн руб.}$$

4. Определим суммарную дополнительную прибыль:

$$\Delta П^Д = \Delta П_1^Д + \Delta П_2^Д = 1,04 + 0,6975 \approx 1,74 \text{ млн руб.}$$

Расчет снижения себестоимости реализуемой продукции

Благодаря функционированию СУ представляется возможным снизить себестоимость реализуемой продукции:

$$\Delta C^Д = \Delta C_{УП}^Д - C_{экс}^{ВЦ}, \qquad (5.12)$$

где $\Delta C_{УП}^Д$ — годовая экономия условно-переменных расходов, определяемых прямым счетом, в условиях функционирования СУ, млн руб.;

$C_{экс}^{ВЦ}$ — эксплуатационные затраты на содержание главного вычислительного центра отрасли, млн руб.

Рассмотрим составляющие $\Delta C_{УП}^Д$, а затем источники получения данных для расчетов:

$$\Delta C_{УП}^Д = \Delta C_М^Д + \Delta C_Э^Д + \Delta C_{ХМ}^Д + \Delta C_З^Д + \\ + \Delta C_Т^Д + \Delta C_{НИР}^Д + \Delta C_Н^Д, \qquad (5.13)$$

где $\Delta C_М^Д$ — экономия затрат на сырье и материалы, млн руб.;

$\Delta C_{Э}^{A}$ — экономия затрат на электроэнергию и топливо, млн руб.;

$\Delta C_{ХМ}^{A}$ — экономия текущих затрат отрасли в результате уменьшения расходов по хранению запасов материалов, млн руб.;

$\Delta C_{З}^{A}$ — экономия затрат на основную и дополнительную заработную плату (с отчислениями на социальное страхование) основных рабочих в результате пересмотра норм выработки, млн руб.;

$\Delta C_{Т}^{A}$ — экономия затрат на транспортные расходы в результате рационального прикрепления поставщиков к потребителям, млн руб.;

$\Delta C_{НИР}^{A}$ — экономия затрат в результате снижения расходов на проведение научно-исследовательских и опытно-конструкторских работ, млн руб.;

$\Delta C_{Н}^{A}$ — экономия затрат в результате замены ручной обработки информации на автоматизированную, млн руб.

Эксплуатационные затраты на содержание ВЦ определяются по формуле

$$C_{экс}^{ВЦ} = З_{Ф}^{ВЦ} + a^{ВЦ} + C_{М}^{ВЦ} + C_{Т}^{ВЦ} + H_{Н}^{ВЦ} + C_{Э}^{ВЦ} + C_{проч}^{ВЦ}, \quad (5.14)$$

где $З_{Ф}^{ВЦ}$ — основная и дополнительная заработная плата персонала, обслуживающего информационную сеть и ВЦ отрасли, с отчислениями на социальное страхование, определяемая из расчетной численности персонала и окладов по утвержденной схеме должностных окладов, млн руб.;

$a^{ВЦ}$ — амортизация оборудования, млн руб.;

$C_{М}^{ВЦ}$ — затраты на материалы, необходимые для функционирования ВЦ, млн руб.;

$C_{Т}^{ВЦ}$ — затраты на текущий и профилактический ремонт технических средств, млн руб.;

$H_{Н}^{ВЦ}$ — накладные расходы на содержание ВЦ, млн руб.;

$C_{Э}^{ВЦ}$ — затраты на производственное потребление электроэнергии техническими средствами, млн руб.;

$C_{проч}^{ВЦ}$ — прочие затраты, млн руб.

Расчеты по формулам (5.12)—(5.14) не будем приводить, но дадим конечные результаты: $\Delta C_{УП}^{A} \approx 9{,}72$ млн руб., а $C_{экс}^{ВЦ} \approx 1{,}74$ млн руб., которые нужны будут в дальнейших расчетах.

Следовательно, изменение себестоимости реализуемой продукции вследствие функционирования СУ (ΔC^{A}) определится согласно формуле (5.12)

$$\Delta C^{A} = 9{,}72 - 1{,}74 = 7{,}98 \text{ млн руб.}$$

Перейдем к подробным расчетам затрат на создание СУ.

Расчет затрат на создание СУ

Рекомендуется проводить расчет затрат отрасли на создание СУ по формуле

$$K_{Д}^{A} = K_{П}^{A} + K_{О}^{A} \pm \Delta O_{О}^{A} + K_{Л}^{A} - K_{выс}^{A}, \quad (5.15)$$

где $K_{П}^{A}$ — производственные затраты, млн руб.;

$K_{О}^{A}$ — затраты на приобретение технических средств и строительство здания главного вычислительного центра отрасли, млн руб.;

$\Delta O_{О}^{A}$ — изменение величины оборотных средств, млн руб.;

$K_{Л}^{A}$ — остаточная стоимость ликвидируемого оборудования, устройств, зданий, сооружений в отрасли, которые при внедрении СУ не нашли применения и реализация которых невозможна, млн руб.;

$K_{выс}^{A}$ — остаточная стоимость высвобождаемого оборудования, устройств, зданий, сооружений в отрасли, которые будут использованы при внедрении СУ или на других участках производства или реализованы на сторону, млн руб.

Проиллюстрируем вычисление всех составляющих $K_{Д}^{A}$.

1. Производственные затраты определяются по формуле

$$K_{П}^{A} = K_{ПП}^{A} + K_{ПР}^{A}, \quad (5.16)$$

где $K_{ПП}^{A}$ — затраты на обследование объекта, разработку технического задания и другие предпроектные работы, млн руб.;

$K_{ПР}^{A}$ — затраты на техническое и рабочее проектирование и внедрение СУ, млн руб.

Рассмотрим методику расчета составляющих формулы (5.16):

$$а)\ K_{ПП}^{A} = Т_{ПР} \cdot З_{ИТР}(1 + Н_{ДС} + Н_{Н}), \tag{5.17}$$

где $Н_{ДС}$ — коэффициент, учитывающий дополнительную заработную плату и отчисления на социальное страхование;

$Н_{Н}$ — коэффициент, учитывающий накладные расходы;

$Т_{ПР}$ — трудоемкость разработки технического задания и других предпроектных работ, чел.-дни;

$З_{ИТР}$ — средняя месячная заработная плата одного инженерно-технического работника, руб.;

$$б)\ K_{ПР}^{A} = З_{ИТР}(Т_{ТП} + Т_{РП} + Т_{ВН})(1 + Н_{ДС} + Н_{Н}), \tag{5.18}$$

где $Т_{ТП}$ — трудоемкость разработки технического проекта, чел.-дни;

$Т_{РП}$ — трудоемкость разработки рабочего проекта, чел.-дни;

$Т_{ВН}$ — трудоемкость внедрения СУ, чел.-дни.

Определение составляющих в формуле (5.18) — весьма сложный процесс. Кроме того, эта формула носит условный характер. Например, в ней не учтена стоимость машинного времени на отладку программ СУ, хотя трудоемкость этой работы в ряде случаев может превзойти суммарную трудоемкость всех трех учитываемых компонент. Указанное упрощение, как и те, на которых мы не останавливались, неизбежно при использовании «лобового» метода, или метода «прямого счета».

Однако, несмотря на эти недостатки, которые могут быть постепенно устранены, рассматриваемая методика позволяет при некотором приближении решать задачу определения экономической эффективности внедрения СУ с единых методологических посылок.

2. Для вычисления затрат на приобретение технических средств и строительство здания ВЦ СУ ($K_{О}^{A}$) необходимо определить три составляющие:

$$K_{О}^{A} = K_{КТС}^{A} + K_{ИП}^{A} + K_{С}^{A}, \tag{5.19}$$

где $K_{КТС}^{A}$ — затраты на приобретение, транспортировку, монтаж и наладку комплекса технических средств ВЦ как серийного, так и несерийного выпуска, млн руб.;

$K_{ИП}^{A}$ — затраты на оборудование информационных и кустовых пунктов и каналов связи СУ, млн руб.;

$K_{С}^{A}$ — затраты на строительство здания ВЦ.

В затраты $K_{ИП}^{A}$ входит стоимость подготовки помещений. Объем затрат зависит от количества пунктов связи, которыми будет располагать данная отрасль, однако стоимость приобретения телетайпов сюда не входит. Они остаются собственностью министерства связи.

Затраты на строительство здания ВЦ определяются по формуле

$$K_{С}^{A} = У_{П}(Б_{Т} + Б_{В} + Б_{Ц}), \tag{5.20}$$

где $У_{П}$ — стоимость 1 м² площади ВЦ, которая определяется по «Единым районным единовременным расценкам», руб.;

$Б_{Т}$, $Б_{В}$, $Б_{Ц}$ — площадь, необходимая соответственно для монтажа технических средств, размещения вспомогательного оборудования, размещения персонала ВЦ.

По формуле (5.20) $K_{С}^{A}$ определяется на этапе разработки технического задания. На этапе технического и рабочего проектирования затраты определяются по проекту сметы проектной организации. На стадии внедрения затраты определяются по утвержденной смете.

3. Определим следующую составляющую формулы (5.15) — $\Delta O_{О}^{A}$. Изменение величины оборотных средств достигается путем сокращения сверхнормативных запасов сырья, материалов и комплектующих изделий. Это может быть достигнуто путем решения в подсистемах СУ следующих задач:

- расчет сводной потребности отрасли в материалах;
- прогнозирование потребности отрасли в материальных ресурсах.

После решения в подсистемах СУ указанных задач можно подсчитать $\Delta O_{О}^{A}$ по формуле

$$\Delta O_{О}^{A} = (О_{Ф} - О_{Н} Г_{В}) Г_{СВ}, \tag{5.21}$$

где $О_{Ф}$, $О_{Н}$ — фактические и нормативные среднегодовые остатки оборотных средств по тем же видам ценностей, млн руб.;

$Г_{В}$ — коэффициент роста выпуска продукции в условиях функционирования СУ;

$Г_{СВ}$ — коэффициент снижения сверхнормативных остатков оборотных средств в связи с решением в СУ двух упомянутых выше задач.

4. Значения двух других составляющих формулы (5.15) — $K_{Л}^{A}$ и $K_{выс}^{A}$ — выявляются при функционировании СУ и зависят от конкретных условий.

Рассмотренную методику проследим при решении следующего примера.

Пример 5.3

Средняя дневная заработная плата одного инженерно-технического работника без учета дополнительной заработной платы, отчислений на социальное страхование и накладных расходов на этапе разработки технического задания условно равна 5 руб.

$Т_{ПР}$ = 10 000 чел.-дней;	$Б_{П}$ = 1500 м^2;
$Н_{ДС}$ = 0,16;	$K_{КТС}^{A}$ = 7 млн руб.;
$Н_{Н}$ = 0,5;	$K_{ИП}^{A}$ = 0,25 млн руб.;
$Т_{ТП}$ = 90 000 чел.-дней;	$K_{Л}^{A}$ = 2,5 млн руб.;
$Т_{РП}$ =160 000 чел.-дней;	$О_{Ф}$ = 12,6 млн руб.;
$Т_{ВН}$ =150 000 чел.-дней;	$О_{Н}$ = 6 млн руб.;
$Ү_{П}$ = 200 руб.;	$Г_{В}$ = 1,2%;
$Б_{Т}$ = 1000 м^2;	$Г_{СВ}$ = 0,2%;
$Б_{В}$ = 200 м^2;	$K_{выс}^{A}$ = 1 млн руб.

Проведите расчет затрат на создание СУ по изложенной методике.

Решение

1. Определим по формуле (5.17) предпроектные затраты:

$$K_{ПП}^{A} = 10\,000 \cdot 5(1 + 0{,}16 + 0{,}5) = 83\,000 \text{ руб.} = 0{,}083 \text{ млн руб.}$$

2. Определим по формуле (5.18) затраты на техническое и рабочее проектирование:

$$K_{ПР}^{A} = 5 \cdot (90\,000 + 160\,000 + 150\,000) \cdot (1 + 0{,}16 + 0{,}5) = $$
$$= 3\,320\,000 \text{ руб.} = 3{,}32 \text{ млн руб.}$$

3. Определим по формуле (5.16) производственные затраты:

$$K_{П}^{A} = 0{,}083 + 3{,}32 = 3{,}403 \text{ млн руб.}$$

4. Рассчитаем по формуле (5.20) затраты на строительство здания ВЦ:

$$K_{С}^{A} = 200(1000 + 200 + 1500) = 540\,000 = 0{,}54 \text{ млн руб.}$$

5. Определим по формуле (5.19) затраты на оборудование и строительно-монтажные работы:

$$K_{О}^{A} = 7 + 0{,}25 + 0{,}54 = 7{,}79 \text{ млн руб.}$$

6. Определим по формуле (5.21) экономию в результате изменения остатков оборотных средств:

$$\Delta O_{О}^{A} = (12{,}6 - 6 \cdot 1{,}2)0{,}2 = 1{,}08 \text{ млн руб.}$$

7. Определим по формуле (5.15) дополнительные затраты отрасли на создание СУ:

$$K_{Д}^{A} = 3{,}403 + 7{,}79 - 1{,}08 + 2{,}5 - 1 \approx 11{,}61 \text{ млн руб.}$$

Задачи на вычисление годового экономического эффекта внедрения СУ и сроков окупаемости

Задачи, решенные выше в данном параграфе, позволили определить ряд характеристик для дальнейшего расчета основных критериев эффективности.

С учетом ранее вычисленных данных решим следующие задачи.

Пример 5.4

Определите годовой прирост прибыли от внедрения СУ $Э_{год}$, если известны следующие данные, млн руб.:

прибыль от реализации продукции до внедрения СУ — 800;

годовой объем реализации продукции до внедрения СУ — 2969,9;

годовой объем реализации продукции после внедрения СУ (пример 5.1) — 3001,92;

дополнительная прибыль в результате ликвидации непроизводительных расходов в отрасли, не входящих в себестоимость выпускаемой продукции (пример 5.2), — 1,74;

ожидаемое снижение себестоимости реализуемой продукции благодаря функционированию СУ — 7,98.

Решение

1. Запишем в принятых обозначениях условие задачи, млн руб.:

$$A_1 = 2969{,}9;\ A_2 = 3001{,}92;$$

$$\Pi_1 = 800;\ \Delta\Pi^A = 1{,}74;$$

$$\Delta C^A = 7{,}98.$$

2. Вычислим годовой прирост прибыли от внедрения СУ по формуле (5.1):

$$Э_{год} = \frac{(3001{,}92 - 2969{,}9)}{2969{,}9} \cdot 800 + 1{,}74 + 7{,}98 = 18{,}36 \text{ млн руб.}$$

Пример 5.5

Рассчитайте годовой экономический эффект от внедрения СУ, используя данные примера 5.4. Кроме того, известно, что нормативный коэффициент экономической эффективности капитальных вложений в отрасли равен 0,4, а дополнительные затраты на создание СУ в соответствии с примером 5.3 равны 11,61 млн руб.

Решение

1. Запишем условие задачи в принятых обозначениях, млн руб.:

$$A_1 = 2969{,}9;\ A_2 = 3001{,}92;$$

$$\Pi_1 = 800;\ \Delta\Pi^A = 1{,}74;$$

$$\Delta C^A = 7{,}98;\ E_Н = 0{,}4;\ K_Д^A = 11{,}61.$$

2. Определим годовой экономический эффект от внедрения СУ по формуле (5.2):

$$Э = 18{,}36 - 0{,}4 \cdot 11{,}61 \approx 13{,}72 \text{ млн руб.}$$

Пример 5.6

Выясните, эффективна ли СУ, и определите расчетный коэффициент эффективности затрат, если известны дополнительные затраты на создание СУ — 11,61 млн руб. (пример 5.3) и годовой прирост прибыли от внедрения СУ — 18,36 млн руб. (пример 5.4).

Решение

1. Определим расчетный коэффициент эффективности затрат:

$$E_Р = \frac{Э_{год}}{K_Д^A} = \frac{18{,}36}{11{,}61} = 1{,}58.$$

2. Определим эффективность внедрения СУ. Для этого сравним нормативный коэффициент капитальных вложений с полученным расчетным значением: $0{,}4 < 1{,}58$.

Следовательно, внедрение СУ вполне себя оправдывает.

Пример 5.7

Определите срок окупаемости СУ, если известны значения дополнительных затрат на создание СУ: $K_Д^A = 11{,}61$ млн руб.; годовой прирост прибыли от внедрения СУ $Э_{год}$ — 18,36 млн руб.

Решение

$$Т = \frac{11{,}61}{18{,}36} \approx 0{,}63 \text{ года.}$$

Следовательно, стоимость внедрения СУ окупится менее чем за 8 месяцев функционирования.

Таким образом, мы закончили рассмотрение методики [64] оценки экономической эффективности отраслевых автоматизированных систем управления.

Решенные условные примеры и задачи убедили нас в том, что проведение расчетов часто является весьма длительным процессом. Подобные расчеты проводятся на ЭВМ.

В данной книге мы провели их вручную. Некоторые специалисты называют рассмотренную методику методикой «прямого счета». Однако если обратить внимание на исходные данные, то можно убедиться в том, что без ЭВМ их рассчитать не удается.

Вследствие оптимизации производственной программы можно получить прирост объема выпуска и реализации про-

10–1447

дукции ΔA_1^A. Известно, что оптимизация планов является достаточно сложной проблемой. Чтобы модели планов были более адекватны реальным процессам, необходимо учитывать и возможные случайные отклонения при реализации планов, что требует применения методов теории вероятностей.

Отметим, что в данном параграфе для получения общего эффекта мы пользовались суммой частных эффектов (составляющих). При определении эффективности затрат на создание СУ делили полученный эффект ($Э_{год}$) на затраты $K_Д^A$.

Примеры, которые были решены в этой главе, показали, насколько разнообразны методы и подходы при определении экономической эффективности СУ. При этом самой сложной задачей было решение вопросов о том, эффективно ли внедрять СУ и как быстро окупятся затраты. Однако с появлением СУ — сложных систем — появилась необходимость разработки методов оценки эффективности сложных систем.

Глава 6

ОСНОВЫ ТЕОРИИ ЭФФЕКТИВНОСТИ СЛОЖНЫХ СИСТЕМ

6.1. НТР и сложные системы

Управление производством и отраслями народного хозяйства в настоящее время осуществляется при помощи автоматизированных систем управления, обеспеченных современными вычислительными машинами. Производственные объединения, а тем более отрасли народного хозяйства, управляемые СУ, являются типичными представителями сложных систем. Задачи, решаемые СУ, можно рассматривать как математические модели анализа и синтеза сложных систем.

Чтобы управлять сложной системой, необходимо составить прогноз поведения управляемого объекта, рассмотреть множество вариантов управляющих воздействий и выбрать из них наиболее эффективные. Изучение прогнозов и процессов управления может быть проведено при помощи современных математических моделей на ЭВМ.

Существование различных типов ЭВМ, разных принципов математического обеспечения ранее приводило к большим затратам на программирование, затрудняло типизацию задач анализа систем и обмен информацией при эксплуатации СУ. Такие разнородные системы невозможно и нецелесообразно было объединять в единую государственную систему. В современных СУ эти существенные недостатки ликвидированы.

Одной из важнейших проблем — проблемой унификации математических схем, используемых при моделировании, и методов машинного анализа моделей — занимается теория сложных систем.

В нашей стране в этой области работы успешно проводились академиком В. М. Глушковым, членом-корреспондентом АН СССР Н. П. Бусленко, а также академиком АН России С. А. Айвазяном. Эти школы являются центрами развития и решения задач для некоторых классов реальных сложных систем. При решении этих задач представляется возможным оценивать эффективность и другие важнейшие характеристики сложных систем.

С усложнением объектов управления и систем управления появляется необходимость более глубокого анализа их свойств. В результате возрастает объем информации, получаемой с ЭВМ при моделировании. Информация при этом оказывается настолько обширной, что затрудняет, а в некоторых случаях исключает практическое осмысление ее человеком. Поэтому используются специальные методы обработки результатов моделирования, которые позволяют представить информацию в удобном и доступном для человека виде.

В основном используются методы многомерной классификации, позволяющие существенно «сжать» количественную информацию. Поведение сложной системы не всегда удается оценить количественными показателями. В этом случае ее можно охарактеризовать только качественно. Это позволило разработать качественную теорию сложных систем. Данное направление развивается на основе обобщения классических постановок задач при применении дифференциальных уравнений, теории динамических систем, теории случайных процессов.

В последнее время в исследовании сложных систем применяется имитационное моделирование.

В настоящее время деление систем на простые и сложные является весьма условным. Разные авторы по-разному определяют понятие сложной системы.

6.2. Понятие сложной системы

Одно из первых определений сложной системы было дано в 1973 г. Н. П. Бусленко в работе [16]. Он писал, что систему надо считать сложной, если она состоит из большого числа взаимосвязанных и взаимодействующих элементов и способна выполнять сложную функцию. Против такого достаточно общего определения трудно возразить. Позднее в работе [17] им было дано определение, характеризующее некоторые основные свойства подобных систем:

«Сложная система является многоуровневой конструкцией из взаимодействующих элементов, объединяемых в подсистемы различных уровней» [17], а математическая модель сложной системы «состоит из математических моделей элементов и математической модели взаимодействия между элементами» [17].

Наконец, в «Кратком экономико-математическом словаре» при пояснении термина «сложная система» вообще не дается определения, а приводятся основные свойства сложных систем[1].

Примеры сложных систем приводятся в работе [16]: энергетические комплексы, телефонные сети крупных городов, информационные системы, производственные процессы крупного предприятия, системы управления полетом в крупных аэропортах, отраслевые системы управления и др.

В качестве **основных свойств сложных систем** можно выделить следующие [16], [17]:

1) большое число взаимосвязанных и взаимодействующих элементов;

2) сложность выполняемой функции для достижения цели функционирования;

3) иерархическую структуру, возможность деления системы на подсистемы;

4) наличие управления, интенсивных потоков информации и разветвленной информационной сети;

5) взаимодействие с внешней средой и функционирование в условиях воздействия случайных факторов.

Первое свойство не требует пояснений.

Второе свойство определяет основную особенность системы. В сложной системе выполняются задачи, которые обеспечивают достижение промежуточных и конечной целей функционирования. Проектируя сложные системы, необходимо прогнозировать их поведение при выполнении этих задач.

Так как на реальные системы воздействует большое число случайных факторов, для прогнозирования поведения сложной системы необходимо использовать теорию вероятностей. Таким образом, параметры моделей прогноза могут быть охарактеризованы законами распределения. Случайные отклонения системы от нормального режима функционирования определяются возмущающими факторами внешней среды и возмущающими факторами, возникающими внутри системы.

[1] *Лопатников Л. И.* Краткий экономико-математический словарь. — М.: Наука, 1979. — С. 256—257.

Внутренними факторами являются ошибки измерительных приборов в пределах допусков, выход из строя отдельных элементов, ошибки людей, работающих в системе, ошибки в управляющей информации, сбои вычислительных устройств. Случайные возмущения иногда могут привести к вынужденному изменению структуры системы.

Нарушение нормального режима функционирования в сложной системе не приводит к нарушению функционирования в целом, но снижает эффективность и качество ее работы. Значит, учет случайных факторов при исследовании сложных систем и определении их эффективности играет большую роль.

Третье свойство заключается в том, что сложная система обладает свойством иерархичности, т. е. возможностью разбиения системы на подсистемы. Цели функционирования подсистем подчинены общей цели функционирования всей системы. Следовательно, сложной системе присуще еще и обобщающее свойство целостности. Данное свойство означает, что изменения, происшедшие с ее элементами, влияют на другие элементы или подсистемы и оказывают влияние на функционирование всей системы. Значит, при изучении сложных систем необходим системный подход, т. е., исследуя какую-то подсистему, мы обязаны учитывать цели функционирования всей сложной системы.

Таким образом, сложная система состоит из отдельных подсистем и является целостным объектом, отдельные части которого функционируют во взаимодействии. С формальной точки зрения любая совокупность элементов системы может считаться подсистемой. В практике исследования выделение подсистемы проводится таким образом, чтобы цели функционирования подсистемы вытекали из целей функционирования системы.

Процесс управления может осуществляться, если подсистема состоит из взаимосвязанных и совместно функционирующих элементов. Что же собой представляет элемент? **Элемент системы** — это совокупность средств, которая при данном исследовании рассматривается как целое и дальнейшему дроблению не подвергается. Внутренняя структура элемента не является предметом изучения. При формализации исследуемого процесса под элементом иногда понимают коллектив людей, оператора, руководителя подразделения и т. д.

Даже если с формальной точки зрения это оказывается удобным, понятно, как велико отличие элемента «человеческий коллектив» от элемента «совокупность технических средств».

Участие человека в управлении СУ часто приводит к неожиданным результатам. Особенно ярко это проявляется в аварийных ситуациях, когда «совокупность технических средств» не в состоянии провести незапрограммированные действия по ликвидации аварийной ситуации, а человек-специалист вполне может справиться с такой задачей.

Расчленение системы на элементы является важным шагом при формальном описании системы. Подсистему можно считать элементом сложной системы.

Благодаря иерархической структуре сложные системы обладают большими преимуществами.

Сложность некоторых объектов исключает их изучение в целом. Тогда они расчленяются на конечное число подсистем с учетом связей между ними. Далее слишком сложные подсистемы делятся на части. Расчленение ведется до подсистем, не подлежащих дальнейшему дроблению на части в данной задаче, т. е. до элементов.

Качество управления во многом зависит от степени централизации управления, которая определяется иерархической структурой управляемой системы. В народном хозяйстве степень централизации управления меняется в зависимости от сложности, объема и важности решаемых задач.

Четвертое свойство говорит о наличии управления в сложной системе. Процесс управления в общем случае включает получение исходной информации о системе и окружающей среде, переработку и преобразование этой информации, выработку управляющего решения, постановку задач дальнейшего функционирования системы и контроль исполнения. Выработка управляющего решения осуществляется на управляющих ЭВМ. Эффективность работы ЭВМ может быть повышена путем подбора лучшей дисциплины диспетчеризации, увеличения объема и рационального распределения памяти.

Совокупность преобразований, которым подвергается информация, называют **оператором переработки информации.** Обычно выделяют оператора первичной обработки информации, который осуществляет все этапы ее преобразования, включая запоминание ее в накопителях.

Оператор вторичной обработки включает подготовку исходных данных для принятия решения. Оператор управления включает принятие решения и выработку управляющих команд. Наконец, оператор последующей обработки информации включает подготовку информации для передачи и передачу ее к управляемым элементам системы.

Существуют системы управления, в которых действует принцип самоорганизации. Примером могут служить системы с встроенными устройствами контроля работы отдельных узлов и элементов с автоматическим удалением из системы элементов, не удовлетворяющих техническим требованиям, и устройством включения вместо них исправных резервных элементов; системы с переключающими устройствами, которые при увеличении, например, числа потребителей электроэнергии могут включать дополнительные агрегаты, чтобы обеспечить нормальное питание во всей сети. При этом структура функционирующей системы меняется: вместо одних элементов включаются другие элементы. В современных вычислительных центрах СУ имеются управляющие программы, которые автоматически перестраивают дисциплину диспетчеризации в зависимости от характера задач, поступающих для решения.

Характерной чертой управления системой являются самонастройка и самообучение. В самонастройке и самообучении кроме технических средств управления большую роль играют люди. Они приводят в действие все средства системы и управляют ими, определяют цели функционирования подсистем, оценивают степень достижения поставленных целей и ставят новые цели. Участие человека в управлении СУ меняет облик всей сложной системы.

С помощью СУ в первую очередь автоматизируются трудные для человека процессы, требующие длительного времени. Это процессы сбора, хранения, обновления и обработки информации. Системы, в которых автоматизируются эти четыре составные части управления, называются автоматизированными системами информационного обеспечения (СИО). Таким образом, СИО — составная часть СУ. Они явились первым этапом на пути создания СУ.

Наиболее ответственной частью управления является прогнозирование поведения системы в зависимости от различных условий функционирования. На основе прогноза составляется план функционирования системы.

Таким образом, эффективность выработки решения в заданное время во многом зависит от эффективной организации работы вычислительного центра, где проводится обработка поступающей информации.

Пятое свойство заключается во взаимодействии с внешней средой и функционировании в условиях воздействия случайных факторов. Это свойство подчеркивает то обстоятельство, что сложные системы в отличие от абстрактных моделей функционируют в реальных условиях, когда на них оказывает влияние большое число случайных факторов, возникающих как вследствие воздействия внешней среды (например, поступление комплектующих элементов от предприятий-смежников на сборочный завод в более поздние сроки, чем предусмотрено планом), так и в результате возмущений внутри самой системы (прибытие нового пополнения работников, не имеющих достаточного опыта работы). Влияние внешних и внутренних случайных факторов сказывается на функционировании элементов и подсистем сложной системы.

6.3. Функциональные характеристики сложных систем

К функциональным характеристикам сложных систем относятся показатели эффективности, надежности, качества управления, помехозащищенности и устойчивости [16]. Основным показателем является эффективность. Она характеризует качество функционирования сложной системы. Определим цели и задачи функционирования сложной системы.

Выбор системы показателей эффективности производится при определении целей и задач сложной системы. Система показателей эффективности должна определять степень приспособленности сложной системы к выполнению поставленных перед нею задач, учитывать основные особенности сложной системы (характер ее взаимодействия с внешней средой, условия возникновения и характер воздействия внутренних возмущающих факторов, структуру системы).

Каждый показатель эффективности (R) определяется множеством параметров, но всегда считают, что основными являются параметры системы $X_1, X_2, ..., X_n$ и параметры внешней среды $Y_1, Y_2, ..., Y_m$. В результате он может быть представлен в общем виде следующей формулой:

$$R = R(X_1, X_2, ..., X_n, Y_1, Y_2, ..., Y_m). \qquad (6.1)$$

Задача вычисления показателей эффективности сложных систем является весьма сложной, но решается на современных ЭВМ. Значения показателей эффективности выражаются действительными числами.

Сложные системы функционируют при воздействии на них большого количества случайных факторов, поэтому результаты их функционирования часто носят случайный характер. В качестве показателей эффективности очень часто пользуются вероятностями некоторых случайных событий или средними значениями (математическими ожиданиями). Так, можно определять среднее число пассажиров, отправляемых из аэропорта в течение суток, среднее число изделий, выпускаемых предприятием в течение недели. Можно также определять вероятность решения задач, поступающих со средней интенсивностью на вычислительный центр, если среднее время их решения определено, и т. д.

Поскольку на систему в разное время действуют различные случайные факторы и условия функционирования могут меняться с течением времени, необходимо всегда указывать, в каких условиях и для какого момента времени проведен расчет показателя эффективности. Для большинства систем не найден общий, единый показатель эффективности. У многих сложных систем, как правило, существует несколько целей функционирования, каждая из которых характеризуется своим показателем эффективности. Желательно использовать такие показатели, которые позволяют согласовать «конфликтующие» между собой тенденции в задачах и целях функционирования системы.

Предположим, что надо оценить эффективность такой сложной системы, как комбинат. В качестве критерия эффективности можно принять приведенные затраты:

$$З_П = C + E_H К, \qquad (6.2)$$

где $З_П$ — приведенные затраты;

C — текущие затраты, включая амортизацию (себестоимость продукции);

К — капитальные вложения;

E_H — нормативный коэффициент эффективности.

Для уменьшения приведенных затрат необходимо экономить сырье и материалы, уменьшать расходы электроэнергии и топлива, снижать фонд заработной платы, уменьшать капитальные вложения до определенного предела и т. д.

Если же в качестве критерия эффективности принять производительность оборудования, то внимание надо сосредоточить на параметрах, которые влияют на этот показатель. Такие параметры, как фонд заработной платы, себестоимость продукции, расход топлива и электроэнергии, отодвигаются на второй план. Эти разнородные тенденции можно согласовать, если ввести некоторые ограничения для одного из критериев. Можно использовать критерий производительности оборудования, при котором приведенные затраты не превышают заданной величины.

Следовательно, приходится либо использовать множество (набор) критериев эффективности, либо общие критерии эффективности с ограничениями.

Для получения показателей эффективности результаты функционирования элементов системы необходимо выражать в количественной форме. Количественное выражение результатов функционирования элементов системы называют *характеристиками функционирования средств системы.*

Показатели эффективности функционирования средств системы получаются путем преобразования характеристик функционирования элементов системы. Показатели эффективности функционирования средств системы называются *частными показателями.* Они не характеризуют роли данной подсистемы в достижении целей всей системы. В связи с этим есть смысл рассматривать не сам частный показатель эффективности, а приращение показателя эффективности системы, полученное вследствие функционирования данной подсистемы.

Рассмотрим приращение показателя эффективности системы [16], которое получается вследствие функционирования i-й подсистемы первого уровня:

$$\Phi_l^{C(1,i)} - \Phi_l^{O(1,i)}, \tag{6.3}$$

где $\Phi_l^{C(1,i)}$ — l-й показатель эффективности функционирования всей системы;

$(1, i)$ — символ, обозначающий первый иерархический уровень рассматриваемой подсистемы (1), и ее номер (i) на этом иерархическом уровне;

l — индекс показателя эффективности, соответствующий l-й цели функционирования;

$$l = \overline{1, k};\ i = \overline{1, n};$$

С — индекс, означающий, что критерий эффективности относится ко всей системе;

$\Phi_l^{O(1,i)}$ — показатель эффективности функционирования средств системы при условии, что i-я подсистема первого уровня совершенно не участвует в достижении l-й цели системы, но все ресурсы системы между остальными подсистемами распределены оптимальным образом.

При таком подходе оказывается возможным выделить вклад каждой системы первого уровня в достижении l-й цели функционирования системы в целом. Меняя l, можно выявить, каким образом данная $(1, i)$-я подсистема участвует в достижении различных целей системы. Точно таким же образом можно определить вклад любой подсистемы второго уровня $(2, ij)$ и т. д. Определенные трудности вызывает размерность критерия. Она может не соответствовать характеру функционирования средств этой подсистемы.

Пусть в качестве системы рассматривается экономика страны, а l-я цель состоит в производстве угля. В качестве $(1, i)$-й подсистемы рассматривается Министерство легкой промышленности РФ. В данной ситуации эффективность средств легкой промышленности не будет измеряться в тоннах угля. Во избежание подобных недоразумений основными показателями эффективности средств $(1, i)$-й подсистемы принято выбирать относительные приращения показателей функциониро-

вания всех средств системы, получаемые вследствие функционирования $(1, i)$-й подсистемы:

$$R_l^{(1,i)} = \frac{\Phi_l^{C(1,i)} - \Phi_l^{O(1,i)}}{\Phi_l^{И(1,i)} - \Phi_l^{O(1,i)}},\ l = \overline{1, k};\ i = \overline{1, n}. \tag{6.4}$$

Необходимо пояснить:

1) $\Phi_l^{И(1,i)}$ — l-й показатель эффективности средств системы при условии, что средства $(1, i)$-й подсистемы идеально участвуют в обеспечении l-й цели системы, а все ресурсы системы распределены оптимальным образом между остальными подсистемами;

2) идеальное участие средств $(1, i)$-й подсистемы в обеспечении l-й цели системы означает, например, что при ресурсах, которые отпускаются Министерству легкой промышленности РФ в соответствии с оптимальным планом функционирования всей экономики страны, оно поставляет продукцию, удовлетворяющую всем требованиям технических условий.

Оценим, в каких пределах будет меняться выбранный критерий $R_l^{(1,i)}$. Из формулы (6.4) можно записать:

$$\Phi_l^{И(1,i)} - \Phi_l^{O(1,i)} \geqslant \Phi_l^{C(1,i)} - \Phi_l^{O(1,i)}, \tag{6.5}$$

поэтому

$$0 \leqslant R_l^{(1,i)} \leqslant 1. \tag{6.6}$$

Вычисляя критерии эффективности, можно сравнивать разные проекты однотипных сложных систем, выбирать лучшие подсистемы из ряда предложенных, провести сравнительную оценку управляющих алгоритмов или систем передачи информации.

Вариант является эффективным, если вся сложная система оказывается лучшей. В экономических исследованиях обычно считается, что экономическая эффективность определяется как отношение полученного эффекта к затратам.

На практике наиболее высокая эффективность сложной системы достигается с большей вероятностью, если все подсистемы оказываются наиболее эффективными с учетом общей эффективности по данному критерию.

На базе показателя эффективности строится критерий надежности сложной системы [16]:

$$\Delta\Phi_{Н} = \Phi_{Н}^{И} - \Phi_{Н}^{*}, \tag{6.7}$$

где $\Delta\Phi_{Н}$ — показатель, характеризующий снижение эффективности системы вследствие возможных отказов ее элементов;

$\Phi_{Н}^{И}$ — показатель эффективности системы при условии, что элементы системы абсолютно надежны;

$\Phi_{Н}^{*}$ — показатель эффективности системы при условии, что отказы будут происходить при критериях, соответствующих заданным характеристикам надежности.

Если окажется, что $\Delta\Phi_{Н}$ мал, то это означает, что отказы мало влияют на эффективность всей системы. Если же их влияние оказывается существенным, то надо выбрать наиболее эффективный метод повышения надежности: повышение надежности элементов, введение профилактики, уменьшающей вероятность появления отказов, изменение схемы отдельных технических подсистем, резервирование ненадежных элементов, улучшение работы ремонтных органов по восстановлению элементов и подсистем и т. д.

Такой же подход на базе показателя эффективности предлагается для оценки сложных систем по устойчивости, помехозащищенности, качеству управления и некоторым другим показателям, существенным для конкретных реальных сложных систем.

При оценке функционирования сложной системы каждый руководитель на любом уровне обязан знать и уметь подбирать соответствующие показатели эффективности и уметь их использовать для повышения качества работы на своем участке. При этом процесс оценки эффективности опирается на исследование операций.

Глава 7
ОСНОВНЫЕ ПРИНЦИПЫ ИССЛЕДОВАНИЯ ОПЕРАЦИЙ, ИСПОЛЬЗУЕМЫЕ В ТЕОРИИ ЭФФЕКТИВНОСТИ

7.1. Схема операции

Рассмотрим простейший пример. Производству задан план выпуска продукции по номенклатуре и в стоимостном выражении. Цель руководства — выполнение плана. Руководство должно провести совокупность действий, направленных на достижение заданной цели. Совокупность действий, направленных на достижение некоторой цели, называется *операцией*.

Оперирующей стороной является совокупность лиц или автоматов, которые в данной операции стремятся к заданной цели. В ряде организаций цель назначается вышестоящим в иерархии управления органом.

Исследователь операции — человек или группа, принадлежащие к оперирующей стороне и добивающиеся той же цели. Исследователю должна быть известна цель операции и все условия проведения операции. Однако часто бывает, что исследователь операции оказывается менее информированным об операции, чем оперирующая сторона, к которой он принадлежит. Это связано с объективными причинами, так как исследователь чаще всего бывает математиком или специалистом по математическому обеспечению ЭВМ, а остальные члены оперирующей стороны — специалистами отрасли или производства, в интересах которого и проводится исследование операций.

Иногда оперирующая сторона не может или не хочет выдать полную информацию исследователю, а в некоторых случаях это происходит из-за отсутствия должного взаимодействия и взаимопонимания между оперирующей стороной и исследователем. В этом случае исследователь не принимает окончательных решений, а помогает выработать их оперирующей стороне.

Оперирующая сторона имеет в своем запасе так называемые *активные средства*, например рабочую силу, сырье, производственное оборудование, финансы и т. п.

Стратегии оперирующей стороны — способы использования активных средств. Работа исследователя операций заключается в сравнении стратегий и оценке их эффективности.

Достижение цели в данной операции зависит от количества активных средств и выбора стратегий. Количество активных средств зависит от факторов, которые контролируются оперирующей стороной, и от факторов, которые оперирующая сторона не может контролировать (неконтролируемых факторов).

Обстановка проведения операции включает неконтролируемые факторы. Например, в сельском хозяйстве для руководства — оперирующей стороны — количество удобрений, порядок сева на разных полях являются контролируемыми факторами, а метеорологические условия — неконтролируемыми факторами. При постановке задачи исследования операций обязательно указывается информированность оперирующей стороны и исследователя об обстановке проведения операции, где главные сведения относятся к неконтролируемым факторам.

Математическая модель операции должна учитывать все компоненты операции, так как адекватность модели будет зависеть от этих компонент. Ход изменения состояний системы, т. е. ход операции, описывается n-фазовыми координатами x_i. Чем больше количество фазовых координат, тем точнее описание операции, но сложнее математическое исследование.

Критерий (показатель) эффективности

$$R = R[x_1(t), x_2(t), \dots, x_n(t); y_1(t), y_2(t), \dots, y_m(t)]$$

характеризует степень соответствия хода операции поставленной цели.

В чем же заключается цель операции с математической точки зрения? Она заключается в стремлении к увеличению (уменьшению) значения критерия эффективности. Критерий эффективности в заданной модели является математическим выражением цели операции.

Полезно отметить разницу между фазовой координатой и критерием эффективности. Так, разность между плановым и фактическим выпуском изделий d может служить критерием, но эта же разность d может быть фазовой координатой, в то

время как критерием эффективности может служить 0 или 1 при невыполнении плана или при его выполнении.

Активные средства определяются количественно и задаются вектором $\mathbf{b} = \{b_1, b_2, \dots, b_n\}$.

Они могут быть ограничены сверху: $b_i \leqslant b_i^0$, где b_i^0 — предельное количество i-х активных средств.

Поскольку неконтролируемые факторы определяют обстановку проведения операции, рассмотрим подробнее их классификацию в зависимости от информированности о них исследователя в момент исследования:

- фиксированные факторы, значения их известны;
- случайные процессы с известными законами распределения (случайные фиксированные факторы);
- неопределенные факторы, известна область распределения фактора или область, внутри которой находятся законы распределения случайного фактора, если точно не известен закон распределения.

Неопределенные факторы также делят на три группы:

- стратегии противника, у которого имеются свои активные средства, т. е. факторы, не зависящие от оперирующей стороны;
- природные неопределенные факторы, т. е. факторы, которые появляются из-за слабой изученности некоторых процессов;
- неопределенные факторы из-за нечеткости знания цели операции или критерия эффективности.

Примеры неопределенности первого типа можно получить в военных действиях или в конкурентной внешней торговле.

Если только известны математическое ожидание и дисперсия случайной величины, то это может иллюстрировать неопределенность второго типа.

Примером неопределенности третьего типа может служить выбор критерия оценки работы предприятий, выпускающих изделия совершенно разных типов [60].

7.2. Объединение операций. Система элементарных действий

На основе теории и практики проведения операций определено, что существуют два вида целей и, следовательно, два вида критериев эффективности [51], [60].

1. **Качественное определение цели**, когда возможны два альтернативных исхода: результаты достигнуты или не достигнуты. При этом критерий эффективности может принимать только два значения (1 или 0):

$$\Phi = \begin{cases} 1 \text{ — результат получен;} \\ 0 \text{ — результат не получен.} \end{cases}$$

2. **Количественное определение цели**, когда стремятся увеличить или уменьшить критерий эффективности операции.

В практике исследований в подавляющем числе случаев применяются количественные критерии.

Результат управления является сложным событием, которое можно представить событием, состоящим из более простых событий. Часто бывает удобно представить общий критерий эффективности через частные критерии. Общим критерием эффективности называют критерии, которые измеряют эффективность системы в целом. Частными называют критерии, измеряющие эффективность некоторой составляющей операции.

При определении общего критерия эффективности можно выделить два случая.

• Общий критерий имеет следующую структуру:

$$\Phi_0 = F(\Phi_1, \Phi_2, ..., \Phi_n),$$

где Φ_n — значение критерия для n-й составляющей операции (n-й частный критерий).

• Суммарный критерий представляется как функция фазовых координат новой операции. Однако он не является функцией частных критериев, как в первом случае. Это значит, что новая объединенная операция имеет свою цель, не связанную с частными целями частных операций. Объединенная операция базируется только на активных средствах частных операций. Этот второй случай иногда является источником ошибок при исследовании операций, но он не относится к процессу получения общего критерия из частных. Значит, когда мы говорим об объединенной операции и получении общего критерия, то имеем в виду только первый случай.

Остановимся на элементарных способах объединения (свертывания) критериев эффективности.

I. Суммирование, или «экономический» способ

Целью объединения операций является максимизация суммарного критерия

$$\Phi_C = \sum_{i=1}^{n} \lambda_i \Phi_i, \tag{7.1}$$

где λ_i — вес частного критерия Φ_i;

Φ_i — значение критерия для i-й составляющей операции (i-й частный критерий).

Если свертывается критерий, зависящий от непрерывного параметра, то

$$\Phi_C = \int \Phi(u) \cdot \lambda(u)\, du; \tag{7.2}$$
$$\lambda(u) \geqslant 0; \int \lambda(u)\, du = 1,$$

где u — переменная.

Если одна из операций такова, что она всегда выполняется, т. е. $\Phi_{n+1} = 1$, то (7.1) можно записать так:

$$\Phi_C = \sum_{i=1}^{n} \lambda_i \Phi_i + \lambda_{n+1}. \tag{7.3}$$

II. Способ перехода к цели первого типа

Этот способ осуществляется путем разбиения векторов Ф на удовлетворительные и неудовлетворительные. Если

$$\left.\begin{aligned} &\Phi_i \geqslant \Phi_i^0,\ 1 \leqslant i \leqslant n,\ \Phi_C = 1; \\ &\Phi_i < \Phi_i^0,\ 1 \leqslant i \leqslant n,\ \Phi_C = 0 \text{ или } -\alpha. \end{aligned}\right\} \tag{7.4}$$

Граничное значение вектора Φ^0 выбирается оперирующей стороной в зависимости от условий решаемых задач. При $n = 1$ оперирующая сторона будет добиваться того, чтобы $\Phi_C \geqslant \Phi_C^0$.

III. Способ последовательного достижения частных целей

К началу выполнения последующей операции должны быть получены абсолютные максимумы (минимумы) критериев эффективности предыдущих i-х частных операций. Этот способ объединения при $\Phi_i \geqslant 0$ можно представить так:

$$\Phi_C = \Phi_i + \sum_{i=1}^{j-1} \sup \Phi_i \; (i \leqslant j-1), \qquad (7.5)$$

где $\sup \Phi_i$ — верхняя грань возможных значений критерия эффективности Φ_i. Данный способ часто используется в экономических задачах.

IV. Логическое объединение целей

Общий критерий и частные критерии относятся к качественному виду (первому) и принимают только значения 0 и 1. При этом используются элементарные действия над критериями.

Можно представить **три случая**:

1) критерий для цели, противоположной данной, определяется по формуле

$$\Phi_C = 1 - \Phi_i; \qquad (7.6)$$

2) суммарная цель состоит в обязательном выполнении всех частных целей

$$\Phi_C = \prod_{i=1}^{n} \Phi_i; \qquad (7.7)$$

3) суммарная цель состоит в выполнении хотя бы одной из частных целей

$$\Phi_C = 1 - \prod_{i=1}^{n} (1 - \Phi_i). \qquad (7.8)$$

Три случая составляют полную систему булевых операций. Если Φ_C и Φ_i принимают значения только 0 и 1, то любая зависимость $\Phi_C = \Phi(\Phi_1, \Phi_2, ..., \Phi_n)$ может быть записана в виде конечного числа последовательных повторений действий 1), 2) и 3).

Примеры можно привести следующие: в сложной системе подсистемы технических средств СУ соединены последовательно; для нормального функционирования системы необходимо, чтобы все подсистемы были в рабочем состоянии (случай 2); однотипные блоки системы соединены параллельно, и для нормального функционирования достаточно одного блока данного типа; если хотя бы один из блоков будет исправен, то все соединение выполнит свои функции (случай 3); если Φ_i характеризует исправное состояние i-й подсистемы, то неисправное состояние может быть охарактеризовано величиной $1 - \Phi_i$ (случай 1).

V. Обобщенное логическое свертывание критериев

Данный способ объединения (свертывания) критериев, т. е. функций $\Phi_C = F(\Phi_i)$, является прямым обобщением предыдущего (четвертого) способа. Так, если у оперирующей стороны существуют антагонистические интересы, то:

а) вместо (7.6) запишем:

$$\Phi_C = -\Phi_i; \qquad (7.9)$$

б) обобщение (7.7) позволяет записать:

$$\Phi_C = \min_{1 \leqslant i \leqslant n} \Phi_i\lambda_i, \; \lambda_i \geqslant 0; \qquad (7.10)$$

в) вместо (7.8) можно записать:

$$\Phi_C = \max_{1 \leqslant i \leqslant n} \Phi_i\lambda_i, \; \lambda_i \geqslant 0. \qquad (7.11)$$

Формулы (7.10) и (7.11) будут соответствовать формулам (7.7) и (7.8), если предположить в частном случае, что Φ_i принимает значения 0 и 1, а $\lambda_i = 1$.

Примером применения операций max и min может служить оценка времени исправной работы неремонтируемых систем.

VI. Случайное и неопределенное свертывание

В данном случае в зависимости от того, какое значение примет i-й неконтролируемый фактор, i-й частный критерий принимается суммарным критерием:

$$\Phi_C = \Phi(i) = \Phi_i. \qquad (7.12)$$

Если частные критерии определяются непрерывной случайной или неопределенной величиной, то общий критерий получит следующее выражение:

$$\Phi_C = \Phi(\alpha) = \Phi_\alpha, \qquad (7.13)$$

где α — случайная или неопределенная величина.

В качестве примера обычно приводят случай, когда оперирующая сторона не в состоянии точно определить коэффициент веса λ_i для частных операций в способах объединения критериев I и V. В этом случае значения $\{\lambda_i\}$ окажутся неопределенными факторами.

Увеличение количества случайных и неопределенных факторов, не контролируемых оперирующей стороной, снижает эффективность стратегий и создает большие трудности при выборе стратегий. В случае (7.13) все критерии равнозначны. Если же вес критериев различен, то рекомендуется вводить коэффициенты веса для частных критериев:

$$\Phi_C = \lambda(\alpha) \cdot \Phi_\alpha.$$

Все шесть рассмотренных элементарных действий над критериями применимы также для тех случаев, когда операция сформулирована не полностью [24]. Частными критериями при этом станут функции

$$[\omega_i(\mathbf{x}, \mathbf{y})],$$

где $\mathbf{x}$ — вектор контролируемых факторов;

$\mathbf{y}$ — вектор неконтролируемых факторов.

Вектор $\omega_i(\mathbf{x}, \mathbf{y})$ является составляющим вектор-функции $\omega(\mathbf{x}, \mathbf{y}) = \{\omega_i(\mathbf{x}, \mathbf{y})\}$ контролируемых и неконтролируемых факторов.

Обычно вектор-функция состоит из всех или части фазовых координат. Однако без вектор-функций нельзя обойтись при не полностью сформулированных моделях операции. Эти модели появляются из-за наличия неопределенных ситуаций. В не полностью сформулированных моделях операции нет единого критерия эффективности, как, например, в полностью сформулированных моделях.

Можно указать на такую дополнительную трудность при исследовании векторов $\omega_i(\mathbf{x}, \mathbf{y})$. Кажется вполне очевидным, что каждую координату вектора следует увеличивать (уменьшать). Однако не всегда эти координаты можно одновременно увеличивать или уменьшать. Тогда появляется неопределенность в решении вопроса о том, каким комбинациям значений координат вектора надо отдать предпочтение перед другими комбинациями. На практике часто осуществляется переход от не полностью сформулированной модели к корректной. Математически это означает свертывание вектор-функции $\omega(\mathbf{x}, \mathbf{y})$ в функцию $\Phi(x, y)$.

7.3. Теоремы о полноте системы элементарных действий над критериями эффективности

Задачей теорем, приводимых в работах [24], [26], является доказательство того, что рассмотренные элементарные действия могут охватить все возможные однозначные зависимости общих критериев от частных.

Первая теорема посвящена точному представлению зависимостей в виде конечного числа элементарных действий. Вторая, третья и четвертая теоремы доказывают возможность приближенного представления с любой заданной точностью. Приведем их без доказательств.

Теорема 7.1. Если однозначная функция

$$\Phi_C = F(\Phi_1, \Phi_2, ..., \Phi_n)$$

и каждая из Φ_i принимает лишь конечное число конечных возможных значений, то зависимость Φ_C от Φ_i может быть представлена в виде конечного числа действий типа IV (7.6)—(7.8) и типа I и II (7.1), (7.2) и (7.4).

Теорема 7.2. Пусть $\Phi_C = F(\Phi_1, \Phi_2, ..., \Phi_n)$ принимает конечное число N значений Φ_{CK}, а Φ_i пусть произвольны, но ограничены. Тогда, каково бы ни было $\varepsilon > 0$, существуют множество M векторов $\{\Phi_i\}$ и функция $F^*(\Phi_1, \Phi_2, ..., \Phi_n)$, составленная из конечного числа действий I, II и IV, такие, что:

1) $F(\Phi_i) = F^*(\Phi_i)$, когда $\{\Phi_i\} \in M$;

2) $F^*(\Phi_i)$ пробегает все N значений Φ_{CK} при $\{\Phi_i\}$, пробегающем M, не принимая иных значений при любых $\{\Phi_i\}$;

3) множество M образует ε-сеть на ограниченном множестве всех $\{\Phi_i\}$, т. е. для любого $\{\Phi_i\}$ найдется $\{\Phi_i'\} \in M$, удаленный от $\{\Phi_i\}$ не более чем на ε.

Теорема 7.3. Если $\Phi_C = F(\Phi_i)$ равномерно непрерывна на некотором параллелепипеде возможных значений $\{\Phi_i\}$, то она с любой степенью точности может быть представлена в виде конечного числа действий типа I, II, IV.

При рассмотрении свертывания типа V было показано, что оно обобщает действия типа IV. Это доказывает и полноту системы действий I, II и V. Однако следующая теорема доказывает более сильное утверждение.

Теорема 7.4. Если $\Phi_C = F(\Phi_i)$ $(i \leqslant n)$ непрерывна на области $-\infty < \Phi_i^0 \leqslant \Phi_i \leqslant \Phi_i'$, то, каково бы ни было ε, найдется такое конечное число коэффициентов a_{ki}, c_{lk} ($l \leqslant l_0$; $k \leqslant k_0 \leqslant n + 2$), что в этой области

$$|F(\Phi_i) - \min_{1 \leqslant l \leqslant l_0} \max_{1 \leqslant k \leqslant k_0} (\sum_{i=1}^{n} a_{lki}\Phi_i + c_{lk})| \leqslant \varepsilon.$$

Приведенные теоремы показывают полноту пяти рассмотренных элементарных способов объединения критериев при $\Phi_C = F(\Phi_i)$.

Как быть, если общий критерий Φ_C зависит не только от частных критериев Φ_i, но и от некоторого неконтролируемого параметра α? В этом случае можно принять фиксированное значение α и для выражения $\Phi_C = F(\Phi_i, \alpha)$ воспользоваться свертыванием типа VI. В результате полнота способов объединения при наличии неконтролируемых факторов также будет доказана. Способ объединения III иллюстрируется в работе [24. — С. 53, 54] при рассмотрении примеров свертывания критериев.

Остановимся на практических рекомендациях исследователю операций. При проведении исследований всегда надо учитывать наличие неопределенных факторов. Это особенно важно в экономических задачах. Увеличение числа стратегий может привести только к бóльшему успеху.

При множестве стратегий появляются бóльшие возможности, чем при простой сумме стратегий, вследствие перераспределения активных средств между частными операциями. Например, в экономических исследованиях можно перераспределить ассигнования, которые должны обеспечить прибыль.

Исследователь операции обычно для осторожности ориентируется на наихудшие значения неконтролируемых факторов. Если это приводит к неудовлетворительным результатам, то исследователь обязан известить оперирующую сторону, которая может либо дать дополнительную информацию, либо принять рискованное решение. Сам исследователь не имеет права на самостоятельное принятие рискованного решения.

Исследователь должен применять принцип получения гарантированного результата. При этом принципе получается логически строгая теория принятия решения. Частным случаем этой теории является теория оптимизации, с помощью которой решают задачи при отсутствии случайных и неопределенных факторов.

Выясним, когда легче и выгоднее проводить исследование операций — в более ранний период или перед утверждением планов? Перед утверждением плана, когда, как правило, имеется готовый станочный парк и определен тип выпускаемой продукции, можно варьировать технологическими процессами. При раннем рассмотрении вариантов можно иметь бóльшую свободу и варьировать и станочным парком, и видом выпускаемой продукции, и технологическими процессами. В этом случае мы имеем меньшее число ограничений.

Итак, ранние исследования математически могут оказаться более простыми, так как увеличение числа ограничений приводит к усложнению задачи. Можно, таким образом, простыми методами получить основное опорное решение, а в более поздний период уточнять результаты в окрестности этого решения. Следовательно, расширение возможностей маневрирования в операции позволяет уменьшить число ограничений на стратегии (технологические процессы) и активные средства (сырье, материалы, станочный парк и др.).

7.4. Критерии эффективности систем управления, построенные на основе системы элементарных действий

В СУ решаются как задачи по определению эффективности внедрения автоматизированной системы управления и ее подсистем, так и задачи определения эффективности функционирования управляемых объектов. Эти задачи затрагивают весьма широкий спектр проблем и требуют большого разнообразия математических методов.

Наиболее часто применяются следующие **математические методы**:

- теория массового обслуживания;
- теория игр;
- теория вероятностей;
- теория информации.

Независимо от широты проблем и разнообразия применяемого математического аппарата остановимся на общих критериях оценки эффективности. В соответствии с двумя видами критериев все задачи в зависимости от цели управления можно подразделить на два класса.

В задачах **первого класса** целью управления является получение заданного эффекта, например заданного значения критерия эффективности:

$$Ф_C = \begin{cases} 1, \text{ если месячный план выполнен предприятием;} \\ 0, \text{ если месячный план не выполнен предприятием.} \end{cases}$$

Такой подход правилен при детерминированной модели управления предприятием, когда результаты анализируются по достоверным данным. Если изучаются случайные процессы управления, то в качестве общего критерия эффективности можно принять вероятность получения заданного результата управления:

$$Ф_C = P(A), \tag{7.14}$$

где $P(A)$ — вероятность выполнения задачи управления;
A — случайное событие, состоящее в выполнении задачи управления.

Если на систему управления действуют возмущающие факторы, которые обозначим через X, то $P(A/X)$ показывает условную вероятность выполнения задачи управления при наличии возмущающих факторов, вносящих неопределенность. В практике исследований крайне редко определяют закон изменения функции $P(A/X)$. Чаще определяют среднее значение, дисперсию, значение при наиболее благоприятных условиях и при наименее благоприятных условиях. Можно гарантировать результат, получающийся при наиболее неблагоприятных условиях функционирования СУ:

$$Ф_C = \min_x P(A/X). \tag{7.15}$$

Этот результат гарантируется с бóльшей вероятностью, чем наилучший или средний результат.

Какой же результат считать оптимальным? Оптимальной будет, конечно, та система, которая обеспечит максимальное значение критерия эффективности $Ф_C^*$:

$$Ф_C^* = \max_Ф Ф_C = \max_Ф \min_x P(A/X). \tag{7.16}$$

В задачах **второго класса** целью управления является получение наилучшего эффекта, оцениваемого экстремальным значением критерия эффективности. Если изучаемая модель носит статистический характер, то целью является получение оптимального значения средней величины эффекта.

Пусть G — случайная величина результата управления, $M\{G\}$ — математическое ожидание G, тогда общий критерий эффективности

$$Ф_C = M\{G\}. \tag{7.17}$$

Гарантированный результат управления, т. е. результат, получаемый при наиболее неблагоприятных условиях, можно оценить при помощи критерия

$$Ф_C = \min_x M\{G/X\}, \tag{7.18}$$

а оптимальный результат — при помощи критерия

$$Ф_C^* = \max_Ф \min_x M\{G/X\}, \tag{7.19}$$

где $M\{G/X\}$ — математическое ожидание случайной величины результата управления при воздействии неопределенных случайных факторов X.

Чем выше значение критерия, определяемого формулой (7.18), тем с лучшей стороны характеризуется оцениваемая система или оцениваемый процесс. Оптимальной будет, естественно, самая высокая оценка $Ф_C^*$ (7.19), полученная по этой формуле.

Чем ниже значение критерия, определяемого формулой (7.17), тем лучше характеризуется оцениваемый процесс или система. Тогда гарантированный результат управления выразится формулой

$$Ф_C = \max_x M\{G/X\}, \tag{7.20}$$

а оптимальный результат управления — формулой

$$Ф_C^* = \min_Ф \max_x M\{G/X\}. \tag{7.21}$$

При оценке производительности труда на однотипных предприятиях отрасли и обнаружении большого разброса между производительностью труда передовых и отстающих предприятий принимаются меры к получению минимального разброса значений производительности труда путем подтягивания отстающих предприятий до уровня передовых. Для этого проводится множество организационно-технических, социологических, кадровых мероприятий. Отстающие предприятия оснащаются более современным передовым оборудованием, организуется обучение специалистов передовым методам работы на лучших предприятиях, принимаются меры по улучшению рационализаторской и изобретательской работы, пересматриваются нормы оплаты труда, открываются школы по повышению квалификации рабочих и т. д.

Таким образом, на практике повышение критерия эффективности достигается в результате определенных затрат, которые нельзя не учитывать.

Критерий эффективности при учете экономических показателей определяется по формуле

$$\Phi_C = F(\Phi, C), \tag{7.22}$$

где Φ_C — оценка результата управления;
C — затраты на получение эффекта Φ.

Общий критерий мы обозначили Φ_C, тогда Φ и C стали частными критериями. Можно принять, что общий критерий зависит не от двух, а от n частных критериев, где каждый i-й частный критерий оценивает i-е свойство системы управления [24], [26]:

$$\Phi_C = F(\Phi_1, \Phi_2, ..., \Phi_n).$$

Тогда общий критерий может быть выражен при помощи известных нам элементарных действий над частными критериями.

Для автоматизированных систем управления рассматриваются **четыре типа элементарных действий**.

1. $$\Phi_C = \int_\alpha^\beta \Phi(x)\lambda(x)\,dx, \text{ если } \lambda(x) \geq 0; \tag{7.23}$$

$$\int_\alpha^\beta \lambda(x)\,dx = 1,$$

где x — реализация непрерывной случайной величины X;
$\Phi(x)$ — функция, характеризующая критерий в зависимости от переменных;
$\lambda(x)$ — функция веса критерия эффективности $\Phi(x)$;
α и β — нижний и верхний пределы изменения случайной величины X.

По этой формуле можно определять средние значения критериев эффективности, зависящих от случайных величин. Если переменная дискретна, то

$$\Phi_C = \sum_{i=1}^{n} \lambda_i \Phi_i, \tag{7.24}$$

где λ_i — вес частного критерия Φ_i;
n — число частных критериев.

В данном случае общий критерий получается сложением частных критериев с учетом их веса.

2. Задача управления будет решена, если будут решены все частные задачи, на которые она подразделяется. Например, план выпуска продукции считается выполненным районом, если все предприятия, входящие в его состав, выполнят свои планы. Если хотя бы одно предприятие не выполнит план, то план района будет считаться невыполненным. Критерий эффективности для такой задачи рассчитывается по формуле (7.7).

Эту задачу можно отнести к первому классу. В общем случае можно воспользоваться формулой (7.10).

3. Общая задача управления считается решенной, если будет решена хотя бы одна из частных задач. Это задача первого класса. В данном случае критерий эффективности рассчитывается по формуле (7.8), в общем случае — по формуле (7.11).

4. Частные цели функционирования сложной системы могут оказаться противоречивыми. Например, не представляется возможным добиться максимальной производительности труда при минимальной себестоимости продукции и т. д. В данной ситуации общая задача управления будет решена, если частные критерии эффективности будут находиться в определенных пределах или не ниже некоторых допустимых пределов согласно формуле (7.4):

$$\Phi_C = \begin{cases} 1, \text{ если } \Phi_i \geq \Phi_i^O; \\ 0, \text{ если } \Phi_i < \Phi_i^O, \ 1 \leq i \leq n. \end{cases} \tag{7.25}$$

Таким образом, в 7.3 и 7.4 показано, что существует ограниченное число элементарных операций при оценке эффективности.

7.5. Оценка эффективности с учетом вероятностей событий

Для иллюстрации возможности использования вероятностей событий в качестве критериев эффективности рассмотрим пример.

Пример 7.1

Пусть завод состоит из пяти цехов. Все цехи работают независимо друг от друга. Выполнение плана заводом считается завершенным, если все цехи выполнили свои планы. Обозначим A_i — наименование цеха, выполнившего план, тогда $P(A_i)$ — вероятность выполнения цехом его плана. Пусть $P(A_1) = 0{,}96$; $P(A_2) = 0{,}97$; $P(A_3) = 0{,}99$; $P(A_4) = 0{,}98$; $P(A_5) = 0{,}97$.

Определим вероятность выполнения плана заводом.

Решение

Поскольку выполнение плана цехом не зависит от условий выполнения плана другими цехами, вероятность выполнения плана заводом можно определить по формуле умножения независимых событий:

$$P(A_1A_2A_3A_4A_5) = P(A_1)P(A_2)P(A_3)P(A_4)P(A_5) =$$
$$= 0{,}96 \cdot 0{,}97 \cdot 0{,}99 \cdot 0{,}98 \cdot 0{,}97 \approx 0{,}88.$$

Рассмотрим более сложный пример, когда работа цеха зависит от результатов работы других цехов завода.

Пример 7.2

На заводе имеется пять цехов. Из них два сборочных и три механических, изготавливающих детали для сборочных цехов. Вероятность выполнения дневного плана: первым механическим цехом — $P(A_3) = 0{,}99$, вторым механическим цехом — $P(A_4) = 0{,}98$, третьим механическим цехом — $P(A_5) = 0{,}97$.

Первый механический цех изготавливает детали только для первого сборочного цеха, второй и третий механические цехи изготавливают детали для второго сборочного цеха. Второй сборочный цех не выполнит дневной план только в том случае, если оба обслуживающих его механических цеха не выполнят планы. Это значит, что эти два механических цеха дублируют друг друга при обеспечении второго сборочного цеха деталями.

Вероятность выполнения плана первым сборочным цехом, если первый механический цех полностью обеспечит его деталями, равна 0,96. Вероятность выполнения плана вторым сборочным цехом, если второй и третий, или один второй, или один третий механические цехи полностью обеспечат его деталями, равна 0,97. Завод выполнит свой дневной план, если его выполнили одновременно первый сборочный цех и второй сборочный цех. Все механические цехи работают независимо друг от друга. Работа одного сборочного цеха независима от работы другого сборочного цеха.

Определите:

1) вероятность выполнения плана первым сборочным цехом $P(A_1)$;

2) вероятность выполнения плана вторым сборочным цехом $P(A_2)$;

3) вероятность выполнения плана заводом.

Запасы деталей и изделий отсутствуют.

Решение

1. Запишем условия задачи в принятых обозначениях (A_i — событие, заключающееся в выполнении плана i-м цехом; $\bar{A}_i$ — событие, заключающееся в невыполнении плана i-м цехом):

$$P(A_1/A_3) = 0{,}96;\ P(A_3) = 0{,}99;\ P(A_2/A_4A_5,\ \bar{A}_4A_5,\ A_4\bar{A}_5) = 0{,}97;$$
$$P(A_4) = 0{,}98;\ P(A_5) = 0{,}97,$$

где $P(A_1/A_3)$ — условная вероятность выполнения плана первым сборочным цехом, если план выполнен первым механическим цехом;

$A_4\bar{A}_5$ — событие, заключающееся в выполнении плана вторым механическим цехом и невыполнении его третьим механическим цехом;

$\bar{A}_4A_5$ — событие, заключающееся в выполнении плана третьим и невыполнении его вторым механическим цехом.

2. Вероятность выполнения плана первым сборочным цехом:

$$P(A_1) = P(A_3) \cdot P(A_1/A_3) = 0{,}99 \cdot 0{,}96 = 0{,}9504 \approx 0{,}95.$$

3. Определим вероятность выполнения плана вторым сборочным цехом. Заметим, что второй сборочный цех выполнит план с вероятностью 0,97 при условии, что план выполнят второй и третий механические цехи (A_4A_3), или второй выполнит, а третий не выполнит ($A_4\bar{A}_5$), или третий выполнит, а второй сборочный цех не выполнит ($\bar{A}_4 A_5$).

Все три перечисленные события могут быть дополнены до полной группы несовместных событий только одним, четвертым событием. Оно заключается в одновременном невыполнении плана вторым и третьим механическими цехами ($\bar{A}_4\bar{A}_5$). Вычислим вероятность этого события:

$$P(\bar{A}_4\bar{A}_5) = [1 - P(A_4)]\,[1 - P(A_5)] = (1 - 0{,}98)\,(1 - 0{,}97) = 0{,}0006.$$

Вычислим вероятность обеспечения второго сборочного цеха деталями механических цехов для выполнения плана:

$$P(C) = 1 - P(\bar{A}_4\bar{A}_5) = 1 - 0{,}0006 = 0{,}9994.$$

Проверим вычисления:

$$P(C) = P(A_4A_5) + P(\bar{A}_4A_5) + P(A_4\bar{A}_5) =$$
$$= 0{,}98 \cdot 0{,}97 + 0{,}02 \cdot 0{,}97 + 0{,}98 \cdot 0{,}03 = 0{,}9994.$$

Тогда вероятность выполнения плана вторым сборочным цехом равна:

$$P(A_2) = P(C)P(A_2/C) = 0{,}9994 \cdot 0{,}97 \approx 0{,}97.$$

4. Определим вероятность выполнения дневного плана заводом:

$$P = P(A_1) \cdot P(A_2) = 0{,}95 \cdot 0{,}97 \approx 0{,}92.$$

После решения примера 7.2 для обобщения проведенной работы по расчету общих критериев эффективности при наличии зависимых случайных событий обратимся к формуле полной вероятности. Данная формула является следствием теоремы сложения вероятностей и теоремы умножения вероятностей.

Представим себе, что необходимо найти вероятность события A_1, которое может произойти при осуществлении одной из гипотез $H_1, H_2, ..., H_n$. Считаем, что все n гипотез (событий) образуют полную группу несовместных событий. При данных условиях вероятность события A вычисляется как

сумма произведений вероятности каждой гипотезы на вероятность события при этой гипотезе [21]. Это и есть словесная формулировка полной вероятности. Формула, соответствующая этой формулировке, записывается так:

$$P(A) = \sum_{i=1}^{n} P(H_i)P(A/H_i), \tag{7.26}$$

где $P(H_i)$ — вероятность гипотезы H_i, $P(A/H_i)$ — условная вероятность появления события A при гипотезе H_i.

Обратимся к критериям эффективности. Пусть $P(A)$ — общий критерий эффективности, тогда

$$\Phi = P(A) = \sum_{i=1}^{n} P_i\Phi_i, \tag{7.27}$$

где Φ_i — частный критерий эффективности, который в соответствии с формулой (7.26) имеет смысл условной вероятности появления события A при гипотезе H_i;

P_i — вероятность гипотезы H_i.

Расчет общего критерия эффективности для сложных систем, у которых отказ подсистемы не приводит к отказу всей системы, а может привести к снижению уровня ее эффективности, описан в работе [16]. Согласно работе [48], в таких системах надо учитывать все возможные состояния системы при отказе ее подсистем в рассматриваемом промежутке времени. Для каждого состояния надо находить частный критерий эффективности системы.

Рассмотрим автоматизированную систему управления, состоящую из n невосстанавливаемых подсистем. Следовательно, каждая подсистема может находиться либо в работоспособном состоянии, либо в состоянии отказа. В заданном интервале времени частные критерии эффективности не зависят от моментов возникновения отказов подсистем. В этом интервале можно отразить все возможные состояния системы на основе полной группы несовместных событий. В данном случае состояние системы определяется состоянием подсистем. Примером одного из состояний системы является работоспособность всех подсистем. Примером другого состояния системы может служить отказ одной подсистемы при нали-

12—1447

чии ($n-1$) работоспособных подсистем. В символах это можно записать с помощью формулы (7.26).

Введем следующие обозначения:

H_0 — гипотеза о том, что все подсистемы работоспособны;
H_1 — гипотеза о том, что i-я подсистема отказала, а ($n-1$) подсистем исправны;
H_{ij} — гипотеза о том, что отказали две подсистемы (i-я и j-я), а ($n-2$) подсистемы работоспособны ($i, j = \overline{1, n}$; $i < j$);
$H_{1, 2, \ldots, n}$ — гипотеза о том, что отказали все подсистемы;
$K_г$ — коэффициент готовности системы, который трактуется как вероятность застать систему в исправном состоянии;
P_0 — вероятность гипотезы H_0;
P_1 — вероятность гипотезы H_1;
P_i — вероятность гипотезы H_i;
q_i — вероятность нахождения i-й подсистемы в состоянии отказа в заданном интервале времени ($p_i + q_i = 1$).

В принятых обозначениях надо различать P_i и p_i:

$$q_i = 1 - p_i,$$

где P_i — вероятность нахождения системы в i-м состоянии, соответствующем гипотезе H_i;
p_i — вероятность нахождения i-й подсистемы в работоспособном состоянии.

Вероятность P_i определяется по следующей формуле:

$$P_i = p_1 p_2 \ldots p_{i-1} q_i p_{i+1} \ldots p_n;$$

вероятность гипотезы H_{ij} —

$$P_{ij} = p_1 p_2 \ldots p_{i-1} q_i p_{i+1} \ldots p_{j-1} q_j p_{j+1} \ldots p_n$$

(вероятности гипотез определяются по правилам умножения вероятностей независимых событий).

Общий критерий эффективности выразится на основе формул полной вероятности так:

$$\Phi = K_г(P_0\Phi_0 + \sum_{i=1}^{n} P_i\Phi_i + \sum_{i,j=1}^{n} P_{ij}\Phi_{ij} +$$
$$+ \sum_{\substack{i,j,k=1 \\ i<j<k}}^{n} P_{ijk}\Phi_{ijk} + \ldots + P_{1, 2, \ldots, n}\Phi_{1, 2, \ldots, n}), \qquad (7.28)$$

где $\Phi_{1, 2, \ldots, n}$ — частный критерий эффективности при отказе всех подсистем.

В данном случае используется комбинация первого и второго действий над частными критериями по классификации Гермейера [24], [26].

Таким образом, для решения практических задач по определению общих критериев эффективности необходимо определить вероятности безотказной работы каждой подсистемы (или вероятность ее отказа) и частные критерии эффективности.

Пример 7.3

Определите эффективность контроля экономической информации при наличии трех видов контроля: синтаксического, семантического и прагматического. Качество контроля одного вида не влияет на качество контроля другого вида. Коэффициент готовности всей системы контроля равен 0,99. При исключении трех видов (уровней) контроля частные критерии эффективности равны нулю. При исключении синтаксического контроля частный критерий его эффективности равен нулю.

Известны вероятности решения задачи при каждом виде контроля:

$$p_{синт} = p_1 = 0{,}99;\ p_{сем} = p_2 = 0{,}96;\ p_{прагм} = p_3 = 0{,}95.$$

Известны частные критерии эффективности контроля:

$$\Phi_0 = 0{,}97;\ \Phi_1 = 0;\ \Phi_2 = 0{,}90;\ \Phi_3 = 0{,}80;\ \Phi_{23} = 0{,}50.$$

Поясним смысл частных критериев эффективности:

$\Phi_0 = 0{,}97$ — частный критерий эффективности при применении всех трех видов контроля;

$\Phi_1 = 0$ — при исключении синтаксического уровня контроля частный критерий эффективности равен нулю, следовательно, $\Phi_{12} = 0$; т. е. при исключении синтаксического и семантического уровней контроля частный критерий эффективности также равен нулю.

Таким образом:

$$\Phi_1 = \Phi_{12} = \Phi_{13} = 0;$$

$\Phi_2 = 0{,}90$ — при исключении семантического уровня контроля частный критерий эффективности равен 0,90;

$\Phi_{23} = 0{,}50$ — при исключении семантического и прагматического уровней контроля частный критерий эффективности равен 0,50.

Решение

1. Упростим общую формулу (7.28) с учетом конкретных условий данной задачи:

$$\Phi = K_\Gamma(P_0\Phi_0 + \sum_{i=1}^{3} P_i\Phi_i + \sum_{i,j=1}^{3} P_i\Phi_{ij}) = K_\Gamma(P_0\Phi_0 + P_1\Phi_1 +$$
$$+ P_2\Phi_2 + P_3\Phi_3 + P_{1,2}\Phi_{1,2} + P_{1,3}\Phi_{1,3} + P_{2,3}\Phi_{2,3}).$$

Учитывая, что $\Phi_1 = \Phi_{1,2} = \Phi_{1,3} = 0$, получим:

$$\Phi = K_\Gamma(P_0\Phi_0 + P_2\Phi_2 + P_3\Phi_3 + P_{2,3}\Phi_{2,3}).$$

Распишем значения P_i:

$$P_0 = p_1p_2p_3;\ P_2 = p_1(1 - p_2)p_3;$$
$$P_3 = p_1p_2(1 - p_3);\ P_{2,3} = p_1(1 - p_2)(1 - p_3).$$

Окончательно

$$\Phi = K_\Gamma[p_1p_2p_3\Phi_0 + p_1(1 - p_2)p_3\Phi_2 + p_1p_2(1 - p_3)\Phi_3 +$$
$$+ p_1(1 - p_2)(1 - p_3)\Phi_{2,3}].$$

2. Определим общий эффект контроля экономической информации:

$$\Phi = 0{,}99\,(0{,}99 \cdot 0{,}96 \cdot 0{,}95 \cdot 0{,}97 + 0{,}99 \cdot 0{,}04 \cdot 0{,}95 \cdot 0{,}90 +$$
$$+ 0{,}99 \cdot 0{,}96 \cdot 0{,}05 \cdot 0{,}80 + 0{,}99 \cdot 0{,}04 \cdot 0{,}05 \cdot 0.50) \approx 0{,}94.$$

Таким образом, на основе вероятностного критерия проведена оценка эффективности одного из этапов обработки экономической информации. Однако в данном примере не учтены затраты на проведение мероприятий по повышению общего критерия эффективности системы контроля.

В настоящее время известно большое количество подходов и моделей, где исследуется общий критерий эффективности с учетом затрат на получение эффекта. Однако в подавляющем большинстве случаев Φ_i не является независимой от C величиной. Поэтому не представляется возможным для выражения (7.22) найти общую зависимость между этими критериями, пригодную для вычислений в самых разных условиях. В работе [87] предлагается пользоваться графическими методами. На графике рекомендуется представлять общий критерий эффективности как функцию затрат. После определения зависимости $\Phi(C)$ можно определить общий критерий эффективности с учетом затрат. Если значения одного из

критериев заданы в какой-то ограниченной области, то можно найти область возможных значений другого критерия.

Очень часто в качестве общего критерия выбирают отношение [87]:

$$\Phi_C = \frac{\Phi(C)}{C}, \qquad (7.29)$$

которое можно рассматривать как среднюю величину критерия эффективности, приходящуюся на единицу стоимости затрат.

Пусть процесс, которым мы управляем, характеризуется максимальным значением критерия $\Phi(C)$. Тогда можно найти максимум среднего эффекта на единицу затрат. Для этого надо в формуле взять первую частную производную по стоимости C и приравнять ее нулю:

$$\frac{\partial \Phi_C}{\partial C} = \frac{1}{C}\frac{\partial \Phi(C)}{\partial C} - \frac{1}{C^2}\Phi_C;$$

$$\frac{\partial \Phi(C)}{\partial C} - \frac{\Phi(C)}{C} = 0;$$

$$\frac{\partial \Phi(C)}{\partial C} = \frac{\Phi(C)}{C}. \qquad (7.30)$$

Выражение (7.30) показывает, что на кривой $\Phi(C)$ есть точка, касательная к которой проходит через начало координат (рис. 7.1).

Следовательно, можно так записать алгоритм определения оптимальных значений рассматриваемых критериев [87].

Шаг 1. Построить функцию $\Phi(C)$.

Шаг 2. Из начала координат провести касательную к кривой $\Phi(C)$.

Шаг 3. Определить координаты, точки касания ($\Phi_{опт}$, $C_{опт}$).

Полученный результат может подсказать проектировщику или эксплуатационнику диапазон изменения затрат, при которых обеспечивается допустимый уровень общего критерия. Это обусловлено тем, что общий критерий можно увеличить как путем увеличения $\Phi(C)$, так и путем уменьшения затрат. Точка L_0 касания прямой из начала координат с кривой $\Phi(C)$ соответствует системе с высоким эффектом управления, но не минимальной по стоимости. Максимальное значение ча-

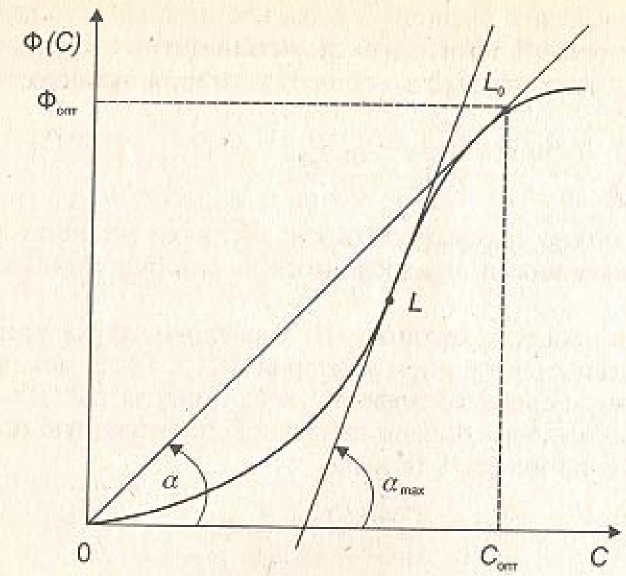


Рис. 7.1. Зависимость $\Phi(C)$ от C

стной производной функции $\Phi(C)$ по C находится в точке L, а не L_0 (точка L_0 соответствует системе с оптимальными критериями).

Пример 7.4

Эксплуатация показала, что коэффициент готовности является важной характеристикой эффективности технических средств СУ. Для повышения K_Γ необходимы затраты средств, представленные в табл. 7.1.

Опыт эксплуатации технических средств двести СУ за трехмесячный период при односменной работе позволил получить следующие значения коэффициента готовности и затрат (табл. 7.1).

Графическим методом определим оптимальные значения коэффициента готовности и затрат.

Решение

1. Построим график зависимости $\Phi_I(C) = K_\Gamma$ от общей стоимости эксплуатации C (рис. 7.2), если известно, что ча-

Таблица 7.1

Стоимость эксплуатации, руб.

K_Γ	Заработная плата обслуживающего персонала	Профилактика и ремонт	Учеба обслуживающего персонала	Замена элементов	Прочие расходы	Общие затраты
0,30	3600	680	—	—	—	4280
0,40	3600	700	—	300	200	4800
0,50	3600	600	400	100	300	5000
0,60	3600	600	800	250	550	5800
0,70	3600	900	1200	300	500	6500
0,80	4000	1000	1800	200	—	7000
0,90	4400	1000	2200	150	250	8000
0,95	5000	1500	2200	300	—	9000
0,96	5000	2000	2200	—	800	10000
0,97	5000	2000	2400	400	2200	12000

стным критерием эффективности является коэффициент готовности.

2. Из начала координат проведем касательную к кривой $\Phi(C)$.

3. Определим координаты точки касания графически. Получаем $K_{\Gamma опт} = 0{,}85$; $C_{опт} = 7500$ руб.

Пример 7.5

По данным предыдущего примера вычислите функцию:

$$\Phi_C = f(K_\Gamma,\ C) = \frac{K_\Gamma}{C}.$$

Полученные результаты представим в табл. 7.2.

Надо определить графически по полученным данным $K_{\Gamma опт}$ и $C_{опт}$ по функции $\Phi_C = \dfrac{K_\Gamma}{C}$ и по функции $\Phi_C = \varphi(C)$.

Решение

1. Строим график $\Phi_C = f(K_\Gamma,\ C)$ (рис. 7.3). Графически в точке $\max f(K_\Gamma,\ C)$ находим значение $K_{\Gamma опт} = 0{,}85$.

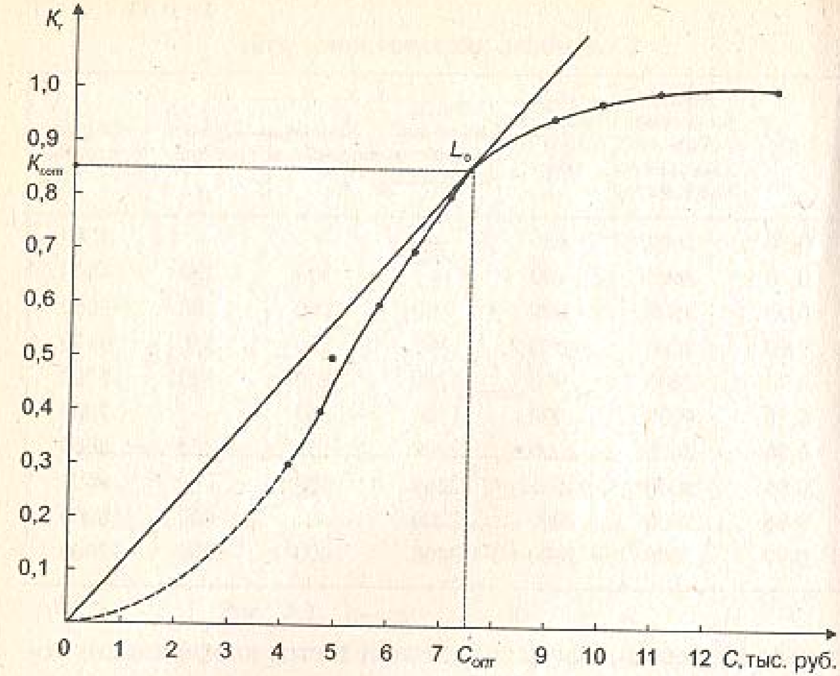


Рис. 7.2. Функция $K_Г = \Phi(C)$

2. Строим график $\Phi_C = \varphi(C)$ (рис. 7.4). Графически в точке max $\varphi(C)$ находим $C_{опт}$ = 7500 руб.

Таким образом, решение, построенное по алгоритму [87], проверено графическим методом. В данном случае мы дифференцированно отобразили зависимость критерия от каждого из его частных критериев.

Таблица 7.2

Функция $\Phi = \dfrac{K_Г}{C}$

$K_Г$	0,3	0,4	0,5	0,6	0,7	0,8	0,9	0,95	0,96	0,97
C, руб.	4280	4800	5000	5800	6500	7000	8000	9000	10000	12000
$10^6\ \Phi_C$	7	8	10	10,3	10,9	11,4	11,2	10,5	9,6	8,1

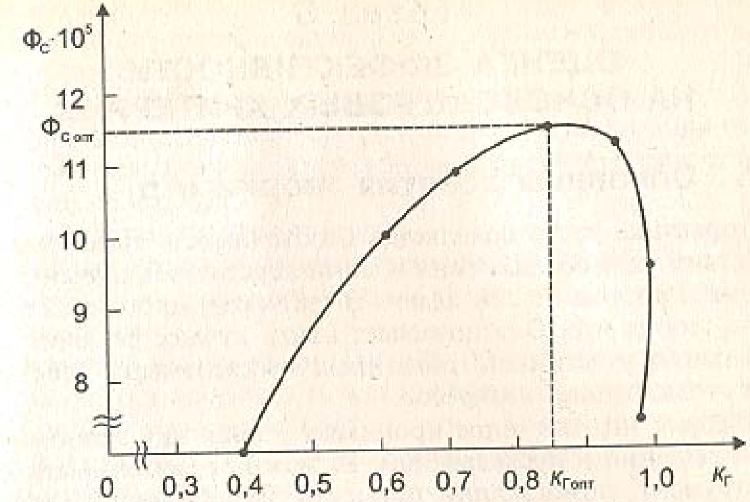


Рис. 7.3. Функция $\Phi_C = f(K_Г, C)$

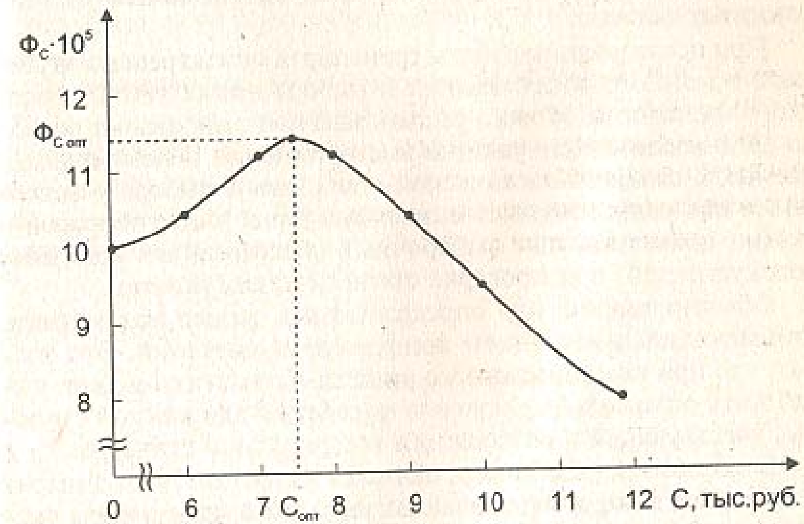


Рис. 7.4. Функция $\Phi_C = \varphi(C)$

Глава 8

ОЦЕНКА ЭФФЕКТИВНОСТИ НА ОСНОВЕ ИГРОВЫХ КРИТЕРИЕВ

8.1. Основные понятия теории игр

На практике часто появляется необходимость согласования действий ряда объединений и министерств в тех случаях, когда их интересы не совпадают. В таких ситуациях может помочь теория игр. Она позволяет найти лучшее решение для поведения участников, обязанных согласовывать действия при столкновении интересов.

Теория игр все шире проникает в практику экономических решений и исследований. Ее можно рассматривать как инструмент, помогающий повысить эффективность плановых и управленческих решений. Это имеет большое значение при решении задач в промышленности, сельском хозяйстве, на транспорте, в торговле, особенно при заключении договоров с иностранными государствами на любом иерархическом уровне. Так можно определить научно обоснованный уровень сезонного снижения розничных цен, оптимальный уровень товарных запасов.

При исследовании работы транспорта можно решать задачи экскурсионного обслуживания и выбора новых линий городского транспорта. Можно решить задачу планирования порядка организации эксплуатации месторождений полезных ископаемых в стране. Классической стала задача выбора участков земли под сельскохозяйственные культуры. Метод теории игр можно применять при выборочных обследованиях конечных совокупностей, при проверке статистических гипотез.

Обычно теорию игр определяют как раздел математики, занимающийся изучением конфликтных ситуаций. Это значит, что при помощи данного раздела математики можно выработать оптимальные правила поведения для каждой стороны, участвующей в разрешении конфликтной ситуации.

В экономике, например, оказался недостаточным аппарат математического анализа, занимающийся определением экстремумов функций. Появилась необходимость изучения оптимальных минимаксных и максиминных решений. Значит,

теорию игр можно рассматривать как новый раздел теории оптимизации, позволяющий решать новые задачи при принятии решений.

Игра — упрощенная формализованная модель реальной конфликтной ситуации. Математически формализация означает, что разработаны определенные правила действия сторон в процессе игры:

- варианты действия сторон;
- исход игры при данном варианте действия;
- объем информации каждой стороны о поведении всех других сторон.

В главе 6 мы отмечали, что одну (играющую) сторону при исследовании операций может представлять коллектив, преследующий некоторую общую цель. Однако разные члены коллектива могут быть по-разному информированы об обстановке проведения игры.

Выигрыш или проигрыш сторон оценивается численно, другие случаи в теории игр не рассматриваются, хотя не всякий выигрыш можно оценить количественно.

Игрок — одна из сторон в игровой ситуации.

Стратегия игрока — правило действия игрока в каждой из возможных ситуаций игры. Существуют игровые системы управления, если процесс управления в них рассматривается как игра.

Платежная матрица (матрица эффективности, матрица игры) — матрица, включающая все значения выигрышей (в конечной игре). Пусть игрок A имеет m стратегий A_i, а игрок B — n стратегий B_j ($i = \overline{1, m}$; $j = \overline{1, n}$). Игра может быть названа игрой $m \times n$. Представим ее матрицу эффективности, сопроводив необходимыми обозначениями:

A_i \ B_j	B_1	B_2	...	B_n	α_i
A_1	a_{11}	a_{12}	...	a_{1n}	α_1
A_2	a_{21}	a_{22}	...	a_{2n}	α_2
...	...	...	...	...	...
A_m	a_{m1}	a_{m2}	...	a_{mn}	α_m
β_j	β_1	β_2	...	β_n	

В данной матрице элементы a_{ij} — значения выигрышей. Элемент a_{ij} может означать и математическое ожидание выигрыша (среднее значение), если выигрыш является случайной величиной.

В теории игр не существует установившейся классификации видов игр. Однако некоторые виды можно выделить.

Если в игре участвуют две стороны, то ее называют *игрой двух игроков.* Если число сторон более двух, то ее относят к *игре n-игроков.* Наибольший интерес вызывают игры двух игроков. Они и математически наиболее глубоко проработаны, и в практических приложениях имеют наиболее обширную библиографию.

В зависимости от количества стратегий игры делятся на конечные и бесконечные. В *конечной игре* каждый из игроков имеет конечное число возможных стратегий. Если же хотя бы один из игроков имеет бесконечное число возможных стратегий, то игра является *бесконечной.*

В зависимости от взаимоотношения сторон игры делятся на бескоалиционные, коалиционные и кооперативные. Если игроки не имеют права вступать в соглашение, образовывать коалицию, то такая игра относится к *бескоалиционной*; если же игроки могут вступать в соглашение, создавать коалицию — к *коалиционной. Кооперативная* игра — это игра, в которой заранее определены коалиции.

Существует классификация по характеру выигрышей. Это игры с нулевой суммой и игры с ненулевой суммой. *Игра с нулевой суммой* предусматривает, что «сумма выигрышей всех игроков в каждой партии равна нулю». Игры двух игроков с нулевой суммой относят к классу *антагонистических игр.* Естественно, выигрыш одного игрока всегда равен проигрышу другого. Примерами игры с нулевой суммой служат многие экономические задачи. В них общий капитал всех игроков перераспределяется между игроками, но не меняется.

В качестве примеров игр с ненулевой суммой можно привести большое количество экономических задач. Так, в результате торговых взаимоотношений стран, участвующих в игре, все участники могут оказаться в выигрыше. Игра, в которой надо вносить взнос за право участия в ней, является *игрой с ненулевой суммой.*

В зависимости от вида функции выигрышей игры подразделяются на матричные, биматричные, непрерывные, выпуклые, сепарабельные и т. д. Поясним некоторые из них.

Матричная игра — конечная игра двух игроков с нулевой суммой. В общем случае ее платежная матрица является прямоугольной. Номер строки матрицы соответствует номеру стратегии, применяемой первым игроком *A.* Номер столбца матрицы соответствует номеру стратегии, применяемой вторым игроком *B.* Выигрыш первого игрока является элементом матрицы. Выигрыш второго игрока равен проигрышу первого игрока. Известно, что матричные игры имеют решения. Они могут быть решены методами линейного программирования. Для этого их надо переформулировать в терминах линейного программирования.

Биматричная игра — конечная игра двух игроков с ненулевой суммой. Выигрыши каждого игрока задаются своей матрицей, в которой строка соответствует стратегии первого игрока, а столбец — стратегии второго игрока. Однако элемент первой матрицы показывает выигрыш первого игрока. Для биматричных игр, так же как и для матричных, имеется обоснованная теория оптимального поведения игроков.

Если функция выигрышей каждого игрока в зависимости от стратегий является непрерывной, то игра считается *непрерывной.* Если функция выигрышей выпуклая, то и игра является *выпуклой.*

Если функция выигрышей может быть разделена на сумму произведений функций от одного аргумента, то игра относится к *сепарабельной.*

В зависимости от количества ходов игры можно разделить на одношаговые и многошаговые. *Одношаговые игры* заканчиваются после одного хода каждого игрока. Так, в матричной игре после одного хода каждого из игроков происходит распределение выигрышей. *Многошаговые игры* бывают позиционными, стохастическими, дифференциальными и др.

При классификации по информированности сторон различают игры с полной информацией и игры с неполной информацией. Если каждый игрок на каждом ходе игры знает все ранее примененные другими игроками на предыдущих ходах стратегии, то игра классифицируется как *игра с полной информацией.*

Если же игроку стратегии предыдущих ходов других игроков известны не все, то такая игра классифицируется как *игра с неполной информацией.*

Далее мы убедимся, что игра с полной информацией имеет решение. Решением будет *седловая точка* при чистых стратегиях.

В зависимости от степени неполноты информации игры подразделяются на:

- статистические (в условиях частичной неопределенности);
- стратегические (в условиях полной неопределенности).

Игры с природой часто относят к *статистическим играм.* В статистической игре имеется возможность получить информацию на основе статистического эксперимента, при котором оценивается распределение вероятностей стратегий природы. С теорией статистических игр тесно связана теория принятия экономических решений.

Получив некоторое представление о существующих подходах к классификации игр, можно остановиться на оценках игры.

Рассмотрим матричную игру, представленную матрицей эффективности $m \times n$, где число строк $i = \overline{1, m}$, а число столбцов $j = \overline{1, n}$. Применим рассмотренный в гл. 6 принцип получения максимального гарантированного результата при наихудших условиях. Игрок A стремится принять такую стратегию, которая должна обеспечить максимальный проигрыш игрока B. Игрок B стремится принять такую стратегию, которая должна обеспечить минимальный выигрыш игрока A.

Рассмотрим подход игрока A. Он должен получить максимальный гарантированный результат при наихудших условиях. Значит, в каждой своей чистой стратегии (строке i) он должен выбрать гарантированный результат в наихудших условиях, т. е. наименьшее значение a_{ij}, которое обозначим

$$\alpha_i = \min_j a_{ij}. \tag{8.1}$$

Для того чтобы этот гарантированный эффект в наихудших условиях был максимальным, надо из всех α выбрать наибольшее значение. Обозначим его α и назовем чистой нижней ценой игры:

$$\alpha = \max \alpha_i = \max \min_j a_{ij}. \tag{8.2}$$

Чистая нижняя цена игры называется «максимин». Таким образом, *максиминная стратегия* — строка матрицы, которой соответствует элемент α. Какие бы стратегии ни применял игрок B, игрок A своей максиминной чистой стратегией гарантировал себе выигрыш, не меньший, чем α. Таково оптимальное поведение первого игрока.

Второй игрок своими оптимальными стратегиями стремится уменьшить выигрыш первого игрока, поэтому он отыскивает

$$\beta_j = \max a_{ij} \tag{8.3}$$

в каждом своем столбце, т. е. определяет максимальный выигрыш игрока A, если B применит свою j-ю чистую стратегию. Из всех своих n j-х чистых стратегий он отыскивает такую стратегию, при которой игрок A получит минимальный выигрыш:

$$\beta = \min_j \beta_j = \min_j \max_i a_{ij}, \tag{8.4}$$

где β — чистая верхняя цена игры.

Чистая верхняя цена игры показывает, какой максимальный выигрыш может себе гарантировать первый игрок, применяя свои стратегии. Если первый игрок применит чистые стратегии, то он может себе гарантировать выигрыш, не меньший, чем α. Второй игрок за счет своих чистых стратегий не допустит, чтобы игрок A мог получить выигрыш, бóльший, чем β.

Таким образом, минимаксная стратегия отображается столбцом матрицы, в котором находится элемент β. Она является оптимальной чистой гарантирующей стратегией игрока B, если он ничего не знает о действиях игрока A.

Чистая цена игры (v) — цена данной игры, если у нее нижняя и верхняя цены игры совпадают:

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = v. \tag{8.5}$$

Пример 8.1

Определите нижнюю и верхнюю цены игры при заданной матрице игры и укажите максиминную и минимаксную стратегии. Представим матрицу игры с обозначениями стратегий β_j, α_i:

$A_i \backslash B_j$	B_1	B_2	B_3	α_i
A_1	1	2	3	1
A_2	4	5	6	4
β_j	4	5	6	

Решение

1. Определим нижнюю цену игры:

$$\alpha_1 = 1;\ \alpha_2 = 4;\ \alpha = 4 \text{ (см. столбец } \alpha_i).$$

2. Определим верхнюю цену игры:

$$\beta_1 = 4;\ \beta_2 = 5;\ \beta_3 = 6;\ \beta = 4 \text{ (см. строку } \beta_j).$$

Таким образом, $\alpha = \beta = 4$, т. е.

$$\max_i \min_j a_{ij} = \min_j \max_i a_{ij} = 4.$$

Значит, $\alpha = \beta = \nu = 4$ — чистая цена игры при стратегиях A_2 и B_1.

Пример 8.2

Определите максиминную и минимаксную стратегии при заданной матрице эффективности:

$A_i \backslash B_j$	B_1	B_2	B_3	B_4
A_1	2	7	6	10
A_2	8	4	9	5

Решение

1. Определим максиминную стратегию:

$$\alpha_1 = 2;\ \alpha_2 = 4;\ \alpha = 4.$$

Максиминная стратегия — строка A_2.

2. Определим минимаксную стратегию:

$$\beta_1 = 8;\ \beta_2 = 7;\ \beta_3 = 9;\ \beta_4 = 10;\ \beta = 7.$$

Минимаксная стратегия — столбец B_2.

Если матрица игры содержит элемент, который является минимальным в своей строке и максимальным в своем столбце, то он является седловой точкой. В этом случае игра является *игрой с седловой точкой.*

Пусть в игре с седловой точкой один игрок придерживается седловой точки, тогда другой игрок получит свой лучший результат, если также будет придерживаться этой точки. Лучшее поведение игрока не должно привести к уменьшению его выигрыша. Зато худшее поведение игрока может привести к уменьшению его выигрыша. В данном случае решением игры являются:

- чистая стратегия первого игрока;
- чистая стратегия второго игрока;
- седловой элемент.

Оптимальные чистые стратегии — это чистые стратегии, образующие седловую точку.

В игре без седловой точки, если игрок A информирован о стратегии, принятой игроком B, он может принять оптимальную стратегию, которая не совпадает с максиминной.

Пример 8.3

Дана матрицы игры

$$\mathbf{A} = \begin{pmatrix} 3 & 5 & 8 & 6 & 11 \\ 8 & 4 & 12 & 7 & 9 \end{pmatrix}.$$

Игроку A стало известно, что игрок B принял минимаксную стратегию. Игрок A должен выбрать оптимальную стратегию при условии, что B_2 — стратегия игрока B ($\beta = 5$).

Решение

1. Определим максиминную стратегию игрока A:

$$\alpha_1 = 3;\ \alpha_2 = 4;\ \alpha = 4.$$

Стратегия $A \to A_2$ — максиминная.

2. Выберем оптимальную стратегию для игрока A. Этой стратегией будет не максиминная A_2, дающая игроку A выигрыш $\alpha = 4$, а та стратегия, которая соответствует $\max a_{ij}$. В этом случае его максимальный гарантированный выигрыш будет равен верхней цене игры $\beta = 5$, поэтому он выберет свою оптимальную стратегию A_1.

Таким образом, рассмотренный пример дает результат, отличный от результата в случае, когда происходит игра с седловой точкой.

Стратегия является *оптимальной*, если ее применение обеспечит игроку наибольший гарантированный выигрыш при любых возможных стратегиях другого игрока.

На примере 8.3 показано, что бывают ситуации, когда игрок A может получить выигрыш, превосходящий максиминный.

Известно, что если игра многократно повторяется в сходных условиях, то в результате можно добиться гарантированного среднего выигрыша, превосходящего для игрока A максиминную стратегию.

8.2. Смешанные стратегии

Если в матричной игре отсутствует седловая точка в чистых стратегиях, то находят верхнюю и нижнюю чистые цены игры. Они показывают, что игрок A не получит выигрыша, превосходящего верхнюю цену игры, и игроку A гарантирован выигрыш, не меньший нижней цены игры. В примере 8.3 по своей оптимальной стратегии A_1, отличной от максиминной, игрок A получил выигрыш, равный верхней цене игры. Это была плата за информированность о стратегии B. Это крайний случай.

Улучшится ли результат игрока A, если информация о действиях противной стороны будет отсутствовать, но игрок будет многократно применять чистые стратегии случайным образом, с определенной вероятностью?

В такой ситуации, оказывается, можно получать выигрыши, в среднем бóльшие нижней цены игры, но мéньшие верхней цены игры.

Смешанная стратегия игрока — это полный набор вероятностей применения его чистых стратегий. Значит, смешанная стратегия является случайной смесью чистых стратегий с определенными вероятностями.

Подведем итоги сказанного и перечислим **условия применения смешанных стратегий**:

1) игра без седловой точки;

2) игроки используют случайную смесь нескольких чистых стратегий;

3) игра многократно повторяется в сходных условиях;

4) ни один из игроков не информирован о данном выборе стратегии другим игроком;

5) допускается осреднение результатов игр.

Применяются следующие обозначения смешанных стратегий.

Для игрока A смешанная стратегия заключается в применении чистых стратегий $A_1, A_2, ..., A_m$ с соответствующими вероятностями $p_1, p_2, ..., p_m$:

$$\mathbf{S}_A = \begin{pmatrix} A_1\ A_2, ..., A_m \\ p_1\ p_2, ..., p_m \end{pmatrix}, \tag{8.6}$$

где

$$\sum_{i=1}^{m} p_i = 1; \tag{8.7}$$

p_i — вероятность применения чистой стратегии A_i.

Для игрока B матрица $\mathbf{S}_B$:

$$\mathbf{S}_B = \begin{pmatrix} B_1\ B_2, ..., B_n \\ q_1\ q_2, ..., q_n \end{pmatrix}, \tag{8.8}$$

где

$$\sum_{i=1}^{n} q_i = 1\ (q_j \geqslant 0); \tag{8.9}$$

q_j — вероятность применения чистой стратегии B_j.

В случае когда $p_i = 1$, для игрока A имеем чистую стратегию

$$\mathbf{S}_A = \begin{pmatrix} A_1, A_2, ..., A_i, ..., A_m \\ 0,\ \ 0, ...,\ \ 1, ...,\ \ 0 \end{pmatrix}. \tag{8.10}$$

Чистые стратегии игрока являются единственно возможными несовместными событиями. В матричной игре, зная матрицу **A** (она относится и к игроку A, и к игроку B), при заданных векторах **p** и **q** можно определить средний выигрыш (математическое ожидание эффекта) игрока A:

$$\mathbf{M}(A, \mathbf{p}, \mathbf{q}) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} p_i q_j, \tag{8.11}$$

где **p** и **q** — векторы, а p_i и q_j — их компоненты.

Первый игрок путем применения смешанных стратегий стремится максимально увеличить свой средний выигрыш. Второй игрок путем применения своих смешанных стратегий стремится этот эффект довести до минимально возможного значения. Игрок A стремится достигнуть:

$$\beta = \min_{\mathbf{q}} \max_{\mathbf{p}} \mathbf{M}(A, \mathbf{p}, \mathbf{q}). \tag{8.12}$$

Игрок B стремится к тому, чтобы

$$\alpha = \max_{\mathbf{p}} \min_{\mathbf{q}} \mathbf{M}(A, \mathbf{p}, \mathbf{q}). \tag{8.13}$$

Обозначим $\mathbf{p}^0$ и $\mathbf{q}^0$ векторы, соответствующие оптимальным смешанным стратегиям. Оптимальными смешанными стратегиями первого и второго игроков называются такие векторы $\mathbf{p}^0$ и $\mathbf{q}^0$, при которых будет выполнено равенство

$$\min \max \mathbf{M}(A, \mathbf{p}, \mathbf{q}) = \max \min \mathbf{M}(A, \mathbf{p}, \mathbf{q}) = \mathbf{M}(A, \mathbf{p}^0, \mathbf{q}^0). \tag{8.14}$$

Цена игры — средний результат игры при использовании оптимальных смешанных стратегий — γ.

Следовательно, решением матричной игры является:

- $\mathbf{p}^0$ — оптимальная смешанная стратегия игрока A;
- $\mathbf{q}^0$ — оптимальная смешанная стратегия игрока B;
- γ — цена игры.

Смешанные стратегии будут оптимальными ($\mathbf{p}^0$ и $\mathbf{q}^0$), если они образуют седловую точку для функции $\mathbf{M}(A, \mathbf{p}, \mathbf{q})$, т. е.

$$\mathbf{M}(A, \mathbf{p}, \mathbf{q}^0) \leqslant \mathbf{M}(A, \mathbf{p}^0, \mathbf{q}^0) \leqslant \mathbf{M}(A, \mathbf{p}^0, \mathbf{q}). \tag{8.15}$$

Существует основная теорема математических игр Неймана.

Теорема 8.1. Для матричной игры с любой матрицей $\mathbf{A}$ величины

$$\alpha = \max_{\mathbf{p}} \min_{\mathbf{q}} \mathbf{M}(A, \mathbf{p}, \mathbf{q})$$

и

$$\beta = \min_{\mathbf{q}} \max_{\mathbf{p}} \mathbf{M}(A, \mathbf{p}, \mathbf{q})$$

существуют и равны между собой:

$$\alpha = \beta = \gamma.$$

Часто данную теорему представляют так: «Всегда имеет место равенство (8.14)».

Следует отметить, что в оптимальных смешанных стратегиях игроку A всегда будет гарантирован средний выигрыш, не меньший, чем цена игры, при любой фиксированной стратегии игрока B (и наоборот для B). Но стоит одному из игроков воспользоваться своей оптимальной смешанной стратегией, его средний выигрыш останется равным цене игры. При этом другой игрок может пользоваться любой смесью активных стратегий.

Активными стратегиями являются стратегии, входящие в состав оптимальных смешанных стратегий с вероятностями, отличными от нуля. Значит, в состав оптимальных смешанных стратегий входят не все существующие стратегии игроков. Все эти преимущества хороши до тех пор, пока не нарушаются правила применения смешанных стратегий.

Представим себе, что противной стороне — игроку B стало известно, какую стратегию использует игрок A при повторениях игры со смешанными стратегиями. Игрок B может выбрать такие стратегии, что игрок A при использовании им смешанных стратегий получит средний выигрыш, меньший его нижней цены игры α. Вместо γ ($\beta > \gamma > \alpha$) он может получить средний выигрыш $\alpha_x < \alpha$.

8.3. Методы решения задач в смешанных стратегиях

Решить игру означает найти цены игры и оптимальные стратегии. Проще всего начать рассмотрение решения матричных игр с простейшей 2×2. Игры с седловой точкой специально рассматриваться не будут. Если получена седловая точка, то это означает, что имеются невыгодные стратегии, от которых следует отказываться. При отсутствии седловой точки можно получить две оптимальные смешанные стратегии. Эти смешанные стратегии записываются так:

$$\mathbf{S}_A = \begin{pmatrix} A_1 & A_2 \\ p_1 & p_2 \end{pmatrix};$$

$$\mathbf{S}_B = \begin{pmatrix} B_1 & B_2 \\ q_1 & q_2 \end{pmatrix}. \tag{8.16}$$

Значит, имеется платежная матрица

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}. \tag{8.17}$$

При этом

$$\left.\begin{aligned} a_{11}p_1 + a_{21}p_2 &= \gamma; \\ a_{12}p_1 + a_{22}p_2 &= \gamma; \\ p_1 + p_2 &= 1; \end{aligned}\right\} \tag{8.18}$$

$$a_{11}p_1 + a_{21}(1 - p_1) = a_{12}p_1 + a_{22}(1 - p_1);$$
$$a_{21}p_1 + a_{21} - a_{21}p_1 = a_{12}p_1 + a_{22} - a_{22}p_1;$$

$$p_1 = \frac{a_{22} - a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})}; \tag{8.19}$$

$$p_2 = 1 - p_1 = \frac{a_{11} - a_{12}}{a_{11} + a_{22} - (a_{12} + a_{21})}. \tag{8.20}$$

Зная p_1 и p_2, находим γ:

$$\gamma = \frac{a_{11}(a_{22} - a_{21})}{a_{11} + a_{22} - (a_{12} + a_{21})} + \frac{a_{21}(a_{11} - a_{12})}{a_{11} + a_{22} - (a_{12} + a_{21})} =$$

$$= \frac{a_{11}a_{22} - a_{12}a_{21}}{a_{11} + a_{22} - (a_{12} + a_{21})}. \tag{8.21}$$

Вычислив γ, находим q_1 и q_2:

$$a_{11}q_1 + a_{12}q_2 = \gamma;$$
$$q_1 + q_2 = 1;$$
$$a_{11}q_1 + a_{12}(1 - q_1) = \gamma;$$

$$\left.\begin{aligned} q_1 &= \frac{\gamma - a_{12}}{a_{11} - a_{12}}; \\ q_2 = 1 - q_1 &= \frac{a_{11} - \gamma}{a_{11} - a_{12}}. \end{aligned}\right\} \tag{8.22}$$

Таким образом, задача решена, так как найдены векторы $\mathbf{q} = \begin{pmatrix} q_1 \\ q_2 \end{pmatrix}$; $\mathbf{p} = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}$ и цена игры γ. Имея матрицу платежей **A**, можно решить эту же задачу графически. При этом методе алгоритм решения весьма прост (рис. 8.1).

1. По оси абсцисс откладывается отрезок единичной длины.

2. По оси ординат откладываются выигрыши при стратегии A_1.

3. На линии, параллельной оси ординат в точке 1, откладываются выигрыши при стратегии A_2.

4. Концы отрезков обозначаются для $a_{11} - B_{11}$, $a_{12} - B_{21}$, $a_{22} - B_{22}$, $a_{21} - B_{12}$ и проводятся две прямые линии $B_{11}B_{12}$ и $B_{21}B_{22}$.

5. Определяется ордината точки пересечения C. Она равна γ. Абсцисса точки C равна p_2 ($p_1 = 1 - p_2$).

Данный метод имеет достаточно широкую область приложения. Это основано на общем свойстве игр $m \times n$: в любой игре $m \times n$ каждый игрок имеет оптимальную смешанную

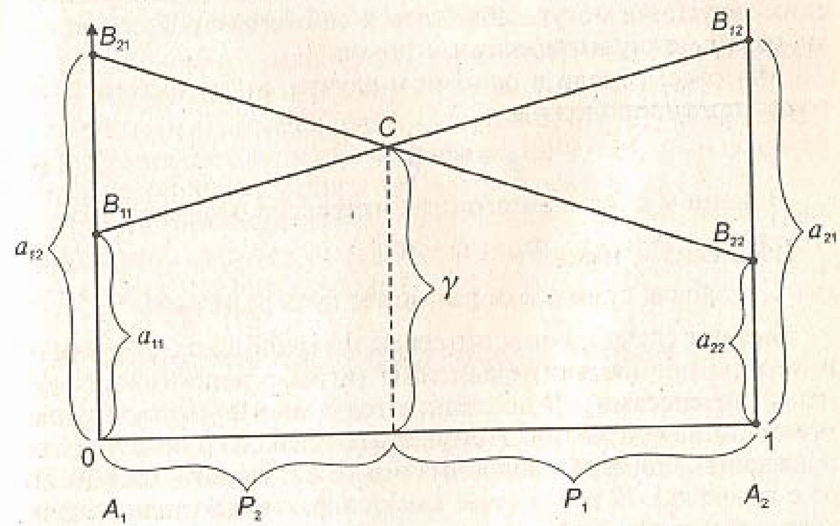


Рис. 8.1. Оптимальная смешанная стратегия

стратегию, в которой число чистых стратегий не больше, чем min (m, n).

Из этого свойства можно получить известное следствие: в любой игре $2 \times n$ и $m \times 2$ каждая оптимальная стратегия $\mathbf{S}_\mathrm{A}^0$ и $\mathbf{S}_\mathrm{B}^0$ содержит не более двух активных стратегий. Значит, любая игра $2 \times n$ и $m \times 2$ может быть сведена к игре 2×2. Следовательно, игры $2 \times n$ и $m \times 2$ можно решить графическим методом.

Когда матрица конечной игры имеет размерность $m \times n$, где $m > 2$ и $n > 2$, то для определения оптимальных смешанных стратегий используется линейное программирование. В работе [118] показано, что любой конечной игре двух лиц с нулевой суммой соответствует некоторая модель линейного программирования вместе с двойственной к ней задачей.

Таким образом, общим методом решения задач при смешанных стратегиях для обоих игроков является построение задачи линейного программирования, которая соответствует данной игре.

Мы остановились на некоторых методах решения задач в теории игр и выбрали простейшие методы, дающие представление о принципиальном подходе к оценке эффективности на основе игровых критериев. На самом деле при практических ситуациях могут наблюдаться самые разнообразные виды связи между интересами игроков.

Мы рассмотрели в основном случаи, когда интересы игроков противоположны:

$$\Phi_B = \varphi(-\Phi_A). \tag{8.23}$$

В данном случае антагонизм интересов означает, что

$$\Phi_B = C - \Phi_A, \tag{8.24}$$

где C — общая сумма, которая может быть разыграна.

Бывают случаи, когда интересы игроков совпадают или их интересы полностью независимы (игры с непротивоположными интересами). В последние годы многие исследователи приступили к развитию этого весьма близкого к практике направления. Примером подобных игр может служить хотя бы игра с природой. В этом случае имеется фактически один игрок, который представляет оперирующую сторону и стремится увеличить критерий эффективности $\Phi(x, y)$, где x — факторы,

которые выбирает оперирующая сторона, а y — факторы, описывающие влияние природы. При этом нельзя считать интересы природы всегда антагонистическими интересам оперирующей стороны.

Другим примером могут служить антагонистические игры с запрещенными ситуациями. Пусть $\Phi_A = f(x_1, x_2)$, а $\Phi_B = -f(x_1, x_2)$. Введем штраф за выход значений факторов x_1 и x_2 за пределы некоторой области D. Тогда игра сводится к игре с непротивоположными интересами:

$$\Phi_A^0 = \begin{cases} f(x_1, x_2) & \text{при } (x_1, x_2) \in D; \\ -\infty & \text{при } (x_1, x_2) \in D; \end{cases} \tag{8.25}$$

$$\Phi_B^0 = \begin{cases} -f(x_1, x_2) & \text{при } (x_1, x_2) \in D; \\ -\infty & \text{при } (x_1, x_2) \in D. \end{cases}$$

Можно привести пример игры с непротивоположными интересами, когда игроки применяют принцип гарантированного результата, а их решения зависят еще и от сил природы:

$$\Phi_A = f(x_1, x_2, y) = -\Phi_B, \tag{8.26}$$

где y — неопределенный природный фактор.

Такое решение может быть принято в системе, в которой на ход игры влияет некоторое третье лицо и поведение его одинаково неясно как первому, так и второму игроку. Эти игроки принимают решение обеспечить себе гарантированный результат по отношению к третьему игроку.

Рассмотрим некоторые практические задачи, в которых используются критерии игр для оценки наиболее эффективного поведения оперирующей стороны.

8.4. Решение экономических задач

Пример 8.4

Предположим, что в некоторой торговой фирме необходимо решить задачу по определению оптимального масштаба сезонного снижения цен.

Имеется 1000 нераспроданных зимних пальто разных размеров и фасонов. Спрос на эти пальто может колебаться в зависимости от характера зимы. Считаем, что зима может быть суровой (B_1) или мягкой (B_2). При мягкой зиме наблюдается слабая зависимость спроса от цены, при суровой — сильная. Торговая фирма может принять четыре стратегии и снизить цену: A_1 — на 10%, A_2 — на 20, A_3 — на 30, A_4 — на 40%.

При мягкой зиме снижение цены на 10% приводит к продаже 200 пальто, на 20% — 300, на 30% — 450, на 40% — 500 пальто. В случае суровой зимы наблюдается повышенный спрос на пальто. При снижении цены на 10% можно продать 300 пальто, на 20% — 700, на 30% — 800, на 40% — 900 пальто.

Для удобства расчетов примем условно, что средняя себестоимость пальто — 150 руб., средняя цена пальто — 250 руб. За функцию потерь (для стратегий A_1, A_2, A_3 и A_4) принимается разность $L(B, A)$ между себестоимостью 1000 нераспроданных пальто и выручкой после продажи их по сниженным ценам.

Определите оптимальную стратегию торговой фирмы, которая гарантирует минимальные потери в данных условиях.

Решение

1. Вычислим функцию потерь $L(B, A)$ (табл. 8.1 и 8.2).

Для удобства выпишем функцию потерь в отдельную таблицу (см. табл. 8.3).

2. Очистим матрицу потерь от доминируемых стратегий. Доминируемой стратегией является строка (столбец), в которой все соответственные члены больше (меньше) членов другой строки при стремлении к минимизации (максимизации) функции. Исходная игра является стратегической, так как нет ни-

Таблица 8.1

Расчет функции потерь $L(B_2, A)$

Стратегия	Процент снижения цены	Новая цена, руб.	Реализация		Себестоимость 1000 пальто, руб.	Потери, руб.
			шт.	руб.		
A_1	10	225	200	45 000	150 000	105 000
A_2	20	200	300	60 000	150 000	90 000
A_3	30	175	450	78 750	150 000	71 250
A_4	40	150	500	75 000	150 000	75 000

Таблица 8.2

Расчет функции потерь $L(B_1, A)$

Стратегия	Процент снижения цены	Новая цена, руб.	Реализация		Себестоимость 1000 пальто, руб.	Потери, руб.
			шт.	руб.		
A_1	10	225	300	67 500	150 000	82 500
A_2	20	200	700	140 000	150 000	10 000
A_3	30	175	800	140 000	150 000	10 000
A_4	40	150	900	135 000	150 000	15 000

каких данных о том, какой будет зима, хотя мы знаем, что зима может характеризоваться двумя возможными состояниями — B_1 или B_2. Эта стратегическая игра обозначается (B, A, L).

В данной игре можно отметить, что стратегии A_1 и A_4 доминируются стратегией A_3 ($71{,}25 < 75 < 105$; $10 < 15 < 82{,}5$). Выбросим эти две стратегии из рассмотрения и получим новую функцию потерь (табл. 8.4).

3. Проведем статистический эксперимент и преобразуем стратегическую игру (B, A, L) в статистическую игру (B, D, K). Статистическая игра имеет характерную особенность, заключающуюся в возможности получения информации на основе некоторого статистического эксперимента. Эксперимент проводится для оценки вероятностей стратегий природы. Этим вопросом необходимо заняться в решаемой задаче. Нам надо выяснить, какой будет эластичность спроса от цены в случае мягкой (B_2) или суровой (B_1) зимы.

Таблица 8.3

Функция потерь $L(B, A)$

Стратегия	Процент снижения цены	Потери, тыс. руб.	
		при мягкой зиме (B_2)	при суровой зиме (B_1)
A_1	10	105,00	82,5
A_2	20	90,00	10,0
A_3	30	71,25	10,0
A_4	40	75,00	15,0

Таблица 8.4

Новая функция потерь $L(B, A)$

Стратегия	Процент снижения цены	Потери, тыс. руб.	
		при мягкой зиме (B_2)	при суровой зиме (B_1)
A_2	20	90,00	10
A_3	30	71,25	10

В результате экспертного оценивания мы узнали условные распределения результатов эластичности спроса в зависимости от B_1 или B_2. Пусть x_1 — высокая эластичность спроса, x_2 — низкая эластичность спроса,

$$p(x_1/B_1) = 0{,}7;$$
$$p(x_1/B_2) = 0{,}4;$$
$$p(x_2/B_1) = 0{,}3;$$
$$p(x_2/B_2) = 0{,}6.$$

Эти результаты надо трактовать следующим образом:

$p(x_2/B_1)$ — условная вероятность низкой эластичности спроса при суровой зиме;

$p(x_2/B_2)$ — условная вероятность низкой эластичности спроса при мягкой зиме и т. д.

Итак, остались две стратегии (см. табл. 8.4) — A_2 и A_3. На основе статистического эксперимента мы получили два возможных результата x_1 и x_2. Этим результатам соответствует одна из двух возможных допустимых стратегий — A_2 или A_3.

Природа также может «применять» смешанную (рандомизированную) стратегию при выборе состояния B в статистической игре. Фирма (оперирующая сторона), учитывая x_1, x_2, A_2 и A_3, имеет множество нерандомизированных функций решения D. Покажем их в табл. 8.5.

Функции d_i можно объяснить так. Например, функция d_3 означает, что надо принять решение A_3, если результат эксперимента будет x_1. Если же по эксперименту получим x_2, то надо принять решение A_2.

Таблица 8.5

Функции решения фирмы

x_i	D			
	d_1	d_2	d_3	d_4
x_1	A_2	A_2	A_3	A_3
x_2	A_2	A_3	A_2	A_3

Теперь с учетом характера зимы можно вычислить функции риска $K(B, d_i)$:

$$K(B_2, d_1) = 90 \cdot 0{,}4 + 90 \cdot 0{,}6 = 90;$$
$$K(B_2, d_2) = 90 \cdot 0{,}4 + 71{,}25 \cdot 0{,}6 = 36 + 42{,}75 = 78{,}75;$$
$$K(B_2, d_3) = 71{,}25 \cdot 0{,}4 + 90 \cdot 0{,}6 = 28{,}5 + 54 = 82{,}5;$$
$$K(B_2, d_4) = 71{,}25 \cdot 0{,}4 + 71{,}25 \cdot 0{,}6 = 71{,}25;$$
$$K(B_1, d_1) = 10 \cdot 0{,}7 + 10 \cdot 0{,}3 = 10;$$
$$K(B_1, d_2) = 10 \cdot 0{,}7 + 10 \cdot 0{,}3 = 10;$$
$$K(B_1, d_3) = 10 \cdot 0{,}7 + 10 \cdot 0{,}3 = 10;$$
$$K(B_1, d_4) = 10 \cdot 0{,}7 + 10 \cdot 0{,}3 = 10.$$

Поясним вычисление функции риска $K(B_2, d_2)$ при мягкой зиме B_2 и функции решения торговой фирмы d_2. Функция решения торговой фирмы d_2 приписывает результату x_1 (большой эластичности спроса) стратегию A_2, а результату x_2 — стратегию A_3 (см. табл. 8.5). При состоянии зимы B_2 стратегии A_2 соответствуют потери 90 тыс. руб., а стратегии A_3 — 71,25 тыс. руб.

Условная вероятность высокой эластичности спроса (x_1) при мягкой зиме $p(x_1/B_2) = 0{,}4$, а вероятность низкой эластичности спроса (x_2) при мягкой зиме $p(x_2/B_2) = 0{,}6$. Следовательно, риск $K(B_2, d_2)$ может быть рассчитан так:

$$K(B_2, d_2) = 90 \cdot 0{,}4 + 71{,}25 \cdot 0{,}6 = 78{,}75 \text{ тыс. руб.}$$

Он характеризует средние потери.

Полученные результаты представим в виде матрицы с обозначениями стратегий (табл. 8.6).

Таким образом, мы сформулировали статистическую игру (B, D, K).

Таблица 8.6

Матрица функции риска $K(B, D)$, тыс. руб.

B_j	D			
	d_1	d_2	d_3	d_4
B_2	90	78,75	82,5	71,25
B_1	10	10,00	10,0	10,00

4. Найдем оптимальную функцию решения для статистической игры (B, D, K). Если бы у нас была априорная информация о распределении состояний природы (зимы), то можно было бы получить байесовскую функцию решения.

Поскольку такой информации нет, то для торговой фирмы оптимальной в игре (B, D, K) будет минимаксная функция. Найдем нижнюю и верхнюю цены игры:

$$\alpha_1 = 71{,}25;\ \alpha_2 = 10;\ \alpha = 71{,}25;$$

$$\beta_1 = 90;\ \beta_2 = 78{,}75;\ \beta_3 = 82{,}5;$$

$$\beta_1 = 71{,}25;\ \beta = 71{,}25.$$

Таким образом, $\alpha = \beta = \nu = 71{,}25$, т. е. имеется седловая точка.

Из табл. 8.5 следует, что принятая торговой фирмой стратегия d_4 имеет два значения: $d_4(x_1) = A_3$ и $d_4(x_2) = A_3$. Следовательно, оптимальной стратегией торговой фирмы будет снижение цен на 30% без учета результатов исследования эластичности спроса. Область D выродилась в отрезок (рис. 8.2). В данном случае максимальный ущерб будет равен 71,25 тыс. руб.

При помощи теории игр можно решать не только задачи в интересах фирм, объединений и отраслей народного хозяйства, но и макроэкономические задачи. Приведем пример выбора оптимального варианта капитальных вложений.

Пример 8.5

В примере 8.4 мы не могли получить предварительную информацию о возможном состоянии погоды и на основе опроса экспертов вынуждены были определять влияние погоды на спрос.

В ряде случаев представляется возможным на основе долгосрочных наблюдений набрать необходимую информацию.

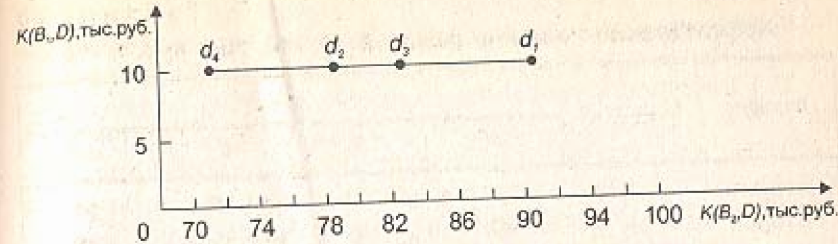


Рис. 8.2. Оптимальная стратегия торговой фирмы

Это позволяет перейти от задачи, решаемой в условиях полной неопределенности, к задаче, решаемой в условиях риска.

Предположим, что перед проектной организацией поставлена задача выбора оптимального плана капитальных вложений для строительства электротехнического комплекса. Выбор может быть проведен из трех возможных вариантов:

- строить в пригороде Санкт-Петербурга;
- строить на Европейском Севере страны;
- строить в Сибири.

Первый вариант характеризуется очень хорошими климатическими условиями, близостью заводов-поставщиков, следовательно, малой зависимостью от состояния транспорта и погоды, но большой удаленностью от основных потребителей продукции.

Второй вариант характерен хорошей привязкой к разветвленной транспортной сети (правда, в этом он уступает первому варианту), зависимостью производства от состояния транспортных путей и от климатических условий, удаленностью от основных потребителей (приблизительно такой же, как и в первом варианте).

Наконец, при третьем варианте влияние климатических условий и состояние транспорта на производство существенно. Однако поставка продукции основным потребителям не связана с дальними перевозками.

Экономический эффект каждого из вариантов капитальных вложений можно определить, если учитывать годовую прибыль, затраты на строительство и эксплуатацию. Результаты экономических расчетов по условным данным представлены в следующей матрице (табл. 8.7).

Таблица 8.7

Эффективность капитальных вложений, млн руб. в год

Вариант	Стратегия	B			
		B_1	B_2	B_3	B_4
Первый	A_1	55	50	45	40
Второй	A_2	60	30	35	25
Третий	A_3	75	75	45	35

Случайные факторы, влияющие на эффективность капитальных вложений, обозначим

$$B = \{B_1, B_2, B_3, B_4\},$$

где B_1 — благоприятные климатические условия и благоприятные условия работы транспорта;

B_2 — благоприятные климатические условия и неблагоприятные условия работы транспорта;

B_3 — неблагоприятные условия климата и благоприятные условия работы транспорта;

B_4 — неблагоприятные климатические условия и неблагоприятные условия для работы транспорта.

Таким образом, оперирующая сторона должна определить оптимальный вариант стратегической игры (A, B, Φ).

Это упрощенный подход, так как не рассматривается весь комплекс факторов, влияющих на эффективность капитальных вложений. Ограничение числа влияющих факторов надо рассматривать как методический прием.

Решение

1. Представим себе, что оперирующая сторона не обладает никакой дополнительной информацией. Тогда решение будет очень простым:

$$\alpha_1 = 40; \alpha_2 = 25; \alpha_3 = 35; \alpha = 40;$$

$$\beta_1 = 75; \beta_2 = 75; \beta_3 = 45; \beta_4 = 40; \beta = 40;$$

$$\alpha = \beta = \nu = 40.$$

Следовательно, надо рекомендовать строить комплекс в Ленинградской области (первый вариант).

Однако проектирующая организация может получить достоверные данные за длительный период о состоянии погоды и загрузке транспорта. Значит, на основе статистических данных можно получить распределение вероятностей состояний B.

2. Определим апостериорное распределение состояния погоды и условий работы транспорта. Так как методы обработки многолетних статистических данных не входят в задачу данной книги, предположим, что в результате этой обработки мы получим следующее апостериорное распределение:

$$p(B_1) = 0{,}13; \ p(B_2) = 0{,}32;$$

$$p(B_3) = 0{,}18; \ p(B_4) = 0{,}37.$$

Теперь можно решить задачу принятия решений в условиях риска. Стратегическая игра (A, B, Φ) преобразуется в статистическую игру. Необходимо найти математические ожидания (млн руб.):

$$\Phi(B, A_1) = 55 \cdot 0{,}13 + 50 \cdot 0{,}32 + 45 \cdot 0{,}18 + 40 \cdot 0{,}37 = 46{,}05;$$

$$\Phi(B, A_2) = 60 \cdot 0{,}13 + 30 \cdot 0{,}32 + 35 \cdot 0{,}18 + 25 \cdot 0{,}37 = 32{,}95;$$

$$\Phi(B, A_3) = 75 \cdot 0{,}13 + 75 \cdot 0{,}32 + 45 \cdot 0{,}18 + 35 \cdot 0{,}37 = 54{,}8.$$

Рассчитанные математические ожидания максимизирует оптимальная байесовская стратегия оперирующей стороны. На ее основе получаем, что наиболее выгодным вариантом будет строительство комплекса в Сибири, где

$$\max_{A_i \in A} \Phi(B, A_i) = \Phi(B, A_3) = 54{,}8 \text{ млн руб.}$$

Как и любой математический аппарат, теория игр имеет свои ограничения. Например, в играх с природой игрок «природа» сознательно не противодействует и относится к нам безразлично. О его поведении можно судить только на основе многолетних статистических данных, если предполагать, что в дальнейшем существенных отклонений от тенденции его поведения в прошлом не будет. Решения следует принимать в расчете на наихудшие условия. Однако заранее известно, что реальный процесс даст более высокие результаты показателя эффективности или эффекта.

Для любой задачи линейного программирования существует эквивалентная задача теории игр.

Некоторые экономические задачи большой размерности, которые нельзя решить существующими методами линейного программирования, были решены с заданной точностью методами теории игр.

Глава 9
ОЦЕНКА ЭФФЕКТИВНОСТИ НА ОСНОВЕ ИНФОРМАЦИОННЫХ КРИТЕРИЕВ

9.1. Основные понятия и определения теории информации

Теория информации широко используется для изучения процессов управления, поэтому остановимся на тех понятиях и определениях, которые связаны с процессом управления.

Преобразование информации является главнейшей проблемой процесса управления. Понятие «информация», которым в настоящее время пользуются, абстрагировано от смыслового содержания информации. Для характеристики информации пользуются количественной мерой Шеннона. Эти количественные характеристики семантического содержания информации позволяют решать множество задач загрузки каналов связи и памяти ЭВМ, выбора систем ЭВМ и их быстрого действия в зависимости от объема обрабатываемой информации и конкретных задач управления.

Теорией информации называется наука, изучающая количественные закономерности получения, передачи, обработки и хранения информации.

Поясним только те понятия и определения, которые применяются при оценке эффективности системы управления или ее подсистем при помощи информационных критериев. Следовательно, рассмотрим вероятностный подход к процессам преобразования информации.

Пусть имеется дискретная случайная величина X, которая может принять только два значения с равной вероятностью. Определим количество информации, содержащейся в данной случайной величине.

Для этого надо воспользоваться известной формулой и считать, что X принимает каждое значение с вероятностью, равной $\frac{1}{2}$:

$$I = -\log\frac{1}{2} = 1 \text{ бит}.$$

Бит является двоичной единицей количества информации. Если дискретная случайная величина может принять n значений с разной вероятностью, то количество информации

$$I_x = -\sum_{i=1}^{n} p_i \log p_i, \qquad (9.1)$$

где p_i — вероятность того, что случайная величина X примет значение x_i:

$$p_i = p(X = x_i).$$

Информационная пропускная способность — максимальное количество информации, которое может быть передано по каналу в единицу времени. Различают *максимальное* и *минимальное* количество информации, необходимой при управлении объектом с заданной точностью. Зная эти значения, можно определить *верхнюю* и *нижнюю* границы объема памяти ЭВМ.

В теории информации одним из основных понятий является энтропия. *Энтропией* системы называется сумма произведений вероятностей различных состояний системы на логарифмы этих вероятностей, взятых с обратным знаком:

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i. \qquad (9.2)$$

Энтропия, как и количество информации, измеряется в двоичных (битах) или десятичных (хартли) единицах в зависимости от того, по какому основанию берется логарифм. На практике удобнее пользоваться битами.

Если случайная величина X принимает n значений и каждое значение равновероятно, то энтропия, являющаяся мерой неопределенности, будет максимальной:

$$H(X) = -n\frac{1}{n}\log\frac{1}{n} = -\log 1 + \log n = \log n.$$

Энтропия системы с равновозможными состояниями равна логарифму числа состояний. Так, если система может находиться в 16 состояниях, ее энтропия $H(X) = \log 16 = 4$. Из формулы (9.2) видно, что энтропия некоторого дискретного процесса в момент времени t зависит только от числа возможных состояний и их вероятностей.

Если каждое состояние характеризуется некоторой координатой процесса, то энтропия не зависит от значения этой координаты.

Если одна из вероятностей достоверна ($p_i = 1$), то все остальные вероятности $p_1, p_2, ..., p_{i-1}, p_{i+1}, ..., p_{n-1}, p_n$ равны нулю и энтропия системы равна нулю.

До сих пор мы рассматривали системы, связанные с дискретными случайными величинами. Мы могли перечислить их возможные значения и соответствующие им вероятности. Пусть имеется простая система, описываемая одной непрерывной случайной величиной X с плотностью распределения $f(x)$. Под X будем понимать координату процесса управления, которая в каждый момент времени t является непрерывной случайной величиной. Тогда процесс назовем *непрерывным случайным процессом.*

Оценим энтропию непрерывной случайной величины. Для этого определим предел точности измерений и непрерывную случайную величину сведем к дискретной величине. Пусть Δx соответствует отрезку, в пределах которого значения координаты X неразличимы. Тогда энтропию системы X, рассматриваемую с точностью до Δx, можно определить приближенно [21]:

$$H_{\Delta x}(X) = -\sum_i f(x_i)\Delta x \log [f(x_i)\Delta x] = -\sum_i f(x_i)\Delta x[\log f(x_i) + \log \Delta x] =$$
$$= -\sum_i f(x_i) \log f(x_i)\Delta x - \log \Delta x \sum_i f(x_i)\Delta x.$$

При достаточно малом Δx можно принять

$$\sum_i f(x_i) \log f(x_i)\Delta x \approx \int_{-\infty}^{\infty} f(x) \log f(x)\, dx,$$

а

$$\sum_i f(x_i)\Delta x \approx \int_{-\infty}^{\infty} f(x)\, dx = 1.$$

Тогда

$$H_{\Delta x}(X) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx - \log \Delta x, \tag{9.3}$$

где Δx — интервал, характеризующий точность измерения;

$f(x)$ — плотность вероятности случайной величины X (или случайной координаты X).

От Δx не зависит $H^*(X) = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx$ и зависит $\log \Delta x$.

С уменьшением Δx этот член будет неограниченно возрастать. С повышением точности будет расти неопределенность:

$$H_{\Delta x}(X) = H^*(X) - \log \Delta x.$$

От точности измерения Δx зависит начало отсчета, при котором энтропия вычисляется.

Доказано, что энтропия процесса с координатой X не меняется, если изменить начало координат. Для сокращения записи обычно принимают $\Delta x = 1$ и не забывают, что если мы хотим сравнивать энтропии координат непрерывного процесса, то начало отсчета должны брать одно и то же. В общем случае энтропию непрерывного процесса можно записать в виде математического ожидания функции:

$$H_{\Delta x}(X) = -\int_{-\infty}^{\infty} f(x) \log [f(x)\Delta x]\, dx = \mathbf{M}\{-\log [f(x)\Delta x]\}. \tag{9.4}$$

Аналогично

$$H^*(X) = \mathbf{M}\{-\log f(x)\}. \tag{9.5}$$

Таким образом, энтропия непрерывного процесса является усредненной характеристикой плотности распределения. Энтропия не является исчерпывающей характеристикой случайной координаты X рассматриваемого процесса. Непрерывные сообщения не имеют абсолютной меры энтропии. Для них вводится понятие относительной энтропии. Эталоном обычно берут непрерывный сигнал X, имеющий равномерный закон распределения в интервале ε:

$$\int_{-\infty}^{\infty} f(x')\, dx = \int_0^{\varepsilon} dx' = \varepsilon.$$

Неопределенность непрерывной величины X можно определить числом, к которому стремится разность энтропий сигналов X и X':

$$H_\varepsilon(x) = \lim_{\Delta x \to 0} [H(x) - H(x')] = -\int_{-\infty}^{\infty} f(x) \log f(x)\, dx - \log \varepsilon.$$

Если считать, что стандартная величина имеет равномерный закон распределения в единичном интервале ($\varepsilon = 1$), то

$$H_{\varepsilon = 1}(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) \, dx. \tag{9.6}$$

Данное выражение является относительной энтропией. Здесь за стандарт принята равномерно распределенная величина в единичном интервале.

Проанализируем экстремальные свойства энтропии некоторых непрерывных процессов.

1. Известен интервал изменения $[a, b]$ непрерывной случайной величины. Известно, что

$$\int_a^b f(x) \, dx = 1. \tag{9.7}$$

Неизвестна плотность распределения случайной величины. Для неизвестной функции $f(x)$ найти максимум энтропии, которая задается выражением:

$$H^*(X) = \int_a^b f(x) \log f(x) \, dx.$$

Для ограниченной на конечном отрезке случайной величины энтропия будет максимальной при равномерном распределении. Таким образом, максимальное значение энтропии

$$H^*(X) = -\int_a^b \frac{1}{b-a} \log \frac{1}{b-a} \, dx = \log (b-a);$$

$$H(X) = \log (b-a) - \log \Delta x.$$

Итак,

$$H(X) = \log \frac{b-a}{\Delta x}, \tag{9.8}$$

где Δx — интервал, на который разбита непрерывная случайная величина X.

2. Случайная величина изменяется в неограниченной области, известны ее дисперсия σ_x^2, математическое ожидание m_x. Определим закон распределения, при котором функционал, равный энтропии, обращается в максимум:

$$H^*(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) \, dx = \max$$

при условиях

$$\int_{-\infty}^{\infty} (x-\bar{x})^2 f(x) \, dx = \sigma_x^2;$$

$$\int_{-\infty}^{\infty} x f(x) \, dx = m_x;$$

$$\int_{-\infty}^{\infty} f(x) \, dx = 1.$$

Тогда

$$H^*(X) = \log(\sigma\sqrt{2\pi}) + \frac{1}{2}\log e = \log(\sqrt{2\pi e}\,\sigma),$$

где e — основание натурального логарифма.

$$H(X) = \log (\sigma\sqrt{2\pi e}) - \log \Delta x = \log \left[\frac{\sqrt{2\pi e}\sigma}{\Delta x}\right]. \tag{9.9}$$

Энтропия непрерывного процесса на неограниченном участке с заданной дисперсией его случайной координаты максимальна при нормальном законе распределения. Она зависит от σ и не зависит от m_x.

3. Случайная величина изменяется на положительной полуоси $0 < x < \infty$. Известно ее математическое ожидание (среднее значение) $a = \frac{1}{\lambda}$. Определим ее экстремальное распределение:

$$H^*(X) = -\int_{-\infty}^{\infty} f(x) \log f(x) \, dx = \max$$

при условии

$$\int_{-\infty}^{\infty} x f(x) \, dx = \frac{1}{\lambda}.$$

Таким распределением будет экспоненциальное:

$$f(x) = \lambda e^{-\lambda x}.$$

Максимальная энтропия

$$H^{*}(X) = -\mathbf{M}[\log (\lambda e^{-\lambda x})] = -\mathbf{M}[\log \lambda - \lambda x \log e] = \log \frac{e}{\lambda};$$

$$H(X) = \log \frac{e}{\lambda} - \log \Delta x. \qquad (9.10)$$

Энтропия непрерывного процесса со случайной координатой, принимающей только положительные значения, при заданном среднем значении $\frac{1}{\lambda}$ максимальна при экспоненциальном законе распределения.

Пример 9.1

На химическом производстве состояние процесса характеризуется тремя координатами: t — температурой, P — давлением и V — объемом прореагировавших газов. Температура распределена с равномерной плотностью в диапазоне $t = 171 - 180$ °С (точность 1° С). Давление распределено по нормальному закону со средним значением $m_p = 120$ Па и средним квадратическим отклонением, равным 4 Па (точность 0,5 Па). Объем реагирующих газов меняется по экспоненциальному закону распределения с параметром $\lambda = \frac{1}{2}$ 1/м^3 (точность 0,1 м^3). Случайные величины параметров процесса считаем независимыми.

Найдите энтропию процесса.

Решение

1. Определим энтропию первой координаты t:

$$H(t) = \log \frac{t_2 - t_1}{\Delta t} = \log \frac{180 - 171}{1} = \log 9.$$

2. Определим энтропию второй координаты P:

$$H(P) = \log \frac{\sigma_p \sqrt{2\pi e}}{\Delta P} = \log \frac{4\sqrt{6{,}283 \cdot 2{,}718}}{0{,}5} \approx \log 33{,}07.$$

Отметим, что при определении энтропии мы не использовали математического ожидания, так как энтропия не зависит от значения математического ожидания при нормальном законе распределения случайной величины.

3. Определим энтропию третьей координаты V:

$$H(V) = \log \frac{e}{\lambda \Delta V} = \log \frac{2{,}718}{0{,}5 \cdot 0{,}1} = \log 54{,}36.$$

4. Определим энтропию химического процесса:

$$H(t, P, V) = \log \frac{t_2 - t_1}{\Delta t} + \log \frac{\sigma_p \sqrt{2\pi e}}{\Delta P} + \log \frac{e}{\lambda \Delta V} =$$
$$= \log (9 \cdot 33{,}07 \cdot 54{,}36) \approx 14 \text{ бит}.$$

Таким образом, общая энтропия системы при трех независимых параметрах управления равна 14 бит. Рассмотрим физический смысл чисел, полученных в последнем расчете. Число 9 показывает, что на участке длиной $t_2 - t_1 = 9$°С уложится ровно девять участков нечувствительности ценой 1°С (при равномерном законе распределения). Число 33,07 показывает, что на общем участке длиной $\sigma_p \sqrt{2\pi e} = 16{,}536$ укладывается 33,07 участка нечувствительности ценой 0,5 Па (при нормальном законе распределения). Наконец, при экспоненциальном законе распределения параметра (объем реагирующих газов) на участке $\frac{e}{\lambda} = 5{,}436$ укладывается 54,36 участка нечувствительности ценой 0,1 м^3.

В примере 9.1 мы считали, что координаты системы управления являются независимыми случайными величинами: t, P, V. Как вести расчеты в том случае, если координаты процесса являются зависимыми случайными величинами? Кроме того, в реальных процессах не всегда представляется возможным измерять все интересующие нас параметры. Часто по информации о некотором зависимом событии приходится судить об информации интересующего нас события. Для этого необходимо использовать условную энтропию.

9.2. Условная энтропия процесса

Имеются две зависимые системы X и Y. Введем следующие обозначения:

$H(X/y)$ — частная условная энтропия непрерывной случайной величины X относительно непрерывной случайной величины y; служит мерой остаточной неопределенности случайной величины X процесса, характеризуемого двумя координатами, после того, как стало известно, что другая координата — случайная величина Y приняла значение $Y = y$. Эта энтропия будет зависима от величины Y:

$$H(X/y) = -\int_{-\infty}^{\infty} f(x/y) \log [f(x/y)]\, dx. \tag{9.11}$$

В формуле (9.11):

$f(x/y)$ — условная плотность вероятности случайной величины X;

$f(x, y)$ — плотность вероятности системы случайных величин (X, Y);

$H(Y/x)$ — частная условная энтропия системы Y при условии, что система X находится в состоянии x (частная условная энтропия Y относительно X).

Запишем в форме математического ожидания $H(Y/x)$:

$$H(Y/x) = M[-\log f(Y/x)]. \tag{9.12}$$

Если частная условная энтропия X относительно Y зависит от величины Y, то можно провести ее усреднение по всем возможным значениям случайной величины Y и получить среднюю условную энтропию X относительно Y:

$$H_y(X) = -M[\log f(X/y)] = M[H(X/y)]. \tag{9.13}$$

Отметим, что для независимых случайных координат процесса X и Y условные энтропии координаты X процесса всегда совпадают с энтропией $H(X)$:

$$H(X/y) = H_y(X) = H(X).$$

В общем случае энтропия случайной величины не может быть меньше ее средней условной энтропии:

$$\left.\begin{aligned} H(X) &\geqslant H_y(X); \\ H(Y) &\geqslant H_x(Y). \end{aligned}\right\} \tag{9.14}$$

В формуле (9.14)

$$H_y(X) = M[H(X/y)] = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_2(y) f(x/y) \cdot \log [f(x/y)]\, dx dy. \tag{9.15}$$

Если учесть, что $f(x, y) = f_2(y) \cdot f(x/y)$, то можно записать:

$$H_y(X) = -\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y) \log [f(x/y)] dx dy,$$

где $f_2(y)$ — плотность вероятности случайной величины Y.

Если случайный процесс характеризуется случайными координатами X и Y, то

$$H(X, Y) = H(X) + H_x(Y) = H(Y) + H_y(X). \tag{9.16}$$

Итак, энтропия двумерного процесса, определяемого двумя зависимыми величинами X и Y, равна сумме энтропий одной из величин и средней условной энтропии другой случайной величины относительно первой. Это положение распространяется на n-мерные процессы:

$$\begin{aligned} H(X_1, X_2, ..., X_n) &= -M[\log f(X_1, X_2, ..., X_n)] = \\ = -M[\log f_1(X_1) &+ \log f_2(X_2/X_1) + ... + \log f_n(X_n/X_1, X_2, ..., X_{n-1})] = \\ &= H(X_1) + H_{x_1}(X_2) + ... + H_{x_1 x_2, ..., x_{n-1}}(X_n). \end{aligned} \tag{9.17}$$

Наконец, если параметры процесса статистически независимы, то приходим к случаю, рассмотренному в примере 9.1:

$$H(X_1, X_2, ..., X_n) = \sum_{i=1}^{n} H(X_i), \; (i = \overline{1, n}). \tag{9.18}$$

Энтропия n-мерного процесса с независимыми координатами равна сумме энтропий каждой координаты процесса. Так, в примере 9.1

$$H(t, P, V) = H(t) + H(P) + H(V).$$

9.3. Определение количества информации в системе с вероятностными параметрами

Предположим, что надо решать задачи информационного плана в системе управления. Для этого рассмотрим безынерционное звено, на входе которого имеется случайная величина Y, а на выходе — случайная величина X. Необходимо определить количество информации о входной величине, содержащееся в выходной величине. В качестве такого безынерционного звена можно рассматривать и систему управления. Если будем наблюдать выходной сигнал X, то полученная информация о входной величине Y изменит энтропию Y, т. е. $H(Y)$.

Условимся за среднее количество информации при наблюдении величины X принимать разность между энтропиями величины Y до и после наблюдения величины X:

$$I_X(Y) = H(Y) - H_X(Y). \tag{9.19}$$

На основании формулы (9.14) можно заключить, что $I_X(Y) \geqslant 0$. При этом среднее количество информации о случайной величине Y, содержащееся в случайной величине X, будет равно нулю, если случайные величины X и Y независимы:

$$I_X(Y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\log \frac{f(x, y)}{f_1(x)f_2(y)}\, dxdy, \tag{9.20}$$

где $f_1(x)$ — плотность вероятности случайной величины X;
$f_2(y)$ — плотность вероятности случайной величины Y;
$f(x, y)$ — плотность вероятности случайного вектора с компонентами X и Y.

Известно, что

$$I_X(Y) = I_y(X). \tag{9.21}$$

Воспользуемся понятием частного количества информации. Пусть случайная величина X приняла данное значение x, т. е. $X = x$. При этом надо определить количество информации $I(Y/x)$ о случайной величине Y, которое и называется *частным количеством информации*:

$$I(Y/x) = \int_{-\infty}^{\infty} f(y/x)\log \frac{f(y/x)}{f_2(y)}\, dy. \tag{9.22}$$

Для определения частного количества информации надо иметь следующие исходные данные:

$f_2(y)$ — плотность вероятности случайной величины Y;
$f(y/x)$ — условная плотность вероятности случайной величины Y.

Для расчета среднего количества информации при входном сигнале Y и наблюдаемом выходном сигнале X надо знать:

$f_1(x)$ — плотность вероятности случайной величины X;
$f_2(y)$ — плотность вероятности случайной величины Y;
$f(x, y)$ — плотность вероятности системы случайных величин (случайного вектора).

Это вполне понятно, так как среднее количество информации является вероятностным усреднением частного количества информации по всем возможным значениям случайной величины X, которую мы наблюдаем и по которой судим о случайной величине Y.

Вызывает определенный интерес среднее количество информации при передаче нормально распределенных случайных величин. Пусть Y — входной сигнал, а X — выходной сигнал. Наблюдаем выходной сигнал X. Определим среднее количество информации по величине Y при наблюдении случайной величины X.

Итак, рассматриваются две случайные величины, подчиненные нормальному закону распределения, т. е. случайный вектор $\begin{pmatrix} X \\ Y \end{pmatrix}$. Этот вектор может быть охарактеризован пятью параметрами:

$$m_x,\ m_y,\ \sigma_x,\ \sigma_y,\ \rho_{xy},$$

где ρ_{xy} — коэффициент корреляции между двумя случайными величинами (компонентами двумерного случайного вектора);
m_x, m_y — математические ожидания;
σ_x^2, σ_y^2 — дисперсии случайных величин X и Y.

Не вдаваясь в подробности вычислений, запишем в форме математического ожидания

$$I_X(Y) = \mathrm{M}\left[\log \frac{f(X, Y)}{f_1(X)f_2(Y)}\right] = -\frac{1}{2}\log\,(1 - \rho^2_{xy}). \tag{9.23}$$

Следовательно, среднее количество информации при входном сигнале Y и наблюдаемом выходном сигнале X, которые распределены нормально, зависит только от одного из пяти перечисленных выше параметров случайного вектора — от ρ_{xy}. Если бы обе случайные величины были независимы, то количество информации равнялось бы нулю, так как $\rho_{xy} = 0$.

Остановимся еще на одном аспекте, приближающем рассмотренную ранее схему (рис. 9.1) к реальным задачам, когда присутствуют возмущающие параметры.

Пусть A — измерительное устройство. Измерение проводится без запаздываний с ошибкой случайного характера Z,

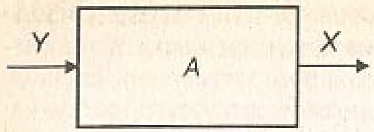


Рис. 9.1. Безынерционное звено системы управления:

Y — входной сигнал;
X — выходной сигнал

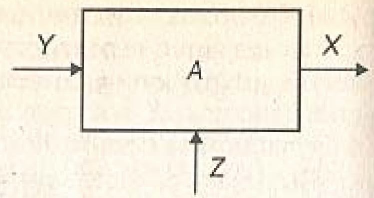


Рис. 9.2. Схема звена системы управления при наличии помехи:

Y — входной сигнал;
X — выходной сигнал;
Z — сигнал помехи

которая подчиняется нормальному закону распределения и не зависит от *Y*. Следовательно, новые условия записываются так:

1) $X = Y + Z$;

2) известны σ_y, σ_z;

3) $\sigma_x^2 = \sigma_y^2 + \sigma^2$;

4) совместный закон распределения (X, Y) нормальный.

Определим среднее количество информации о величине *Y* по наблюдению на выходе измерительного прибора при наличии возмущения в звене (рис. 9.2).

В данном случае коэффициент корреляции удобно выразить через дисперсии случайных величин:

$$\rho_{xy} = \frac{\sigma_y}{\sqrt{\sigma_y^2 + \sigma_z^2}}.$$

Тогда среднее количество информации определится по формуле (9.23):

$$I_x(Y) = -\frac{1}{2}\log\left(1 - \frac{\sigma_y^2}{\sigma_y^2 + \sigma_z^2}\right) = -\frac{1}{2}\log\frac{\sigma_z^2}{\sigma_y^2 + \sigma_z^2} =$$

$$= \frac{1}{2}\log\left(1 + \frac{\sigma_y^2}{\sigma_z^2}\right). \qquad (9.24)$$

Таким образом, среднее количество информации на выходе измерительного устройства будет уменьшаться с ростом дисперсии ошибки измерения.

9.4. Выбор систем управления и типов ЭВМ по информационным критериям эффективности

Рассмотрим примеры применения информационных критериев и пути возможного перехода к экономическим критериям. Так, проектировщику достаточно знать, что его система, имеющая большую информационную пропускную способность, будет более эффективно работать, чем эксплуатируемая система с меньшей пропускной способностью. Данный информационный критерий для него является достаточным. Экономисту этого знать мало. Ему надо знать, сколько необходимо затратить средств на систему и какой эффект получит производство в результате применения этой системы, определить срок окупаемости или сравнить полученный коэффициент капитальных вложений с нормативным и т. д.

Итак, если специалист в состоянии определить эффективность не только по информационным, техническим или технологическим критериям, но и по экономическим, то его решение будет, как правило, более обоснованным.

При проектировании отдельных элементов системы не всегда удается выходить на экономические критерии, хотя по техническим и информационным показателям эффективность можно определить. В некоторых случаях экономический критерий не представляется возможным определить. В таких ситуациях отсутствие экономической оценки не следует считать незавершенностью работы.

Пример 9.2

На химическом производстве дистанционное управление технологическим процессом производится вручную путем регулирования со щита управления температуры *t*, давления *P* и содержания примесей *q*. Разброс координат процесса подчиняется нормальному закону распределения со средними квадратическими отклонениями $\sigma_t = 24 \cdot 10^{-1}$ °C; $\sigma_p = 32 \cdot 10^{-1}$ Па; $\sigma = 16\%$. От уровня разброса всех трех параметров при регулировке процесса зависит эффективность химической реакции. При более точном управлении тремя параметрами наблюдается повышение процента выхода конечного продукта реакции. Результат химической реакции ощутим при повышении информационной пропускной способности по всем трем параметрам управ-

Таблица 9.1

Статистические характеристики процесса управления

Условия управления	$\sigma_{t_i}\cdot 10^{-1}$, °С	$\sigma_{p_i}\cdot 10^{-1}$, Па	σ_{q_i}, %
Без СУ ТП	32	32	16
СУ ТП-1	4	8	1
СУ ТП-2	2	4	2

ления процессом. Для повышения эффективности получения данного продукта реакции можно применить одну из двух систем СУ ТП. Параметры систем СУ ТП-1 и СУ ТП-2 представлены в табл. 9.1

Все параметры являются независимыми величинами и подчиняются нормальному закону распределения. Процесс регулирования при всех трех системах управления продолжается в течение 5 с (τ).

Надо выбрать систему, обеспечивающую максимальную информационную пропускную способность C канала управления при условии, что энтропия меняется равномерно в течение периода управления. Для этого необходимо определить относительную эффективность по информационной пропускной способности.

Решение

1. Рассмотрим возможный алгоритм решения. Для вычисления C_i надо знать I_i; I можно вычислить как разность энтропий при ручном управлении и при управлении с помощью СУ ТП$_j$ ($j=\overline{1,2}$).

Итак, для определения относительной эффективности по информационному критерию можно предложить следующий алгоритм:

$$\begin{tikzcd} & H_2 \arrow[r] & I_2 \arrow[r] & C_2 \arrow[dr] & \\ H_1 \arrow[ur] \arrow[dr] & & & & \Phi \\ & H_3 \arrow[r] & I_3 \arrow[r] & C_3 \arrow[ur] & \end{tikzcd}$$

В данном алгоритме индекс 1 относится к ручному управлению, индекс 2 — к управлению СУ ТП-1, а индекс 3 — к управлению СУ ТП-2.

2. Найдем энтропию при ручном управлении процессом:

$$H(t_1, P_1, q_1) = H(t_1) + H(P_1) + H(q_1) =$$
$$= \log\left(\sigma_{t_1}\sqrt{2\pi e}\right) + \log\left(\sigma_{P_1}\sqrt{2\pi e}\right) + \log\left(\sigma_{q_1}\sqrt{2\pi e}\right).$$

3. Определим энтропию при управлении СУ ТП-1:

$$H(t_2, P_2, q_2) = H(t_2) + H(P_2) + H(q_2) =$$
$$= \log\left(\sigma_{t_2}\sqrt{2\pi e}\right) + \log\left(\sigma_{P_2}\sqrt{2\pi e}\right) + \log\left(\sigma_{q_2}\sqrt{2\pi e}\right).$$

4. Найдем энтропию при управлении процессом СУ ТП-2:

$$H(t_3, P_3, q_3) = H(t_3) + H(P_3) + H(q_3) =$$
$$= \log\left(\sigma_{t_3}\sqrt{2\pi e}\right) + \log\left(\sigma_{P_3}\sqrt{2\pi e}\right) + \log\left(\sigma_{q_3}\sqrt{2\pi e}\right).$$

5. Вычислим среднее количество информации при замене ручного управления регулировкой при помощи СУ ТП-1 (I_2) и СУ ТП-2 (I_3):

$$I_2 = H(t_1, P_1, q_1) - H(t_2, P_2, q_2) = \log\left(\sigma_{t_1}\sqrt{2\pi e}\right) - \log\left(\sigma_{t_2}\sqrt{2\pi e}\right) +$$
$$+ \log\left(\sigma_{P_1}\sqrt{2\pi e}\right) - \log\left(\sigma_{P_2}\sqrt{2\pi e}\right) + \log\left(\sigma_{q_1}\sqrt{2\pi e}\right) - \log\left(\sigma_{q_2}\sqrt{2\pi e}\right) =$$
$$= \log\frac{\sigma_{t_1}}{\sigma_{t_2}} + \log\frac{\sigma_{P_1}}{\sigma_{P_2}} + \log\frac{\sigma_{q_1}}{\sigma_{q_2}} = \log\frac{32}{4} + \log\frac{32}{8} + \log\frac{32}{16} = 9 \text{ бит}.$$

$$I_3 = H(t_1, P_1, q_1) - H(t_3, P_3, q_3) = \log\frac{\sigma_{t_1}}{\sigma_{t_3}} + \log\frac{\sigma_{P_1}}{\sigma_{P_3}} + \log\frac{\sigma_{q_1}}{\sigma_{q_3}} =$$
$$= \log\frac{32}{2} + \log\frac{32}{4} + \log\frac{16}{2} = 10 \text{ бит}.$$

6. Информационную пропускную способность каналов при управлении СУ ТП-1 и СУ ТП-2 вычислим:

$$C_2 = \frac{I_2}{\tau} = 1{,}8 \text{ бит/с};$$

$$C_3 = \frac{I_3}{\tau} = 2 \text{ бит/с}.$$

7. Определим критерий отношения эффекта СУ ТП-2 к эффекту СУ ТП-1:

$$\Phi = \frac{C_3}{C_2} = \frac{2}{1{,}8} \approx 1{,}11.$$

Следовательно, по критерию «информационная пропускная способность» конструктору целесообразнее выбрать систему СУ ТП-2 с информационной пропускной способностью 2 бит/с.

Пример 9.3

Воспользуемся данными примера 9.2 и дополнительной информацией о процессе. Научно-исследовательским институтом отрасли установлено, что при повышении информационной пропускной способности от 1 бит/с до 2 бит/с на каждую десятую долю бит/с в среднем приходится 1% повышения выхода продукта. Повышение выхода продукта на 1% предприятию дает годовой прирост прибыли 400 тыс. руб. Система СУ ТП-1 стоит 100 тыс. руб., ее монтаж — 50 тыс., наладка — 25 тыс. руб., а СУ ТП-2 стоит 200 тыс. руб., ее монтаж — 40 тыс., наладка — 15 тыс. руб. Эксплуатационные расходы в течение года в каждой из систем равны 10 тыс. руб.

Рассчитаем коэффициент эффективности затрат и сроки окупаемости при внедрении СУ ТП-2 вместо СУ ТП-1, если нормативный коэффициент экономической эффективности капитальных вложений равен 0,5.

Решение

1. Определим дополнительные капитальные вложения при внедрении СУ ТП-2 вместо СУ ТП-1:

$$K_{д} = 200 + 40 + 15 - (100 + 50 + 25) = 80 \text{ тыс. руб.}$$

2. Дополнительный годовой прирост прибыли $Э_{год}$ при внедрении СУ ТП-2 составит:

$$C_3 - C_2 = 2 - 1{,}8 = 0{,}2 \text{ бит/с.}$$

Это повысит на 2% выход продукции и на 800 тыс. руб. (2 · 400) годовой прирост прибыли, т. е.

$$Э_{год} = 800 \text{ тыс. руб.}$$

3. Вычислим расчетный коэффициент эффективности затрат:

$$E_{р} = \frac{Э_{год}}{K_{д}} = \frac{800}{80} = 10 \ (E_{р} > E_{н}).$$

4. Определим срок окупаемости дополнительных затрат:

$$T = \frac{K_{д}}{Э_{год}} = 0{,}1 \text{ года.}$$

Дополнительные затраты окупятся за 1,2 месяца. Следовательно, с экономической точки зрения выбор конструктора СУ ТП-2 по информационному критерию является выгодным для предприятия.

Другой тип задач, связанный с выбором специализированной ЭВМ для управления технологическим процессом на основе информационного критерия, требует обращения к ранее описанному материалу. Обратимся к схеме, представленной на рис. 9.2. Найдем максимальное и минимальное количество информации, которое необходимо системе для управления объектом с заданной точностью. Это значит, что следует найти среднее количество информации, которое надо передать на вход управляющего устройства, чтобы обеспечить заданную точность управления объектом.

Для повышения точности измерения повторяют m раз ($i = \overline{1, m}$). Предельное значение m^* определит тот номер измерения, при котором будет обеспечена заданная точность управления.

Представим себе, что имеется дискретная система управления (рис. 9.3).

Среднее количество информации X, поступающее к выходу системы за m интервалов дискретности T, равно разности энтропии выходного сигнала $H_0(X)$ в начальный момент времени и энтропии $H_m(X)$ в момент времени mT:

$$I_x^m(X) = H_0(X) - H_m(X). \tag{9.25}$$

По этой формуле можно определить и минимальное количество информации на входе управляющего устройства при отсутствии потерь в системе. Однако в этом случае

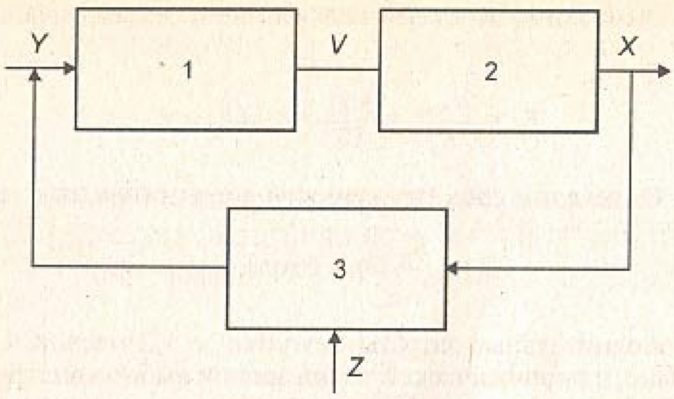


Рис. 9.3. Схема системы управления:

1 — управляющее устройство;
2 — исполнительное устройство;
3 — измерительное устройство

$$I_y^m(Y) = H_0(Y) - H_m(Y),$$

т. е. это суммарное среднее количество информации в выходном сигнале на входе управляющего устройства ($I_y^m = I_x^m$).

Максимальное количество информации на входе управляющего устройства определится по формуле (9.24) при суммировании по всем (m + 1) измерениям:

$$I_y^m(X) = \sum_{i=0}^{m} \frac{1}{2} \log \left(1 + \frac{\sigma_{x_i}^2}{\sigma_{z_i}^2}\right). \qquad (9.26)$$

Данная формула верна при следующих условиях:

- значения выходного сигнала независимы в момент измерения;
- значения выходного сигнала распределены по нормальному закону;
- помеха Z аддитивна и нормально распределена.

Отметим, что $\sigma_{x_i}^2$ — дисперсия выходного сигнала X в момент i-го измерения, $\sigma_{z_i}^2$ — дисперсия помехи Z.

Если таких параметров управления будет не один, а n, то задача решится путем суммирования по $j = \overline{1, n}$.

Для минимального количества информации

$$I_x^m(X) = \Sigma\, [H_0(X_j) - H_n(X_j)]_j, \qquad (9.27)$$

где X_j — выходной сигнал j-го параметра;
j за квадратной скобкой — индекс параметра, по которому берется разность, записанная в квадратной скобке.

Для максимального количества информации

$$I_y^m(X) = \sum_{j=1}^{n} \sum_{i=0}^{m} \left[\frac{1}{2} \log \left(1 + \frac{\sigma_{x_{ji}}^2}{\sigma_{z_j}^2}\right)\right], \qquad (9.28)$$

где $\sigma_{z_j}^2$ — дисперсия помехи Z_j j-го параметра;
$\sigma_{x_{ji}}^2$ — дисперсия выходного сигнала X_j j-го параметра в момент i-го измерения.

По формулам (9.27) и (9.28) определяются верхние и нижние границы емкости памяти специализированных ЭВМ для данного СУ ТП. Разумеется, если ЭВМ будет использовать комбинацию измерения из K первых измерений или K наибольших или наименьших измерений, то формулу надо будет преобразовать соответствующим образом. В данном случае подразумевается, что ЭВМ использует все (m + 1) измерения для каждого параметра или координаты процесса управления.

Пример 9.4

Надо выбрать тип специализированной ЭВМ дискретного действия для управления технологическим процессом на основе определения верхней и нижней границ емкости ее памяти. Каждый выходной сигнал должен измеряться четыре раза для обеспечения заданной точности. Интервал дискретности равен 1 с. Координаты параметров и ошибки их измерения предполагаются независимыми и нормально распределенными. Измеряются четыре параметра:

t — температура;
P — давление;
V — объем реагирующих веществ;
q — процент примесей.

Таблица 9.2

Характеристики средних квадратических отклонений

Координата	№ параметра	Единица измерения	Точность управления координатой, σ_i	Точность измерения координаты, σ_{Δ_i}	Средние квадратические отклонения выходных координат в момент времени, σ_{ij}*			
					0	1	2	3
t	1	0,1 °C	8	4	63,9	45,1	31,7	23,7
P	2	0,1 Па	8	4	63,9	31,8	23,7	22,3
V	3	0,1 м^3	4	2	32,0	22,5	15,9	11,1
q	4	0,1%	4	2	31,9	15,9	11,1	7,8

* σ_{ij} означает среднее квадратическое отклонение по i-му параметру, полученному в i-й момент времени (i = 1, 2, 3, 4; j = 0, 1, 2, 3).

Характеристики точности управления и измерения и средние квадратические отклонения параметров системы в четыре дискретных момента времени приведем в табл. 9.2.

Требуется определить емкость памяти ЭВМ путем вычисления минимального и максимального количества информации на входе управляющего устройства, у которого управление проводится по четырем координатам.

Решение

1. Определим минимальное количество информации о выходных координатах на входе управляющего устройства:

$$I_x^m(t, P, V, q) = H_0(t) - H_m(t) + H_0(P) - H_m(P) + H_0(V) - H_m(V) +$$

$$+ H_0(q) - H_m(q) = \log\frac{\sigma_{t_0}}{\sigma_1} + \log\frac{\sigma_{P_0}}{\sigma_2} + \log\frac{\sigma_{V_0}}{\sigma_3} + \log\frac{\sigma_{q_0}}{\sigma_4} =$$

$$= \log\frac{63{,}9}{8} + \log\frac{63{,}9}{8} + \log\frac{32}{4} + \log\frac{31{,}9}{4} \approx 12 \text{ бит.}$$

2. Максимальное количество информации, получаемое на входе управляющего устройства при $(m + 1)$-значении каждого выходного сигнала, следующее:

$$I_y^m(t, P, V, q) = \frac{1}{2}\log\left(1 + \frac{\sigma_{10}^2}{\sigma_{\Delta_1}^2}\right) + \frac{1}{2}\log\left(1 + \frac{\sigma_{11}^2}{\sigma_{\Delta_1}^2}\right) + \frac{1}{2}\log\left(1 + \frac{\sigma_{12}^2}{\sigma_{\Delta_1}^2}\right) +$$

$$+ \frac{1}{2}\log\left(1 + \frac{\sigma_{13}^2}{\sigma_{\Delta_1}^2}\right) + \frac{1}{2}\log\left(1 + \frac{\sigma_{20}^2}{\sigma_{\Delta_2}^2}\right) + \frac{1}{2}\log\left(1 + \frac{\sigma_{21}^2}{\sigma_{\Delta_2}^2}\right) +$$

$$+ \frac{1}{2}\log\left(1 + \frac{\sigma_{22}^2}{\sigma_{\Delta_2}^2}\right) + \frac{1}{2}\log\left(1 + \frac{\sigma_{23}^2}{\sigma_{\Delta_2}^2}\right) + \ldots + \frac{1}{2}\log\left(1 + \frac{\sigma_{43}^2}{\sigma_{\Delta_4}^2}\right) =$$

$$= \frac{1}{2}\log\left(1 + \frac{63{,}9^2}{4^2}\right) + \frac{1}{2}\log\left(1 + \frac{45{,}1^2}{4^2}\right) + \frac{1}{2}\log\left(1 + \frac{31{,}7^2}{4^2}\right) +$$

$$+ \frac{1}{2}\log\left(1 + \frac{23{,}7^2}{4^2}\right) + \frac{1}{2}\log\left(1 + \frac{63{,}9^2}{4^2}\right) + \frac{1}{2}\log\left(1 + \frac{31{,}8^2}{4^2}\right) +$$

$$+ \frac{1}{2}\log\left(1 + \frac{23{,}7^2}{4^2}\right) + \frac{1}{2}\log\left(1 + \frac{22{,}3^2}{4^2}\right) + \ldots + \frac{1}{2}\log\left(1 + \frac{7{,}8^2}{4^2}\right) \approx$$

$$\approx 4 + 3{,}5 + 3 + 2{,}5 + 4 + 3 + 2{,}5 + 2{,}5 + 4 + 3{,}5 +$$

$$+ 3 + 2{,}5 + 4 + 3 + 2{,}5 + 2 = 49{,}5 \text{ бита.}$$

Таким образом, установлено, что для эффективного решения задачи управления по четырем управляемым координатам процесса с заданной точностью σ^2 по каждому параметру при ошибках измерения показаний параметров, указанных в табл. 9.2, σ_{Δ_i} и независимости и нормальности распределения координат и ошибок, при использовании четырех замеров для управления следует выбрать тип ЭВМ с нижней границей емкости памяти, более близкой к 12 бит, и верхней границей, более близкой к 49,5 бита.

Рассмотренные в главе задачи решались в интересах СУ. Далее целесообразно рассмотреть класс задач, решаемых на СУ в интересах управляемых объектов. Для примера рассмотрим расчет эффективности на основе критериев теории массового обслуживания.

Глава 10

ОЦЕНКА ЭФФЕКТИВНОСТИ НА ОСНОВЕ КРИТЕРИЕВ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

10.1. Основные определения и задачи теории массового обслуживания

Теория массового обслуживания впервые была использована при эксплуатации телефонных сетей. Большую роль в развитии теории сыграл датский ученый А. К. Эрланг (1908—1922 гг.). Методы теории массового обслуживания позволили рассчитывать эффективность обслуживания абонентов в зависимости от числа используемых каналов связи. Достижения математических основ теории массового обслуживания стали достоянием ученых и практиков, работающих в разных областях производства и обслуживания.

Модели данной теории могут быть использованы весьма широко: при организации торговли, эксплуатации станочного парка предприятий, расчете пропускной способности аэропортов, подготовке технических средств СУ к работе, оценке эффективности систем управления по их пропускной способности, а также эффективности функционирования вычислительных центров и организации вычислительного процесса на современных ЭВМ и т. п.

В настоящее время теория массового обслуживания широко используется при оценке эффективности функционирования разнообразных сложных систем в отраслях народного хозяйства, объединениях и на предприятиях.

Развитию методов теории массового обслуживания способствовали работы А. Я. Хинчина, А. Н. Колмогорова, Б. В. Гнеденко, Н. П. Бусленко, И. Н. Коваленко, Б. А. Севастьянова и других ученых, которые не только расширили область приложения марковских моделей, но и внесли большой вклад в развитие немарковских моделей *системы массового обслуживания* [16], [17], [27], [107].

Работа любой *системы массового обслуживания* (СМО) состоит в удовлетворении поступающего на нее потока требований или заявок, т. е. в обслуживании. Средство, которое осуществляет это обслуживание, называется *обслуживающим устройством* (аппаратом, прибором, каналом обслуживания).

Поток — это последовательность событий, происходящих в какие-то моменты времени. *Входящий поток* — поток требований, нуждающихся в обслуживании и поступающих в обслуживающую систему. *Выходящий поток* — поток требований, покидающих обслуживающую систему. При этом в выходящем потоке могут быть как обслуженные требования, так и заявки, покидающие СМО необслуженными. Выходящий поток одной системы может стать входящим потоком другой системы массового обслуживания.

Всякая система массового обслуживания включает входящий поток, очередь на обслуживание, обслуживающие устройства, выходящий поток. В частном случае, при рассмотрении СМО с потерями, длина очереди равна нулю.

Однако ни одна характеристика критерия эффективности функционирования системы не может быть определена без знания входящего потока. Например, поток отказов элементов ЭВМ влечет за собой поток заявок на ремонт. Этот же поток отличается от потока требований на профилактическое обслуживание.

Заявки поступают в случайные моменты времени последовательно. Обслуживание осуществляется какое-то случайное время. Можно найти математическое ожидание и дисперсию поступающего потока заявок. Если же время обслуживания постоянно, то можно считать, что математическое ожидание равно времени обслуживания, а дисперсия равна нулю. На такой ситуации останавливаются редко и чаще рассматривают случайное время обслуживания.

После окончания обслуживания аппарат освобождается и готов к приему следующей заявки.

Важнейшей характеристикой СМО является ее пропускная способность. Под пропускной способностью понимают среднее число заявок, обслуживаемых системой за единицу времени. Она зависит как от характера потока заявок, так и от качества системы. Различают номинальную и фактическую пропускную способность.

Номинальную пропускную способность можно пояснить так. Пусть время обслуживания одной заявки равно τ и заявки поступают на обслуживание регулярно также с интервалом τ. При этих условиях один аппарат может обслужить за единицу времени $\frac{1}{\tau}$ заявок. Если же в системе n аппаратов, то обслужено будет $\frac{n}{\tau}$ заявок. Это отношение определяет номинальную пропускную способность.

Фактически на практике заявки поступают случайным образом, поскольку моменты поступления отдельных заявок зависят от большого числа случайных факторов и не согласованы ни между собой, ни с длительностью обслуживания. Это приводит к накоплению заявок в некоторые промежутки времени и появлению малого их числа в другие такие же промежутки времени.

При поступлении большого числа заявок могут появиться очереди заявок, а при поступлении малого числа заявок обслуживающие аппараты могут простаивать в ожидании очередной заявки. Следовательно, фактическая пропускная способность будет ниже номинальной.

Чтобы *построить модель на основе теории массового обслуживания, необходимо решить ряд основных вопросов*:

- определить, какой элемент выполняет роль прибора (канала, аппарата) обслуживания;
- рассчитать параметры входного потока заявок;
- определить параметры потока обслуживания;
- оценить возможность образования очереди и ее характеристики;
- выяснить, какие заявки пользуются при обслуживании приоритетом и характер приоритетов;
- определить структуру системы обслуживания.

Основной задачей в изучении систем массового обслуживания является оценка возможности образования очередей у каналов обслуживания и определение порядка обслуживания в системах.

Таким образом, можно считать, что теория массового обслуживания занимается построением математических моделей, связывающих условия работы систем массового обслуживания с показателями эффективности этих систем. В качестве показателей эффективности используются различные величины или функции, которые далее будут рассмотрены для систем разного типа.

Системой массового обслуживания называется совокупность однородных обслуживающих устройств, каждое из которых в данный момент времени способно обслуживать только одну из поступивших заявок.

В зависимости от порядка обслуживания различают три типа систем:

- системы с потерями;
- системы с ожиданием;
- смешанные системы.

В *системах с потерями* заявки, поступившие в момент времени, когда все каналы заняты, получают отказ и больше не влияют на ход обслуживания.

В *системах с ожиданием* заявки, поступившие в момент времени, когда все каналы заняты обслуживанием, становятся в очередь и ожидают, пока не освободится один из каналов. Как только очередной прибор освободится, одна из заявок, стоящих в очереди, немедленно принимается на обслуживание. Например, телевизор, поступивший в ателье для ремонта, когда все мастера заняты, будет поставлен в очередь и после обслуживания вернется в эксплуатацию исправным.

Смешанные системы включают свойства систем первого и второго типа при наличии каких-то промежуточных условий. Например, заявки, заставшие все каналы занятыми, могут не покидать систему и стать в очередь ограниченной длины или будут ждать ограниченное время, а по прошествии этого времени покинут систему. Например, хозяин телевизора, узнав, что в ателье его очередь на ремонт наступит не ранее чем через 15 дней, покидает это ателье и передает свою заявку на ремонт в другое ателье, где очередь подойдет раньше.

Как только заявка попала в очередь, она может быть обслужена в порядке поступления заявки по принципу «первым пришел — первым обслужен». Система с таким порядком обслуживания называется *системой без приоритетов*.

Пусть оборудование цеха поступает на ремонт в мастерскую. В первую очередь будут обслужены те агрегаты, которые наиболее существенно влияют на ход технологического процесса. При такой организации обслуживания рассматриваемая система становится системой массового обслуживания с *приоритетом*.

Приоритетов может быть несколько. Пусть в системе имеются три типа заявок: с высшим приоритетом, с приоритетом и без приоритета. В первую очередь будут обслужены заявки с высшим приоритетом. Заявки с приоритетом пойдут на обслуживание только тогда, когда в очереди не будет ни одной заявки с высшим приоритетом. Заявки без приоритета поступят на обслуживание, если в очереди не будет ни одной приоритетной заявки.

Можно приоритетное обслуживание рассматривать как обслуживание нескольких потоков требований, имеющих определенные преимущества. Пусть высший номер потока обладает бо́льшим преимуществом. В потоке 1 — десять заявок, в потоке 2 — четыре заявки и в потоке 3 — две заявки. Вначале будут обслужены заявки потока 3. Из этого потока они будут поступать в порядке «первым пришел — первым обслужен». Когда все заявки потока 3 будут обслужены, начнут поступать заявки потока 2 в порядке очередности поступления. Затем после обслуживания всех заявок из потока 2 в порядке очередности начнут поступать заявки из потока 1.

Порядок поступления заявок может быть построен таким образом, что на обслуживание в первую очередь поступают заявки, появившиеся последними. Такой принцип называется «последним пришел — первым обслужен». Итак, высший приоритет приобретает заявка, меньше других ожидавшая в очереди. Это наблюдается при обработке быстро стареющей экономической информации на ЭВМ.

Приоритет может быть *абсолютным*. В этом случае при появлении заявки с более высоким приоритетом прекращается обслуживание заявки с более низким приоритетом или без приоритета независимо от того, какова степень законченности обслуживания вытесненной заявки.

Приоритет может быть *относительным*. В этом случае при поступлении заявки с более высоким приоритетом начатое обслуживание продолжается до конца, а поступившая заявка будет обслуживаться после окончания обслуживания очередной заявки.

Различают системы с *многофазовым* обслуживанием. Например, последовательно начинается обслуживание (изготовление) измерительного прибора: в специальном цехе изготавливается корпус — первая фаза обслуживания, затем прибор собирается в монтажном цехе — вторая фаза, наконец, он поступает в лабораторию контроля и регулировки — третья фаза. После этого прибор готов к упаковке и отправке. В такой системе выходной поток обслуженных требований первой фазы является входным потоком второй фазы обслуживания и т. д.

Наблюдаются системы массового обслуживания открытые и замкнутые. В *открытых* системах характеристики потока заявок не зависят от того, сколько каналов СМО занято, в *замкнутой* системе — зависят. В открытой системе количество заявок, поступающих в среднем в единицу времени в систему, не зависит от самой системы. Поток заявок формируется вне системы. Например, можно считать, что поток заявок на обслуживание (ремонт) телевизоров в Москве не зависит от того, сколько телевизоров находится в ателье. Другим примером может служить поток заявок на наладку станков в цехе. Все станки за достаточно короткий промежуток времени проходят через обслуживающие устройства — бригаду наладчиков. Таким образом, входящий поток требований формируется из выходящего потока. Это характерный пример замкнутой системы массового обслуживания, когда общее количество источников заявок ограничено и интенсивность входящего потока зависит от того, сколько станков находится в системе массового обслуживания.

Кроме рассмотренной классификации систем массового обслуживания, различают системы упорядоченные, неупорядоченные, с поступлением групповых заявок, с приборами разной производительности, с накопителем требований и др.

В *неупорядоченной* системе любую заявку может обслуживать любой из освободившихся аппаратов. Представим себе СМО, где все аппараты пронумерованы. Данный свободный аппарат может обслуживать поступившее требование только тогда, когда все приборы с меньшим номером заняты обслуживанием. Такая система называется *упорядоченной*. Иногда упорядочивают не отдельные аппараты, а группы аппаратов, но суть не меняется. В этом случае упорядочение обобщено на группы приборов.

В системах с *поступлением групповых заявок* заявки поступают группами. При этом число требований в каждой группе может быть постоянным или случайным.

В СМО с *приборами разной производительности* разные приборы обрабатывают однотипные заявки с разной производительностью. Например, за двумя одинаковыми станками стоят разные рабочие. Один из них имеет большой опыт работы, а второй рабочий работает с меньшей производительностью. Если один рабочий закончил деталь, то он берется за изготовление следующей детали, не обращая внимания на то, сколько деталей изготовил его сосед. Значит, детали обслуживаются независимо от того, кто из рабочих освободился — тот, кто обслуживает с меньшей производительностью, или тот, кто работает с большей производительностью.

В системах с *накопителем требований* требования накапливаются в нем до окончания обслуживания предыдущей партии требований. Как только ранее накопившаяся партия будет обслужена, находящиеся в накопителе требования поступят на обслуживание.

Каждый из типов систем массового обслуживания имеет специфические критерии оценки эффективности, кроме общих показателей, присущих любой СМО. В настоящее время для ВЦ и ВЦ СУ характерны потоки с требованиями, включающими повторные задания. Так появился новый класс СМО.

10.2. Поток требований на обслуживание

Простейший поток и его свойства

Рассмотрим потоки требований (событий), обладающие некоторыми простыми свойствами.

• **Стационарный поток.** Поток событий считается стационарным, если вероятность попадания того или иного числа событий на участок времени длиной τ (рис. 10.1) зависит только от длины участка и не зависит от того, где именно на оси $0t$ расположен этот участок. На рис. 10.1 последовательность точек $t_1, t_2, ..., t_n$ на числовой оси соответствует моментам появления событий. Примером может служить поток отказов в определенный ограниченный период эксплуатации технических средств СУ или поток вызовов на автоматиче-

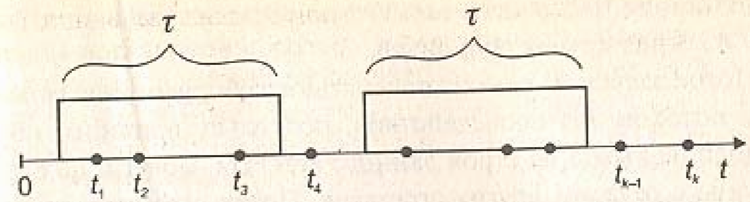


Рис. 10.1. Стационарный поток

скую телефонную станцию в определенное ограниченное время суток.

• **Поток без последействия.** Поток событий является потоком без последействия, если для любых неперекрывающихся отрезков времени число событий, попадающих на один из них, не зависит от числа событий, попадающих на другие отрезки (рис. 10.2).

В таком потоке число событий, наступающих за время τ_i после момента T, не зависит от того, сколько заявок было до наступления момента T, т. е. условная вероятность поступления k заявок за интервал времени $(T, T + \tau_i)$, равна безусловной вероятности поступления требований за этот интервал.

Примером может служить детская погремушка, в которой находится один шарик. В погремушке имеется одно отверстие с диаметром, несколько большим, чем у шарика. Вероятность выпадания шарика будет возрастать с увеличением числа встряхиваний погремушки. Однако если мы погремушку тысячу раз встряхивали и при этом шарик не выпал, а затем после некоторого перерыва ее снова начнем встряхивать, то при этой новой серии встряхиваний вероятность появления шарика не зависит от того, сколько раз погремушку встряхивали ранее.

Наблюдается ли такая идеальная картина при эксплуатации аппаратуры? Лишь иногда, в большинстве же случаев не

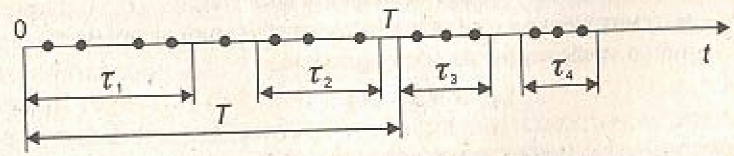


Рис. 10.2. Поток без последействия

наблюдается. Часто один элемент при отказе приводит к выходу из строя других элементов.

Поток заявок на ремонт однотипного оборудования может быть потоком без последействия, поскольку причины, обусловившие выход из строя данного агрегата, могут и не быть связаны с отказом других агрегатов. Поток агрегатов, покидающих ремонтный орган, не может считаться потоком без последействия, так как моменты окончания ремонта нескольких агрегатов могут зависеть от работы одной бригады.

• **Ординарный поток.** Поток событий является ординарным, если вероятность попадания на элементарный участок Δt двух или более событий пренебрежимо мала по сравнению с вероятностью попадания одного события. Это условие можно записать в математических символах так:

$$\lim_{\Delta t \to 0} \frac{P_{>1}(\Delta t)}{\Delta t} = 0, \qquad (10.1)$$

где $P_{>1}(\Delta t)$ — вероятность появления в интервале времени $(t, t + \Delta t)$ более одной заявки.

Запись (10.1) обозначает, что вероятность появления более одной заявки является величиной высшего порядка малости по сравнению с длиной интервала.

• **Простейший поток.** Такой поток одновременно обладает свойствами стационарности, ординарности и отсутствия последействия. Простейший поток подчиняется закону распределения Пуассона с постоянным параметром распределения

$$M_\tau[k] = a = \lambda\tau, \qquad (10.2)$$

где λ — плотность потока требований, т. е. это математическое ожидание числа требований за единицу времени;

a — математическое ожидание числа требований за время τ;

k — число требований на обслуживание.

$$M_\tau = 1[k] = \lambda. \qquad (10.3)$$

Простейший поток часто называют *стационарным пуассоновским потоком.*

Интенсивность, или плотность, потока требований есть величина, обратная среднему времени между соседними требованиями в простейшем потоке:

$$\lambda = \frac{1}{t_{cp}}. \qquad (10.4)$$

Вероятность того, что за время τ произойдет ровно k событий, по закону Пуассона равна:

$$P_k(\tau) = \frac{a^k}{k!} e^{-a}. \qquad (10.5)$$

Определим вероятность непоявления требования $P_0(\Delta t)$ за бесконечно малый промежуток времени Δt, затем вероятность появления одного требования, двух требований и т. д. (рис. 10.3).

По формуле (10.5)получим:

$$P_0(\Delta t) = \frac{(\lambda\Delta t)^0}{0!} e^{-\lambda\Delta t} = e^{-\lambda\Delta t}, \qquad (10.6)$$

где $P_0(\Delta t)$ — вероятность непоявления требований за время Δt.

Разлагая функцию $e^{-\lambda\Delta t}$ в ряд, получаем:

$$P_0(\Delta t) = 1 - \frac{\lambda\Delta t}{1!} + \frac{(\lambda\Delta t)^2}{2!} - \frac{(\lambda\Delta t)^3}{3!} \pm ... = 1 - \lambda\Delta t + 0(\Delta t),$$

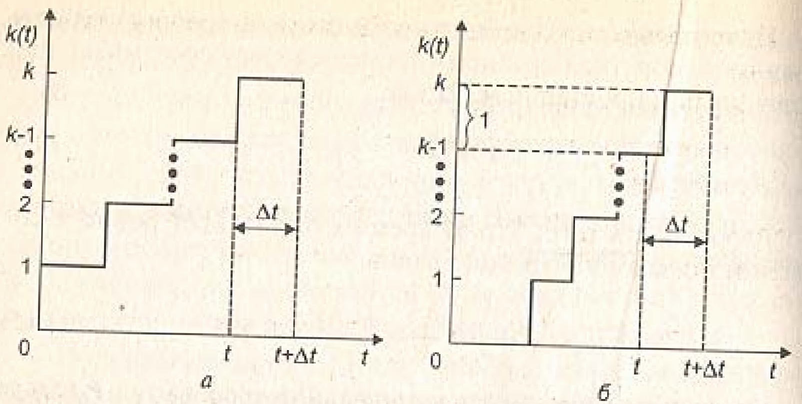
где через $0(\Delta t)$ обозначена сумма всех членов более высокого порядка малости, чем Δt, называемая остатком.

Эта сумма является знакочередующимся рядом, сумма которого по абсолютной величине не превосходит значения первого отброшенного $\frac{(\Delta t\lambda)^2}{2!}$, а этот член по своей величине будет порядка $(\Delta t)^2$.

Пренебрегая величинами более высокого порядка малости $0(\Delta t)$, можно записать:

$$P_0(\Delta t) = 1 - \lambda\Delta t. \qquad (10.7)$$

Так как простейший поток обладает свойством ординарности, то вероятность появления двух и более требований за время Δt бесконечно мала по сравнению с вероятностью по-



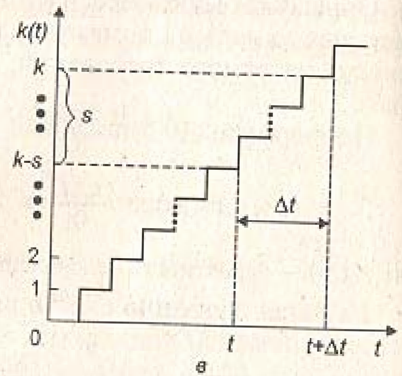


Рис. 10.3. К определению вероятности появления: *а*) 0; *б*) 1; *в*) $s > 1$ требований за время Δt:
$а - P_0(\Delta t) = 1 - \lambda \cdot \Delta t + 0(\Delta t)$;
$б - P_1(\Delta t) = \lambda \cdot \Delta t + 0(\Delta t)$;
$в - P \geqslant s\,(\Delta t) = 0(\Delta t),\ s \geqslant 2$

явления одного требования, поэтому с точностью до бесконечно малых второго порядка запишем:

$$P_0(\Delta t) + P_1(\Delta t) \approx 1.$$

Следовательно, вероятность появления одного требования за время Δt:

$$P_1(\Delta t) \approx \lambda \Delta t. \tag{10.8}$$

Пример 10.1

Решим задачу из практики хранения стабилизаторов напряжения для ЭВМ. Интенсивность отказов, или плотность потока требований, в течение периода хранения $\lambda = 10^{-4}$ 1/сутки.

Определим, какова вероятность того, что за время $\Delta t = 5$ месяцам:

1) не появится ни одного требования на замену стабилизатора;

2) появится одно требование;

3) появятся два требования.

Решение

1. Найдем вероятность того, что ни один стабилизатор за 5 месяцев не откажет:

$$P_0(150) = \frac{(10^{-4} \cdot 150)^0}{0!} e^{-0{,}015} \approx 0{,}985.$$

2. Определим вероятность появления одного требования за 5 месяцев:

$$P_1(150) = \frac{(10^{-4} \cdot 150)^1}{0!} e^{-0{,}015} \approx 0{,}015.$$

3. Вычислим вероятность появления двух требований за 150 суток:

$$P_2(150) = \frac{(10^{-4} \cdot 150)^2}{0!} e^{-0{,}015} \approx 0{,}00011.$$

Следовательно, в практике эксплуатации в течение 5 месяцев вероятность того, что один стабилизатор необходимо будет заменить на исправный, в 136 раз меньше, чем вероятность появления двух отказов. При расчетах обычно такими малыми величинами вероятности появления более одного события за время Δt пренебрегают.

Следует отметить одно важное свойство простейшего потока: закон распределения промежутка времени T между соседними событиями является экспоненциальным с постоянным параметром λ:

$$P\{T \geqslant t\} = P_0(t). \tag{10.9}$$

Здесь $P\{T \geqslant t\}$ — вероятность того, что на участке времени длиной t, начинающемся в момент появления одного события потока, не появится ни одно другое событие. Это случай появления нуля событий на участке длиной t, а вероятность этого случая равна вероятности непоявления событий:

$$P_0(t) = e^{-\lambda t};$$

$$\frac{1}{\lambda} = m_T = \sigma_T; \; v_T = 1, \tag{10.10}$$

где m_T — математическое ожидание случайной величины T;
σ_T — среднее квадратическое отклонение;
v_T — коэффициент вариации.

• **Нестационарный пуассоновский поток.** Это поток однородных событий, ординарный и без последействия, но не стационарный, с переменной плотностью $\lambda(t)$. *Мгновенная плотность* потока характеризуется пределом отношения среднего числа событий, приходящегося на элементарный участок времени $(t;\ t + \Delta t)$, к длине этого участка Δt, когда последняя стремится к нулю:

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{m(t+\Delta t) - m(t)}{\Delta t} = m'(t), \tag{10.11}$$

где $m(t)$ — математическое ожидание числа событий на участке $(0;\ t)$;
$m(t + \Delta t)$ — математическое ожидание числа событий на участке $(0, t + \Delta t)$.

Для такого потока число событий, попадающих на участок длины τ, начинающийся в точке t_0, подчиняется закону Пуассона:

$$P_k(\tau,\ t_0) = \frac{a^k}{k!} e^{-a} \; (k = 0, 1, 2, ...), \tag{10.12}$$

где a — математическое ожидание числа событий на участке от t_0 до $t_0 + \tau$:

$$a = \int_{t_0}^{t_0+\tau} \lambda(t)\,dt. \tag{10.13}$$

Величина a зависит не только от длины участка τ, но и от его положения на оси $0t$.

Несмотря на то, что структура нестационарного пуассоновского потока несколько сложнее, чем простейшего, он удобен в практике, так как главное свойство простейшего потока — отсутствие последействия — в нем сохранено. Следовательно, если зафиксируем на оси $0t$ произвольную точку t_0, закон распределения времени T, отделяющего эту точку t_0 от

ближайшего по времени будущего события, не зависит от того, появлялись ли ранее другие события.

Простейший поток является частным случаем нестационарного пуассоновского потока.

Простейший поток в практике эксплуатации

Предположение о том, что входящий поток требований является простейшим, значительно облегчает математические выкладки. В то же время на основании накопленного практического опыта можно заключить, что реальный поток требований не отвечает основным предпосылкам простейшего потока.

Например, предположение, что поток телефонных вызовов на станцию считается простейшим, может быть нарушено в течение одних суток. Так, поток вызовов меняется в течение суток: днем количество звонков в единицу времени больше, а ночью — меньше. Если в данные сутки от некоторого министерства поступит срочное распоряжение подчиненным организациям, то этот единственный звонок может вызвать большую серию телефонных звонков, являющуюся последействием, зависящим от ранее поступившего звонка [21].

Не совсем строго выполняется требование ординарности. Следовательно, может показаться, что выводы, полученные на основании предположения о простейшем потоке, должны значительно отличаться от практических результатов. Однако глубокое исследование практических результатов, проведенное А. Я. Хинчиным, показало, что в действительности дело обстоит как раз наоборот: опытные данные согласуются с выводами построенной теории, как правило, лучше, чем это можно было бы ожидать по принципиальным соображениям [107]. Этим объясняется то, что простейший поток среди потоков играет особую роль.

При суммировании (взаимном наложении) большого числа ординарных, стационарных потоков с практически любым последействием получается поток, сколь угодно близкий к простейшему. Только при этом должно соблюдаться условие, чтобы складываемые потоки на суммарный поток оказывали приблизительно равноценное влияние. На практике оказывается достаточным сложить 4—5 потоков, чтобы получить поток, с которым можно оперировать как с простейшим потоком.

Если имеется сложная система, состоящая из большого числа элементов, каждый из которых может отказать с малой вероятностью за единицу времени независимо от состояния других элементов, то число элементов сложной системы, отказавших за промежуток времени $(0, t)$, представляет собой случайный процесс, который во многих случаях хорошо описывается стационарным пуассоновским процессом.

Возможно, что в систему обслуживания будет поступать нестационарный, неординарный поток с последействием. В этом случае систему обслуживания целесообразно рассчитывать при входящем простейшем потоке с интенсивностью (плотностью потока), соответствующей максимально возможной плотности реального потока. Если систему массового обслуживания рассчитать таким образом, то она будет эффективно справляться с обслуживанием реального потока.

Эффективность системы при обслуживании реального потока, отличающегося от стационарного пуассоновского потока, будет выше расчетной эффективности для принятого простейшего потока.

Если не следовать изложенным советам, то окажется, что подавляющее большинство задач теории массового обслуживания становится очень сложным в том случае, когда поток отличается от стационарного пуассоновского потока. Несмотря на эти трудности, многие ученые вынуждены заниматься и непуассоновскими потоками, идти на усложнение математического аппарата, когда это диктуется практикой.

Остановимся на практическом примере расчета параметра простейшего потока требований.

Пример 10.2

В цехе имеется склад инструментов. В случайные моменты времени рабочие приходят за инструментом. Наша задача заключается в том, чтобы определить параметр потока рабочих, следующих на склад, и характер потока. Для решения этой задачи будем фиксировать количество рабочих, приходящих за каждые 15 мин. Таких наблюдений организуем 100. Результаты запишем в таблицу (см. табл. 10.1, гр. 1, 2).

Решение

1. Определим среднее число обращений на склад за 15 мин. и за 1 мин.:

Таблица 10.1

Исследование потока обращений за инструментом в цехе

Число обращений за инструментом за 15-минутный интервал, k_i	Число интервалов с одинаковым числом обращений, n_i	Вероятность числа обращений в интервале по закону Пуассона, P_i	Математическое ожидание числа интервалов с заданным числом обращений, m_i	Число обращений за инструментом за 15-минутный интервал, k_i	Число интервалов с одинаковым числом обращений, n_i	Вероятность числа обращений в интервале по закону Пуассона, P_i	Математическое ожидание числа интервалов с заданным числом обращений, m_i
1	2	3	4	1	2	3	4
5	0	0,001	0,1	16	13	0,099	9,9
6	1	0,002	0,2	17	9	0,093	9,3
7	1	0,006	0,6	18	8	0,083	8,3
8	1	0,012	1,2	19	7	0,069	6,9
9	2	0,021	2,1	20	6	0,055	5,5
10	2	0,034	3,4	21	5	0,042	4,2
11	4	0,049	4,9	22	3	0,031	3,1
12	5	0,066	6,6	23	2	0,021	2,1
13	7	0,081	8,1	24	1	0,014	1,4
14	10	0,093	9,3	25	1	0,009	0,9
15	12	0,099	9,9				

$$a = \sum_{i=1}^{21} k_i \cdot \frac{n_i}{100} = 6 \cdot 0{,}01 + 7 \cdot 0{,}01 + 8 \cdot 0{,}01 + 9 \cdot 0{,}02 + $$
$$+ 10 \cdot 0{,}02 + 11 \cdot 0{,}04 + 12 \cdot 0{,}05 + 13 \cdot 0{,}07 + 14 \cdot 0{,}10 +$$
$$+ 15 \cdot 0{,}12 + 16 \cdot 0{,}13 + 17 \cdot 0{,}09 + 18 \cdot 0{,}08 + 19 \cdot 0{,}07 +$$
$$+ 20 \cdot 0{,}06 + 21 \cdot 0{,}05 + 22 \cdot 0{,}03 + 23 \cdot 0{,}02 + 24 \cdot 0{,}01 +$$
$$+ 25 \cdot 0{,}01 = 15{,}98 \approx 16;$$

$$a = \lambda t;\ \lambda = \frac{a}{t} = \frac{15{,}98}{15} \approx 1{,}07\ \frac{\text{обр.}}{\text{мин.}}.$$

2. Проверим, является ли поток заявок на инструмент простейшим. Для этого подсчитаем значения вероятностей числа обращений в кладовую цеха при законе Пуассона с параметром $a = 16$ (значения вероятностей определяются по таблице распределения Пуассона при $a = 16$) и математическое ожидание числа интервалов с заданным числом обращений в кладовую (см. табл. 10.1, гр. 3 и 4) (например, вероятность 9 обращений в кладовую за 15 мин. равняется: $P_9 = 0{,}021$, а $m_9 = 0{,}021 \cdot 100 = 2{,}1$).

Определим значение χ^2:

$$\chi^2 = \sum_{i=1}^{21} \frac{(n_i - m_i)^2}{m_i};$$

$$\chi^2 = \frac{(0-0{,}1)^2}{0{,}1} + \frac{(1-0{,}2)^2}{0{,}2} + \frac{(1-0{,}6)^2}{0{,}6} + \frac{(1-1{,}2)^2}{1{,}2} + \frac{(2-2{,}1)^2}{2{,}1} +$$
$$+ \frac{(2-3{,}4)^2}{3{,}4} + \frac{(4-4{,}9)^2}{4{,}9} + \frac{(5-6{,}6)^2}{6{,}6} + \frac{(7-8{,}1)^2}{8{,}1} + ... + \frac{(2-2{,}1)^2}{2{,}1} +$$
$$+ \frac{(1-1{,}4)^2}{1{,}4} + \frac{(1-0{,}9)^2}{0{,}9} = 0{,}1 + 3{,}2 + 0{,}27 + 0{,}03 + 0{,}005 +$$
$$+ 0{,}58 + 0{,}16 + 0{,}39 + 0{,}15 + ... + 0{,}005 + 0{,}11 + 0{,}01 \approx 6{,}7.$$

Определим вероятность того, что полученное распределение подчинено пуассоновскому распределению.

Число степеней свободы равно числу интервалов минус 2, т. е. 19. По таблице χ^2 распределения определяем, что при числе степеней свободы, равном 19, и значении $\chi^2 = 6{,}7$ вероятность того, что полученные статистические данные о работе инструментальной кладовой цеха совпадают с пуассоновским потоком, $P \approx 0{,}99$.

Таким образом, экспериментально определено, что закон распределения потока требований на инструмент является пуассоновским, а его параметр $a = 16$, или $\lambda = 1{,}07 \ \frac{1}{\text{мин.}}$.

10.3. Время обслуживания

Время обслуживания показывает, сколько времени тратится на обслуживание одного требования одним аппаратом. Оно характеризует и пропускную способность обслуживающего аппарата. Считают, что если обслуживание требования в системе закончено, то заявка является полностью удовлетворенной. Время обслуживания может быть детерминированной величиной и случайной. В общем случае время обслуживания рассматривается как случайная величина, подчиненная некоторому закону распределения. Если время обслуживания $T_{обс}$ подчинено показательному закону распределения, то все расчеты упрощаются.

Показательный закон распределения времени обслуживания

Если обозначить $t_{обс}$ — среднее время обслуживания одним прибором одного требования, а μ — интенсивность обслуживания, то

$$\mu = \frac{1}{t_{обс}}; \tag{10.14}$$

$$t_{обс} = \mathrm{M}[T_{обс}], \tag{10.15}$$

где $T_{обс}$ — время обслуживания отдельного требования;
$\mathrm{M}[T_{обс}]$ — математическое ожидание времени обслуживания.

Если предположим, что $T_{обс}$ распределяется по показательному закону, то вероятность того, что требования будут обслужены за время $t > T_{обс}$, выразится функцией

$$P_{обс}(t) = P(T_{обс} < t) = 1 - e^{-\mu t}. \tag{10.16}$$

У показательного закона есть одно важное свойство: вероятность завершения обслуживания за время t не зависит от того, сколько времени обслуживание уже продолжалось до этого. Для каждого прибора обслуживания в текущий момент времени t это свойство позволяет учитывать только, занят прибор или свободен, и не учитывать, сколько времени уже продолжалась его занятость.

Из рис. 10.4 видно, что с увеличением μ вероятность окончания обслуживания поступивших заявок растет. Следовательно, при организации обслуживания по технологическим процессам следует отдавать предпочтение тому прибору, у которого при всех прочих равных условиях μ больше.

Отметим, что вероятность того, что требование будет обслужено за бесконечно малый промежуток времени Δt, будет равна:

$$P_{обс}(\Delta t) = 1 - e^{-\mu t}.$$

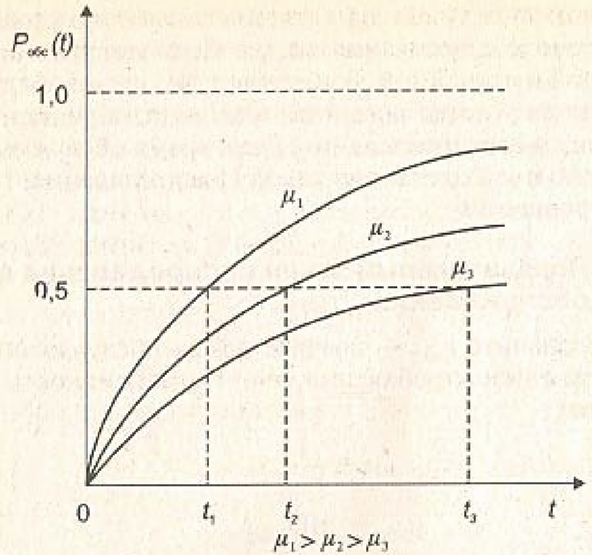


Рис. 10.4. Вероятность окончания обслуживания

Разлагая функцию $e^{-\mu\Delta t}$ в ряд, получим:

$$P_{обс}(\Delta t) = 1 - [1 - \frac{\mu\Delta t}{1!} + \frac{(\mu\Delta t)^2}{2!} - \frac{(\mu\Delta t)^3}{3!} \pm ...] = \mu\Delta t + 0(\Delta t).$$

Следовательно, с точностью до бесконечно малых второго порядка можно записать:

$$P_{обс}(\Delta t) \approx \mu\Delta t. \tag{10.17}$$

Предположение о показательном законе распределения времени обслуживания заявки на первый взгляд кажется искусственным, но удобным для аналитических расчетов. При показательном законе распределения предполагается, что большая часть требований обслуживается быстро. Однако на практике это часто бывает не так.

Плотности распределения времени обслуживания определяются по *формуле Эрланга*:

$$f_k(t) = \frac{(\mu k)^k t^{k-1}}{(k-1)!} e^{-\mu k t}, \tag{10.18}$$

где k — число обслуженных заявок;

μ — интенсивность потока обслуживания.

В формуле Эрланга $f_k(t)$ рассматривается как плотность распределения суммы k независимых случайных величин с показательным законом распределения.

В работе [27] показано, что если в СМО поступило одно требование и его одновременно начинают обслуживать n однотипных аппаратов, время обслуживания каждого аппарата подчинено показательному закону с параметром μ, то общая интенсивность обслуживания будет равна $n\mu$. Значит, среднее время обслуживания уменьшится в n раз, а дисперсия — в n^2 раз.

Общий закон обслуживания также показательный. Если же все n приборов имеют разную производительность, но для каждого прибора время обслуживания подчинено показательному закону, то закон обслуживания всеми приборами будет показательным с общим параметром $\mu = \sum_{i=1}^{n} \mu_i$:

$$P(T_{обс} < t) = 1 - e^{-(\mu_1 + \mu_2 + ... + \mu_n)t}, \tag{10.19}$$

где μ_i — интенсивность потока обслуживания i-м аппаратом.

Рассмотрим процесс обслуживания при простейшем входящем потоке и показательном законе распределения времени обслуживания как *случайный процесс Маркова*. Процесс обслуживания относится к классу случайных процессов. Он полностью определяется:

- моментами окончания операций по обслуживанию, выполняемых в момент t_0;
- моментами поступления новых заявок;
- длительностью обслуживания заявок, поступивших позднее момента t_0.

Предположим, что:

1) время выполнения операций распределено по показательному закону (время окончания операций по обслуживанию заявок не зависит от того, сколько времени эти операции выполнялись до момента t_0, т. е. предыстория процесса нас не интересует);

2) поток заявок простейший (требования, поступившие до момента t_0, не влияют на число и порядок поступления новых заявок благодаря наличию условий стационарности и отсутствию последействия).

Из двух условий следует, что весь процесс обслуживания заявок, поступивших после момента t_0, не зависит от того, как шло обслуживание до момента t_0. В теории случайных процессов процессы без последействия, для которых будущее развитие зависит только от достигнутого в данный момент t_0 состояния и не зависит от того, как происходило развитие процесса до этого момента времени, называются марковскими процессами.

На основании изложенного можно сделать заключение: процесс обслуживания в случае простейшего входящего потока и показательного закона распределения времени выполнения операций по обслуживанию может рассматриваться как процесс Маркова.

В исследованиях процессов Маркова важную роль играют переходные вероятности. *Вероятностью перехода* (переходной вероятностью) называется условная вероятность $P_{ik}(t)$ того, что через время t_1 будет занято k аппаратов, если вначале (в момент времени t_0) было занято i аппаратов.

Обозначим $P_i(t_0)$ — вероятность того, что в момент времени t_0 занято i аппаратов. Если $i = 0, 1, 2, ..., n$, то вероятность того, что в момент времени $(t_0 + t_1)$ в системе занято k аппаратов, определится из формулы полных вероятностей:

$$P_k(t_0 + t_1) = \sum_{i=0}^{n} P_i(t_0) \cdot P_{ik}(t_1). \qquad (10.20)$$

Это выражение может быть сформулировано следующим образом: если в момент времени t_0 в системе занято i аппаратов с вероятностью $P_i(t_0)$ $(i = 0, 1, 2, ..., n)$, а условная вероятность перехода к k занятым аппаратам через промежуток времени t_1 равна $P_{ik}(t_1)$ $(i = 0, 1, 2, 3, ..., n)$, то полная вероятность того, что в момент времени $(t_0 + t_1)$ занято k аппаратов, равна сумме произведений $P_i(t_0) \cdot P_{ik}(t_1)$ по всем возможным i $(i = 0, 1, 2, 3, ..., n)$.

10.4. Критерии эффективности

Эффективность функционирования систем массового обслуживания определяется при помощи большого числа различных критериев. Рассмотрим некоторые из них.

1. Вероятность потери требования в системе массового обслуживания. Эта вероятность равна вероятности того, что в

системе с потерями все аппараты заняты обслуживанием. Ее обозначают либо P_n, либо $P_{отк}$.

2. Вероятность того, что обслуживанием в системе занято ровно k аппаратов из общего числа n аппаратов. Эту величину обозначают P_k, где $k = 0, 1, 2, ..., n$.

3. Среднее число занятых аппаратов или каналов обслуживания:

$$N_З = \sum_{k=1}^{n} kP_k. \qquad (10.21)$$

4. Коэффициент занятости обслуживающих аппаратов:

$$K_З = \frac{N_З}{n}. \qquad (10.22)$$

Эта величина может характеризовать среднюю долю времени, которую аппарат занят обслуживанием.

5. Среднее число аппаратов, свободных от обслуживания:

$$N_0 = \sum_{k=0}^{n-1} (n-k) P_k. \qquad (10.23)$$

6. Коэффициент простоя аппарата:

$$K_П = \frac{N_0}{n}. \qquad (10.24)$$

Коэффициент характеризует среднюю долю времени, которую обслуживающий аппарат простаивает в ожидании очередной заявки на обслуживание. Разумеется,

$$N_0 + N_З = n; \qquad (10.25)$$

$$K_З + K_П = 1. \qquad (10.26)$$

7. Вероятность отказа в обслуживании заявки P_n (см. критерий 1).

8. Относительная пропускная способность:

$$П_0 = 1 - P_n. \qquad (10.27)$$

9. Абсолютная пропускная способность:

$$П = \lambda П_0. \qquad (10.28)$$

10. Закон распределения. Системы с ожиданием характеризуются законом распределения времени ожидания в очереди до начала обслуживания. Чаще вместо этой наиболее полной характеристики времени ожидания определяют среднее время ожидания в очереди до начала обслуживания.

11. Среднее время ожидания требований в очереди до начала обслуживания, если в системе имеется n аппаратов:

$$t_{\text{ож}} = \mathrm{M}[T_{\text{ож}}] = \int_0^\infty t\,\mathrm{d}P_1\ (T_{\text{ож}} > t). \tag{10.29}$$

12. Вероятность того, что время пребывания в очереди продлится больше определенной величины:

$$P_1(T_{\text{ож}} > t) = \sum_{k=0}^{n} P_k P_k\ (T_{\text{ож}} > t), \tag{10.30}$$

где $P_k(T_{\text{ож}} > t)$ — условная вероятность того, что время ожидания $T_{\text{ож}} > t$ при условии, что в момент поступления требования в систему в ней уже обслуживалось k требований.

Вероятность того, что время пребывания требования в очереди продлится меньше определенной величины:

$$P_2(T_{\text{ож}} < t) = \sum_{k=n}^{\infty} P_k P_k\ (T_{\text{ож}} < t), \tag{10.31}$$

где $P_k(T_{\text{ож}} < t)$ — условная вероятность того, что время ожидания $T_{\text{ож}} < t$ при условии, что в момент поступления требования в систему в ней уже обслуживалось k требований.

13. Средняя длина очереди:

$$M_{\text{ож}} = \sum_{k=n+1}^{\infty} (k-n)\,P_k, \text{ при } k \geqslant n. \tag{10.32}$$

14. Среднее число заявок, находящихся в системе обслуживания:

$$M = \sum_{k=1}^{\infty} k\,P_k. \tag{10.33}$$

Разумеется, из формул (10.21), (10.32) и (10.33) следует, что

$$M = M_{\text{ож}} + N_3. \tag{10.34}$$

15. Вероятность того, что число требований в очереди, ожидающих начала обслуживания, больше m:

$$P > m = \sum_{k=m+1}^{\infty} P_k. \tag{10.35}$$

В перечисленных критериях не учитывается экономический фактор, который часто играет важную роль при выборе оптимальных параметров систем массового обслуживания. Поэтому целесообразно обратиться к следующим критериям, учитывающим экономические показатели [52], [70].

16. Функции стоимости потерь в системе.

А. Функция для системы с отказами

$$G_{\Pi} = (q_k N_3 + q_{\text{у}} P_n \lambda + q_{nk} N_0)\,T, \tag{10.36}$$

где T — рассматриваемый интервал времени;
q_k — стоимость эксплуатации каждого (одного) аппарата системы в единицу времени;
$q_{\text{у}}$ — стоимость убытков в результате ухода (потери) требований из системы.

Б. Функция для системы с ожиданием

$$G_{\Pi} = (q_{\text{ож}} M_{\text{ож}} + q_{nk} N_0 + q_k N_3)\,T, \tag{10.37}$$

где $q_{\text{ож}}$ — стоимость потерь, связанных с простоем требований в очереди в единицу времени;
q_{nk} — стоимость единицы времени простоя прибора системы.

В. Функция для смешанной системы

$$G_{\Pi} = (q_{nk} N_0 + q_{\text{ож}} M_{\text{ож}} + q_{\text{у}} P_n \lambda + q_k N_3)\,T. \tag{10.38}$$

17. Критерий экономической эффективности системы массового обслуживания

$$E = P_{\text{обс}} \lambda C T - G_{\Pi}, \tag{10.39}$$

где C — экономический эффект, полученный при обслуживании каждого требования;
$P_{\text{обс}}$ — вероятность обслуживания требования.

Воспользуемся рассмотренными показателями эффективности для отдельных видов СМО.

10.5. Система обслуживания с потерями

Задачей является получение критериев эффективности системы: вероятности отказа в обслуживании, относительной пропускной способности системы, абсолютной пропускной способности системы, частных и производных критериев эффективности системы.

Рассмотрим *условия обслуживания в системе:*

- каждый канал может обслуживать только одно требование;
- время обслуживания подчиняется экспоненциальному закону распределения со средним временем $t_{\text{обс}} = \frac{1}{\mu}$ (μ — среднее число требований, обслуженных в единицу времени);
- на вход системы поступает простейший поток требований;
- требование принимается на обслуживание немедленно любым из свободных аппаратов;
- требование теряется, если все каналы заняты в момент его появления в системе.

Система может находиться в следующих состояниях:

x_0 — все каналы свободны;

x_1 — один канал занят, $n - 1$ каналов свободны;

x_2 — два канала заняты, $n - 2$ каналов свободны;

x_k — занято k каналов, $n - k$ каналов свободны;

x_n — заняты все n каналов.

Определим $P_0(t + \Delta t)$ — вероятность того, что в момент времени $t + \Delta t$ система будет в состоянии «0». Это сложное событие может произойти при свершении двух несовместимых событий.

Первое событие заключается в том, что в момент времени t все каналы были свободны, а за время Δt не поступило ни одного требования. Само первое событие состоит из двух независимых событий, вероятности которых определяются так:

а) $P_0(t)$ — вероятность того, что в момент времени t все каналы были свободны;

б) $e^{-\lambda\Delta t} = 1 - \lambda\Delta t + 0(\Delta t)$ — вероятность того, что за время Δt не поступило ни одного требования. Отсюда вероятность

первого события может быть записана по теореме умножения вероятностей независимых событий:

$$P_0(t)e^{-\lambda\Delta t} = P_0(t)[1 - \lambda\Delta t + 0(\Delta t)]. \qquad (10.40)$$

Второе событие заключается в том, что в момент времени t один канал был занят, но за время Δt он освободился, за это же время Δt не поступило ни одного нового требования. Значит, само второе событие состоит из трех независимых событий, вероятности которых определяются так:

а) $P_1(t)$ — вероятность того, что в момент времени t был занят один канал;

б) $(1 - e^{-\mu\Delta t})$ — вероятность того, что за время Δt будет завершено обслуживание одного требования;

в) $e^{-\lambda\Delta t}$ — вероятность того, что за время Δt не поступит нового требования.

Таким образом, вероятность второго события определится как произведение трех вероятностей:

$$P_1(t)(1 - e^{-\mu\Delta t})e^{-\lambda\Delta t} = P_1(t)[1 - 1 + \mu\Delta t](1 - \lambda\Delta t) =$$
$$= P_1(t)\mu\Delta t - P_1(t)\lambda(\Delta t)^2\mu + 0(\Delta t) = P_1(t)\mu\Delta t + 0(\Delta t). \qquad (10.41)$$

Отметим, что произведение $P_1(t)\lambda(\Delta t)^2\mu$ по крайней мере на порядок меньше Δt, поэтому оно вошло в слагаемое, обозначенное $0(\Delta t)$.

Вероятность события $P_0(t + \Delta t)$, зависящего от двух несовместимых *первого* и *второго событий*, определится по теореме сложения:

$$P_0(t + \Delta t) = P_0(t)(1 - \lambda\Delta t) + P_1(t)\mu\Delta t + 0(\Delta t). \qquad (10.42)$$

При $\Delta t \to 0$ можно записать, что

$$\frac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t).$$

Для любых других событий при $1 \leqslant k < n$ учитывается, что процесс марковский, поэтому по аналогии можно записать:

$$\frac{dP_k(t)}{dt} = \lambda P_{k-1}(t) - (\lambda + k\mu)P_k(t) + (k + 1)\mu P_{k+1}(t), \qquad (10.43)$$

где $P_{k-1}(t)$ — вероятность нахождения в системе ровно $k - 1$ заявок;
$P_k(t)$ — вероятность нахождения в системе ровно k заявок;
$P_{k+1}(t)$ — вероятность нахождения в системе ровно $k + 1$ заявок;
$k\mu$ — среднее число обслуживаемых в единицу времени требований при работе k аппаратов;

$(k+1)\mu$ — среднее число обслуживаемых требований при работе $k+1$ каналов.

Запишем для случая $k = n$, не забывая при этом, что в системе не может быть $n + 1$ требований, так как одно требование получит отказ и будет потеряно в системе:

$$\frac{dP_n(t)}{dt} = \lambda P_{n-1}(t) - n\mu P_n(t). \tag{10.44}$$

Начальные условия, свидетельствующие о том, что все аппараты свободны при $t = 0$, записываются так:

$$\left.\begin{aligned} P_0(0) &= 1; \\ P_k(0) &= 0. \end{aligned}\right\} \tag{10.45}$$

Есть еще одно условие: сумма вероятностей полной группы несовместных $n + 1$ событий равна 1. Оно называется нормирующим условием:

$$\sum_{k=0}^{n} P_k(t) = 1. \tag{10.46}$$

В установившемся стационарном процессе можно получить предельные значения вероятностей для стационарного решения системы уравнений при

$$P_k(t) \to P_k = \text{const};$$

$$\frac{dP_k(t)}{dt} \to 0 \ (k = 0, 1, 2, ..., n). \tag{10.47}$$

При упомянутых условиях полученная система дифференциальных уравнений преобразуется в систему однородных алгебраических уравнений:

$$\left.\begin{aligned} -\lambda P_0 + \mu P_1 &= 0; \\ &\cdots\cdots\cdots \\ \lambda P_{k-1} - (\lambda + k\mu)P_k + (k+1)\mu P_{k+1} &= 0 \ (1 \leqslant k < n); \\ &\cdots\cdots\cdots \\ \lambda P_{n-1} - n\mu P_n &= 0. \end{aligned}\right\} \tag{10.48}$$

Далее

$$\sum_{k=0}^{n} P_k = 1. \tag{10.49}$$

Совместное решение системы (10.48) при условии (10.49) дает формулу Эрланга:

$$P_k = \frac{\left(\frac{\lambda}{\mu}\right)^k \cdot \frac{1}{k!}}{\sum_{i=0}^{n} \left(\frac{\lambda}{\mu}\right)^i \cdot \frac{1}{i!}}, \tag{10.50}$$

где $\gamma = \frac{\lambda}{\mu}$ — нормированная интенсивность входного потока;

γ — среднее число требований, поступающих в систему за среднее время обслуживания одного требования.

Вероятность отказа системы определяется из формулы (10.50) при $k = n$:

$$P_{\text{отк}} = P_n = \frac{\left(\frac{\lambda}{\mu}\right)^n \cdot \frac{1}{n!}}{\sum_{i=0}^{n} \left(\frac{\lambda}{\mu}\right)^i \cdot \frac{1}{i!}}. \tag{10.51}$$

Относительная пропускная способность определится как вероятность того, что хотя бы один канал в момент поступления заявки свободен:

$$\Pi_0 = 1 - P_n. \tag{10.52}$$

Абсолютная пропускная способность

$$\Pi = \lambda\Pi_0 = \lambda(1 - P_n). \tag{10.53}$$

Среднее число занятых аппаратов и коэффициент занятости рассчитываются так:

$$\left.\begin{aligned} N_З &= \sum_{k=1}^{n} k P_k; \\ K_З &= \frac{N_З}{n}. \end{aligned}\right\} \tag{10.54}$$

Среднее число аппаратов, свободных от обслуживания,

$$\left.\begin{aligned} N_0 &= \sum_{k=0}^{n-1} (n-k) P_k; \\ K_П &= \frac{N_0}{n}. \end{aligned}\right\} \tag{10.55}$$

Пример 10.3

Вычислительный центр СУ может одновременно производить расчет оптимального недельного плана для трех предприятий на трех имеющихся ЭВМ. Среднее время работы ВЦ для одного предприятия равно 3 ч. Интенсивность поступления заявок на расчет оптимальных планов равна 0,25 1/ч. Если ВЦ одновременно занят вычислением оптимальных планов для трех предприятий, он отказывает вновь поступившему заказу, и предприятие, получившее отказ в обслуживании, вынуждено обращаться в другой ВЦ. Найдем характеристики эффективности ВЦ по расчету еженедельных оптимальных планов: P_n, $П_0$, П, считая процесс работы ВЦ установившимся.

Решение

1. Запишем условия задачи в принятых обозначениях:

$$n = 3;\ \lambda = 0{,}25\ 1/ч;\ t_{обс} = 3\ ч.$$

2. $$P_n = \frac{\gamma^n \frac{1}{n!}}{\sum\limits_{i=0}^{n} \gamma^i \frac{1}{i!}} = \frac{\frac{(0{,}25 \cdot 3)^3}{3!}}{\sum\limits_{i=0}^{3} (0{,}75)^i \frac{1}{i!}} \approx 0{,}033\ (\gamma = \frac{0{,}25}{1/3} = 0{,}75).$$

3. $$П_0 = 1 - P_n = 0{,}967.$$

4. $$П = \lambda П_0 = 0{,}25 \cdot 0{,}967 \approx 0{,}242\ 1/ч.$$

Пропускная способность ВЦ СУ равна 0,242 оптимального плана в час, или 1,94 плана в сутки при односменной восьмичасовой работе.

Вероятность отказа равна 3,3% заявок. Относительная пропускная способность показывает, что ВЦ может обслужить 96,7% заявок.

Пример 10.4

Воспользуемся условиями примера 10.3 и определим потери ВЦ из-за отказов предприятиям, простоя ЭВМ и расходы на работу ЭВМ. Выясним экономическую эффективность работы ВЦ при условии, что других задач он не будет выполнять. Известно, что простой ЭВМ в течение 1 ч рабочего времени приносит убытков 100 руб. Каждая заявка на расчет недельного плана приносит 300 руб. дохода в час. Стоимость 1 ч работы ЭВМ обходится ВЦ в 30 руб.

Решение

1. Запишем условия задачи в принятых обозначениях:

$$П_0 = 0{,}967;\ q_k = 30\ руб.;\ q_у = 300\ руб.;$$

$$q_{пк} = 100\ руб.;$$

$$P_n = 0{,}033;$$

$\lambda = 0{,}25$ 1/ч; $T = 25$ рабочих дней (в течение месяца).

Вычислим $G_П$, N_0, $N_З$, E, $K_З$, $K_П$.

2. Определим N_0 и $K_П$ по формуле (10.55):

$$N_0 = \sum_{k=0}^{2} (3-k) P_n = 3P_0 + 2P_1 + P_2.$$

Для определения P_0 в формуле (10.50) предполагаем $k = 0$:

$$P_0 = \frac{1}{\sum\limits_{i=0}^{3} \gamma^i \frac{1}{i!}} = \frac{1}{1 + 0{,}75 + 0{,}28125 + 0{,}07032} \approx 0{,}48.$$

Далее для определения P_1 и P_2 полагаем в формуле (10.50) $k = 1$ и $k = 2$ соответственно:

$$P_1 = \frac{0{,}75}{2{,}102} \approx 0{,}36;$$

$$P_2 = \frac{\frac{1}{2} \cdot 0{,}75^2}{2{,}102} \approx 0{,}13;$$

$$N_0 = 1{,}44 + 0{,}72 + 0{,}13 = 2{,}29;$$

$$K_П = \frac{2{,}29}{3} \approx 0{,}76.$$

3. Найдем $N_З$ и $K_З$:

$$N_З = \sum_{k=1}^{3} k P_k = P_1 + 2P_2 + 3P_3 = 0{,}36 + 0{,}26 + 0{,}099 \approx 0{,}72$$

(заметим, что $P_3 = P_n = 0{,}033$);

$$K_З = \frac{0{,}72}{3} = 0{,}24.$$

4. Вычислим средние потери ВЦ по формуле (10.36):

$$G_{\Pi} = (30 \cdot 0{,}72 + 300 \cdot 0{,}033 \cdot 0{,}25 + 100 \cdot 2{,}29)25 \cdot 8 =$$
$$= (21{,}6 + 2{,}48 + 229) \cdot 200 = 253{,}08 \cdot 200 = 50\ 616 \text{ руб.}$$

Проведем анализ цифр последней суммы в скобках: 21,6 руб. — удельная стоимость работы ЭВМ; 2,48 руб. — часовые потери ВЦ из-за отказа заявкам; 229 руб. — потери ВЦ в результате простоя ЭВМ (эти потери самые большие, потому что в среднем из трех ЭВМ 2,29 простаивают).

Выясним экономическую эффективность работы ВЦ при таком режиме обслуживания.

5. Определим E:

$$E = P_{\text{обс}}\lambda CT - G_{\Pi} = P_0\lambda q_y T - G_{\Pi} = 0{,}967 \cdot 0{,}25 \cdot 300 \cdot 200 -$$
$$- 50\ 616 = -36\ 111 \approx -36\ 100 \text{ руб.}$$

Следовательно, при заданных условиях ВЦ будет нуждаться в дополнительной дотации, равной 36 100 руб. в месяц. Его работу нельзя признать эффективной.

Таким образом, можно сделать следующие **выводы**:

$P_n = 0{,}033$, т. е. 3,3% заявок на расчеты недельных планов предприятий получают отказ;

$P_0 = 0{,}48$, т. е. 48% рабочего времени все три ЭВМ простаивают;

$N_3 = 0{,}72$, т. е. в среднем работает 0,72 ЭВМ;

$K_3 = 0{,}24$, т. е. 24% рабочего времени работает каждая ЭВМ;

$N_0 = 2{,}29$, т. е. 2,29 ЭВМ в среднем простаивают в ожидании заявок;

$K_{\Pi} = 0{,}76$, т. е. 76% рабочего времени каждая ЭВМ простаивает, ожидая заявок;

$\Pi_0 = 0{,}967$, т. е. 96,7% заявок, поступивших на ВЦ, будут им обслужены;

$\Pi = 0{,}24$, т. е. 0,24 заявки обслуживается в течение одного часа;

$G_{\Pi} = 50\ 616$ руб., т. е. потери из-за отказов в обслуживании, простоев ЭВМ и расходы на работу ЭВМ составляют эту сумму;

$E = -36\ 100$ руб., т. е. превышение расходов над доходами составляет эту сумму.

Общий вывод таков — на данном ВЦ машинное время используется неэффективно, ВЦ надо догрузить другими задачами, которые должны выполняться в перерывах между поступлениями заявок, если их считать приоритетными в данных условиях.

Рассмотрим далее эффективность работы ВЦ, если от него потребовать, чтобы он отказывал не более чем 0,01% заявок.

Пример 10.5

Пусть $\lambda = 0{,}25$ 1/ч; $\mu = \frac{1}{3}$ 1/ч; $n = 3$. Условия работы ЭВМ на ВЦ остаются теми же, что и в примерах 10.3 и 10.4, но ВЦ не может допустить отказа более чем 1% заявок. Определим характеристики эффективности работы ВЦ, как это было сделано в примерах 10.3 и 10.4.

Решение

Воспользуемся данными решений в примерах 10.3 и 10.4. Разумеется, что для выполнения задачи есть один путь: увеличивать число ЭВМ в ВЦ до тех пор, пока $P_n \leqslant 0{,}01$. Предположим, что мы увеличим число ЭВМ на одну, затем на две машины и т. д.

1. Предположим, что на ВЦ поставлена четвертая ЭВМ, и определим P_n. Для этого воспользуемся готовыми данными из предыдущих задач, но для определения P_0 посчитаем $(0{,}75)^4 \frac{1}{4!} = 0{,}013$.

Затем найдем сумму

$$\sum_{i=0}^{4} (0{,}75)^i \frac{1}{i!} = \frac{1}{P_0} \approx 2{,}12.$$

Отсюда

$$P_0 = \frac{1}{2{,}12} \approx 0{,}47;$$

$$P_4 = 0{,}013 \cdot 0{,}47 \approx 0{,}006 < 0{,}01.$$

Следовательно, условие задачи выполняется. Отказано будет только 0,6% заявок. Таким образом, нет необходимости изучать случаи, когда $n > 4$.

2. Определим P_k при $n = 4$:

$$P_1 = 0{,}75 \cdot 0{,}47 \approx 0{,}35; \quad P_2 = 0{,}28 \cdot 0{,}47 \approx 0{,}13;$$

$$P_3 = 0{,}07 \cdot 0{,}47 \approx 0{,}03; \quad P_4 \approx 0{,}006.$$

3. Найдем N_0 и $K_П$ по формуле (10.55):

$$N_0 = \sum_{k=0}^{3} (4-k)\,P_k = 4P_0 + 3P_1 + 2P_2 + P_3 =$$

$$= 4 \cdot 0{,}47 + 3 \cdot 0{,}35 + 2 \cdot 0{,}13 + 0{,}03 \approx 3{,}22;$$

$$K_П = \frac{3{,}22}{4} \approx 0{,}81.$$

4. Вычислим $N_з$ и $K_з$:

$$N_з = \sum_{k=1}^{4} k\,P_k = P_1 + 2P_2 + 3P_3 + 4P_4 =$$

$$= 0{,}35 + 2 \cdot 0{,}13 + 3 \cdot 0{,}03 + 4 \cdot 0{,}006 \approx 0{,}72;$$

$$K_з = \frac{0{,}72}{4} \approx 0{,}18.$$

5. Найдем П и $П_0$:

$$П_0 = 1 - P_n = 1 - 0{,}006 \approx 0{,}994;$$

$$П = П_0 \cdot 1 = 0{,}994 \cdot 0{,}25 \approx 0{,}25\ 1/ч.$$

6. Потери СМО по формуле (10.36) составят:

$$G_П = (30 \cdot 0{,}72 + 300 \cdot 0{,}006 \cdot 0{,}25 + 100 \cdot 3{,}22)200 = 68\,810\ руб.$$

7. Подсчитаем потери СМО с учетом доходов от выполнения заявок по формуле (10.39):

$$E = 0{,}994 \cdot 0{,}25 \cdot 300 \cdot 200 - 68\,810 = -53\,900\ руб.$$

Представим характеристики эффективности в табл. 10.2 и сравним их с результатами в примере 10.4.

Следовательно, по сравнению с примером 10.4 почти на 6% уменьшился коэффициент занятости, более чем на 5% увеличился коэффициент простоя (с учетом округления). Если раньше простаивало в среднем 2,29 ЭВМ, то теперь простаивает 3,22 ЭВМ. Существенно возросла относительная пропускная способность: с 0,967 до 0,994. Однако это увеличение достигнуто очень дорогой ценой. Доходы возросли на 405 руб.

Таблица 10.2

Сравнение характеристик эффективности СМО

№ примера	n	P_n	P_0	$N_з$	$K_з$	N_0	$K_П$	$П_0$	П, 1/ч	$G_П$, руб.	E, руб.
10.5	4	0,006	0,47	0,72	0,18	3,22	0,81	0,994	0,25	68 810	−53 900
10.4	3	0,033	0,48	0,72	0,24	2,29	0,76	0,967	0,24	50 616	−36 100

(14 910 − 14 505), а убытки возросли на 18194 руб. (68 810 − − 50 616) в месяц.

Таким образом, если предварительно оценивать экономическую эффективность нового требования о допуске 1% отказов в расчетах планов вместо 3,3%, то можно заранее оценить, во что это обойдется и как будет обеспечена загрузка ЭВМ на ВЦ.

Критиковать плохую организацию работы всегда легче, чем предложить способ улучшения работы. По-видимому, мы бы могли по-другому построить СМО, если бы заявкам не было отказов в обслуживании. Можно было бы принять бóльший поток заявок. Часть из них могла стоять в очереди на обслуживание. Рассмотрим другой тип обслуживания — СМО с ожиданием.

10.6. Эффективность работы систем, построенных по принципу СМО с ожиданием

В СМО с ожиданием заявки, заставшие все аппараты занятыми, не покидают СМО, а становятся в очередь в ожидании обслуживания. Таким образом, можно воспользоваться формулой (10.48) для общего случая, когда число требований равно k, а число аппаратов равно 1.

При этом

$$\lambda P_{k-1} - (\lambda + \mu) P_k + \mu P_{k+1} = 0 \text{ при } k > 1; \qquad (10.56)$$

$$-\lambda P_0 + \mu P_1 = 0 \text{ при } k = 1, \qquad (10.57)$$

откуда

$$P_1 = \frac{\lambda}{\mu} P_0 = \gamma P_0, \qquad (10.58)$$

где $\frac{\lambda}{\mu} = \gamma$ — коэффициент загрузки.

Если $\gamma < 1$, то режим обслуживания в системе с ожиданием является устойчивым, если $\gamma > 1$, очередь на обслуживание неограниченно возрастает.

Рассмотрим случай, когда $k > 1$. Пусть $k = 2$. Тогда

$$\lambda P_0 - (\lambda + \mu)P_1 + \mu P_2 = 0,$$

откуда

$$P_2 = (1 + \gamma)P_1 - \gamma P_2;$$

$$P_2 = (1 + \gamma)\gamma P_0 - \gamma P_0 = \gamma^2 P_0.$$

Получили рекуррентные соотношения:

$$P_1 = \gamma P_0;$$

$$P_2 = \gamma^2 P_0;$$

....

$$P_k = \gamma^k P_0.$$

Если $k \to \infty$, то $\sum\limits_{i=0}^{\infty} P_i = 1$, т. е.

$$(1 + \gamma + \gamma^2 + ... + \gamma^k + ...)P_0 = 1.$$

Сумма геометрической прогрессии с первым членом, равным 1, знаменателем прогрессии γ равна $\frac{1}{1-\gamma}$. Следовательно, $1 = \frac{P_0}{1-\gamma}$, откуда

$$\left.\begin{aligned} P_0 &= 1 - \gamma; \\ P_k &= (1 - \gamma)\gamma^k. \end{aligned}\right\} \qquad (10.59)$$

Среднее число заявок, находящихся в системе массового обслуживания, равно:

$$M = \sum_{k=0}^{\infty} k P_k = \sum_{k=0}^{\infty} k\gamma^k(1-\gamma) = (1-\gamma)(\gamma + 2\gamma^2 + 3\gamma^3 + ...) =$$
$$= \gamma(1-\gamma)(1 + 2\gamma + 3\gamma^2 + ...).$$

Рассмотрим последнюю скобку:

$$1 + 2\gamma + 3\gamma^2 + ... = \frac{d}{d\gamma}(\gamma + \gamma^2 + \gamma^3 + ...) = \frac{d}{d\gamma}\left(\frac{\gamma}{1-\gamma}\right) = \frac{1}{(1-\gamma)^2}.$$

Таким образом,

$$M = \frac{\gamma(1-\gamma)}{(1-\gamma)^2} = \frac{\gamma}{1-\gamma}. \qquad (10.60)$$

Пусть l — длина очереди. Так как число аппаратов $k = 1$, то $l = k - 1$. Поток пуассоновский, значит, стационарный, поэтому

$$\frac{M}{\mu} = \frac{M_{\text{ож}}}{\lambda};\ M_{\text{ож}} = \frac{1}{\mu} M\lambda = \gamma M = \frac{\gamma^2}{1-\gamma}, \qquad (10.61)$$

где $\frac{M}{\mu}$ — время обслуживания среднего числа единиц в системе.

Среднее время ожидания в очереди

$$t_{\text{ож}} = \frac{M_{\text{ож}}}{\lambda} = \frac{\gamma^2}{\gamma(1-\gamma)} = \frac{\gamma}{\mu(1-\gamma)}, \qquad (10.62)$$

где $t_{\text{ож}}$ — время ожидания среднего числа единиц, или среднее время ожидания в очереди.

Пример 10.6

В цехе имеется одна инструментальная кладовая, которую обслуживает один кладовщик, получающий заработную плату, например, 100 руб. Средняя заработная плата рабочего условно равна 250 руб., $\lambda = 1{,}1 \frac{1}{\text{мин.}}$, $\mu = 1{,}28 \frac{1}{\text{мин.}}$.

Определим:

среднее число рабочих, ожидающих в очереди;

среднее число рабочих в системе обслуживания;

потери в результате простоя кладовщика.

Рассматриваемый период равен 25 рабочим дням в течение месяца при восьмичасовом рабочем дне.

Решение

1. Определим среднее число рабочих, ждущих инструмента в очереди:

$$M_{\text{ож}} = \frac{\gamma^2}{1-\gamma} = \frac{\left(\frac{1,1}{1,28}\right)^2}{1-\frac{1,1}{1,28}} = \frac{0,86^2}{0,14} \approx 5,28.$$

2. Рассчитаем среднее время ожидания рабочего в очереди:

$$t_{\text{ож}} = \frac{\gamma}{\mu(1-\gamma)} = \frac{0,86}{1,28 \cdot 0,14} \approx 4,8 \text{ мин.}$$

Таким образом, каждый рабочий при обращении в кладовую в среднем теряет 4,8 мин. рабочего времени.

3. Определим среднюю долю времени простоя кладовщика:

$$N_0 = \sum_{k=0}^{0}(1-0)P_0 = P_0 = 1-\gamma = 1-0,86 = 0,14.$$

Кладовщик простаивает 14% своего рабочего времени.

4. Вычислим потери кладовщика. Кладовщик простаивает 14% рабочего времени. В течение месяца его простой обойдется в $C_{\text{к}} = 100 \cdot 0,14 = 14$ руб.

5. Определим потери рабочих из-за простоя в очереди. Попробуем это сделать логически. Общее число рабочих, приходящих в кладовую в течение месяца, равно:

$$1 \cdot 60 \cdot 200 = 1,1 \cdot 60 \cdot 200 = 13\,200 \text{ обращений.}$$

При каждом обращении рабочий простаивает в среднем 4,8 мин. Таким образом, потери рабочих цеха в течение месяца составят:

$$13\,200 \cdot 4,8 = 63\,360 \text{ мин.};$$

$$63\,360 : 60 \text{ с} = 1056 \text{ чел.-ч};$$

$$1056 : 8 \text{ ч} = 132 \text{ чел.-дня.}$$

Сколько же человеко-месяцев потеряно? Разделим 132 на 25 и получим 5,28 чел.-мес. Случайно ли эта цифра совпала с $M_{\text{ож}}$? Нет! В среднем стоят в очереди за инструментом 5,28 человека.

Значит, в течение дня теряется 5,28 чел.-дня, а в течение месяца — 5,28 чел.-мес., отсюда

$$C_{\text{р}} = 5,28 \cdot 250 = 1320 \text{ руб.}$$

6. Определим общие потери кладовщика и рабочих цеха из-за простоев:

$$C = C_{\text{к}} + C_{\text{р}} = 14 + 1320 = 1334 \text{ руб.}$$

Таким образом, цех несет большие потери из-за простоя рабочих в очереди. Целесообразно в штаты цеха добавить дополнительных кладовщиков при заданных условиях работы.

Рассмотрим более сложный случай, когда число однородных обслуживающих устройств равно n. Как и ранее (при $n = 1$), нет ограничений на очередь. При этом можно записать:

$$\left.\begin{aligned} -\lambda P_0 + \mu P_1 = 0 \text{ при } k = 0; \\ \lambda P_{k-1} - (\lambda + k\mu)P_k + (k+1)\mu P_{k+1} = 0 \text{ при } 1 \leqslant k < n; \\ \lambda P_{n-1} - (\lambda + n\mu)P_k + n\mu P_{k+1} = 0,\ k \geqslant n. \end{aligned}\right\} \quad (10.63)$$

Отсюда

$$P_k = P_0\frac{\gamma^k}{k!} \text{ при } 1 \leqslant k < n; \quad (10.64)$$

$$P_k = P_0\frac{\gamma^k}{n!\,n^{k-n}} \text{ при } k \geqslant n. \quad (10.65)$$

Выпишем для наглядности значения P_k:

$$P_1 = \gamma P_0;$$

$$P_2 = \frac{\gamma^2}{2!}P_0; \ldots$$

$$P_n = \frac{\gamma^n}{n!}P_0;$$

$$P_{n+1} = \frac{\gamma^{n+1}}{n!} \cdot \frac{P_0}{n};$$

$$P_{n+2} = \frac{\gamma^{n+2}}{n!} \cdot \frac{P_0}{n^2}; \ldots$$

$$P_{n+l} = \frac{\gamma^{n+l}}{n!} \cdot \frac{P_0}{n^l},$$

где l — длина очереди ($l = k - n$; $k > n$).

Определим P_0.

Имеем

$$\sum_{k=0}^{\infty} P_k = 1,$$

т. е.

$$P_0\left(1+\gamma+\frac{\gamma^2}{2!}+\ldots+\frac{\gamma^n}{n!}+\ldots+\frac{\gamma^{n+1}}{n!}\cdot\frac{1}{n}+\ldots+\frac{\gamma^{n+l}}{n!}\cdot\frac{1}{n^l}+\ldots\right)=1;$$

$$P_0=\left[1+\gamma+\frac{\gamma^2}{2!}+\ldots+\frac{\gamma^n}{n!}\left(1+\frac{\gamma}{n}+\frac{\gamma^2}{n^2}+\ldots+\frac{\gamma^l}{n^l}+\ldots\right)\right]^{-1}=$$

$$=\left[\sum_{k=0}^{n-1}\frac{\gamma^k}{k!}+\frac{\gamma^n}{n!}\cdot\frac{n}{n-\gamma}\right]^{-1}. \qquad (10.66)$$

При $\frac{\gamma}{n} < 1$ ряд в круглых скобках сходится. Сумма членов этого ряда равна:

$$\frac{1}{1-\frac{\gamma}{n}} = \frac{n}{n-\gamma}.$$

Можно P_0 переписать:

$$P_0=\left[\sum_{k=0}^{n-1}\frac{\gamma^k}{k!}+\frac{\gamma^n}{(n-1)!}\cdot\frac{1}{n-\gamma}\right]^{-1}.$$

Средняя длина очереди может быть определена как сумма произведений возможного числа рабочих в очереди на соответствующие вероятности:

$$M_{\text{ож}} = 1\cdot P_{n+1} + 2P_{n+2} + \ldots + lP_{n+l} + \ldots =$$

$$=\frac{\gamma^{n+1}}{nn!}P_0\left[1+\frac{2\gamma}{n}+\ldots+l\left(\frac{\gamma}{n}\right)^{l-1}+\ldots\right]=\frac{\gamma^{n+1}}{nn!}\frac{P_0}{\left(1-\frac{\gamma}{n}\right)^2}, \qquad (10.67)$$

так как по аналогии с формулой (10.60)

$$1+\frac{2\gamma}{n}+3\left(\frac{\gamma}{n}\right)^2+\ldots+l\left(\frac{\gamma}{n}\right)^{l-1}+\ldots=\frac{1}{\left(1-\frac{\gamma}{n}\right)^2}.$$

Среднее время ожидания в очереди:

$$t_{\text{ож}}=\frac{M_{\text{ож}}}{\lambda}=\frac{\gamma\gamma^2}{nn!\gamma\mu}\cdot\frac{P_0}{\left(1-\frac{\gamma}{n}\right)^2}=\frac{\gamma^n}{n\mu n!}\cdot\frac{P_0}{\left(1-\frac{\gamma}{n}\right)^2} \text{ при } \gamma/n<1. \qquad (10.68)$$

Среднее число заявок в системе обслуживания:

$$M = M_{\text{ож}} + \gamma. \qquad (10.69)$$

Среднее число простаивающих аппаратов можно определить по формуле:

$$N_0=\sum_{k=0}^{n-1}(n-k)P_k. \qquad (10.70)$$

Для условий задачи 10.6 выявим, при каком числе кладовщиков потери в цехе будут минимальными.

Пример 10.7

Выясним, при каком числе кладовщиков потери в цехе будут минимальными, если оставить одну очередь, но обслуживание организовать так, чтобы работали 2, 3, ..., n^0 кладовщиков, где n^0 — число кладовщиков, обеспечивающих минимальные потери от простоя кладовщиков и ожидания рабочих в очереди, если $\lambda = 1{,}1\ \frac{1}{\text{мин.}}$, $\mu = 1{,}28\ \frac{1}{\text{мин.}}$.

Решение

1. Оптимизацию надо проводить относительно суммы потерь:

$$C = C_р \cdot M_{\text{ож}} + C_к N_0 = C_р f(n) + C_к\psi(n).$$

Примем $n = 2, 3, 4, \ldots$

2. Определим P_0 для $n = 2, 3, 4$:

$$n=2,\ \frac{\gamma}{n}=\frac{1{,}1}{1{,}28\cdot 2}=0{,}43;\ 0{,}43<1;$$

$$P_0 = \left[1 + 0{,}86 + \frac{(0{,}86)^2 \cdot 2}{2!(2-0{,}86)}\right]^{-1} \approx 0{,}40;$$

$$n = 3,\ \frac{\gamma}{n} = \frac{0{,}86}{3} \approx 0{,}29;\ 0{,}29 < 1;$$

$$P_0 = \left[1 + 0{,}86 + \frac{0{,}86^2}{2!} + \frac{0{,}86^3}{3!} \cdot \frac{3}{3-0{,}86}\right]^{-1} \approx 0{,}42;$$

$$n = 4,\ \frac{\gamma}{n} = \frac{0{,}86}{4} = 0{,}215;\ 0{,}215 < 1;$$

$$P_0 = \left[1 + 0{,}86 + \frac{0{,}86^2}{2!} + \frac{0{,}86^3}{3!} + \frac{0{,}86^4}{4!} \cdot \frac{4}{4-0{,}86}\right]^{-1} \approx 0{,}42.$$

3. Простои рабочих в очереди для $n = 2, 3, 4$ составят:

$$M_{\text{ож}} = \frac{\gamma^{n+1}}{nn!} \cdot \frac{P_0}{\left(1 - \frac{0{,}86}{2}\right)^2};$$

$$n = 2;\ M_{\text{ож}} = \frac{0{,}86^3}{2 \cdot 2!} \cdot \frac{0{,}40}{\left(1 - \frac{0{,}86}{2}\right)^2} \approx 0{,}195;$$

$$n = 3;\ M_{\text{ож}} = \frac{0{,}86^4}{3 \cdot 3!} \cdot \frac{0{,}42}{\left(1 - \frac{0{,}86}{3}\right)^2} \approx 0{,}025;$$

$$n = 4;\ M_{\text{ож}} = \frac{0{,}86^5}{4 \cdot 4!} \cdot \frac{0{,}42}{\left(1 - \frac{0{,}86}{4}\right)^2} = 0{,}003.$$

4. Простои кладовщиков в ожидании требований на инструмент рассчитаем так:

$$n = 2,\ N_0 \approx 1{,}14;$$
$$n = 3,\ N_0 \approx 2{,}14;$$
$$n = 4,\ N_0 \approx 3{,}14.$$

5. Рассчитаем потери цеха при заданном потоке заявок на инструмент и разном числе кладовщиков (1, 2, 3, 4) и определим оптимальное число кладовщиков, руб.:

$$C_1 = 5{,}28 \cdot 250 + 0{,}14 \cdot 100 = 1334;$$
$$C_2 = 0{,}195 \cdot 250 + 1{,}14 \cdot 100 = 162{,}75;$$
$$C_3 = 0{,}025 \cdot 250 + 2{,}14 \cdot 100 = 220{,}25;$$
$$C_4 = 0{,}003 \cdot 250 + 3{,}14 \cdot 100 = 314{,}75.$$

Итак, цех будет иметь минимальные потери при работе двух кладовщиков. При одном кладовщике значительны потери из-за простоя рабочих, а при увеличении числа кладовщиков более двух растут потери из-за простоя кладовщиков.

Если в инструментальную кладовую добавить одного кладовщика, выигрыш в месяц составит:

$$1334 - (162{,}85 + 100) = 1071{,}25 \text{ руб.}$$

Таким образом, при помощи критериев эффективности теории массового обслуживания можно выбрать экономически наиболее выгодную организацию работы по обеспечению инструментом рабочих цеха.

Рассмотрим далее замкнутую систему с очередью ограниченной длины.

Пример 10.8

В системе имеется m источников заявок, n аппаратов, известны λ и μ; в систему поступает простейший поток заявок, и обслуживание проводится по показательному закону распределения времени окончания обслуживания. Любой освободившийся аппарат немедленно приступает к обслуживанию очередной заявки. В системе массового обслуживания может быть $(m + 1)$ состояний:

- x_0 — все аппараты свободны;
- x_1 — один аппарат занят, $(n - 1)$ свободны;
- x_n — все n аппаратов заняты;
- x_{n+1} — все n аппаратов заняты, одна заявка стоит в очереди;
- x_{n+l} — все n аппаратов заняты, l заявок стоят в очереди;
- x_m — все n аппаратов заняты, $(m - n)$ стоят в очереди.

Пусть наладчик обслуживает пять станков. Среднее время обслуживания равно 12 мин. Станки требуют наладки в среднем раз в 2 ч. Наладчик получает в месяц, например, 220 руб. Простой одного станка в течение месяца приводит к потере 12 500 руб. (в месяце 25 рабочих дней по 8 ч).

Определите критерии эффективности системы массового обслуживания: P_0, $N_З$, $M_{ож}$, M, П, $П_0$, $G_П$, $K_З$, $K_П$.

Для решения этой задачи необходимы формулы, естественно отличающиеся от предыдущих, так как изменились условия: заявки ждут теперь не неограниченное время, а ориентируются по длине очереди. Очереди, большей $(m - n)$, не может быть. Формулы для расчета P_k должны учитывать возможности образования очереди ограниченной длины. Следовательно, для вероятностей P_k должны быть два типа формул:

$$P_k = \frac{m!\gamma^k}{k!(m-k)} P_0,\ 1 < k \leqslant n; \tag{10.71}$$

$$P_k = \frac{m!\gamma^k}{n^{k-n}n!(m-k)},\ n < k \leqslant m. \tag{10.72}$$

$$P_0 = \left[\sum_{k=0}^{n} \frac{m!\gamma^k}{k!(m-k)!} + \sum_{k=n+1}^{m} \frac{m!\gamma^k}{n^{k-n}n!(m-k)!}\right]^{-1}. \tag{10.73}$$

Для расчета средней длины очереди используем формулу (10.72), а для расчета среднего числа простаивающих каналов обслуживания — формулу (10.71).

Решение

1. Выпишем исходные данные:

$$\lambda = 0{,}5\ \frac{1}{ч};\ \mu = 5\ \frac{1}{ч}\ (\gamma = 0{,}1);$$

$$q_{nk} = 220 \text{ руб./мес.} = 1{,}1 \text{ руб./ч};$$

$$q_{ож} = 62{,}5 \text{ руб./ч} = 12\ 500 \text{ руб./мес.}$$

Проведем основное решение, пользуясь табл. 10.3.

Заполнение второго столбца проводится по мере расчета отношений:

$$\frac{P_k}{P_0} = \frac{m!\gamma^k}{k!(m-k)!},\ 1 \leqslant k \leqslant n;$$

$$\frac{P_k}{P_0} = \frac{m!\gamma^k}{n^{k-n}n!(m-k)!},\ n < k \leqslant m.$$

В нашем случае это

$$\frac{P_0}{P_0} = 1;\ \frac{P_1}{P_0} = \frac{5!(0{,}1)^1}{1!(5-1)!} = \frac{120\cdot 0{,}1}{24} = 0{,}5;$$

$$\frac{P_2}{P_0} = \frac{5!(0{,}1)^2}{(5-2)!} = \frac{120\cdot 0{,}01}{6} = 0{,}2 \text{ и т. д.}$$

Просуммируем значения второго столбца:

$$\sum_{k=0}^{n} \frac{P_k}{P_0} = \frac{1}{P_0}\sum_{k=0}^{n} P_k = \frac{1}{P_0}.$$

Значит,

$$P_0 = \left(\sum_{k=0}^{n} \frac{P_k}{P_0}\right)^{-1} = \frac{1}{1{,}7732} \approx 0{,}564.$$

2. Проведем анализ результатов. Сумма членов четвертого столбца $M \approx 0{,}64$. Следовательно, из пяти станков 0,64 станка находится в наладке или в очереди, т. е. в нерабочем состоянии. Сумма членов шестого столбца $M_{ож} \approx 0{,}22$. Следовательно, в среднем 0,22 станка из пяти ждет, пока к нему подойдет наладчик.

Таблица 10.3

Расчет характеристик СМО

k	$\frac{P_k}{P_0}$	P_k	kP_k	$k-n$	$(k-n)P_k$
0	1,0000	0,5640000	0,0000	0	0,00000
1	0,5000	0,2820000	0,2820	0	0,00000
2	0,2000	0,1128000	0,2256	1	0,11280
3	0,0600	0,0338400	0,1015	2	0,06768
4	0,0120	0,0067680	0,0271	3	0,02030
5	0,0012	0,0006768	0,0034	4	0,01700
Итого	1,7732	1,0000848	0,3696		0,21778

Для того чтобы выяснить, какую долю рабочего времени простаивает каждый станок, разделим M на 5 и получим, что 12,8% рабочего времени каждый станок не выдает продукцию. В очереди на наладку каждый станок находится 4,4% (0,22 : 5) рабочего времени.

Определим, сколько времени простаивает наладчик: N_0 = 0,564. Это значение совпадает с P_0 (формула 10.70). Значит, 56,4% рабочего времени в среднем наладчик простаивает и занят 43,6% рабочего времени.

3. Определим $П_0$ и П. Формально можно вычислить $П_0$:

$$П_0 = 1 - P_{отк} = 1 - P_5 = 1 - 0{,}00068 \approx 1.$$

В данной системе практически заявки не получают отказа, так как больше пяти станков в очереди не может быть. Однако вероятность отказать одновременно всем пяти станкам, отличная от нуля, существует. В нашей задаче она равна 0,00068.

Точно так же можно формально рассчитать

$$П = \lambda П_0 = 0{,}5 \cdot 0{,}99932 \approx 0{,}5.$$

Абсолютная пропускная способность равна 0,5 заявки в час.

4. Рассчитаем экономические потери в течение месяца $C_м$ и часа $C_ч$, учитывая потери от простоя наладчика и простоя станков в очереди, считая, что наладка является неизбежным элементом технологического процесса, руб.:

$$C_м = q_{ож} M_{ож} + q_{nk} N_0 = 12\ 500 \cdot 0{,}22 + 220 \cdot 0{,}564 \approx 2874;$$

$$C_ч = 62{,}5 \cdot 0{,}22 + 1{,}1 \cdot 0{,}564 = 14{,}37.$$

Проверим предыдущий расчет $C_м$:

$$C_м = C_ч \cdot 8 \cdot 25 = 14{,}37 \cdot 200 = 2874 \text{ руб.}$$

Величина $M_{ож.}$ показывает, какое количество станков в среднем простаивает в очереди. Если мы рассчитываем на 1 ч, то это потери в течение часа, а если за рабочий день, то в наших условиях это будет в 8 раз больше, и т. д.

Возможно, что было бы выгоднее поставить второго наладчика.

Глава 11

ОПРЕДЕЛЕНИЕ ЭФФЕКТИВНОСТИ ОРГАНИЗАЦИИ РАБОТЫ ВЫЧИСЛИТЕЛЬНОГО ЦЕНТРА НА ОСНОВЕ КРИТЕРИЕВ ТЕОРИИ МАССОВОГО ОБСЛУЖИВАНИЯ

11.1. Введение и постановка задачи

Автоматизированные системы управления первой очереди в основном были информационного типа. На СУ второй очереди решаются весьма сложные задачи анализа экономических и технологических процессов и оптимизации решений.

Потребности в решении более трудоемких по времени и по памяти ЭВМ задач постоянно растут. Следовательно, растут потребности СУ в памяти и производительности ЭВМ. Возможности ЭВМ в решении этих растущих потребностей ограничены. В такой ситуации возникает естественная проблема поиска внутренних резервов. Таким резервом является *повышение эффективности организации вычислительного процесса.*

Для ЭВМ третьего и особенно четвертого поколения объем внешней памяти для хранения программ и банков данных не знает ограничений. Однако на практике специалисты сталкиваются с большими затратами времени на поиск информации на внешних накопителях и на переписъ ее в оперативную память. Эти растущие затраты времени становятся естественным ограничением объема памяти ЭВМ.

Развитие современных ЭВМ показало, что эффективное быстродействие (производительность) в соответствии с законом Гроша растет медленнее, чем идет увеличение объема памяти.

Производительность при всех прочих равных условиях может оказаться более существенным ограничением, чем объем памяти. Отсюда при организации вычислительного процесса необходимо серьезное внимание обращать на рациональное использование производительности ЭВМ, особенно для технологических систем. Для информационных систем наиболее важным будет объем памяти.

В дальнейшем не будем останавливаться на выборе критериев качества ЭВМ. Считаем, что система решает свою главную целевую задачу с заданным и удовлетворяющим нас качеством. Будем стараться провести анализ и минимизировать затраты на решение основной целевой задачи системы.

Задача может быть решена, но затраты времени превзойдут допустимое время, в течение которого целесообразно получить конечный результат. Такую задачу нет смысла решать. Следовательно, интерес представляют производительность ЭВМ и затраты времени, необходимые на решение управленческой задачи.

Если же задачи приняты для решения, то нас интересует загрузка системы этими задачами, т. е. темп поступления задач на ЭВМ и длительность их решения, возможное время ожидания задачи в очереди на решение и вероятность пропуска решения задачи.

Для анализа методов организации вычислительного процесса необходимо использовать методы: теории вероятностей, теории расписаний и теории массового обслуживания.

На эффективность использования вычислительных ресурсов управляющих ЭВМ в реальном масштабе времени влияют [57], [58]:

- параметры входных потоков;
- параметры потоков на выдачу сообщений;
- параметры обслуживания заявок i-х типов;
- общая загрузка ЭВМ;
- общая загрузка систем передачи данных внешним абонентам;
- загрузка заявками i-х типов систем передачи данных внешним абонентам;
- загрузка заявками j-х типов ЭВМ;
- структура памяти для заявок i-х типов;
- объем памяти для заявок разных типов;
- дисциплины распределения вычислительных ресурсов и использования памяти при приеме и выдаче сообщений.

Управляющая ЭВМ является потребителем информации внешних потоков сообщений и источником информации для внешних абонентов. Поэтому анализируются потоки заявок, процессы обслуживания, исследуются структура памяти и дисциплина распределения ресурсов ЭВМ на двух этапах:

- прием и обработка информации в ЭВМ;
- подготовка и выдача информации внешним абонентам.

На первом этапе ЭВМ выступает накопителем и обслуживающим аппаратом (каналом обслуживания), на втором – источником и накопителем информации перед обслуживанием. Различие функций на двух этапах требует различных критериев оценки эффективности дисциплин на этих этапах.

На основе теории массового обслуживания можно решить многие классы сложных задач. Для иллюстрации этой возможности целесообразно перечислить задачи, которые имеют наиболее близкую связь с практикой организации современного вычислительного процесса. Это задачи, которые позволяют оценить [57], [58] эффективность:

- методов диспетчеризации с квантованием обслуживания при неограниченной буферной памяти;
- приоритетных методов оперативной диспетчеризации при неограниченном времени ожидания;
- приоритетных и бесприоритетных методов диспетчеризации при ограниченной буферной памяти;
- методов выдачи сообщений из ЭВМ внешним абонентам;
- методов детерминированного планирования вычислительного процесса.

Системы организации вычислительного процесса можно разделить на следующие типы:

- *первый тип* — полностью стохастические;
- *второй тип* — стохастические с известными априорными стохастическими характеристиками;
- *третий тип* — детерминированные или квазидетерминированные.

Эти типы содержат достоверную информацию о заявках.

В системах третьего типа бывает проще составлять расписание работы на длительный период. В системах второго типа используют стохастические характеристики. Наиболее сложными являются задачи первого типа.

11.2. Критерии эффективности методов организации вычислительного процесса

Критерии эффективности дисциплин диспетчеризации по штрафам

Если существует возможность выбора и изменения дисциплины диспетчеризации, то появляется необходимость сопоставления получаемого выигрыша с затратами вычислительных ресурсов ЭВМ. Если бы эти ресурсы были неограниченными, то мы получили бы идеальную эффективность использования ресурсов ЭВМ. На практике же эффективность получается меньшей. Мера снижения эффективности может быть вычислена через **функционал потерь**.

Считают, что *снижение эффективности определяется в первую очередь*:

- потерями из-за ожидания заявок на включение программ до начала решения конкретных задач;
- потерями из-за стирания сообщений в буферной памяти, которая ограничена в реальных системах;
- потерями из-за задержки в решении периодических задач и вследствие увеличения периода их решения.

Пусть заявка обесценивается пропорционально времени задержки, тогда обозначим через $C_\tau^{(s)}$ величину среднего суммарного штрафа вследствие ожидания заявок до начала их обработки.

Функционал ($C_{\tau_{\text{ож}}}^{(s)}$) определяется по формуле:

$$C_{\tau_{\text{ож}}}^{(s)} = \sum_{i=1}^{m} \alpha_i \lambda_i \tau_{\text{ож}\ i}^{(s)}, \tag{11.1}$$

где α_i — штраф за единицу времени ожидания заявки i-го типа;
λ_i — интенсивность потока заявок;
m — число типов заявок;
$\tau_{\text{ож}\ i}^{(s)}$ — средняя длительность ожидания в очереди i-х заявок;
s — характеристика дисциплины приоритетов и системы организации вычислительного процесса.

Однако $\tau_{\text{ож}\ i}^{(s)}$ не учитывает ожидания после прерывания обслуживания, а учитывает только среднее время ожидания начала обслуживания.

Если обслуживание прерывается, то используется другая формула:

$$\tau_i^{(s)} = \tau_{\text{ож}\ i}^{(s)} + \tau_{\Pi i}^{(s)} + \tau_{\text{O}i}, \tag{11.2}$$

где $\tau_i^{(s)}$ — среднее время пребывания заявки в системе до полного завершения обслуживания;
$\tau_{\Pi i}^{(s)}$ — среднее время нахождения заявки в очереди после прерывания;
$\tau_{\text{O}i}$ — среднее время обслуживания заявки.

В общем случае функционал штрафа

$$C_\tau^{(s)} = \sum_{i=1}^{m} \alpha_{\text{ож}\ i} \lambda_i \tau_{\text{ож}\ i}^{(s)} + \sum_{i=1}^{m} \alpha_{\Pi i} \lambda_i \tau_{\Pi i}^{(s)} + \sum_{i=1}^{m} \alpha_{\text{O}i} \lambda_i \tau_{\text{O}i}. \tag{11.3}$$

Если же все коэффициенты принимаются равными ($\alpha_{\text{ож}\ i} = \alpha_{\Pi i} = \alpha_{\text{O}i} = \alpha_i$), то функционал примет вид:

$$C_\tau^{(s)} = \sum_{i=1}^{m} \alpha_i \lambda_i \tau_i^{(s)}. \tag{11.4}$$

Пример 11.1

В рассматриваемом периоде на ВЦ СУ решаются следующие задачи:

тип 1 — перспективное планирование;

тип 2 — недельное планирование;

тип 3 — корректировка технологических процессов в режиме реального времени.

Принята система диспетчеризации, при которой задачи типа 3 имеют приоритет перед задачами типа 2 и 1, а задачи типа 2 имеют приоритет перед задачами типа 1.

На основе обработки статистических данных установлено, что характеристики имеют значения, представленные в табл. 11.1.

Если прервано обслуживание заявки, то после ожидания решение данной задачи продолжается.

Вычислим эффективность дисциплины по штрафам по сравнению с дисциплиной «первым пришел — первым обслужен».

Решение

1. Найдем среднее время пребывания заявок в системе при приоритетном обслуживании по формуле (11.2), ч:

$$\tau_1^{(s)} = 0{,}6 + 0{,}6 + 5 = 6{,}2;$$

Таблица 11.1

Результаты обработки статистических данных

Тип задачи	Приоритетная система					Система без приоритетов				
	α_i, руб.	λ_i, 1/ч	$\tau^{(s)}_{\text{ож } i}$, ч	$\tau^{(s)}_{\Pi i}$, ч	τ_{Oi}, ч	α_i, руб.	λ_i, 1/ч	$\tau^{(s)}_{\text{ож } i}$, ч	$\tau^{(s)}_{\Pi i}$, ч	τ_{Oi}, ч
1	1,0	0,01	0,60	0,60	5,0	1,0	0,01	0,4	0,50	5,0
2	2,5	0,10	0,30	0,20	1,0	2,5	0,10	0,2	0,15	1,0
3	50,0	5,00	0,01	0,01	0,1	50,0	5,00	0,2	0,20	0,1

$$\tau_2^{(s)} = 0{,}3 + 0{,}2 + 1 = 1{,}5;$$

$$\tau_3^{(s)} = 0{,}01 + 0{,}01 + 0{,}1 = 0{,}12.$$

2. Определим среднее время пребывания заявок в системе при бесприоритетном обслуживании, ч:

$$\tau_1 = 0{,}4 + 0{,}5 + 5 = 5{,}9;$$
$$\tau_2 = 0{,}2 + 0{,}15 + 1 = 1{,}35;$$
$$\tau_3 = 0{,}2 + 0{,}2 + 0{,}1 = 0{,}5.$$

3. Найдем потери при приоритетном обслуживании по формуле (11.4):

$$C_\tau^{(s)} = 1 \cdot 0{,}01 \cdot 6{,}2 + 2{,}5 \cdot 0{,}1 \cdot 1{,}5 + 50 \cdot 5 \cdot 0{,}12 \approx 30{,}44 \text{ руб.}$$

4. Вычислим потери при бесприоритетном обслуживании по формуле (11.4):

$$C_\tau^{(q)} = 1 \cdot 0{,}01 \cdot 5{,}9 + 2{,}5 \cdot 0{,}1 \cdot 1{,}35 + 50 \cdot 5 \cdot 0{,}5 \approx 125{,}40 \text{ руб.}$$

5. Определим относительный критерий эффективности диспетчеризации обработки информации:

$$\Phi^{(s,\,q)} = \frac{C_\tau^{(q)}}{C_\tau^{(s)}} = \frac{125{,}40}{30{,}44} \approx 4{,}12.$$

6. Рассчитаем приращение эффекта в результате применения приоритетного обслуживания:

$$125{,}40 - 30{,}44 \approx 95 \text{ руб.}$$

Критерий эффективности дисциплин по эквивалентной производительности ЭВМ

На практике далеко не всегда представляется возможным производить оценку эффективности по суммарной величине штрафа. В таких случаях применяют относительный коэффициент эффективности. Обычно в качестве эталона принимают простейшую дисциплину обслуживания заявок (q). Критерием эффективности s-й дисциплины диспетчеризации может стать отношение

$$\Phi^{(s,\,q)} = \frac{C_\tau^{(q)}}{C_\tau^{(s)}}. \tag{11.5}$$

Этот коэффициент характеризует уменьшение средней величины суммарного штрафа при s-й дисциплине по сравнению с простейшей q-й дисциплиной. В примере (11.1) это коэффициент $\Phi^{(s,\,q)}$.

Если использование приоритетной дисциплины уменьшает суммарный штраф в $\Phi^{(s,\,q)}$ раз, то, сохранив q-ю дисциплину обслуживания заявок, можно добиться этого же результата, применив ЭВМ с бо́льшей производительностью. Другими словами, критерием эффективности приоритетных дисциплин может служить относительное увеличение производительности ЭВМ, которое обеспечит ту же величину суммарного штрафа за ожидание при простейшей q-й дисциплине обслуживания, что и при s-й приоритетной дисциплине:

$$Э^{(s,\,q)} = \frac{\tau_i^{(s)}}{\tau_i^{(q)}}, \tag{11.6}$$

где $Э^{(s,\,q)}$ — относительный критерий эффективности по производительности;

$\tau_i^{(s)}$ — среднее время решения всех задач при s-й дисциплине диспетчеризации;

$\tau_i^{(q)}$ — среднее время решения всех задач при простейшей дисциплине q.

При учете всех возможных вариантов получаются весьма сложные формулы, которые не всегда позволяют построить целесообразные алгоритмы расчетов.

Оговорим некоторые условия. Имеется m разных типов подпрограмм. Заявки на решение задач по этим подпрограммам образуют m простейших потоков с интенсивностями λ_i. Пусть $\tau_{Oi}^{(s)}$ — среднее время решения задачи i-го типа, а ν — коэффициент вариации, т. е. отношение среднего квадратического отклонения к математическому ожиданию. Примем, что простейшей дисциплиной будет обслуживание в порядке поступления задач (заявок). Эффекты от применения дисциплины обслуживания сопоставляются с учетом $\tau_i^{(s)}$ — полного времени пребывания заявок в системе массового обслуживания. Штрафы не зависят от этапа ожидания, т. е. $\alpha_{\text{ож}\, i} = \alpha_{\Pi i} = \alpha_{Oi}$. Время ожидания не ограничено. По условию стационарности процесса функционирования ЭВМ (составляющие потоки и общий поток — простейшие) определится

$$\gamma = \sum_{i=1}^{m} \gamma_i = \sum_{i=1}^{m} \lambda_i \tau_i < 1, \tag{11.7}$$

где γ — общая загрузка системы;

γ_i — загрузка системы i-м потоком заявок.

При условии (11.7) можно определить среднее время ожидания в очереди [52], [58]:

$$\tau_{\text{ож}}^{(q)} = \frac{\frac{1}{2}\sum_{i=1}^{m} \gamma_i^{(q)} \tau_i^{(q)} (1+\nu_i^2)}{1-\gamma^{(q)}}. \tag{11.8}$$

Величина штрафа определяется так [58]:

$$C_\tau^{(q)}(\tau_i^{(s)}) = \sum_{k=1}^{m} \alpha_k \lambda_k \left[\frac{1}{1-\gamma^{(q)}} \cdot \frac{1}{2}\sum_{i=1}^{m} \gamma_i^{(q)} \tau_i^{(q)} (1+\nu_i^2) + \tau_k^{(q)}\right], \tag{11.9}$$

где $\tau_k^{(q)}$ — среднее время обслуживания заявки k-го типа при простейшей q-й дисциплине (без приоритетов).

Вместо $\tau_i^{(q)}$ подставим значение $\frac{\tau_i^{(s)}}{Э^{(s)}}$, вместо γ_i возьмем $\frac{\lambda_i \tau_i^{(s)}}{Э^{(s)}}$ и получим:

$$C_\tau^{(s)}(\tau_i^{(s)}) = \sum_{k=1}^{m} \alpha_k \lambda_k \left[\frac{\frac{1}{2}\sum_{i=1}^{m} \frac{\gamma_i^{(s)} \tau_i^{(s)}}{Э^{(s)^2}}(1+\nu_i^2)}{1-\frac{\gamma^{(s)}}{Э^{(s)}}} + \frac{\tau_k^{(s)}}{Э^{(s)}}\right]. \tag{11.10}$$

Решение в явном виде для систем с приоритетами при неограниченном ожидании по эквивалентному изменению производительности рассчитывается по формуле [58]:

$$Э_\tau^{(s)} = \frac{\gamma^{(s)}+\Delta}{2} + \sqrt{\frac{(\gamma^{(s)}+\Delta)^2}{4} + \frac{C_\tau^{(q)}}{C_\tau^{(s)}}(1-\gamma^{(s)}) - \Delta}, \tag{11.11}$$

где

$$\Delta = \frac{\sum_{k=1}^{m} \alpha_k \lambda_k \tau_k^{(s)}}{C_\tau^{(s)}(\tau_i^{(s)})}. \tag{11.12}$$

Значение относительного критерия эффективности по производительности рассчитывается по разработанному алгоритму на ЭВМ [57], [58].

Если же учесть только пребывание заявки в очереди до начала обслуживания $\tau_{\text{ож}\, i}$, т. е. при $\Delta = 0$, то получим:

$$Э_\tau^{(s)} = \frac{\gamma}{2} + \sqrt{\frac{\gamma^2}{4} + \frac{C_\tau^{(q)}}{C_\tau^{(s)}}(1-\gamma)}. \tag{11.13}$$

Перед корнем всегда берется «+», и, естественно, выигрыш по производительности появится при $C_\tau^{(q)} > C_\tau^{(s)}$.

Пример 11.2

Определим эффективность дисциплины по эквивалентному изменению производительности по сравнению с бесприоритетной дисциплиной при отсутствии прерывания в обслуживании, если же обслуживание начато, то оно ведется до конца, несмотря на приход приоритетной заявки (относительный приоритет).

Имеются две системы обслуживания: q — «первый пришел — первым обслужен» и s — в первую очередь обслуживаются задачи корректировки технологического процесса (тип 3), затем недельного планирования (тип 2) и, наконец, перспективного планирования (тип 1), куда относятся задачи планирования на срок более одной недели.

Возьмем статистические данные примера 11.1, но не будем учитывать прерывание. Оставим среднее время ожидания до начала обслуживания и время «чистого» обслуживания.

Определим эффективность дисциплины по эквивалентному изменению производительности с учетом пребывания заявок в очереди до начала обслуживания.

Представим исходные статистические данные после обработки в табл. 11.2.

Решение

1. Определим общую загрузку системы:

$$\gamma = \gamma^{(s)} = \gamma^{(q)} = \sum_{i=1}^{3} \lambda_i \tau_{O\,i} = 0{,}01 \cdot 5 + 0{,}1 \cdot 1 + 5 \cdot 0{,}1 = 0{,}65.$$

2. Рассчитаем штраф в бесприоритетной системе по формуле (11.4):

$$C_\tau^{(q)} = 1 \cdot 0{,}01 \cdot 0{,}4 + 2{,}5 \cdot 0{,}1 \cdot 0{,}2 + 50 \cdot 5 \cdot 0{,}2 = 50{,}054 \text{ руб.}$$

Таблица 11.2

Результаты обработки статистических данных

Тип задачи	Приоритетная система				Система без приоритетов			
	α_i, руб.	λ_i, 1/ч	$\tau^{(s)}_{\text{ож}\,i}$, ч	$\tau_{O\,i}$, ч	α_i, руб.	λ_i, 1/ч	$\tau^{(s)}_{\text{ож}\,i}$, ч	$\tau_{O\,i}$, ч
1	1,0	0,01	0,60	5,0	1,0	0,01	0,4	5,0
2	2,5	0,10	0,30	1,0	2,5	0,10	0,2	1,0
3	50,0	0,50	0,01	0,1	50,0	5,00	0,2	0,1

3. Штраф в системе с установленными приоритетами по формуле (11.4) составил:

$$C_\tau^{(s)} = 1 \cdot 0{,}01 \cdot 0{,}6 + 2{,}5 \cdot 0{,}1 \cdot 0{,}3 + 50 \cdot 5 \cdot 0{,}01 = 2{,}581 \text{ руб.}$$

4. Определим эффективность дисциплины диспетчеризации по сравнению с бесприоритетной системой по формуле (11.13):

$$Э_\tau^{(s)} = \frac{0{,}65}{2} + \sqrt{\frac{0{,}65^2}{4} + \frac{50{,}054}{2{,}581}(1 - 0{,}65)} \approx 2{,}95.$$

Следовательно, введение приоритетов обслуживания задач на ВЦ эквивалентно почти трехкратному выигрышу по производительности.

11.3. Квантование обслуживания при неограниченной буферной памяти

Квантование времени обслуживания

При реализации алгоритмов полностью стохастических задач, когда априорно неизвестно время обслуживания заявок, можно учесть среднее время обслуживания всей совокупности программ управляющей системы. Можно планировать использование простейших приоритетных дисциплин ($s = 1$): «первым пришел — первым обслужен», «последним пришел — первым обслужен», дисциплину со случайным выбором заявок из очереди.

В первую очередь бывает более выгодно получить решение (окончить обслуживание) задач, которые требуют малого времени работы ЭВМ. Подобный режим применяется в системах с разделением времени.

Если отсутствует информация о необходимом времени решения задач, которые поступают на вычислительный центр в случайные моменты времени, то рациональным бывает решение задач по частям, в течение небольших интервалов времени, называемых **квантами**. При данном подходе можно существенно сократить время ожидания заявок, требующих малой длительности обслуживания (реализации). Задачи, требующие длительного времени обслуживания, будут получать вычислительные ресурсы по частям и их обслуживание задержится.

Различают два основных типа систем обслуживания с квантованием по времени:

- циклическая дисциплина обслуживания;
- многоуровневая дисциплина обслуживания.

При *циклической дисциплине обслуживания* к процессору имеется одна очередь. Заявки, приходящие случайным образом, поступают в конец очереди, которая считается неограниченной. Каждой заявке отводится квант времени обслуживания Q. Заявка, обслуженная за $t \leqslant Q$, уходит из системы, а система приступает к обслуживанию следующей заявки, если в очереди есть хотя бы одна заявка. Если заявка оказалась за квант времени Q не обслуженной полностью, то ее обслуживание прерывается и она становится в конец очереди. Когда она повторно подойдет к процессору, ее обслуживание начнется с этапа, на котором оно было прервано. Время, ранее затраченное на обслуживание, не теряется, и обработка заявки с самого начала не повторяется. В крайнем случае при $Q \to \infty$ циклическая дисциплина вырождается в простейшую.

Многоуровневая дисциплина обслуживания является логическим продолжением циклической дисциплины обслуживания, но число неограниченных очередей заявок $N \geqslant 2$. Каждая очередь имеет свой номер. Номер очереди соответствует определенному приоритету. Высший приоритет принадлежит наименьшему номеру.

Пусть новая заявка поступает в конец очереди с наименьшим номером, т. е. с наивысшим приоритетом. Когда она поступит в процессор, то будет обслуживаться в течение кванта времени. Если она оказалась не полностью обслуженной, то она пойдет в конец очереди с бóльшим на единицу номером и поступит на дальнейшее обслуживание только тогда, когда дойдет ее очередь и не будет ни одной заявки в другой очереди с более высоким приоритетом.

Внутри данной очереди всегда соблюдается порядок «первым пришел — первым обслужен». Последняя очередь с номером N имеет особенность: все ее заявки обрабатываются до конца, т. е. она обладает квантом неограниченной длительности. Для каждой очереди, вообще говоря, может быть свой квант обслуживания.

Бесприоритетное обслуживание «первым пришел — первым обслужен»

Пусть на входе имеется простейший поток заявок с интенсивностью λ и загрузкой $\gamma = \frac{\lambda}{\mu} < 1$. Среднее значение времени ожидания завершения начатого обслуживания одной заявки равно:

$$\tau_З = \frac{\frac{1}{2}\lambda\tau_О^2(1+\nu^2)}{1-\gamma}, \qquad (11.14)$$

где $\tau_О$ — среднее время обслуживания заявок, $\tau_О = \frac{1}{\mu}$,
μ — параметр потока обслуживания;
γ — коэффициент загрузки;
ν — коэффициент вариации времени обслуживания.

В формуле (11.14) $\tau_О$ и ν берутся в квадрате, т.е. $\tau_О^2$ и ν^2.

$$\nu = \frac{\sigma(\tau_0)}{1/\mu} = \frac{\sigma(\tau_0)}{\tau_0}, \qquad (11.15)$$

где $\sigma(\tau_0)$ — среднее квадратическое отклонение времени обслуживания.

Воспользуется формулой Полячека—Хинчина [20], [52] для определения средней длины очереди $M_{ож}$ и среднего числа заявок в системе M:

$$M_{ож} = \frac{\gamma^2 + \lambda^2\sigma^2(\tau_0)}{2(1-\gamma)}; \qquad (11.16)$$

$$M = \gamma + \frac{\gamma^2 + \lambda^2\sigma^2(\tau_0)}{2(1-\gamma)}. \qquad (11.17)$$

Среднее время ожидания при произвольном законе распределения времени обслуживания можно выразить через время ожидания $\tau_{пост}$ при постоянном времени обслуживания и его коэффициент вариации:

$$\tau_{ож} = \tau_{пост}(1+\nu^2). \qquad (11.18)$$

Формулой Полячека—Хинчина можно воспользоваться для определения среднего времени пребывания заявки в очереди $\tau_{ож}$ и среднего времени пребывания заявки в системе массового обслуживания τ при простейшем потоке заявок с интенсивностью λ и произвольным распределением времени

19—1447

обслуживания с математическим ожиданием $\tau_0 = \frac{1}{\mu}$ и коэффициентом вариации v:

$$\tau_{ож} = \frac{\gamma^2 + \lambda^2\sigma^2(\tau_0)}{2\lambda(1-\gamma)}; \qquad (11.19)$$

$$\tau = \frac{\gamma}{\lambda} + \frac{\gamma^2 + \lambda^2\sigma^2(\tau_0)}{2\lambda(1-\gamma)} = \frac{M}{\lambda}. \qquad (11.20)$$

Как было показано в работе [52], простое выражение для времени ожидания заявок в очереди будет получено при экспоненциальном законе распределения времени обслуживания:

$$P(\tau) = 1 - \gamma e^{-(\mu - \lambda)\tau}. \qquad (11.21)$$

Распределение количества заявок в системе согласно формуле (10.59)

$$P_k = (1-\gamma)\gamma^k. \qquad (11.22)$$

Это выражение называется **геометрическим законом.**

При многомерных простейших потоках с интенсивностью λ_i суммарный поток будет пуассоновским. Его интенсивность

$$\Lambda = \sum_{i=1}^{m} \lambda_i. \qquad (11.23)$$

При произвольных случайных потоках суммарный поток приближается к простейшему. В практических задачах обычно пользуются суммарным простейшим потоком и получают характеристики с некоторым приближением.

В первую очередь нас интересует продолжительность ожидания заявок в очереди при многомерном потоке. На нее наибольшее влияние оказывают различия во времени обслуживания заявок разных типов, которые характеризуются дисперсией распределения времени обслуживания суммарного потока. Для повышения эффективности обслуживания при бесприоритетной дисциплине желательно так подобрать алгоритмы, чтобы длительности обслуживания разных заявок были более или менее близкими. Рассмотрим практическую задачу.

Пример 11.3

Имеется набор программ для решения 20 задач. Интенсивности поступления потоков задач и время работы программ указаны в табл. 11.3.

Таблица 11.3

Характеристики потоков заявок и процесса обслуживания

№ задачи	λ_i, 1/ч	τ_{0i}, ч	№ задачи	λ_i, 1/ч	τ_{0i}, ч
1	0,11	1,0	11	0,02	2,5
2	0,05	1,1	12	0,03	2,6
3	0,09	1,2	13	0,01	2,7
4	0,20	1,3	14	0,05	2,8
5	0,06	1,6	15	0,02	2,9
6	0,02	1,7	16	0,03	3,1
7	0,01	1,8	17	0,03	3,2
8	0,01	1,9	18	0,02	3,6
9	0,03	2,0	19	0,03	3,8
10	0,02	2,1	20	0,02	3,9

Определим пропускную способность системы массового обслуживания для решения 20 представленных задач. Оценим, какая очередь появится в этой системе — ограниченная или неограниченная.

Если заданная система не в состоянии обслужить указанный поток заявок, то ее целесообразно разбить на две системы. В первую систему включим 10 задач (программ решения этих задач), во вторую — оставшиеся 10 задач. По полученным системам определим среднюю длину очереди, среднее время ожидания в очереди и среднее время пребывания заявок в системе.

Решение

1. Определим пропускную способность заданной системы:

$$\gamma_0 = \sum_{i=1}^{20} \lambda_i \tau_{0i} = 0{,}11 \cdot 1 + 0{,}05 \cdot 1{,}1 + \ldots + 0{,}03 \cdot 3{,}8 + 0{,}02 \cdot 3{,}9 = 1{,}608.$$

Величина $\gamma_0 > 1$, следовательно, система не в состоянии обслужить поток в 20 задач. В ней будет образовываться возрастающая неограниченная очередь.

Таблица 11.4

Схема расчета

№ задачи	τ_{0i}, ч	$\tau_{0i} - m_{\tau_{01}}$	$(\tau_{0i} - m_{\tau_{01}})^2$	№ задачи	τ_{0i}, ч	$\tau_{0i} - m_{\tau_{02}}$	$(\tau_{0i} - m_{\tau_{02}})^2$
1	1,0	−0,57	0,3249	11	2,5	−0,61	0,3721
2	1,1	−0,47	0,2209	12	2,6	−0,51	0,2601
...	...	...	...	...	...	...	...
9	2,0	0,43	0,1849	19	3,8	0,69	0,4761
10	2,1	0,53	0,2809	20	3,9	0,79	0,6241

2. Распределим задачи по предложенным в условии двум системам и определим их пропускные способности:

$$\gamma_1 = 0{,}11 \cdot 1 + 0{,}5 \cdot 1{,}1 + ... + 0{,}03 \cdot 2{,}0 + 0{,}02 \cdot 2{,}1 = 0{,}802;$$

$$\gamma_2 = 0{,}02 \cdot 2{,}5 + 0{,}03 \cdot 2{,}6 + ... + 0{,}03 \cdot 3{,}8 + 0{,}02 \cdot 3{,}9 = 0{,}806.$$

Приблизительное равенство пропускных способностей двух систем подтверждает, что с этой точки зрения деление проведено удачно. Кроме того, $\gamma_1 < 1$, $\gamma_2 < 1$, следовательно, очереди будут конечной длины.

3. Интенсивность суммарных входящих потоков новых систем по формуле (11.23), 1/ч составит:

$$\Lambda_1 = 0{,}11 + 0{,}05 + ... + 0{,}03 + 0{,}02 = 0{,}60;$$

$$\Lambda_2 = 0{,}02 + 0{,}03 + ... + 0{,}03 + 0{,}02 = 0{,}26.$$

4. Рассчитаем дисперсии каждой системы. Расчет проведем в табл. 11.4

Отсюда

$$m_{\tau_{01}} = 1{,}57; \; \sigma^2_{\tau_{01}} = 0{,}156;$$

$$m_{\tau_{02}} = 3{,}11; \; \sigma^2_{\tau_{02}} = 0{,}254.$$

5. Найдем среднюю длину очереди до начала обслуживания каждой системы по формуле (11.16):

$$M_{\text{ож }1} = \frac{0{,}802^2 + 0{,}60^2 \cdot 0{,}156}{2(1 - 0{,}802)} \approx 1{,}77;$$

$$M_{\text{ож }1} = \frac{0{,}806^2 + 0{,}26^2 \cdot 0{,}254}{2(1 - 0{,}806)} \approx 1{,}72.$$

6. Определим среднее время пребывания заявки в системе и среднее время ожидания ее в очереди, ч:

$$\tau_{\text{ож }1} = \frac{M_{\text{ож }1}}{\Lambda_1} = \frac{1{,}77}{0{,}60} = 2{,}95;$$

$$\tau_{\text{ож }2} = \frac{M_{\text{ож }2}}{\Lambda_2} = \frac{1{,}72}{0{,}26} \approx 6{,}62.$$

Среднее время пребывания в системах можно определить, используя формулу Полячека—Хинчина:

$$\tau_{C_i} = \frac{M}{\Lambda} = \left[\gamma + \frac{\gamma^2 + \Lambda^2 \sigma^2(\tau_0)}{2(1 - \gamma)}\right] \Lambda^{-1};$$

$$\tau_{C_1} = \frac{M_{\text{ож }1}}{\Lambda_1} + \frac{\gamma_1}{\Lambda_1} = 2{,}95 + \frac{0{,}802}{0{,}60} \approx 4{,}29 \text{ ч};$$

$$\tau_{C_2} = \frac{M_{\text{ож }2}}{\Lambda_2} + \frac{\gamma_1}{\Lambda_1} = 6{,}62 + \frac{0{,}806}{0{,}26} = 9{,}72 \text{ ч}.$$

Необходимо заметить, что расчеты по формулам для определения условного среднего времени ожидания в очереди при учете времени переключения достаточно сложны. По этой причине они проводятся по разработанным алгоритмам и программам на ЭВМ [57], [58]. Оценка эффективности на основе критериев теории массового обслуживания позволяет выработать рекомендации по существенному улучшению работы ЭВМ и их эксплуатации на ВЦ.

Глава 12

ОЦЕНКА ЭФФЕКТИВНОСТИ ФУНКЦИОНИРОВАНИЯ ТЕХНИЧЕСКИХ СРЕДСТВ СИСТЕМ УПРАВЛЕНИЯ В ЗАВИСИМОСТИ ОТ УРОВНЯ ОБУЧЕННОСТИ ОБСЛУЖИВАЮЩЕГО ПЕРСОНАЛА

12.1. Постановка вопроса

Современные СУ являются большими системами, в которых значительную роль играет человек — руководитель, оператор, техник или инженер, обслуживающий технические средства СУ [16]. В СУ, управляющих технологическими процессами, неправильная работа оператора может привести к аварии или катастрофе. В СУ, управляющих отраслью, принятие решений, выработанных специалистами, может оказать существенное воздействие на эффективность развития и функционирования отрасли.

Технические средства СУ включают линии связи, ВЦ, узлы связи, коммутационные блоки, пульты управления и ряд других устройств. От надежности функционирования технических средств во многом зависит эффективность СУ.

Вмешательство человека, с одной стороны, способствует повышению надежности сложной системы. С другой стороны, сам человек может внести неисправности, допустить ошибки. Итак, технические средства и человека можно формально рассматривать как последовательно соединенные элементы единой системы. При таком подходе не учитывается, что человек, включаясь в работу системы, преобразует ее, вносит элемент разумной деятельности, превращает технические средства в орудия для обеспечения высокой эффективности своей деятельности.

Какие же требования надо предъявить к этим двум последовательно соединенным элементам, чтобы человеко-машинная (сложная) система, представляемая СУ, за определенный промежуток времени выполнила поставленную перед ней задачу с наиболее высокой эффективностью?

Человек в этот промежуток времени должен выполнять свои функции без ошибок, технические средства СУ должны

функционировать безотказно. При таком положении будет выполнена поставленная задача функционирования СУ.

Для обеспечения функционирования СУ человек должен освоить технику, знать технологию правильной и безопасной работы, изучить ориентиры и связи, приводящие к аварийным ситуациям. В случае отказа, аварийной ситуации он должен суметь устранить неисправность и возвратить систему в рабочее состояние, обеспечивающее безопасную работу.

Чтобы человек (оператор, руководитель и т. д.) справлялся со всеми этими сложными задачами, необходимо предварительно ввести систему обучения специалиста.

Рассмотрим влияние только обучения на эффективность работы специалистов. Как отмечает крупнейший представитель советской психологической школы по вопросам обучения специалистов технических систем академик А. Н. Леонтьев, при этом необходимо решить две проблемы:

- приспособления технических средств СУ к способностям и возможностям человека (эта проблема решается при проектировании СУ);
- приспособления человека к работе в спроектированной реальной системе.

Нас интересует вторая проблема.

Человек должен правильно выполнять требуемые действия и обладать необходимой ориентировочной основой действий [99], [100].

Ориентировочная основа действий — это система условий, на которую человек опирается, выполняя действия. В каждом действии человека есть ориентировочная, исполнительная и контрольная часть. В зависимости от типа ориентировочной основы будет различаться качество действия.

Теоретически можно представить восемь различных типов ориентировочной основы действия [99], [100]. Остановимся на характеристике первых трех основных типов по данной классификации.

Первый тип («проб и ошибок») характеризуется неполным составом ориентиров. Ориентиры представлены в частном виде и выделяются самим субъектом путем слепых проб. Процесс формирования действия идет очень медленно и с большим числом ошибок. Сформулированное действие оказывается чувствительным к малейшим изменениям условий выполнения.

Второй тип характеризуется наличием всех условий, необходимых для правильного выполнения действия. Эти условия даются в готовом виде, пригодном для ориентировки только в данном случае. Формирование действия идет быстро и безошибочно. Сформированное действие более устойчиво, чем при первом типе. Сфера переноса действия ограничена сходством конкретных условий.

Третий тип ориентировочной основы характеризуется полным составом ориентиров, представленных в обобщенном виде. Этот тип присущ целому классу явлений.

В каждом конкретном случае ориентировочная основа действия составляется субъектом (обучаемым) самостоятельно с помощью общего метода, который ему выдается. Действие формируется быстро, проводится безошибочно. Обеспечивается большая устойчивость процесса переноса действия на другие задачи данного класса.

На практике наиболее широкое распространение получил второй тип ориентировочной основы действия, несмотря на то, что третий тип превосходит его по эффекту. Это объясняется тем, что для работы по третьему типу надо быть хорошо подготовленным специалистом и самому составлять инструкции по работе с техническими средствами. При втором типе надо научиться работать по инструкции и описанию, которые выдаются обучаемому.

Таким образом, для того чтобы исполнителем (обучаемым) не была допущена ошибка, ему следует до начала работы знать и научиться выделять ориентировочную, исполнительную и контрольную часть действия:

- необходимые и достаточные условия, на которые исполнитель должен ориентироваться до начала действия;
- необходимые и достаточные условия, обеспечивающие нормальный ход действия;
- признаки, характеризующие законченность действия;
- признаки, обеспечивающие контроль в ходе выполнения действия.

Исполнитель должен уметь выполнять трудовой прием.

Итак, существует множество объективных условий каждого действия ($L_{\text{ОУД}}$). Если исполнитель ориентируется на $L_{\text{ОУД}}$, то он может правильно выполнить действие. Исполнитель же фактически ориентируется на множество условий действия,

которое стало его достоянием. Это множество условий действия составляет ориентировочную основу действия и обозначается $L_{\text{ООД}}$. Отсюда условие безошибочной работы

$$L_{\text{ОУД}} = L_{\text{ООД}}. \tag{12.1}$$

Причина ошибок отразится в

$$L_{\text{ОУД}} \neq L_{\text{ООД}}. \tag{12.2}$$

До начала процесса обучения необходимо создать пособия, подготовить преподавателей, которые дадут возможность изучить СУ и работать с ее техническими средствами.

Между $L_{\text{ОУД}}$ и $L_{\text{ООД}}$ должно существовать промежуточное множество выделенных условий действия $L_{\text{ВУД}}$.

Принцип обучения на полной ориентировочной основе может быть формально записан так:

$$L_{\text{ОУД}} = L_{\text{ВУД}} = L_{\text{ООД}}. \tag{12.3}$$

При идеальном процессе обучения добиваемся равенства:

$$L_{\text{ВУД}} = L_{\text{ООД}}. \tag{12.4}$$

При идеальной организации учебного процесса наблюдается равенство

$$L_{\text{ВУД}} = L_{\text{ОУД}}. \tag{12.5}$$

Состав информации, подлежащей изучению обучаемым, существенно зависит от уровня развертывания ориентировочной основы действия в документации. Можно излагать сведения об объекте и его функционировании в общем виде или с подробностью на уровне блока, подсистемы или элемента блока или прибора.

12.2. Уровни обученности и их влияние на критерий надежности функционирования

При работе сложных систем типа СУ возможны аварийные ситуации и отказы. В практике эксплуатации принято не устранять неисправности внутри блока или прибора на месте, а заменять блок или прибор на исправный целиком.

В зависимости от подготовленности специалистов при такой замене различают три уровня обученности специалистов:

- исполнительский;
- функциональный;
- конструктивный.

На каждом из этих уровней в технической документации и в инструкциях по работе обеспечивается полная ориентировочная основа действия, позволяющая работать безошибочно и безопасно.

Исполнительский уровень обученности

Требования при обучении на этом уровне заключаются в формировании полной ориентировочной основы действия обучаемого только:

- по изменению состояния технической системы (считается, что техническая система перешла в новое состояние, если хотя бы один ее элемент перешел в другое состояние);
- по контролю ее состояния;
- по переходу от операции к операции.

Целью каждого действия для обучаемого на первом уровне является получение определенных, предусмотренных технической документацией показаний на элементах орудий контроля. Исполнитель ориентируется только на признаки состояния элементов орудий управления и орудий контроля. Он не знает принципов действия, технической системы, а также какими конструктивными элементами производятся переходы в новые состояния управляемой и управляющей систем.

Исполнители, подготовленные на первом уровне обученности, работая безошибочно, обеспечивают надежность работы системы «человек — техника» в лучшем случае не выше критерия надежности технических устройств:

$$P_С(t)_I \leqslant P_{ТУ}(t), \qquad (12.6)$$

где $P_С(t)_I$ — вероятность безотказной работы системы, обслуживающий персонал которой обучен на первом уровне;
$P_{ТУ}(t)$ — вероятность безотказной работы технических устройств СУ.

По формуле (12.6) можно оценить верхнюю границу значений $P_С(t)_I$, так как $P_{ТУ}(t)$ — известная величина.

Однако надо учитывать, что в случае отказа исполнитель не в состоянии отыскать неисправность и возвратить систему в исходное состояние, если эти действия не описаны в инструкции. Перенос ориентировочной основы действий на другие работы весьма ограничен.

Преимуществом исполнительского (первого) уровня обученности является малый объем информации, который требуется усвоить, малое время обучения, дешевизна организации учебного процесса, отсутствие необходимости создания новой технической и рабочей документации, отличной от представленной конструктором по существующим ГОСТам, для эксплуатации технических средств СУ.

Функциональный (технологический) уровень обученности

Целью каждого действия для обучаемого на этом уровне является изменение технологического состояния блоков (подсистем), приборов автоматизированной системы управления и управляемого объекта. Вспомним, что на исполнительском уровне целью действия является получение определенных показаний элементов орудий контроля. Теперь эти показания лишь характеризуют функциональное состояние определенного блока системы.

При обучении на этом уровне у исполнителя формируется полная ориентировочная основа действия, раскрываемая до уровня конструктивного блока, прибора, подсистемы, которые могут быть целиком заменены на исправные блоки в случае их отказа.

При этом должны быть указаны:

- параметр, который изменяется в результате действия и характеризует исходное и конечное состояние объекта управления;
- конструктивные элементы орудий управления и контроля и их особенности;
- элементы орудий управления и контроля, находящиеся в цепи прохождения управляющего сигнала от технических средств СУ до управляемого объекта при данном действии.

Исполнитель на данном уровне обученности представляет принцип действия и сущность процессов, протекающих в технической системе, сам дифференцирует признаки, характеризующие условия действия, по их существенности и может наиболее целесообразно построить свою работу. Каждая связь между действиями и признаками, определяющими условия их проведения, формируется как специфическая, поэтому ориентировочная основа действия лучше и длительнее сохраняется в памяти обучаемого.

Расширяются возможности переноса ориентировочной основы действия на другие работы, и обеспечивается безошибочная и безопасная работа.

Исполнитель способен быстро среагировать на аварийную ситуацию и восстановить систему до рабочего состояния. Он способен на уровне до блока, узла, прибора осуществить поиск и обнаружить неисправности, заменить неисправный прибор или узел на исправный из ЗИПа[1]. При данном уровне обученности конструктору не требуется перерабатывать всю существующую эксплуатационную документацию.

Основной недостаток второго уровня — наличие неполной ориентировочной основы действия в существующей конструкторской и эксплуатационной документации. Она пополняется в процессе обучения преподавателем. Однако надо иметь в виду, что при переиздании или разработке новой документации целесообразно отражать все необходимые и достаточные признаки действия на полной ориентировочной основе.

Надо отметить, что исполнитель не умеет производить поиск и идентификацию неисправности на уровне элементов внутри узла, блока, прибора. Для овладения работой на данном уровне надо изучить конструкцию, функциональные блочные схемы системы, принципы работы технических устройств на уровне узла, блока, прибора, научиться рабочим приемам.

Функция вероятности безотказной работы системы на втором уровне обученности может быть выражена формулой:

$$P_{\text{C}}(t)_{\text{II}} \leqslant P_{\text{ТУ}}(t) + [1 - P_{\text{ТУ}}(t)]P_{\text{B}}(t)_{\text{З}}, \qquad (12.7)$$

где $P_{\text{B}}(t)_{\text{З}}$ — вероятность замены исполнителем прибора, блока, узла, вышедшего из строя, на исправный из ЗИПа за время t.

[1]ЗИП — сокращенное обозначение для набора запасных инструментов, приборов и материалов, придаваемое заводом-поставщиком либо каждому изделию, либо ремонтному органу на группу изделий.

Можно отметить, что

$$P_{\text{C}}(t)_{\text{II}} \geqslant P_{\text{C}}(t)_{\text{I}}, \qquad (12.8)$$

где $P_{\text{C}}(t)_{\text{II}}$ — можно трактовать как вероятность безотказной работы технических средств СУ, обслуживающий персонал которых подготовлен на функциональном уровне.

Конструктивный уровень обученности

Этот уровень предполагает наличие полной ориентировочной основы действий за счет установления связей между конструктивными элементами технической системы внутри узла, блока, прибора. Умение проводить действия сформировано у исполнителя с ориентировкой на принципиальные электрические, кинематические и другие схемы, на процессы, протекающие в технической системе при каждом действии по изменению ее состояния. При этом уровне обученности исполнитель способен устранить неисправность на уровне элемента внутри прибора или другого конструктивно делимого узла системы. Уровень надежности системы, обслуживаемой такими исполнителями, может быть выражен функцией вероятности безотказной работы:

$$P_{\text{C}}(t)_{\text{III}} \leqslant P_{\text{ТУ}}(t) + [1 - P_{\text{ТУ}}(t)]P_{\text{B}}(t)_{\text{З}} + \\ + \{1 - P_{\text{ТУ}}(t) - [1 - P_{\text{ТУ}}(t)]P_{\text{B}}(t)_{\text{З}}\}P_{\text{B}}(t)_{\text{Э}}, \qquad (12.9)$$

где $P_{\text{C}}(t)_{\text{III}}$ — вероятность безотказной работы системы, если обслуживающий персонал обучен на конструктивном уровне;

$P_{\text{B}}(t)_{\text{Э}}$ — вероятность восстановления обслуживающим персоналом элементов, вышедших из строя за время t, в узлах, блоках, приборах, не входящих в состав ЗИПа.

Существенным недостатком данного уровня является необходимость при обучении осваивать в десятки, сотни раз больший объем информации, чем на предшествующем уровне. Необходимо также перерабатывать издаваемую конструктором эксплуатационную документацию. Однако без некоторого числа специалистов, обученных на данном уровне, пока не удается эксплуатировать большие системы, к которым относят и СУ.

Таким образом, первый уровень обученности допустим в очень редких случаях, второй уровень используется в подавляющем числе случаев, третий — необходим для ограниченной группы специалистов, обслуживающих технические средства СУ, для ремонтников-наладчиков.

Обследование обслуживающего персонала технических средств СУ показало, что для их успешного функционирования необходимо, чтобы основной состав был обучен на втором уровне. Для повышения эффективности работы СУ следует включать в каждую рабочую смену одного-двух специалистов, подготовленных на конструктивном уровне обученности.

Целесообразно оценить эффективность (по критерию надежности) при переходе от одного уровня обученности к другому.

12.3. Оценка эффективности систем управления по критерию надежности с учетом обученности обслуживающего персонала

Сравнивая формулы (12.6), (12.7) и (12.9), можно записать:

$$P_C(t)_I \leq P_C(t)_{II} \leq P_C(t)_{III}. \qquad (12.10)$$

Для оценки эффективности перехода от одного уровня обученности к другому введем следующие обозначения:

Φ_i^j — эффективность перехода от уровня i к уровню j ($j > i$; $i = 1, 2, 3$; $j = 2, 3$);

$\Delta\Phi_i^j$ — выигрыш в эффективности при переходе от уровня i к уровню j.

Определим критерии эффективности при переходе от первого уровня обученности ко второму уровню. При этом, учитывая лишь верхнюю границу надежности, запишем:

$$\Phi_I^{II} = \frac{P_C(t)_{II}}{P_C(t)_I} = \frac{P_{ТУ}(t) + [1 - P_{ТУ}(t)]P_B(t)_З}{P_{ТУ}(t)} =$$

$$= 1 + P_B(t)_З\left(\frac{1}{P_{ТУ}(t)} - 1\right) = 1 + \varepsilon P_B(t)_З, \qquad (12.11)$$

где

$$\varepsilon = \frac{1}{P_{ТУ}(t)} - 1. \qquad (12.12)$$

Определим далее

$$\Delta\Phi_I^{II} = \Phi_I^{II} - \Phi_I^I = \varepsilon P_B(t)_З, \qquad (12.13)$$

так как $\Phi_I^I = 1$.

Таким образом, чем ниже надежность технических средств СУ и выше способность специалистов восстанавливать вышедшие из строя узлы, блоки и приборы путем замены на исправные из ЗИПа, тем эффективнее переход ко второму уровню обученности. При этом под способностью восстанавливать надо понимать не только технику перестановки исправных приборов вместо неисправных, а главным образом способность осуществлять поиск, обнаружение, идентификацию неисправности.

Определим эффективность перехода от первого уровня обученности к третьему:

$$\Phi_I^{III} = \frac{P_C(t)_{III}}{P_C(t)_I} = \frac{1}{P_{ТУ}(t)}\{P_{ТУ}(t) + [1 - P_{ТУ}(t)]P_B(t)_З +$$

$$+ \{1 - P_{ТУ}(t) - [1 - P_{ТУ}(t)]P_B(t)_З P_B(t)_Э\} = 1 + \varepsilon P_B(t)_З + \varepsilon P_B(t)_З -$$

$$- \varepsilon P_B(t)_З P_B(t)_Э = \varepsilon\{P_B(t)_З + P_B(t)_Э[1 - P_B(t)_З]\} + 1; \qquad (12.14)$$

$$\Delta\Phi_I^{III} = \Phi_I^{III} - \Phi_I^I = \varepsilon\{P_B(t)_З + P_B(t)_Э[1 - P_B(t)_З]\}. \qquad (12.15)$$

Из формулы (12.15) видно, что выигрыш в эффективности при переходе с первого уровня обученности на третий уровень будет тем выше, чем ниже надежность технических средств СУ, выше вероятность восстановления приборов и узлов, входящих в состав ЗИПа, элементов, не входящих в состав ЗИПа. Следовательно, специалисты, подготовленные на третьем уровне, должны уметь производить поиск, обнаружение и устранение неисправностей как блоков и узлов ЗИПа, так и конструктивных элементов внутри агрегата, блока и прибора.

Выведем формулу эффективности перехода от второго (функционального) уровня к третьему (конструктивному):

$$\Phi_{II}^{III} = \frac{P_C(t)_{III}}{P_C(t)_{II}} = \{P_{ТУ}(t) + [1 - P_{ТУ}(t)]P_B(t)_З\}^{-1}\{P_{ТУ}(t) +$$

$$+ [1 - P_{ТУ}(t)]P_B(t)_З + \{1 - P_{ТУ}(t) - [1 - P_{ТУ}(t)]P_B(t)_З\} \cdot P_B(t)_Э\} =$$

$$= 1 + \{P_{ТУ}(t) + [1 - P_{ТУ}(t)]P_B(t)_З\}^{-1} \cdot \{[1 - P_{ТУ}(t)]P_B(t)_Э -$$

$$- [1 - P_{ТУ}(t)]P_B(t)_З P_B(t)_Э\} = 1 + \frac{\varepsilon P_B(t)_Э[1 - P_B(t)_З]}{1 + \varepsilon P_B(t)_З}. \qquad (12.16)$$

Выигрыш в эффективности

$$\Delta\Phi_{II}^{III} = \Phi_{II}^{III} - \Phi_{II}^{II} = \frac{\varepsilon P_B(t)_з[1 - P_B(t)_з]}{1 + \varepsilon P_B(t)_з}. \qquad (12.17)$$

Эффективность при переходе на конструктивный уровень с функционального уровня будет тем выше, чем ниже надежность технических средств СУ, выше способность специалистов восстанавливать элементы, не входящие в ЗИП, и выше доля приборов, не восстанавливаемых из запаса.

Пример 12.1

На ВЦ имеются две машины, обслуживающие абонентов отрасли. Для решения поступающих задач необходимо, чтобы обе ЭВМ были в рабочем состоянии и выполняли поступающие задания одновременно. За среднее время решения задач вероятность безотказной работы одной ЭВМ $P(t)$ равна 0,95, вероятность восстановления неисправностей путем замены отказавших на исправные элементы из запаса равна 0,8, вероятность восстановления элементов, не входящих в запас, равна 0,75.

Определим эффективность перехода от первого уровня обученности ко второму, от первого — к третьему и от второго — к третьему уровню.

Решение

1. Запишем условия задачи в принятых обозначениях:

$$N = 2;\ P(t) = 0{,}95;\ P_B(t)_З = 0{,}8;\ P_B(t)_Э = 0{,}75.$$

2. Вычислим вероятность безотказной работы технических устройств:

$$P_{ТУ}(t) = [P(t)]^N = 0{,}95^2 \approx 0{,}90.$$

3. Определим эффективность перехода от первого уровня обученности ко второму уровню по формуле (12.11):

$$\Phi_I^{II} = \frac{P_C(t)_{II}}{P_C(t)_I} = 1 + \left(\frac{1}{0{,}90} - 1\right) \cdot 0{,}8 = 1 + 0{,}0888 \approx 1{,}089.$$

4. Найдем эффективность перехода от первого уровня обученности к третьему уровню по формуле (12.14):

$$\Phi_I^{III} = 1 + \left(\frac{1}{0{,}90} - 1\right)(0{,}8 + 0{,}75 \cdot 0{,}2) \approx 1{,}105.$$

5. Рассчитаем эффективность перехода от второго уровня обученности к третьему уровню по формуле (12.16):

$$\Phi_{II}^{III} = 1 + \frac{0{,}111 \cdot 0{,}75 \cdot 0{,}2}{1 + 0{,}111 \cdot 0{,}8} \approx 1{,}015.$$

Пример 12.2

Воспользуемся условиями и решением примера 12.1 и добавим следующие данные: для перехода ВЦ на третий уровень обученности надо из двенадцати человек троих подготовить на третьем уровне, а всех остальных специалистов подготовить на втором уровне, первоначально все они получили подготовку на первом уровне. На подготовку на втором уровне надо затратить 120 руб. за человека, подготовка на третьем уровне обходится в 1800 руб.

При обученности на первом уровне ВЦ, благодаря решению задач на двух ЭВМ, в среднем получает прибыль 24 тыс. руб. в месяц.

При существующем положении ВЦ в среднем выплачивает бригаде обслуживания ежемесячно 4000 руб., при обученности на втором уровне — 1200 руб., при обученности на третьем уровне — 250 руб. Специалисту, подготовленному на втором уровне, увеличивается заработная плата на 10 руб., а на третьем — на 30 руб. в месяц.

Определим годовой экономический эффект перехода: ко второму уровню обученности; к третьему уровню обученности. Определим сроки окупаемости.

Решение

1. Определим годовой экономический эффект при переучивании всех специалистов с исполнительского уровня на функциональный, руб.:

$$C_1^I = 24\,000 \cdot 12 \cdot 1{,}089 = 313\,632;$$

$$C_1 = 313\,632 - 1200 \cdot 12 - 120 \cdot 12 - 10 \cdot 12 \cdot 12 = 296\,352;$$

$$C_2^I = 24\,000 \cdot 12 = 288\,000;$$

$$C_2 = 288\,000 - 4000 \cdot 12 = 240\,000;$$

$$\Delta C_I^{II} = C_1 - C_2 = 296\,352 - 240\,000 = 56\,352.$$

2. Найдем годовой эффект при переходе к третьему уровню обученности, т. е. при подготовке девяти специалистов на втором уровне, а трех специалистов на третьем уровне, руб.:

$$C_1^{\mathrm{I}} = 24\,000 \cdot 12 \cdot 1{,}05 = 318\,240;$$

$$C_1 = 318\,240 - 250 \cdot 12 - 9 \cdot 120 - 3 \cdot 1800 - 10 \cdot 12 \cdot 9 - 30 \cdot 12 \cdot 3 = 306\,600;$$

$$C_2^{\mathrm{I}} = 24\,000 \cdot 12 = 288\,000;$$

$$C_2 = 288\,000 - 4000 \cdot 12 = 240\,000;$$

$$\Delta C_{\mathrm{I}}^{\mathrm{III}} = C_1 - C_2 = 306\,600 - 240\,000 = 66\,600.$$

3. Определим срок окупаемости дополнительных затрат при переподготовке на второй уровень обученности:

$$T_{\mathrm{I}}^{\mathrm{II}} = \frac{K_{\text{Д}}}{Э_{\text{год}}} = \frac{1200 \cdot 12 + 120 \cdot 12 + 10 \cdot 12 \cdot 12}{56\,352} \approx 0{,}31.$$

4. Срок окупаемости при переподготовке девяти человек на второй уровень, а трех человек на третий уровень составит:

$$T_{\mathrm{I}}^{\mathrm{III}} = \frac{250 \cdot 12 + 9 \cdot 120 + 3 \cdot 1800 + 9 \cdot 10 \cdot 12 + 3 \cdot 30 \cdot 12}{66\,600} \approx 0{,}175.$$

12.4. Оценка эффективности систем управления технологическими процессами в химическом производстве с учетом результатов подготовки и работы специалистов

Обратимся к задаче, которая решалась в интересах повышения эффективности функционирования систем управления технологическими процессами (СУТП), используемой в химическом производстве.

Известно, что некоторые компоненты высококачественных лаков и красок обладают токсичностью, поэтому их производство автоматизировано. Для управления технологическими процессами используются типовые автоматизированные системы управления. Специалистов для работы на подобных системах готовят как техникумы, так и вузы. Обучение будущих

специалистов практической работе проводится до второго или третьего уровня обученности.

Однако, несмотря на высокий уровень подготовки специалистов, принимаемые меры по повышению безопасности работы, совершенствование автоматизированных систем управления и повышение надежности их технических средств, на предприятиях иногда складываются аварийные ситуации.

Анализ их причин показал, что в бóльшей части они происходят по вине обслуживающего персонала. Следовательно, на основе экономического анализа можно сформулировать одну проблему: снижение вероятности появления аварийной ситуации. Эта проблема требует решения двух самостоятельных задач:

- снижение вероятности появления аварийной ситуации по вине операторов СУТП;
- снижение вероятности появления аварийной ситуации из-за отказов технических средств СУ.

Более сложной является первая задача, но ее решение позволяет получить более существенный эффект. Остановимся на ней.

Известно, что успехи в обучении сказываются на работе операторов СУ наиболее существенно в течение первого года после окончания вуза. В дальнейшем это влияние несколько нивелируется. Поэтому в качестве исходной информации о степени подготовки специалистов были использованы оценки по соответствующим изучаемым в вузе предметам. Если оценок по одному предмету было несколько, то учитывалась средняя оценка.

На производстве выпускник вуза или техникума не допускается к работе, пока не сдаст ряда зачетов и экзаменов. При этом время от момента прибытия на работу до допуска к самостоятельной деятельности на рабочем месте является весьма существенной характеристикой — оценкой, выраженной в днях. В дальнейшем специалист периодически оценивается еще по трем показателям:

- работа по специальности;
- умение руководить подчиненными;
- производственная дисциплина.

В рассматриваемой задаче исходные признаки:

1) специальность до поступления в техникум;
2) электротехника;

3) основы ремонта блоков технических средств СУТП;

4) конструкция блоков технических средств СУТП;

5) техника безопасности;

6) электрооборудование технических средств СУТП;

7) эксплуатация оборудования СУТП;

8) организация работы на СУТП;

9) дополнительная подготовка на производстве (время до момента допуска к самостоятельной работе);

10) выполнение функциональных обязанностей на производстве;

11) результаты руководства подчиненными;

12) производственная дисциплина.

По этим признакам была проведена оценка каждого из 77 наблюдаемых специалистов. Полученная матрица исходных данных размерности 77 × 12 была обработана методом главных компонент. Было извлечено семь наиболее значимых главных компонент, которые объясняли 83% общей дисперсии процесса. Остальные пять главных компонент объясняли только 17% общей дисперсии процесса.

Представим доли объясняемой дисперсии первых семи главных компонент (v_r и v_r/n) и суммарный вклад набора главных компонент (γ_Σ), где n — значение общей дисперсии процесса, r — номер главной компоненты.

Из табл. 12.1 видно, что наибольший вклад в общую дисперсию процесса вносит первая главная компонента. Анализ составляющих главной компоненты показал, что она может быть названа «Характеристика способностей к учебе и работе по избранной специальности». Вторая главная компонента была названа «Характеристика способностей к руководству», но ее вклад в общую дисперсию изучаемого процесса составляет только 16%, в то время как первая главная компонента объясняла 25%.

Таким образом, первая главная компонента оказалась весьма важной для определения эффективности обучения человека специальности и выполнения им практической работы на производстве.

Целесообразно выяснить, является ли первая главная компонента основной характеристикой только данной выборки специалистов или общей характеристикой любого учебного процесса. Для ответа на данный вопрос была поставлена серия

Таблица 12.1

Оценки вкладов главных компонент в общую дисперсию

Критерии	Главная компонента						
	f_1	f_2	f_3	f_4	f_5	f_6	f_7
v_r	3,02	1,94	1,34	1,17	0,90	0,88	0,71
v_r/n, %	25,17	16,17	11,17	9,75	7,50	7,33	5,92
γ_Σ, %	25,17	41,34	52,51	62,25	69,76	77,09	83,01

экспериментов в вузах. Приведем оценки вкладов первой и второй главных компонент в общую дисперсию процесса и характеристику $k = V_1/V_2$ (табл. 12.2).

Эксперименты показали, что вес первой главной компоненты колеблется в пределах 33—40%. В других аналогичных экспериментах вклады первой главной компоненты составляли 32—36%. Следовательно, первая главная компонента может быть использована для выделения группы специалистов, способных с наибольшей вероятностью обеспечить безаварийную работу СУ в рассматриваемом химическом производстве.

Такая задача была решена для группы, состоявшей из 106 человек. После года их самостоятельной работы по 20 признакам (16 вузовских оценок и 4 оценки их работы на СУ) были получены индивидуальные оценки первой главной компоненты для каждого специалиста. Затем всю группу из 106 человек разбили на две подгруппы. В первую подгруппу во-

Таблица 12.2

Оценки вкладов главных компонент

Критерии	Число студентов		
	106	56	60
V_1	6,62	7,50	7,90
V_2	1,70	1,60	1,90
k	3,89	4,69	4,16
V_1/n, %	33,10	37,50	39,50

шли специалисты с бóльшими значениями главной компоненты, во вторую подгруппу — с меньшими значениями главной компоненты. Каждая из подгрупп включала конкретные фамилии выпускников. По данным присланных с производства анкет были выявлены лица, получившие в течение рассматриваемого года работы оценку «2» (по пятибалльной системе). Это были выпускники вуза, создавшие условия для аварийной ситуации или допустившие ее. Оказалось, что 70% лиц с оценкой «2» относятся ко второй подгруппе, а 30% — к первой.

Поскольку подгруппа содержит ранжированные значения оценок, ее можно еще раз разделить пополам. После этого обнаружилось, что оставшиеся 30% лиц, допустивших условия аварийной ситуации, попали в «худшую» половину первой подгруппы. Из «лучшей» половины первой подгруппы были отобраны специалисты для выдвижения на наиболее ответственные должности. Наблюдения за ними в течение последующих пяти лет показали, что ни у них, ни в подразделениях, которыми они руководили, не было ни одной аварийной ситуации.

Таким образом, можно полагать, что путем отбора на научной основе операторов СУ существует возможность повысить вероятность безаварийной работы рассмотренной человеко-машинной системы.

* * *

Нами рассмотрены математико-статистические основы современной теории эффективности, решены практические задачи. В книге дан синтез различных подходов к решению общей задачи: оценке эффективности, оценке экономической эффективности на основе действующих методик «прямого счета» и с использованием современных методов прикладной математики. При составлении задач автор старался более четко представить методологию и принципы подхода.

Описанные методы оценки эффективности можно применить при решении задач для разных отраслей, предприятий, их подразделений, несмотря на существенные различия между ними. Это обусловлено наличием общих математических подходов и ограниченного числа классов решаемых задач.

В условиях перехода экономики на рыночные отношения существенно меняются информационные запросы управляющих структур по объему, составу, достоверности и оперативности информации. Принятие обоснованных решений должно опираться на тщательный анализ информации, позволяющий понять закономерности, взаимосвязи, зависимости между различными показателями. Поэтому в настоящее время статистические методы все шире применяются в деятельности плановых, аналитических, маркетинговых отделов производственных предприятий, объединений, торговых и страховых компаний, банков, правительственных учреждений.

Производство, экономика постоянно ставят перед нами новые, более сложные задачи, которые предстоит решать читателю. Было бы полезно познакомиться с работами по использованию многомерного анализа [3], [36], [44], функций случайных аргументов, марковских и полумарковских процессов для оценки эффективности [27], [32], [47], [107].

Для рассмотрения широкого комплекса многомерных статистических методов (кластерного анализа, дискриминантного анализа, канонических корреляций, факторного анализа и др.) в качестве дополнительной литературы можно рекомендовать учебники, изданные за последние несколько лет.

Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики: Учебник. — М.: ЮНИТИ, 1998.

Дубров А. М. и др. Многомерные статистические методы: Учебник / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. — М.: Финансы и статистика, 2000.

Многомерный статистический анализ в экономике: Учеб. пособие / Л. А. Сошникова, В. Н. Тамашевич, Г. Н. Уебе и др.; Под ред. В. Н. Тамашевича. — М.: ЮНИТИ, 1999.

Внедрение многомерных статистических методов в экономическую практику неразрывно связано с использованием современного программного обеспечения: широкопрофильных пакетов прикладных программ (ППП) по статистической обработке данных.

Большую известность получили такие современные статистические системы, как SAS, STATISTICA, SPSS и др. Они создают пользователю среду, в которой обработка данных становится увлекательным исследованием, позволяют получать многовариантные решения с использованием компьютерных технологий и современных методов.

При этом пользователь освобожден от рутинной, трудоемкой работы (по проведению расчетов, построению таблиц, графиков и т. д.), на его долю остается творческая, исследовательская работа по постановке задачи, выбору метода анализа, оценке качества полученных моделей, интерпретации результатов. Поэтому для грамотного применения современных ППП по обработке данных необходима глубокая подготовка в области математико-статистических методов.

В качестве дополнительной литературы, посвященной использованию современных статистических ППП, можно рекомендовать следующие книги:

Боровиков В. П. Программа STATISTICA для студентов и инженеров. — 2-е изд. — М.: КомпьютерПресс, 2001.

Боровиков В. П., Боровиков И. П. STATISTICA — Статистический анализ и обработка данных в среде Windows. — М.: Информационно-издательский дом «Филинъ», 1997.

Боровиков В. П., Ивченко Г. И. Прогнозирование в системе STATISTIKA в среде Windows: Основы теории и интенсивная практика на компьютере: Учеб. пособие. — М.: Финансы и статистика, 2000.

Если у читателя появилось желание продолжить знакомство с подходами общей теории эффективности, многомерными статистическими методами, их практическим приложением, то автор будет считать свою задачу выполненной.

ПРИЛОЖЕНИЯ

Приложение 1

Элементы линейной алгебры, используемые в методе главных компонент

Основные определения

Матрицей называется прямоугольная таблица чисел. Числа таблицы называются ее *элементами*. Они образуют *строки* и *столбцы матрицы*. Индексом i обычно обозначают *номер строки*, а индексом j — *номер столбца*. Матрица, имеющая N строк и n столбцов, называется *матрицей размера* $N \times n$.

Записывается матрица любым из указанных способов:

$$\mathbf{A} = (a_{ij}) = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \cdots & a_{Nn} \end{pmatrix};$$

$$i = 1, 2, ..., N;\ j = 1, 2, ..., n.$$

Если $N \neq n$, то матрица называется *прямоугольной*, если $N = n$, то матрица называется *квадратной*. Диагональ квадратной матрицы, идущая от верхнего левого угла к нижнему правому углу, называется *главной диагональю*. Ее элементы имеют одинаковые индексы i и j. Если все элементы с неодинаковыми индексами равны нулю, то матрица называется *диагональной*:

$$\mathbf{A} = \begin{pmatrix} a_{11} & 0 & \dots & 0 \\ 0 & a_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

Диагональные матрицы, в которых все диагональные элементы равны, называются *скалярными матрицами*:

$$\mathbf{B} = \begin{pmatrix} b_{11} & 0 & \dots & 0 \\ 0 & b_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b_{nn} \end{pmatrix};$$

$$b_{11} = b_{22} = \dots = b_{nn} = b.$$

Если у скалярной матрицы все диагональные элементы равны единице, то матрица называется *единичной*:

$$\mathbf{E} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 \end{pmatrix}.$$

Матрица, все элементы которой равны нулю, называется *нулевой*. Если $N = 1$, то матрица называется *вектором-строкой*, если $n = 1$, то матрица называется *вектором-столбцом*.

Вектор-строка $\mathbf{a}' = (a_1\ a_2 \dots a_n)$, вектор-столбец

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \dots \\ a_n \end{pmatrix},$$

или $\mathbf{a} = \{a_1\ a_2 \dots a_n\}$.

Элементы матриц могут быть как действительными, так и мнимыми числами. При проведении компонентного анализа используются матрицы с действительными числами, поэтому другие матрицы не будут рассматриваться.

Треугольная матрица — это квадратная матрица, у которой все элементы, стоящие по одну из сторон главной диагонали, равны нулю.

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{3n} \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{nn} \end{pmatrix}.$$

Диагональные матрицы являются одновременно и треугольными.

Квадратная матрица называется *матрицей перестановок*, если в каждой ее строке и в каждом ее столбце содержится один элемент, равный единице, а все остальные элементы нулевые. Существует $n!$ матриц-перестановок порядка n, одной из которых является единичная матрица, например:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};\ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix};\ \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix};\ \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix};\ \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 1 & \dots & 0 \end{pmatrix}.$$

Блочной называется матрица, элементы которой разбиты на некоторое число подматриц: $\mathbf{A} = (\mathbf{A}_{k;\,i})$, где $\mathbf{A}_{k;\,i}$ — блоки матрицы $\mathbf{A}$.

Пусть $\mathbf{A}$ — матрица порядка $m \times n$:

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{pmatrix},$$

тогда $\mathbf{A}_{11}$ — матрица порядка $m_1 \times n_1$;

$\mathbf{A}_{12} - m_1 \times (n - n_1)$;

$\mathbf{A}_{21} - (m - m_1) \times n_1$;

$\mathbf{A}_{22} - (m - m_1) \times (n - n_1)$.

Если матрицу можно разбить на подматрицы так, чтобы только на ее главной диагонали стояли ненулевые квадратные матрицы, то ее называют *блочно-диагональной матрицей*.

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_{11} & 0 & \dots & 0 \\ 0 & \mathbf{A}_{22} & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \mathbf{A}_{qq} \end{pmatrix}.$$

Блочно-треугольная матрица — это матрица, которую можно разбить на подматрицы так, чтобы по одну сторону ее главной диагонали, составленной из подматриц, стояли нули. Блочнодиагональная матрица является одновременно и блочно-треугольной матрицей.

Если определитель квадратной матрицы равен нулю, то матрица называется *особенной* (*вырожденной*); если ее определитель не равен нулю, то матрица называется *неособенной* (*невырожденной*). Всякая неособенная квадратная матрица **A** имеет *обратную матрицу* $\mathbf{A}^{-1}$. Определитель матрицы является числом. Это число записывается в виде таблицы. Определителем квадратной матрицы порядка n называется число, равное алгебраической сумме:

$$|\mathbf{A}| = a_{11}A_{11} + a_{21}A_{21} + \ldots + a_{n1}A_{n1},$$

где A_{n1} — определитель $(n-1)$-го порядка, получаемый из исходного путем вычеркивания первого столбца и i-й строки и умножения на $(-1)^{i+1}$.

Определитель n-го порядка вычисляется через определители более низких порядков. Определитель матрицы **A** записывается так:

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & \ldots & a_{nn} \end{vmatrix}.$$

Если в определителе n-го порядка выбрать элемент a_{ij} и вычеркнуть столбец j и строку i, то оставшиеся столбцы и строки образуют определитель $(n-1)$-го порядка, который называют *минором заданного определителя*, соответствующим a_{ij}. Он записывается $\mathbf{M}_{ij}$. Алгебраическим дополнением A_{ij} элемента a_{ij} определителя называется соответствующий ему минор, взятый со знаком плюс или минус. Минору элемента a_{ij} приписывается знак плюс, если сумма $i + j$ номеров столбца и строки четная, и знак минус, если сумма номеров нечетная. Таким образом, $A_{ij} = \mathbf{M}_{ij}(-1)^{i+j}$.

Определитель n-го порядка можно вычислить, разложив его по элементам любого столбца или любой строки. Разложим его по элементам i-й строки:

$$|\mathbf{A}| = a_{i1}A_{i1} + a_{i2}A_{i2} + \ldots + a_{in}A_{in},$$

или по элементам j-го столбца:

$$|\mathbf{A}| = a_{1j}A_{1j} + a_{2j}A_{2j} + \ldots + a_{nj}A_{nj}.$$

Итак, определитель равен сумме произведений элементов i-й строки на их алгебраические дополнения или сумме произведений элементов j-го столбца на их алгебраические дополнения. Сумма произведений элементов столбца (строки) определителя на алгебраические дополнения соответствующих элементов другого столбца (строки) равна нулю.

Определитель диагональной матрицы равен произведению ее элементов, стоящих на диагонали:

$$|\mathbf{A}| = \begin{vmatrix} a_{11} & 0 & \ldots & 0 \\ 0 & a_{22} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & a_{nn} \end{vmatrix} = a_{11} \cdot a_{22} \cdot \ldots \cdot a_{nn}.$$

Следовательно, определитель единичной матрицы $|\mathbf{E}| = 1$.

В отличие от матрицы в определителе существует равноправие столбцов и строк, поэтому *определитель не изменится, если его столбцы заменить строками, не меняя их порядка.*

Пример П.1.1

$$\begin{vmatrix} 1 & 2 \\ 3 & 4 \end{vmatrix} = \begin{vmatrix} 1 & 3 \\ 2 & 4 \end{vmatrix} = -2.$$

Перестановка местами двух столбцов (строк) означает умножение определителя на -1.

Если в определителе есть столбец (строка), элементы которого равны соответствующим элементам другого столбца (строки), то он равен нулю.

Если все элементы k-го столбца (строки) определителя являются суммой двух слагаемых, то определитель можно представить суммой двух определителей, у которых элементами k-го столбца (строки) являются соответственно первые и вторые слагаемые элементов k-го столбца исходного определителя; остальные элементы остаются такими же, как у исходного определителя.

Если все элементы одного из столбцов (строки) определителя равны нулю, то и он равен нулю.

Чтобы умножить определитель на некоторое число, достаточно умножить на это число элементы одной строки.

Если у всех элементов некоторого столбца (строки) есть общий множитель, то его можно вынести за знак определителя.

Если ко всем элементам какого-нибудь столбца (строки) определителя прибавить соответствующие элементы другого столбца (строки), умноженные на одно и то же число, то значение определителя не изменится.

Если все элементы k-го столбца (строки) определителя, кроме одного a_{ik}, равны нулю, то такой определитель равен произведению этого элемента на его алгебраическое дополнение: $|\mathbf{A}| = a_{ik}A_{ik}$.

Если одна из строк (столбцов) определителя является линейной комбинацией его других строк (столбцов), то определитель равен нулю.

Пример П.1.2

$$\begin{vmatrix} 4 & 8 \\ 3 & 6 \end{vmatrix} = 4 \cdot 6 - 3 \cdot 8 = 0.$$

Если элементы двух столбцов определителя пропорциональны, то он равен нулю.

Определитель произведения двух квадратных матриц равен произведению определителей этих матриц: $|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|$.

Пример П.1.3

Образуем определитель следующим образом: запишем две его строки, состоящие из разных и соответственно непропорциональных элементов. Третью строку составим как линейную комбинацию этих двух строк (умноженных соответственно на 2 и на 3). Задача заключается в проверке равенства нулю полученного определителя.

Решение

1. Составляем две строки:

$$\begin{matrix} 1 & 2 & 3 \\ 4 & 5 & 6. \end{matrix}$$

2. Составляем третью строку:

$$\begin{array}{r} + \begin{matrix} 2 & 4 & 6 \\ 12 & 15 & 18 \end{matrix} \\ \hline \begin{matrix} 14 & 19 & 24 \end{matrix} \end{array}$$

3. Вычисляем значение полученного определителя:

$$\begin{vmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 14 & 19 & 24 \end{vmatrix} = 14(-1)^{3+1} \begin{vmatrix} 2 & 3 \\ 5 & 6 \end{vmatrix} + 19(-1)^{3+2} \begin{vmatrix} 1 & 3 \\ 4 & 6 \end{vmatrix} +$$

$$+ 24(-1)^{3+3} \begin{vmatrix} 1 & 2 \\ 4 & 5 \end{vmatrix} = 14 \cdot (-3) + 19 \cdot 6 - 24 \cdot 3 = 0.$$

Пример П.1.4. Вычислить определитель, применяя свойства определителей.

Решение

$$|\mathbf{A}| = \begin{vmatrix} 1 & 2 & 1 & 2 \\ 2 & -2 & 3 & -1 \\ 2 & -3 & 2 & 1 \\ 3 & 4 & -3 & 6 \end{vmatrix} = \begin{vmatrix} 1 & 4 & -2 & 3 \\ 1 & 0 & 0 & 0 \\ 2 & 1 & -4 & 3 \\ 3 & 10 & -12 & 9 \end{vmatrix} =$$

$$= 1(-1)^{2+1} \begin{vmatrix} 4 & -2 & 3 \\ 1 & -4 & 3 \\ 10 & -12 & 9 \end{vmatrix} = 1(-1) \cdot 42 = -42,$$

или

$$|\mathbf{A}| = 1(-1)^{2+1} \begin{vmatrix} 4 & 14 & -9 \\ 1 & 0 & 0 \\ 10 & 28 & -21 \end{vmatrix} = (-1)(-1)^{2+1} \cdot 1 \begin{vmatrix} 14 & -9 \\ 28 & -21 \end{vmatrix} =$$

$$= -294 + 252 = -42.$$

Столбец, содержащий элемент $a_{21} = 1$ (выделен жирным шрифтом), умножаем на 2 и складываем со вторым столбцом, затем умножаем на -3 и складываем с третьим столбцом, наконец, умножаем на единицу и складываем с четвертым столбцом. Таким образом, получается строка, состоящая из нулей, кроме элемента $a_{21} = 1$. Определитель равен произведению этого элемента на его алгебраическое дополнение.

Можно произвести проверку, доведя определитель третьего порядка до произведения одного его элемента с его алгебраическим дополнением второго порядка. Повторим операции по приведению всех элементов второй строки, кроме одного, к нулю. Ответы, естественно, совпадают.

Действия над матрицами

Две матрицы называются *равными*, если они одинакового порядка (имеют одинаковое число строк и столбцов) и их соответствующие элементы равны: $\mathbf{A} = \mathbf{B}$, если $N_A = N_B$; $n_A = n_B$ и $a_{ij} = b_{ij}$.

Суммой (разностью) двух прямоугольных матриц одинакового порядка называется *матрица того же порядка*, элементами которой являются суммы (разности) соответствующих элементов слагаемых:

$$\mathbf{A} + \mathbf{B} = \mathbf{C};$$

$$N_A = N_B;\ n_A = n_B;\ c_{ij} = a_{ij} + n_{ij};$$

$$N_C = N_A = N_B;\ n_A = n_B = n_C.$$

Пример П.1.5

$$\mathbf{A} = \begin{pmatrix} 2\ 5\ 6 \\ 3\ 4\ 5 \\ 1\ 2\ 3 \end{pmatrix};\ \mathbf{B} = \begin{pmatrix} 0\ 1\ 3 \\ 7\ 4\ 2 \\ 3\ 2\ 1 \end{pmatrix};$$

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} 2+0 & 5+1 & 6+3 \\ 3+7 & 4+4 & 5+2 \\ 1+3 & 2+2 & 3+1 \end{pmatrix} = \begin{pmatrix} 2 & 6 & 9 \\ 10 & 8 & 7 \\ 4 & 4 & 4 \end{pmatrix}.$$

Известно, что

$$\mathbf{A} + (\mathbf{B} + \mathbf{C}) = (\mathbf{A} + \mathbf{B}) + \mathbf{C};$$

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A};\ \mathbf{A} + 0 = \mathbf{A}.$$

Если α и β скалярны, то можно записать:

$$(\alpha + \beta)\mathbf{A} = \alpha\mathbf{A} + \beta\mathbf{A};$$

$$\alpha(\mathbf{A} + \mathbf{B}) = \alpha\mathbf{A} + \alpha\mathbf{B};$$

$$1\mathbf{A} = \mathbf{A};$$

$$\alpha(\beta\mathbf{A}) = (\alpha\beta)\mathbf{A}.$$

Умножение матрицы **A** на матрицу **B** возможно только в том случае, когда число столбцов матрицы **A** равно числу строк матрицы **B**. Произведением двух матриц порядка $N \times n$ и $n \times k$ является матрица **C** порядка $N \times k$, элемент которой c_{ij} равен сумме произведений элементов i-й строки первой матрицы на соответствующие элементы j-го столбца второй матрицы:

$$\begin{pmatrix} a_{11}\ a_{12}\ \dots\ a_{1n} \\ a_{21}\ a_{22}\ \dots\ a_{2n} \\ \dots\dots\dots\dots \\ a_{N1}\ a_{N2}\ \dots\ a_{Nn} \end{pmatrix} \cdot \begin{pmatrix} b_{11}\ b_{12}\ \dots\ b_{1k} \\ b_{21}\ b_{22}\ \dots\ b_{2k} \\ \dots\dots\dots\dots \\ b_{n1}\ b_{n2}\ \dots\ b_{nk} \end{pmatrix} = \begin{pmatrix} c_{11}\ c_{12}\ \dots\ c_{1k} \\ c_{21}\ c_{22}\ \dots\ c_{2k} \\ \dots\dots\dots\dots \\ c_{N1}\ c_{N2}\ \dots\ c_{Nk} \end{pmatrix};$$

$$c_{ij} = \sum_{l=1}^{n} a_{il} b_{lj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \dots + a_{in}b_{nj};$$

$$1 \leqslant i \leqslant N;\ 1 \leqslant j \leqslant k.$$

Пример П.1.6

$$\begin{pmatrix} 1 & 2 & 3 \\ 4 & -2 & 5 \end{pmatrix} \cdot \begin{pmatrix} 4 & 2 \\ 1 & 3 \\ 1 & 6 \end{pmatrix} =$$

$$= \begin{pmatrix} 1\cdot 4 + 2\cdot 1 + 3\cdot 1 & 1\cdot 2 + 2\cdot 3 + 3\cdot 6 \\ 4\cdot 4 + (-2)\cdot 1 + 5\cdot 1 & 4\cdot 2 + (-2)\cdot 3 + 5\cdot 6 \end{pmatrix} = \begin{pmatrix} 9 & 26 \\ 19 & 32 \end{pmatrix}.$$

В частных случаях, когда перемножаются вектор-строка и вектор-столбец:

$$c_1 = \mathbf{a}'\mathbf{b} = (a_1, a_2, \dots, a_n) \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} = a_1b_1 + a_2b_2 + \dots + a_nb_n;$$

$$\mathbf{C} = \mathbf{b}\mathbf{a}' = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_n \end{pmatrix} (a_1, a_2, \dots, a_n) = \begin{pmatrix} b_1a_1 & b_1a_2 & \dots & b_1a_n \\ b_2a_1 & b_2a_2 & \dots & b_1a_n \\ \dots & \dots & \dots & \dots \\ b_na_1 & b_na_2 & \dots & b_na_n \end{pmatrix}.$$

В первом случае получен скаляр, во втором — матрица.

Строка или столбец матрицы являются векторами, следовательно, матрица-строка и матрица-столбец — это вектор-строка и вектор-столбец соответственно. Значит, внутреннее, или скалярное, произведение двух векторов — это произведение вектора-строки на вектор-столбец. Оно равно скаляру.

Пример П.1.7

$$\mathbf{a}' = (1, 2, 3);\ \mathbf{b} = \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix};$$

$$\mathbf{a}'\mathbf{b} = (1, 2, 3) \times \begin{pmatrix} 2 \\ 3 \\ 4 \end{pmatrix} = 2 + 6 + 12 = 20.$$

Значит, $\mathbf{a}'\mathbf{b} = \sum_{j=1}^{n} a_j b_j$.

Векторы называются *ортогональными*, если их скалярное произведение равно нулю.

Пример П.1.8

$$\mathbf{a}' = (6, 8);\ \mathbf{b} = \begin{pmatrix} 4 \\ -3 \end{pmatrix};$$

$$\mathbf{a}'\mathbf{b} = (6, 8) \times \begin{pmatrix} 4 \\ -3 \end{pmatrix} = 24 - 24 = 0.$$

Векторы **a** и **b** ортогональны.

Если образовать произведение вектора **b**, представленного в виде вектора-столбца порядка n, на вектор-строку $\mathbf{b}'$, то получим матрицу рассеяния. Она будет квадратной симметричной (симметрической) порядка **n** (это внешнее произведение вектора на себя самого).

Пример П.1.9. Дан вектор-столбец.

$$\mathbf{b} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

Определите матрицу рассеяния.

Решение

1. Находим, что $\mathbf{b}' = (1, 2, 3, 4)$.

2.

$$\mathbf{b}\mathbf{b}' = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix} \times (1, 2, 3, 4) = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 6 & 8 \\ 3 & 6 & 9 & 12 \\ 4 & 8 & 12 & 16 \end{pmatrix}.$$

Внешнее произведение двух векторов порядка n — это произведение вектора-столбца на вектор-строку, равное квадратной матрице порядка n.

В общем случае произведение двух матриц зависит от порядка сомножителей, поэтому различают умножение матрицы справа или слева $\mathbf{AB} \neq \mathbf{BA}$. В частном случае равенство возможно. Матрицы **A** и **B**, для которых выполняется равенство $\mathbf{AB} = \mathbf{BA}$, называются *перестановочными*. Если **A** и **B** квадратные матрицы одного и того же порядка, то всегда существуют произведения **AB** и **BA**. Произведения матриц обладают следующими свойствами:

$$\alpha[\mathbf{AB}] = [\alpha\mathbf{A}]\mathbf{B} = \mathbf{A}[\alpha\mathbf{B}],$$

где α — скаляр;

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC});\ \mathbf{C}(\mathbf{A} + \mathbf{B}) = \mathbf{CA} + \mathbf{CB};$$

$$(\mathbf{A} + \mathbf{B})\mathbf{C} = \mathbf{AC} + \mathbf{BC};\ \mathbf{AE} = \mathbf{EA} = \mathbf{A}.$$

Скалярная матрица, т. е. диагональная матрица, у которой $b_{11} = b_{22} = \dots = b_{nn} = b$.

$$\mathbf{B} = \begin{pmatrix} b & 0 & \dots & 0 \\ 0 & b & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & b \end{pmatrix}$$

может быть представлена в виде произведения $\mathbf{B} = b\mathbf{E}$.

Произведение квадратной матрицы **A** на скалярную матрицу **B** того же размера коммутативно (перестановочно):

$\mathbf{AB} = \mathbf{A}b\mathbf{E} = b\mathbf{AE} = \mathbf{BA}$.

Известно, что $|\mathbf{AB}| = |\mathbf{BA}| = |\mathbf{A}| \times |\mathbf{B}|$.

Рассмотрим, как производится запись линейных преобразований при помощи матриц.

Пусть переменные $\mathbf{f}(f_1, f_2, ..., f_m)$ необходимо линейно преобразовать в переменные $\mathbf{y}(y_1, y_2, ..., y_n)$. Это может быть записано в виде системы линейных уравнений:

$$\left.\begin{aligned} y_1 &= a_{11}f_1 + a_{12}f_2 + \dots + a_{1m}f_m; \\ y_2 &= a_{21}f_1 + a_{22}f_2 + \dots + a_{2m}f_m; \\ &\dots\dots\dots\dots\dots\dots\dots\dots \\ y_n &= a_{n1}f_1 + a_{n2}f_2 + \dots + a_{nm}f_m. \end{aligned}\right\}$$

Данное линейное преобразование записывается так:

$$\mathbf{Af} = \mathbf{y}, \tag{П.1.1}$$

где **A** — матрица линейного преобразования.

Развернем это выражение, чтобы показать вид матриц

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix} \begin{pmatrix} f_1 \\ f_2 \\ \dots \\ f_m \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix}.$$

Известны тождественные преобразования, переводящие переменные в себя:

$$\left.\begin{aligned} {}_1\mathbf{F}_1 + {}_0\mathbf{F}_2 &= \mathbf{F}_1; \\ {}_0\mathbf{F}_1 + {}_1\mathbf{F}_2 &= \mathbf{F}_2. \end{aligned}\right\}$$

Матрицей данного преобразования будет единичная матрица

$$\mathbf{E} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Если в матрице **A** порядка $N \times n$ заменить строки соответствующими столбцами, то полученная матрица **A′** порядка $n \times N$ называется *транспонированной по отношению к исходной матрице.*

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{N1} & a_{N2} & \dots & a_{Nn} \end{pmatrix}; \mathbf{A}' = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{N1} \\ a_{12} & a_{22} & \dots & a_{N2} \\ \dots & \dots & \dots & \dots \\ a_{1n} & a_{2n} & \dots & a_{Nn} \end{pmatrix}.$$

Если над матрицей **A** дважды произвести транспонирование, то матрица не изменится: $\mathbf{A}'' = \mathbf{A}$. Нетрудно доказать, что

$$(\mathbf{A} + \mathbf{B})' = \mathbf{A}' + \mathbf{B}'; \ (\alpha\mathbf{A})' = \alpha\mathbf{A}'; \ |\mathbf{A}| = |\mathbf{A}'|.$$

Последнее выражение справедливо, так как определитель, как известно, не меняет своей величины при замене строк столбцами и наоборот.

Транспонированная матрица произведения матриц равна произведению транспонированных матриц, взятых в обратном порядке:

$$(\mathbf{AB})' = \mathbf{B}'\mathbf{A}';$$

$$(\mathbf{ABCD})' = \mathbf{D}'\mathbf{C}'\mathbf{B}'\mathbf{A}'.$$

Квадратная матрица **A** называется *симметричной*, если $\mathbf{A} = \mathbf{A}'$, следовательно, $a_{ij} = a_{ji}$. Квадратная матрица называется *кососимметричной*, если $\mathbf{A} = -\mathbf{A}'$, следовательно, $a_{ij} = -a_{ji}$. Элементы ее главной диагонали равны нулю, например:

$$\mathbf{A} = \begin{pmatrix} 0 & 4 \\ -4 & 0 \end{pmatrix}.$$

Произведение матрицы **A** на ее транспонированную **A′** есть симметричная матрица **R**: $\mathbf{AA}' = \mathbf{R}$.

Рангом матрицы называется максимальный порядок минора, не равного нулю. Значит, ранг матрицы **A** равен r, если:

а) имеется по крайней мере один минор матрицы **A** порядка r, не равной нулю;

б) все миноры матрицы **A** порядка $r + 1$ и выше равны нулю.

Отсюда вытекает целесообразное правило вычисления ранга матрицы: переходить от миноров первого порядка к минорам больших порядков; если найден минор k-го порядка, не равный нулю, следует вычислять миноры $(k + 1)$-го порядка, получающиеся окаймлением минора k-го порядка. Вычисления продолжать, пока среди них не найдется минор, не равный нулю. После этого процедура продолжается. Если же все миноры $(k + 1)$-го порядка окажутся равными нулю, то ранг матрицы равен k.

Пример П.1.10. Определите ранг матрицы

$$\mathbf{A} = \begin{pmatrix} 1 & 2 & 5 & 1 & 2 \\ 3 & 6 & 6 & 2 & 3 \\ 2 & 4 & 1 & 1 & 1 \end{pmatrix}.$$

Решение

1. Определим минор первого порядка $|1| = 1 \neq 0$.

2. Определим миноры второго порядка, окаймляющие минор первого порядка.

$$\begin{vmatrix} 1 & 2 \\ 3 & 6 \end{vmatrix} = 0; \quad \begin{vmatrix} 1 & 5 \\ 3 & 6 \end{vmatrix} = -9 \neq 0.$$

3. Вычислим миноры третьего порядка, окаймляющие минор второго порядка, не равный нулю.

$$\begin{vmatrix} 1 & 5 & 2 \\ 3 & 6 & 6 \\ 2 & 1 & 4 \end{vmatrix} = 0; \quad \begin{vmatrix} 1 & 5 & 1 \\ 3 & 6 & 2 \\ 2 & 1 & 1 \end{vmatrix} = 0; \quad \begin{vmatrix} 1 & 5 & 2 \\ 3 & 6 & 3 \\ 2 & 1 & 1 \end{vmatrix} = 0.$$

Остальные миноры третьего порядка можно не вычислять, так как из миноров второго и вычисленных миноров третьего порядка видно, что и все остальные миноры третьего порядка будут также равны нулю. Следовательно, заданная матрица имеет ранг, равный двум.

Пример П.1.11. Определите ранг матрицы **A** методом Жордана—Гаусса:

$$\mathbf{A} = \begin{pmatrix} 3 & 2 & 5 & 1 & 8 \\ 6 & 0 & 6 & 6 & 12 \\ 6 & 2 & 8 & 4 & 14 \\ 0 & 4 & 4 & -4 & 4 \end{pmatrix}.$$

Для решения примера П.1.11 можно воспользоваться элементарными преобразованиями, которые не меняют ранга матрицы:

- перемена местами любых двух строк или столбцов;
- умножение любой строки или столбца на произвольное действительное число, отличное от нуля;
- сложение одной строки (столбца), умноженной на произвольное действительное число, с другой строкой (столбцом);
- вычеркивание строки (или столбца), являющейся линейной комбинацией других строк (столбцов);
- вычеркивание строки (столбца), все элементы которой состоят из нулей.

В результате проведенных преобразований следует получить единицы на главной диагонали матрицы. Число единиц в диагональной (единичной) матрице, полученной после преобразований, соответствует рангу матрицы.

Решение

1. Разделим на общие делители первый, второй и пятый столбцы (соответственно на 3, 2 и 2):

$$\begin{pmatrix} 3 & 2 & 5 & 1 & 8 \\ 6 & 0 & 6 & 6 & 12 \\ 6 & 2 & 8 & 4 & 14 \\ 0 & 4 & 4 & -4 & 4 \end{pmatrix} \to \begin{pmatrix} 1 & 1 & 5 & 1 & 4 \\ 2 & 0 & 6 & 6 & 6 \\ 2 & 1 & 8 & 4 & 7 \\ 0 & 2 & 4 & -4 & 2 \end{pmatrix} \to$$

2. Разделим на 2 вторую и четвертую строки и получим нули в первом столбце путем умножения элементов первой строки на -1 и -2 и сложения с элементами соответственно второй и третьей строк. В четвертой строке на месте первого элемента уже был нуль, поэтому она осталась, как и первая строка, без изменений:

$$\rightarrow\begin{pmatrix}1&1&5&1&4\\1&0&3&3&3\\2&1&8&4&7\\0&1&2&-2&1\end{pmatrix}\rightarrow\begin{pmatrix}1&1&5&1&4\\0&-1&-2&2&-1\\0&-1&-2&2&-1\\0&1&2&-2&1\end{pmatrix}\rightarrow$$

Проанализируем полученную матрицу. При умножении четвертой строки на -1 получаем одинаковые три строки: вторую, третью и четвертую.

3. Можем вычеркнуть две строки, оставив одну из трех равных строк. Затем первый столбец умножаем на -1, -5, -1 и -4 и складываем соответственно со вторым, третьим, четвертым и пятым столбцами. Вторую строку умножаем на -1, и второй ее элемент стал положительной единицей. Тогда получаем:

$$\rightarrow\begin{pmatrix}1&1&5&1&4\\0&-1&-2&2&-1\end{pmatrix}\rightarrow\begin{pmatrix}1&0&0&0&0\\0&1&2&-2&1\end{pmatrix}.$$

При помощи второго столбца получаем нули в третьем, четвертом и пятом столбцах. Вычеркиваем нулевые столбцы и получаем:

$$\begin{pmatrix}1&0&0&0&0\\0&1&2&-2&1\end{pmatrix}\rightarrow\begin{pmatrix}1&0\\0&1\end{pmatrix}.$$

Значит, ранг матрицы **A** равен двум.

Дефектом матрицы порядка $N \times n$ называется разность между наименьшим из чисел N и n и рангом матрицы r. Так, у матрицы из примера П.1.11 дефект равен 1, так как $\min(N, n) = \min(3, 5)$, а $r = 2$, то $\min(3, 5) - 2 = 1$.

Следом матрицы **A** называется сумма ее диагональных элементов. Известно, что

$$\mathrm{tr}(\mathbf{A} + \mathbf{B}) = \mathrm{tr}(\mathbf{A}) + \mathrm{tr}(\mathbf{B});\ \mathrm{tr}(\mathbf{AB}) = \mathrm{tr}(\mathbf{BA}).$$

Это равенство справедливо и для случая, когда существуют оба произведения и **A** и **B** — прямоугольные матрицы.

Пример П.1.12. *Расчет коэффициентов ковариации и корреляции.*

Определите взаимозависимость себестоимости 1 т угля (**y**) и фондоотдачи (**x**) на основании данных табл. П.1.1.

Таблица П.1.1

Исходные данные по восьми бассейнам

Себестоимость, y	20	25	20	15	10	30	10	20
Фондоотдача, x	30	20	40	35	45	25	50	30

Решение

Будем считать, что приведенные в примере данные являются выборкой из двумерной, нормально распределенной совокупности.

Для оценивания взаимосвязи признаков воспользуемся формулой для парного коэффициента корреляции:

$$r_{yx} = \frac{\mathrm{cov}(\mathbf{y},\mathbf{x})}{\mathbf{S_y}\cdot\mathbf{S_x}},$$

где $\mathrm{cov}(\mathbf{y}, \mathbf{x})$ — коэффициент ковариации;
$\mathbf{S_y}, \mathbf{S_x}$ — выборочные средние квадратические отклонения признаков.

Вспомогательные вычисления представим в табл. П.1.2, в последней строке получим необходимые для дальнейших расчетов суммы.

1. Определим значения выборочных дисперсий исследуемых признаков:

$$\mathbf{S_x^2} = \frac{\sum_{i=1}^{n}\mathbf{x}_i^2}{n} - (\bar{\mathbf{x}})^2,$$

где $\bar{\mathbf{x}} = \frac{\sum_{i=1}^{n}\mathbf{x}_i}{n}$;

$$\mathbf{S_x^2} = \frac{10175}{8} - \left(\frac{275}{8}\right)^2 = 89{,}90;$$

$$\mathbf{S_y^2} = \frac{\sum_{i=1}^{n}\mathbf{y}_i^2}{n} - (\bar{\mathbf{y}})^2,$$

Таблица П.1.2

Вспомогательные вычисления

№ шахты	x	y	x^2	y^2	xy
1	30	20	900	400	600
2	20	25	400	625	500
3	40	20	1600	400	800
4	35	15	1225	225	525
5	45	10	2025	100	450
6	25	30	625	900	750
7	50	10	2500	100	500
8	30	25	900	400	600
Σ	**275**	**150**	**10175**	**3150**	**4725**

где $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$.

$$S_y^2 = \frac{3150}{8} - \left(\frac{150}{8}\right)^2 = 42{,}19.$$

Следовательно, выборочные средние квадратические отклонения признаков соответственно равны:

$$S_x = \sqrt{89{,}90} = 9{,}48;\ S_y = \sqrt{42{,}19} = 6{,}50.$$

2. Числитель выражения для парного коэффициента корреляции представляет собой коэффициент ковариации.

Определим его значение с помощью следующей формулы:

$$\operatorname{cov}(y, x) = \overline{xy} - \bar{x} \cdot \bar{y},$$

где $\overline{xy} = \dfrac{\sum_{i=1}^{n} x_i y_i}{n}$.

Следовательно,

$$\overline{xy} = \frac{4725}{8} = 590{,}63;$$

$$\bar{x} = \frac{275}{8} = 34{,}38;$$

$$\bar{y} = \frac{150}{8} = 18{,}75;$$

$$\operatorname{cov}(y, x) = 590{,}63 - 34{,}38 \cdot 18{,}75 = -54.$$

3. Рассчитаем значение парного коэффициента корреляции:

$$r_{yx} = \frac{\operatorname{cov}(y,x)}{S_y \cdot S_x} = \frac{-54}{9{,}48 \cdot 6{,}50} = -0{,}876.$$

Дополнительная проверка показала значимость коэффициента корреляции, следовательно, между признаками выявлена сильная отрицательная (обратная) статистическая взаимосвязь.

Приложение 2

Процентные точки выборочного коэффициента r при $\rho = 0$

ν	Значения Q $(2Q)$, %					
	5(10)	2,5(5)	1(2)	0,5(1)	0,25(0,5)	0,05(0,1)
1	0,9877	0,9969	0,9995	0,99988	0,99997	0,99999
2	0,9000	0,9500	0,9800	0,9900	0,9950	0,99900
3	0,8054	0,8783	0,9343	0,9587	0,9740	0,99911
4	0,7293	0,8114	0,8822	0,9172	0,9417	0,9741
5	0,6694	0,7545	0,8329	0,8745	0,9056	0,9509
6	0,6215	0,7067	0,7887	0,8343	0,870	0,9249
7	0,5822	0,6664	0,7498	0,7977	0,836	0,898
8	0,5494	0,6319	0,7155	0,7646	0,805	0,872
9	0,5214	0,6021	0,6851	0,7348	0,776	0,847
10	0,4973	0,5760	0,6581	0,7079	0,750	0,823
11	0,4762	0,5529	0,6339	0,6835	0,726	0,801
12	0,4575	0,5324	0,6120	0,6614	0,703	0,780
13	0,4409	0,5139	0,5923	0,6411	0,683	0,760
14	0,4259	0,4973	0,5742	0,6226	0,664	0,742
15	0,4124	0,4821	0,5577	0,6055	0,647	0,725
16	0,4000	0,4683	0,5425	0,5897	0,631	0,708
17	0,3887	0,4555	0,5285	0,5751	0,616	0,693
18	0,3783	0,4438	0,5155	0,5614	0,602	0,679
19	0,3687	0,4329	0,5034	0,5487	0,589	0,665
20	0,3598	0,4227	0,4921	0,5368	0,576	0,652
25	0,3233	0,3809	0,4451	0,4869	0,524	0,597
30	0,2960	0,3494	0,4093	0,4487	0,484	0,554
35	0,2746	0,3246	0,3810	0,4182	0,452	0,519
40	0,2573	0,3044	0,3578	0,3932	0,425	0,490
45	0,2428	0,2875	0,3384	0,3721	0,403	0,465
50	0,2306	0,2732	0,3218	0,3541	0,384	0,443
60	0,2108	0,2500	0,2948	0,3248	0,352	0,408
70	0,1954	0,2319	0,2737	0,3017	0,327	0,380
80	0,1829	0,2172	0,2565	0,2830	0,307	0,357
90	0,1726	0,2050	0,2422	0,2673	0,290	0,338
100	0,1638	0,1946	0,2301	0,2540	0,276	0,321

Приложение 3

Соответствие между r и $Z = \frac{1}{2}\ln\frac{1+r}{1-r}$

Z	0,01	0,02	0,03	0,04	0,05
0,0	0,0100	0,0200	0,0300	0,0400	0,0500
0,1	0,1096	0,1194	0,1293	0,1391	0,1489
0,2	0,2070	0,2165	0,2260	0,2355	0,2449
0,3	0,3004	0,3095	0,3185	0,3275	0,3364
0,4	0,3885	0,3969	0,4053	0,4136	0,4219
0,5	0,4699	0,4777	0,4854	0,4930	0,5005
0,6	0,5441	0,5511	0,5580	0,5649	0,5717
0,7	0,6107	0,6169	0,6231	0,6291	0,6351
0,8	0,6696	0,6751	0,6805	0,6858	0,6911
0,9	0,7211	0,7259	0,7306	0,7352	0,7398
1,0	0,7658	0,7699	0,7739	0,7779	0,7818
1,1	0,8041	0,8076	0,8110	0,8144	0,8178
1,2	0,8367	0,8397	0,8426	0,8455	0,8483
1,3	0,8643	0,8668	0,8692	0,8717	0,8741
1,4	0,8875	0,8896	0,8917	0,8937	0,8957
1,5	0,9069	0,9087	0,9104	0,9121	0,9138
1,6	0,9232	0,9246	0,9261	0,9275	0,9289
1,7	0,9366	0,9379	0,9391	0,9402	0,9414
1,8	0,94783	0,94884	0,94983	0,95080	0,95175
1,9	0,95709	0,95792	0,95873	0,95953	0,96032
2,0	0,96473	0,96541	0,96609	0,96675	0,96739
2,1	0,97103	0,97159	0,97215	0,97269	0,97323
2,2	0,97622	0,97668	0,97714	0,97752	0,97803
2,3	0,98049	0,98087	0,98124	0,98161	0,98197
2,4	0,98399	0,98431	0,98462	0,98492	0,98522
2,5	0,98688	0,98714	0,98739	0,98764	0,98788
2,6	0,98924	0,98945	0,98966	0,98987	0,99007
2,7	0,99118	0,99136	0,99153	0,99170	0,99185
2,8	0,99278	0,99292	0,99306	0,99320	0,99333
2,9	0,99408	0,99420	0,99131	0,99443	0,99454

Продолжение

Z	0,06	0,07	0,08	0,09	0,10
0,0	0,0599	0,0699	0,0798	0,0898	0,0997
0,1	0,1586	0,1684	0,1781	0,1877	0,1974
0,2	0,2548	0,2636	0,2729	0,2821	0,2913
0,3	0,3452	0,3540	0,3627	0,3714	0,3800
0,4	0,4301	0,4382	0,4462	0,4542	0,4621
0,5	0,5080	0,5154	0,5227	0,5299	0,5370
0,6	0,5784	0,5850	0,5915	0,5980	0,6044
0,7	0,6411	0,6469	0,6527	0,6584	0,6640
0,8	0,6963	0,7014	0,7064	0,7114	0,7163
0,9	0,7443	0,7487	0,7531	0,7574	0,7616
1,0	0,7857	0,7895	0,7932	0,7969	0,8005
1,1	0,8210	0,8243	0,8275	0,8306	0,8337
1,2	0,8511	0,8538	0,8565	0,8591	0,8617
1,3	0,8764	0,8787	0,8810	0,8832	0,8854
1,4	0,8977	0,8996	0,9015	0,9033	0,9051
1,5	0,9154	0,9170	0,9186	0,9201	0,9217
1,6	0,9302	0,9316	0,9329	0,9341	0,9354
1,7	0,9425	0,9436	0,9447	0,9458	0,9468
1,8	0,95268	0,95359	0,95449	0,95537	0,95624
1,9	0,96109	0,96185	0,96259	0,96331	0,96403
2,0	0,96803	0,96865	0,96926	0,96986	0,97045
2,1	0,97375	0,97426	0,97477	0,97526	0,97574
2,2	0,97846	0,97888	0,97929	0,97970	0,98010
2,3	0,98233	0,98267	0,98301	0,98335	0,98367
2,4	0,98551	0,98579	0,98607	0,98635	0,98661
2,5	0,98812	0,98835	0,98858	0,98881	0,98903
2,6	0,99026	0,99045	0,99064	0,99083	0,99101
2,7	0,99202	0,99218	0,99233	0,99284	0,99263
2,8	0,99346	0,99359	0,99372	0,99384	0,99396
2,9	0,99464	0,99475	0,99485	0,99495	0,99505

КРАТКИЙ СЛОВАРЬ ТЕРМИНОВ

Анализ дисперсионный — статистический метод качественного решения задачи измерения связи. Устанавливает структуру связи между результативным признаком и факторными признаками.

Анализ кластерный — совокупность многомерных статистических методов, предназначенных для формирования относительно "отдаленных" друг от друга групп "однородных" объектов по информации о расстояниях или связях между ними. Используется для анализа структуры совокупности социально-экономических показателей.

Анализ последовательный — методы анализа статистических данных, характерной чертой которых является то, что число производимых наблюдений не фиксируется заранее и объем выборки является случайной величиной, зависящей от значений наблюдений. В этом методе на каждом этапе решается вопрос, прекращать или продолжать выбор.

Анализ факторный — раздел многомерного статистического анализа, объединяющий математико-статистические методы снижения размерности исследуемого многомерного признака.

Варимакс-метод — наиболее распространенный способ ортогонального поворота системы координат, относящийся к проблеме вращения в факторном анализе. Заключается в выборе углов поворота m-мерной системы координат и служит для проведения содержательной интерпретации. Основан на том, что изменение матрицы факторных нагрузок не приводит к изменению редуцированной по этой матрице выборочной корреляционной матрицы исходных признаков.

Время ожидания — 1) время между возникновением требования в системе массового обслуживания и моментом начала обслуживания этого требования; 2) время, в течение которого центральный процессор находится в состоянии ожидания.

Дисперсия — числовая характеристика степени разброса значений случайной величины. Дисперсия постоянной величины равна нулю. Постоянный множитель можно выносить за знак дисперсии, возведя его в квадрат, $D(cx) = c^2D(x)$. Дисперсия суммы двух и более взаимно независимых случайных величин равна сумме дисперсий этих величин:

$$D(x_1 + x_2 + ... + x_n) = D(x_1) + ... + D(x_n).$$

Дисперсия разности двух независимых случайных величин равна сумме их дисперсий.

Дисперсия генеральная — математическое ожидание квадрата отклонения случайной величины от ее математического ожидания. Обозначается D_x или σ_x^2.

Индекс (лат. *index* — показатель) — относительная величина среднего измерения двух состояний одного и того же явления, состоящего из совокупности элементов (товаров, услуг, ресурсов и т. д.).

Компоненты главные — линейные комбинации случайных величин, характеристические векторы ковариационной матрицы. Первой главной компонентой исследуемой генеральной совокупности наблюдений называют такую нормированную линейную комбинацию n исходных признаков, которая обладает наибольшей дисперсией. В статистике используются для определения линейных комбинаций величин с большими дисперсиями.

Коэффициент ковариации — числовая характеристика совместного распределения двух случайных величин. Парный коэффициент ковариации получается в результате деления коэффициента ковариации на произведение средних квадратических отклонений соответствующих случайных величин x_1 и x_2.

Коэффициент корреляции множественный — показатель тесноты линейной статистической связи между одной из случайных величин и совокупностью случайных величин.

Коэффициент корреляции парный — числовая характеристика, выражающая взаимосвязь двух случайных величин, степень их линейной зависимости.

Матрица корреляционная — симметрическая полуопределенная матрица **R** из коэффициентов корреляции элементов k-мерного случайного вектора с ненулевыми дисперсиями. На ее главной диагонали стоят единицы. Выборочная М. К. состоит из оценок корреляции. Используется в многомерном статистическом анализе.

Наблюдение статистическое — планомерный, научно-организованный сбор данных о явлениях и процессах социально-экономической жизни путем регистрации по заранее разработанной программе наблюдения их существенных признаков.

Надежность ЭВМ — способность электронно-вычислительной машины определенное время работать безотказно. Показатели ЭВМ: среднее время безотказной работы, среднее время поиска, локализации и восстановления отказавшего элемента, среднее количество отказов в единицу времени.

Обработка данных — систематизированная последовательность технологических операций, при которой изменяет свое значение хотя бы один из показателей, характеризующих состояние данных. При обработке статистической информации используются сортировка, выборка, корректировка и т. д.

Основа выборки — совокупность единиц, подлежащих изучению, и система ее определения — описание вида единиц, из которых состоит совокупность, и изложение правил включения или не включения любой частоты единицы в состав данной совокупности.

Оценка достаточная — одно из важнейших понятий теории оценивания параметров. Оценка $T = T(x_1, x_2, ..., x_n)$ называется достаточной для параметра θ, если условная плотность выборки при $T(x_1, x_2, ..., x_n) = t$ не зависит от этого параметра.

Оценка статистическая — функция выборочных наблюдений для приближения замены параметра распределения. Например, для нормального распределения случайной величины средняя арифметическая является оценкой математического ожидания.

Оценка точечная — значение неизвестного параметра генеральной совокупности, исчисляемое как функция выборки, не зависящая от оцениваемого параметра.

Оценка эффективная — несмещенная оценка параметра θ_n^*, обладающая минимальной дисперсией среди всего возможного класса оценок этого параметра. Эффективность является решающим свойством, определяющим качество оценки.

Плотность потока событий — основная характеристика потока событий. Определяется как среднее число событий, поступающих за единицу времени. Плотность потока событий широко используется в теории массового обслуживания и теории надежности.

Регрессия — функция $f(x_1, x_2, ..., x_n)$, описывающая зависимость условного математического ожидания зависимой переменной V от заданных фиксированных значений независимых переменных.

Ряд вариационный — упорядоченная по величине последовательность выборочных значений $x_1 \leqslant x_2 \leqslant ... x_n$, полученная в результате преобразования выборки $(x_1, x_2, ..., x_n)$ из генеральной совокупности x с распределением $F(x)$. Член вариационного ряда называется i порядковой статистикой, а i — ранг статистики $x(i)$. Одинаковые члены можно нумеровать в любом порядке. Вариационный ряд используется при исследовании статистических данных с помощью непараметрических и робастных методов.

Пример. При проверке десяти контрольных работ получены оценки: 4, 4, 3, 5, 2, 4, 4, 5, 5, 3. Вариационный ряд имеет вид: 2, 3, 3, 4, 4, 4, 4, 5, 5, 5.

Ряд динамики (динамический ряд, временной ряд) — хронологическая последовательность значений определенного статистического показателя или признака единицы совокупности, называемых уровнями ряда. Уровни рядов динамики бывают абсолютными, относительными или средними величинами. Является важнейшей информацией для изучения и количественной характеристики процесса рассматриваемого объекта.

Себестоимость — выраженные в денежной форме текущие затраты предприятия на производство и реализацию продукции, а также все затраты по использованию в процессе производства и реализации продукции природных ресурсов, сырья, топлива, энергии и др. Отражает результаты деятельности предприятия.

Средняя величина — одна из важнейших категорий статистической науки, широко распространенная форма статистических показателей. Выражает величину признака, отнесенную к единице совокупности и абстрагированную от индивидуальных особенностей отдельных единиц.

Статистика математическая — раздел математики, посвященный математическим методам систематизации, обработки и использования статистических данных для получения научных и практических выводов. Важной частью современной математической статистики стали методы планирования экспериментов, статистического контроля качества и надежности.

Статистика экономическая — общественная наука и отрасль практической статистики, обеспечивающая на основе системы показателей количественную характеристику происходящих в экономике явлений и процессов, выявление основных пропорций, тенденций и закономерностей экономического развития.

Статистическое изучение связи признаков — предполагает использование ряда методов: сопоставление параллельных рядов признаков по единой совокупности, графический; разложение составных показателей, индексный, дисперсионный и др. Для получения количественных характеристик степени взаимосвязи признаков применяется регрессионный и корреляционный анализ.

Структура простая — понятие введено в факторный анализ Терстоуном:

1) каждая сторона матрицы факторной структуры должна содержать хотя бы один нулевой элемент;

2) в каждом столбце матрицы факторной структуры должно быть не менее m нулей, где m — число общих факторов;

3) для каждой пары столбцов матрицы факторной структуры найдется несколько признаков, которые имеют нулевую нагрузку от одного фактора и отличные от нуля в другом;

4) при числе факторов от четырех и более достаточно невелика доля параметров, имеющих в любой паре столбцов одновременно нулевые коэффициенты;

5) для любой пары столбцов факторного отображения найдется мало признаков, соответствующие элементы которых в обоих столбцах отличны от нуля.

Структура факторная — матрица коэффициентов корреляции между признаками и факторами. Коэффициенты корреляции признаков с характерными факторами равны весовым коэффициентам при характерных факторах в модели факторного анализа.

Фактор генеральный — это общий фактор, связанный значимыми коэффициентами веса со всеми переменными.

Фактор общий — это фактор, присущий более чем одной переменной.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. *Абезгауз Г. Г. и др.* Справочник по вероятностным расчетам / Г. Г. Абезгауз, А. П. Тронь, Ю. Н. Копенкин. — М.: ВИМО, 1966.
2. *Абчук В. А. и др.* Справочник по исследованию операций / В. А. Абчук, Ф. А. Матвейчук, Л. П. Томашевский. — М.: ВИМО, 1979.
3. *Айвазян С. А. и др.* Классификация многомерных наблюдений / С. А. Айвазян, З. И. Бежаева, О. В. Староверов. — М.: Статистика, 1974.
4. *Айвазян С. А. и др.* Прикладная статистика: Классификация и снижение размерности / С. А. Айвазян, В. М. Бухштабер, И. С. Енюков. — М.: Финансы и статистика, 1989.
5. *Айвазян С. А., Мхитарян В. С.* Прикладная статистика и основы эконометрики. — М.: ЮНИТИ, 1998.
6. *Андерсон Т.* Введение в многомерный статистический анализ / Пер. с англ. — М.: ГИФМЛ, 1963.
7. *Андрукович П. Ф.* Применение метода главных компонент в практических исследованиях // Межфакультетская лаборатория статистических методов. — Вып 36. — М.: Изд-во МГУ, 1979.
8. *Бабин М. П.* Управление качеством продукции в приборостроении. — М.: Машиностроение, 1976.
9. *Беллман Р., Заде Л.* Принятие решений в расплывчатых условиях // Вопросы анализа и процедуры принятия решений. — М.: Мир, 1975.
10. *Боярский А. Я.* Математика для экономистов. — М.: Госстатиздат, 1961.
11. *Боярский А. Я.* Теоретические исследования по статистике. — М.: Статистика, 1974.
12. *Большев Л. Н., Смирнов Н. В.* Таблицы математической статистики. — М.: Наука, 1965.

13. *Бронштейн О. И., Духовный И. М.* Модели приоритетного обслуживания в информационно-вычислительных системах. — М.: Наука, 1976.
14. *Брудник О. С.* Экономические основы надежности АСУП. — М.: Машиностроение, 1975.
15. *Бро Г. Г., Шнайдман Л. М.* Математические методы экономического анализа на предприятии. — М.: Экономика, 1976.
16. *Бусленко Н. П. и др.* Лекции по теории сложных систем / Н. П. Бусленко, В. В. Калашников, И. Н. Коваленко. — М.: Сов. радио, 1979.
17. *Бусленко Н. П.* Моделирование сложных систем. — М.: Наука, 1978.
18. *Венецкий И. Г., Венецкая В. И.* Основные математико-статистические понятия и формулы в экономическом анализе. — М.: Статистика, 1979.
19. *Вентцель Е. С.* Введение в исследование операций. — М.: Сов. радио, 1964.
20. *Вентцель Е. С.* Исследование операций: Задачи, принципы, методология. — М.: Наука, 1980.
21. *Вентцель Е. С.* Теория вероятностей. — М.: ФМГ, 1964.
22. *Воробьев Н. Н.* Теория игр: Лекции для экономистов-кибернетиков. — Л.: ЛГУ, 1974.
23. *Гантмахер Ф. Р.* Теория матриц. — М.: Наука, 1967.
24. *Гермейер Ю. Б.* Введение в теорию исследования операций. — М.: Наука, 1971.
25. *Гермейер Ю. Б.* Игры с непротивоположными интересами. — М.: Наука, 1976.
26. *Гермейер Ю. Б.* Методологические и методические основы исследования операций и теории игр. — М.: Изд-во МГУ, 1987.
27. *Гнеденко Б. В., Коваленко И. Н.* Введение в теорию массового обслуживания. — М.: Наука, 1966.
28. *Гнеденко Б. В.* Курс теории вероятностей. — М.: Наука, 1965.
29. *Гнеденко Б. В. и др.* Математические методы теории надежности / Б. В. Гнеденко, Ю. К. Беляев, А. Д. Соловьев. — М.: Наука, 1965.
30. *Грень Е.* Статистические игры и их применение: Пер. с польск. — М.: Статистика, 1975.

31. *Дедков В. К., Северцев Н. А.* Основные вопросы эксплуатации сложных систем. — М.: Высшая школа, 1976.
32. *Джейсуол Н.* Очереди с приоритетами: Пер. с англ. — М.: Мир, 1973.
33. *Дружинин Н. К.* Развитие основных идей статистической науки. — М.: Статистика, 1979.
34. *Дубров А. М. и др.* Многомерные статистические методы: Учебник / А. М. Дубров, В. С. Мхитарян, Л. И. Трошин. — М.: Финансы и статистика, 2000.
35. *Дубров А. М. и др.* Моделирование рисковых ситуаций в экономике и бизнесе: Учеб. пособие / А. М. Дубров, Б. А. Лагоша, Е. Ю. Хрусталев. — М.: Финансы и статистика, 1999.
36. *Дубров А. М.* Обработка статистических данных методом главных компонент. — М.: Статистика, 1978.
37. *Дубров А. М.* Последовательный анализ в статистической обработке информации. — М.: Статистика, 1976.
38. *Дюбин Г. Н., Суздаль В. Г.* Введение в прикладную теорию игр. — М.: Наука, 1981.
39. *Езекиэл М., Фокс К.* Методы анализа корреляций и регрессий: Пер. с англ. — М.: Статистика, 1976.
40. *Ефимов Н. В.* Квадратичные формы и матрицы. — М.: Наука, 1967.
41. *Жевержеев В. Ф. и др.* Специальный курс высшей математики для втузов / В. Ф. Жевержеев, Л. А. Кальницкий, Н. А. Сапогов. — М.: Высшая школа, 1970.
42. *Жуковская В. М., Мучник И. Б.* Факторный анализ в социально-экономических исследованиях. — М.: Статистика, 1976.
43. *Заде Л.* Понятие лингвистической переменной и его применение к принятию приближенных решений. — М.: Мир, 1975.
44. *Иберла К.* Факторный анализ: Пер. с нем. — М.: Статистика, 1980.
45. *Интрилигатор М.* Математические методы оптимизации и экономическая теория: Пер. с англ. — М.: Прогресс, 1975.
46. *Кильдишев Г. С., Аболенцев Ю. И.* Многомерные группировки. — М.: Статистика, 1978.

47. *Коваленко И. Н.* Анализ редких событий при оценке эффективности и надежности систем. — М.: Сов. радио, 1980.
48. *Козлов Б. А., Ушаков И. А.* Справочник по расчету надежности. — М.: Сов. радио, 1975.
49. *Кокс Д., Смит У.* Теория очередей: Пер. с англ. — М.: Мир, 1980.
50. *Колемаев В. А. и др.* Теория вероятностей и математическая статистика / В. А. Колемаев, О. В. Староверов, Б. В. Турундаевский. — М.: Высшая школа, 1990.
51. *Колмогоров А. Н.* Основные понятия теории вероятностей. — М.: Наука, 1974.
52. *Кофман А., Крюон Р.* Массовое обслуживание: Теория и приложение: Пер. с франц. — М.: Мир, 1975.
53. *Кофман А., Фор Р.* Займемся исследованием операций: Пер. с франц. — М.: Мир, 1966.
54. *Крушевский А. В.* Теория игр. — Киев: Вища школа, 1977.
55. *Кузин Л. Т.* Основы кибернетики // Математические основы кибернетики. — Т. 1. — М.: Энергия, 1973.
56. *Липаев В. В.* Проектирование автоматического обеспечения АСУ: Системотехника, архитектура, технология. — М.: Сов. радио, 1979.
57. *Липаев В. В.* Распределение ресурсов в вычислительных системах. — М.: Статистика, 1979.
58. *Липаев В. В., Яшков С. Ф.* Эффективность методов организации вычислительного процесса в АСУ. — М.: Статистика, 1975.
59. *Лоули Д., Максвелл А.* Факторный анализ как статистический метод: Пер. с англ. — М.: Мир, 1967.
60. *Лукомский Я. И.* Теория корреляции и ее применение к анализу производства. — М.: Госстатиздат, 1961.
61. *Маленво Э.* Статистические методы эконометрии: Пер. с франц. — М.: Статистика, 1975.
62. *Мальцев А. И.* Основы линейной алгебры. — М.: Наука, 1970.
63. Методика определения экономической эффективности автоматизированных систем управления предприятиями и производственными объединениями. — М.: Статистика, 1979.

64. Методика оценки экономической эффективности отраслевых автоматизированных систем управления (ОАСУ) в промышленных министерствах, всесоюзных и республиканских промышленных объединениях. — М.: Экономика, 1976.
65. *Михельсон В. С.* Элементы вычислительной математики. — М.: Высшая школа, 1966.
66. Многомерный статистический анализ в социально-экономических исследованиях. — М.: Наука, 1974.
67. *Мова В. В. и др.* Организация приоритетного обслуживания в АСУ / В. В. Мова, Л. А. Пономаренко, А. М. Калиновский и др. — Киев: Техника, 1977.
68. *Модин А. А. и др.* Справочник проектировщика АСУП / А. А. Модин, Е. Г. Яковенко, Е. П. Погребной. — М.: Экономика, 1974.
69. *Мот Ж.* Статистические предвидения и решения на предприятии: Пер. с франц. — М.: Прогресс, 1966.
70. *Новиков О. А., Петухов С. И.* Прикладные вопросы теории массового обслуживания. — М.: Сов. радио, 1969.
71. *Овчаров Л. А.* Прикладные задачи теории массового обслуживания. — М.: Машиностроение, 1969.
72. *Окунь Я.* Факторный анализ: Пер. с польск. — М.: Статистика, 1974.
73. *Оуэн Г.* Теория игр: Пер. с англ. — М.: Мир, 1971.
74. *Оуэн Д. Б.* Сборник статистических таблиц. — М.: АН СССР, 1966.
75. *Пасхавер И. С.* Закон больших чисел и закономерности массового процесса. — М.: Статистика, 1966.
76. *Патрик Э.* Основы теории распознавания образов: Пер. с англ. — М.: Сов. радио, 1980.
77. *Плохинский Н. А.* Алгоритмы биометрии. — М.: Изд-во МГУ, 1967.
78. *Плохинский Н. А.* Биометрия. — Новосибирск: Изд-во СО АН СССР, 1961.
79. *Половко А. М.* Основы теории надежности. — М.: Наука, 1964.
80. *Половко А. М.* Сборник задач по теории надежности. — М.: Сов. радио, 1972.
81. Применение методов корреляции в экономических исследованиях. — М.: Наука, 1969.

82. *Пупков К. А.* Статистический расчет нелинейных систем автоматического управления. — М.: Машиностроение, 1965.
83. *Рао С. Р.* Линейные статистические методы и их применение: Пер. с англ. — М.: Наука, 1968.
84. *Решетова З. А., Колошина И. П.* Психологические вопросы построения учебных программ // Теория поэтапного формирования умственных действий и управление процессом учения. — М.: Изд-во МГУ, 1967.
85. *Риордан Дж.* Вероятностные системы обслуживания: Пер. с англ. — М.: Связь, 1966.
86. *Розенберг В. Я., Прохоров А. И.* Что такое теория массового обслуживания. — М.: Сов. радио, 1965.
87. *Росин М. Ф., Булыгин В. С.* Статистическая динамика и теория эффективности систем управления. — М.: Машиностроение, 1981.
88. *Саати Т.* Математические модели конфликтных ситуаций: Пер. с англ. — М.: Мир, 1977.
89. *Саати Т.* Элементы теории массового обслуживания и ее приложения: Пер. с англ. — М.: Сов. радио, 1965.
90. *Самойленко С. И.* Системы обработки информации. — М.: Наука, 1975.
91. *Саульев В. К.* Математические модели теории массового обслуживания. — М.: Статистика, 1979.
92. *Синавина В. С.* Оценка качества функционирования АСУ. — М.: Экономика, 1973.
93. *Сирл С., Госман У.* Матричная алгебра в экономике: Пер. с англ. — М.: Статистика, 1974.
94. *Сиськов В. И.* Экономико-статистические исследования качества продукции. — М.: Статистика, 1971.
95. *Смоляк С. А., Титаренко Б. П.* Устойчивые методы оценивания. — М.: Статистика, 1980.
96. *Солодов А. В.* Теория информации и ее применение к задачам автоматизированного управления и контроля. — М.: Наука, 1967.
97. Справочник по надежности / Под ред. Б. Р. Левина: Пер. с англ. — Т. 1. — М.: Мир, 1969.
98. Статистическое изучение эффективности общественного производства // Ученые записки по статистике / Под ред. В. Е. Адамова. — Т. 30. — М.: Наука, 1977.

99. *Талызина Н. Ф.* Теоретические проблемы программирования обучения. — М.: Знание, 1969.
100. *Талызина Н. Ф.* Управление процессом усвоения знаний. — М.: Изд-во МГУ, 1975.
101. *Тернер Д.* Вероятность, статистика и исследование операций: Пер. с англ. — М.: Статистика, 1976.
102. *Титаренко Б. П.* Статистическое оценивание в условиях «засоренности». — Вып. 3. — М.: Изд-во ЦЭМИ АН СССР, 1970.
103. *Фаддеев Д. К., Фаддеева В. Н.* Вычислительные методы линейной алгебры. — М.: Физматгиз, 1960.
104. *Френкель А. А.* Математические методы анализа динамики и прогнозирования производительности труда. — М.: Экономика, 1972.
105. *Хазен Э. М.* Методы оптимальных статистических решений и задачи управления. — М.: Сов. радио, 1963.
106. *Харман Г.* Современный факторный анализ: Пер. с англ. — М.: Статистика, 1972.
107. *Хинчин А. Я.* Работы по математической теории массового обслуживания. — М.: ГИФМЛ, 1963.
108. *Червоный А. А.* Методы определения и контроля надежности больших систем. — М.: Энергия, 1976.
109. *Червоный А. А. и др.* Надежность сложных систем / А. А. Червоный, В. И. Лукьященко, Л. В. Котин. — М.: Машиностроение, 1972.
110. *Четвериков В. Н., Воробьев Г. Н.* Автоматизированные системы управления предприятиями. — М.: Высшая школа, 1979.
111. *Шварц Г.* Выборочный метод: Пер. с нем. — М.: Статистика, 1978.
112. *Шор Я. Б.* Статистические методы анализа и контроля качества и надежности. — М.: Сов. радио, 1962.
113. *Шураков В. В.* Надежность программного обеспечения систем обработки данных. — М.: Статистика, 1981.
114. Эффективность. Качество: Популярный словарь-справочник / Д. В. Валовой, А. П. Вавилов, Г. Е. Лапшина и др. — М.: Знание, 1979.
115. *Bartlett M. S.* A Note on the Multiplying Factor for Various χ^2 Approximations // J. Royal Statistical Society. — 1954. — № 16. — Ser. B.

116. *Clift Norman.* Orthogonal rotation to congruence // Psychometrika. — 1966. — V. 31. — № 1.
117. *Cowles T. C.* Commission Monograph. — № 13. — New York: Wiley, 1951.
118. *Dantzig G. B.* A proof of the equivalence of the programming and the game problem // Activity Analysis of Production and Allocation; Ed. by Koopmans. — 1960. — V. 1, 8.
119. *Dantzig G. B., Wolfe P.* Decomposition principle for linear programs // Oper. Res. — 1960. — V. 18.
120. *Harris Ch., Knoell D.* The oblique solution factorial Analysis // J. Educational Psychol. — 1948. — V. 39. — № 7.
121. *Holzinger K., Harman H.* Factor analysis. — Chicago, 1941.
122. *Huber P. I.* Robust estimation of a location parameter // The Annals of Mathematical Statistics. — 1964. — V. 35. — № 1.
123. *Jennrich R. J., Sampson P. F.* Rotation for simple loadings // Psychometrika. — 1966. — V. 31. — № 3.
124. *Jowett G.* Factor analysis // Appl. Statistical. — 1958. — V. 7. — № 2.
125. *Kendall M., Smith B.* Factor analysis // J. Royal Statistical Society. — 1950. — Ser. B. — V. 12. — № 1.
126. *Neumann J.* A numerical method to determine optimum strategy // Nav. Res. Logist. Quart. — 1954. — V. 1. — № 2.
127. *Schwidetzky I.* Faktoren des Schädelbaus bei der Vorspanfschen Bevolkerung der Kanarischen Inseln // Homo. — 1959. — V. 10. — № 4.
128. *Sixte F., Wender K.* Der Zusammenhang zwischen multimensionalem Skalieren und Faktorenanalyse // Biometr. — 1964. — Z. 6. — № 4.
129. *Thurstone L.* Multiple-factor analysis. — Chicago, 1947.
130. *Thurstone L.* Factorial analysis of body measurements // Amer. J. Phys. Anthrop. — 1947. — V. 5. — № 1.
131. *Tiit E., Parring A., Mols T.* Tõenäosusteooria ja matemaatiline statistika. — Tallinn: Valgus, 1977.

ОГЛАВЛЕНИЕ

Учебное издание

Дубров Абрам Моисеевич

КОМПОНЕНТНЫЙ АНАЛИЗ И ЭФФЕКТИВНОСТЬ В ЭКОНОМИКЕ

Заведующая редакцией *Л. А. Табакова*
Редактор *Е. В. Стадниченко*
Художественный редактор *Ю. И. Артюхов*
Технический редактор *В. Ю. Фотиева*
Корректоры *Р. А. Скруль, Т. М. Колпакова*
Обложка художника *О. В. Толмачева*

ИБ № 4327

Сдано в набор 06.09.2001
Подписано в печать 25.03.2002. Формат 60×88 1/16
Гарнитура "Таймс". Печать офсетная. Усл. п. л. 21,56
Уч.-изд. л. 18,35. Тираж 3000 экз. Заказ 1447. "С"045

Издательство «Финансы и статистика»
101000, Москва, ул. Покровка, 7
Телефон (095) 925-35-02, факс (095) 925-09-57
E-mail: mail@finstat.ru http://www.finstat.ru

ГУП «Великолукская городская типография»
Комитета по средствам массовой информации Псковской области,
182100, Великие Луки, ул. Полиграфистов, 78/12
Тел./факс: (811-53) 3-62-95
E-mail: VTL@MART.RU