

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Уфимский государственный авиационный технический университет»  
Кафедра технической кибернетики

## ИССЛЕДОВАНИЕ МОДЕЛЕЙ И МЕТОДОВ АНАЛИЗА РИСКОВ В УПРАВЛЕНИИ КАЧЕСТВОМ

Лабораторный практикум  
по дисциплине  
«Системы поддержки принятия решений  
в процессах управления качеством»



Уфа 2021

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«Уфимский государственный авиационный технический университет»  
Кафедра технической кибернетики

## ИССЛЕДОВАНИЕ МОДЕЛЕЙ И МЕТОДОВ АНАЛИЗА РИСКОВ В УПРАВЛЕНИИ КАЧЕСТВОМ

Лабораторный практикум  
по дисциплине  
«Системы поддержки принятия решений  
в процессах управления качеством»

© УГАТУ

Уфа 2021

Авторы-составители: Л. Р. Черняховская, С. В. Сильнова,  
Г. И. Рыжов

Исследование моделей и методов анализа рисков в управлении качеством : лабораторный практикум по дисциплине «Системы поддержки принятия решений в процессах управления качеством» [Электронный ресурс] / Уфимск. гос. авиац. техн. ун-т ; [авт.-сост. : Л. Р. Черняховская, С. В. Сильнова, Г. И. Рыжов]. – Уфа : УГАТУ, 2021. – URL: [https://www.ugatu.ru/media/uploads/MainSite/Ob%20universitete/Izdateli/El\\_izd/2021-80pdf](https://www.ugatu.ru/media/uploads/MainSite/Ob%20universitete/Izdateli/El_izd/2021-80pdf).

Представленные теоретические сведения, модели и методы анализа рисков необходимы для формирования у магистрантов знаний и компетенций для проведения идентификации, анализа и оценки рисков в процессах управления качеством, а также приобретения навыков их применения в управлении организационными системами.

Предназначен для направления 27.04.02 Управление качеством.

Рецензент канд. техн. наук, доцент Н.О. Никулина

При подготовке электронного издания использовались следующие программные средства:

- Adobe Acrobat – текстовый редактор;
- Microsoft Word – текстовый редактор.

Авторы-составители: *Черняховская Лилия Рашитовна*  
*Сильнова Светлана Валерьевна*  
*Рыжов Геннадий Иванович*

Редактирование и верстка: *О. А. Соколова*

Программирование и компьютерный дизайн: *А. П. Меркулова*

Подписано к использованию: 11.06.2021  
Объем 2,14 Мб

ФГБОУ ВО «Уфимский государственный авиационный технический университет»  
450008, Уфа, ул. К. Маркса, 12.  
Тел.: +7-908-35-05-007  
e-mail: rik@ugatu.su

Все права на размножение, распространение в любой форме остаются за разработчиком.  
Нелегальное копирование, использование данного продукта запрещено.

## Введение

Лабораторный практикум регламентирует порядок подготовки и проведения лабораторных работ по дисциплине «Системы поддержки принятия решений в процессах управления качеством» и предназначен для направления подготовки магистров 27.04.02 Управление качеством.

Практикум содержит порядок выполнения лабораторных работ, задания, контрольные вопросы.

Задания для проведения лабораторных работ носят прикладной характер.

Лабораторные работы адаптированы к учебному процессу и ориентированы на формирование у студентов компетенций, которые в соответствии с изучаемыми дисциплинами, будут способствовать получению практических навыков применения теоретических и практических знаний в исследовании реальных производственных процессов.

Практическая значимость лабораторного практикума обусловлена тем, что практикум предназначен дать опыт решения практических задач:

- усвоения методов разработки системы управления безопасностью пищевых продуктов ХАССП;
- освоения методики построения и анализа диаграммы Парето, на основе заданных статистических данных;
- получения практических навыков, необходимых для построения регрессионных моделей;
- получения практических навыков, необходимых для проведения кластерного анализа.

# Лабораторная работа № 1

## АНАЛИЗ РИСКОВ В СООТВЕТСТВИИ СО СТАНДАРТОМ ХАССП

### 1. Цель работы

Целью работы является изучение и усвоения методов разработки системы управления безопасностью пищевых продуктов ХАССП.

### 2. Краткие теоретические сведения

Согласно положениям Технического регламента таможенного союза ТР ТС 021/2011 «О безопасности пищевой продукции» с 15 февраля 2015 года при осуществлении процессов производства (изготовления) пищевой продукции, связанных с требованиями безопасности такой продукции, изготовителями должны разрабатываться, внедряться и поддерживаться процедуры, основанные на принципах ХАССП (англ. *HACCP – Hazard Analysis and Critical Control Points*, анализ рисков и критические контрольные точки) – системы управления безопасностью пищевых продуктов.

Сущность системы ХАССП состоит в том, что процесс изготовления продукции от закупки сырья до потребления готовых изделий делится на стадии с контролем на промежуточных этапах. После каждой последующей стадии риск получить «на выходе» некачественный продукт «уменьшается».

Система ХАССП предусматривает выявление и оценку всех видов опасностей, включая биологические (микробиологические), химические и физические, и выявление всех возможных опасных факторов, которые могут присутствовать в производственных процессах. Опасные факторы, приведенные для групп пищевой продукции в Санитарных правилах и нормах, система ХАССП включает в перечень учитываемых факторов в первую очередь и без изменения.

По каждому потенциальному фактору система ХАССП проводит анализ риска с учетом вероятности появления опасного фактора и значимости его последствий. По результатам анализа система ХАССП определяет и документирует предупреждающие действия, которые устраняют риски или снижают их до допустимого уровня.

К предупреждающим действиям относят:

- контроль параметров технологического процесса производства;
- термическую обработку;
- применение консервантов;
- использование металлодетектора;
- периодический контроль концентрации вредных веществ;
- мойку и дезинфекцию оборудования, инвентаря, рук и обуви и др.

В основе системы ХАССП лежит семь принципов, последовательная реализация которых позволяет разработать, внедрить и успешно управлять ХАССП на предприятии:

– **принцип 1:** идентификация потенциального риска или рисков (опасных факторов), которые сопряжены с производством продуктов питания, начиная с получения сырья (разведения или выращивания) до конечного потребления, включая все стадии жизненного цикла продукции (обработку, переработку, хранение и реализацию) с целью выявления условий возникновения потенциального риска (рисков) и установления необходимых мер для их контроля;

– **принцип 2:** выявление критических контрольных точек в производстве для устранения (минимизации) риска или возможности его появления, при этом рассматриваемые операции производства пищевых продуктов могут охватывать поставку сырья, подбор ингредиентов, переработку, хранение, транспортирование, складирование и реализацию;

– **принцип 3:** определение критических пределов для каждой критической контрольной точки;

– **принцип 4:** разработка системы мониторинга, позволяющая обеспечить контроль критических контрольных точек на основе планируемых мер или наблюдений;

– **принцип 5:** разработка корректирующих действий и применение их в случае отрицательных результатов мониторинга;

– **принцип 6:** разработка процедур проверки, которые должны регулярно проводиться для обеспечения эффективности функционирования системы ХАССП;

– **принцип 7:** документирование всех процедур системы, форм и способов регистрации данных, относящихся к системе ХАССП.

Внедрение системы ХАССП дает ряд преимуществ предприятию:

- гарантированное производство качественной продукции;
- повышение конкурентоспособности продукции;
- улучшение экономических показателей деятельности предприятия (рост прибыли, уменьшение затрат) за счет роста производительности;
- повышение степени доверия со стороны представителей надзорных органов и, следовательно, уменьшение количества инспекционных проверок.

Выделяют двенадцать этапов внедрения принципов ХАССП.

**Этап 1.** Определение области применения системы ХАССП.

**Этап 2.** Подбор и назначение группы ХАССП.

**Этап 3.** Описание продукции.

**Этап 4.** Определение предполагаемой области применения продукции.

**Этап 5.** Составление группой ХАССП блок-схемы производственных процессов и, при необходимости, планов производственных помещений.

Цель блок-схемы – представление производственного процесса в виде четкой, простой последовательности шагов, из которых состоит процесс. Область, включенная в блок-схему, должна охватывать все стадии производственного процесса, находящиеся под непосредственным контролем предприятия (т.е. все технологические операции от поступления ингредиентов до поставки продукции потребителю). Так, блок-схема может включать этапы, осуществляемые до и после стадий обработки, происходящих на предприятии, например, распределение, розничная продажа и обработка потребителем, а также некоторые процессы, осуществляемые поставщиком сырья. Состав блок-схемы устанавливается в техническом задании. Блок-схема течения процесса не должна быть сложной (как технические чертежи). Блок-схема наиболее информативна. Простая схема предприятия часто полезна при изучении и оценке продукта и производственного процесса.

Поскольку блок-схема технологического процесса используется в качестве основы при проведении анализа возможных опасностей, она должна содержать достаточно технических элементов и деталей



для более точного представления технологического процесса, и производственных условий. Наличие такой информации обеспечит более тщательное рассмотрение каждого этапа и выявление всех биологических, химических, физических опасностей, связанных с этапом, а также позволит правильно судить о вероятности возникновения той или иной опасности и условий, приводящих к ней.

Таким образом, на блок-схеме, планах или приложениях к ним должны быть приведены дополнительные сведения:

- элементы сырья и упаковки изделия, включая нормы и необходимые условия хранения;

- режимы переработки и условия хранения на этапах процесса, что необходимо учитывать при анализе микробиологических опасностей, поскольку важно оценить возможность роста патогенных микроорганизмов;

- контролируемые параметры технологического процесса, периодичность и объем контроля (схемы производственного контроля);

- инструкции о процедурах уборки, дезинфекции и дезаэрации, а также гигиене персонала;

- план размещения оборудования;

- техническое обслуживание и мойка оборудования и инвентаря, а также виды оборудования и составные части оборудования, которые трудно очистить и которые могут стать источником загрязнения;

- петли возврата, доработки и переработки продукции, а также возможные задержки сырья, полуфабриката на каком-либо этапе;

- пункты санитарной обработки, расположение туалетов, умывальников, хозяйственно-бытовых зон;

- пункты возможных перекрестных загрязнений от сырья, обрабатываемой и конечной продукции, добавок, смазочных материалов, хладагентов, персонала, упаковки, поддонов и контейнеров;

- система вентиляции;

- условия распределения и хранения в пунктах розничной продажи (если это предусмотрено в техническом задании) и др. Стиль блок-схемы – это выбор самих разработчиков системы ХАССП, здесь нет определенных ограничений и правил. Приведем несколько примеров построения блок-схем, рис. 1.1.



Рис. 1.1. Блок-схема технологического процесса производства мороженого с использованием модульного подхода

На рис. 1.1 и рис 1.2 представлены примеры диаграмм технологического процесса производства мороженого в виде блок-схем с использованием модульного подхода.



Рис. 1.2. М2: Неосновные компоненты – приемка и хранение

**Этап 6.** Проверка (верификация) информации.

**Этап 7.** Определение перечня потенциально опасных факторов и предупреждающих действий (устранение причин потенциально опасных факторов), табл. 1.1.

Таблица 1.1

Таблица анализа рисков в производственном процессе

Наименование операции	Учитываемый опасный фактор	Контролируемые признаки	Предупреждающие действия

**Этап 8.** Определение критических контрольных точек – места проведения контроля для идентификации опасного фактора и (или) управления риском. Критические контрольные точки определяются по методу «Дерево принятия решения» на основе анализа каждого учитываемого опасного фактора при последовательном рассмотрении всех операций, включенных в блок-схему производственного процесса.

Критическая контрольная точка – это этап или процедура, где необходимо применение контроля, чтобы предотвратить, устранить или уменьшить опасность до приемлемого уровня. Примерами критических контрольных точек могут служить: температурная обработка, охлаждение, проверка ингредиентов на присутствие остатков химических веществ, контроль за составом продукта, проверка продукта на загрязнение металлами. Так, процесс пастеризации с заданной температурой и временем воздействия является критической контрольной точкой, поскольку в течение этого процесса определенные микроорганизмы должны быть уничтожены. Процесс охлаждения готового изделия – критичный с точки зрения безопасности, поскольку для предотвращения развития опасных микроорганизмов необходимо соблюдать определенное время охлаждения и нужную температуру. Критические контрольные точки должны быть тщательно изучены, а все данные по ним – задокументированы.

Не существует ограничений для количества критических контрольных точек, это зависит от сложности и вида продукции, производственного процесса, подвергающихся анализу. Критические контрольные точки, определенные для продукта на одной производственной линии, могут отличаться от критических контрольных точек для такого же продукта на другой. Это объясняется тем, что опасные факторы и точки для их контроля могут изменяться в зависимости от:

- планировки завода;
- рецептуры;
- протекания процессов;
- оборудования;
- выбранных ингредиентов;
- санитарных и вспомогательных программ;
- географического расположения предприятия.

Критическая контрольная точка может характеризовать сырье, место, методику, процедуру или стадию процесса при следующих условиях:

- отсутствие конкретных загрязняющих веществ в сырье;
- определенная операция по очистке;
- разделение установок для сырья и продуктов, подвергавшихся кулинарной обработке;
- хлорирование охлаждающей воды в контейнерах и т.д.

Критические контрольные точки могут быть определены путем несложных рассуждений и заключений рабочей группы ХАССП с использованием собранной информации о процессе, возможных опасностях, контрольных и предупреждающих воздействиях. Однако из-за разногласий в месторасположении точек может быть определено больше критических контрольных точек, чем необходимо. Например, при неудачном определении металла в сырье, можно положиться на датчик, который находится в конце линии и обеспечить безопасность продукта. Но слишком большое количество критических контрольных точек может дать прямо противоположный эффект и стать причиной возникновения опасности в продукте. То есть не будет обеспечено должного внимания к действительно критичным этапам производственного процесса.

Для точного определения критических контрольных точек разработан инструмент – дерево принятия решений. Это диаграмма, которая описывает ход логических рассуждений при изучении опасности на каждом этапе производственного процесса. Отвечая последовательно на вопросы дерева принятия решений, команда ХАССП принимает решение о целесообразности установления критических контрольных точек на данном этапе.

Применение дерева принятия решений должно быть гибким, с учетом того, где происходит процесс: в производстве, на этапе заготовки сырья, переработки, хранения, реализации и др. Члены рабочей группы должны использовать дерево принятия решений в описанной ниже последовательности, руководствуясь при этом здравым смыслом. Следует отметить, что этот метод не может применяться во всех ситуациях.

Существуют и другие версии дерева принятия решений с различающимися формулировками, но с общим подходом к определению местоположения критических контрольных точек.

Рассмотрим более подробно варианты дерева принятия решений, которые учитывают особенности анализа процесса и сырья.

Для определения **критических контрольных точек процесса** необходимо ответить на вопросы последовательно по каждому этапу, где выявлены значимые опасные факторы, и по каждому установленному опасному фактору. На рис. 1.3 изображено дерево принятия решений для анализа опасностей процесса и приведены пояснения хода логических рассуждений.

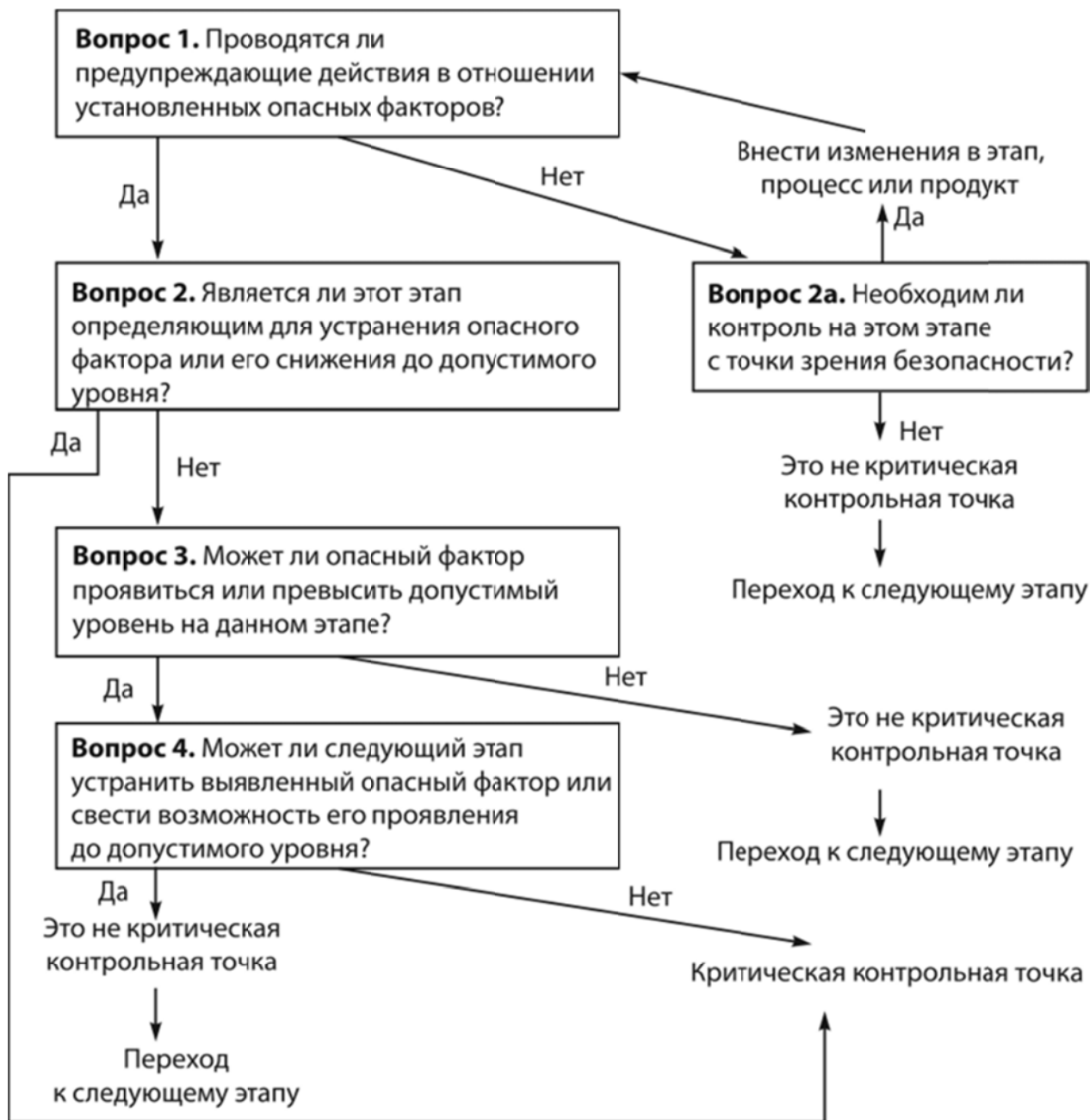


Рис. 1.3. Дерево принятия решений по критическим контрольным точкам процесса

### **Вопрос 1.** Существует ли опасность на этом этапе?

Несмотря на очевидность ответа, этот вопрос помогает рабочей группе сосредоточиться на определенном этапе процесса. В принципе данный вопрос необязательный, хотя он полезен, если прошло достаточно времени после проведения анализа потенциальных опасностей. Например, идентифицированная ранее во время наибольшего проявления опасность при последующем рассмотрении не является таковой – это может быть проблемой качества. Если все-таки опасность подтверждена, то следует перейти к **вопросу 2**.

**Вопрос 2.** Проводятся ли предупреждающие действия в отношении установленных опасных факторов?

Если предупреждающие действия проводятся, то группа переходит к рассмотрению **вопроса 3**. Если они не проводятся, то группа должна определить, необходимо ли организовать на этом этапе контроль для обеспечения безопасности продукта, т.е. ответить на **вопрос 2а**. Если контроль необязателен, то установление критической контрольной точки на данном этапе не требуется и можно переходить к изучению следующего этапа производственного процесса, начиная отвечать на вопросы дерева принятия решения сначала.

Если контроль необходим, группа должна подготовить предложения по внесению изменений в этап или процесс, чтобы контроль безопасности пищевого продукта стал возможным. Например, если на данном этапе возможно присутствие бактерии сальмонеллы и существующий способ нагрева недостаточно эффективен для ее уничтожения, то необходимо либо изменить параметры процесса, либо пересмотреть метод контроля. Когда необходимые модификации контроля проведены, следует снова ответить на вопросы дерева принятия решений.

Важно, чтобы любые необходимые изменения, разработанные командой ХАССП, были своевременно внедрены.

**Вопрос 3.** Является ли этот этап определяющим для устранения опасного фактора или его снижения до допустимого уровня?

Отвечая на этот вопрос, рабочая группа должна учесть технические показатели продукта (например, *pH*, *Aw*, концентрация консервантов и т.д.) и процесса. Ключевым моментом в решении этого вопроса является то, что здесь рассматривается именно шаг процесса, а не мера контроля. Если члены команды неверно

рассматривают меру контроля, то это приводит к установлению ненужной критической контрольной точки. Этот вопрос составлен для определения шагов процесса, с помощью которых возможно управление определенными видами опасности. Смысл вопроса заключается в выявлении, связан ли данный этап непосредственно с уничтожением опасности.

Например, пастеризация молока при температуре  $71,7^{\circ}\text{C}$  в течение 15с предназначена для контроля за вегетативными патогенными микроорганизмами, в то время как хранение сырья не защищает от заражения.

Разработчики должны тщательно рассмотреть информацию об анализе опасности наряду с блок-схемой производственного процесса, чтобы правильно ответить на этот вопрос. Важно рассмотреть этапы, на которых производится смешивание ингредиентов: если ингредиенты не смешаны должным образом, то применяемые методы контроля могут быть не эффективны, т.е. дальнейшая обработка не устраняет опасности, как в случае термической обработки.

Если группа считает, что ответ должен быть положительным и на данном этапе существует критическая контрольная точка, то следует отвечать на вопросы дерева принятия решений сначала для следующего этапа процесса. Если ответ на этот вопрос отрицательный, группе необходимо перейти к **вопросу 4**.

**Вопрос 4.** Может ли опасный фактор проявиться или превысить допустимый уровень на данном этапе?

Рабочая группа должна использовать данные из технологической схемы и информацию, полученную при изучении производственной линии, чтобы определить, не может ли изучаемый опасный фактор находиться в производственной среде (например, персонал, оборудование, стены, полы, система канализации, сырье), которая в этом случае способна вызвать загрязнение продукта.

Таким образом, чтобы дать объективный ответ на этот вопрос, необходимо знать:

- не осуществляется ли процесс в условиях, которые могут содержать опасный фактор;
- имеет ли упаковка продукта важное значение для предотвращения загрязнения на этой стадии;



- возможно ли перекрестное загрязнение от другого продукта или сырья;
- возможно ли загрязнение или повторное загрязнение от персонала;
- нет ли в оборудовании какого-либо пространства, где может накапливаться и застаиваться продукт, вызывая увеличение опасного фактора;
- не могут ли время и температурные условия хранения продукта в нерасфасованном виде вызвать нарастание опасного фактора в продукте;
- существуют ли какие-либо другие факторы или условия, приводящие к увеличению опасности загрязнения на данном этапе и приближению к недопустимым уровням.

В тех случаях, когда при определении пределов недопустимых уровней отсутствует уверенность, в каком случае загрязнение начинает представлять опасность, важно, чтобы команда ХАССП собрала совет перед принятием решения.

Если ответ на **вопрос 4** отрицательный, на данном этапе не существует критических контрольных точек, и необходимо вернуться к началу дерева принятия решений для исследования следующей опасности на других этапах.

Если ответ на **вопрос 4** положительный, следует перейти к **вопросу 5**.

**Вопрос 5.** Может ли следующий этап устранить выявленный опасный фактор или свести возможность его появления до допустимого уровня?

Здесь рабочая группа должна рассмотреть все следующие этапы производственной блок-схемы, а также приготовление продукта потребителем и определить, устранит ли один из них опасный фактор или сведет возможность его возникновения до допустимого уровня. Таким образом, это минимизирует число этапов процесса, которые являются критическими контрольными точками и подлежат контролю для обеспечения безопасности изделия. Если ответ на вопрос положительный, то данный этап не является критической контрольной точкой и не будет учитываться в последующих действиях, а группа должна перейти к анализу следующего этапа. Например, правильное приготовление сырых мясных изделий потребителем уменьшит возможность возникновения микробиоло-

гических опасностей. Также обнаружение металлических частиц в готовом изделии на стадии упаковки, которые могли попасть вместе с сырьем или на ранних стадиях процесса, сможет предотвратить попадание такой продукции к потребителю. Если ответ на этот вопрос отрицательный, то данный этап процесса является критической контрольной точкой для рассматриваемой опасности.

Важно отметить следующее. Хотя этот вопрос позволяет свести к минимуму число критических контрольных точек, это может строго не соблюдаться во всех случаях. В вышеупомянутом примере обнаружения металлических посторонних частиц единственная критическая контрольная точка, которая является абсолютно критической, – это обнаружение металлов на стадии упаковки готового изделия. Однако с коммерческой точки зрения раннее обнаружение металлов или других опасностей предпочтительнее, поскольку позволяет свести возможные потери изделия к минимуму, поэтому могут быть введены дополнительные профилактические контрольные точки. Также необходимо принимать во внимание то, насколько технически возможно и эффективно проведение контроля на данном этапе процесса. Примером этого может служить обнаружение осколков костей в мясе с помощью рентгена, где в данном случае может применяться только один рентгеновский датчик. На этом месте необходимо установить критическую контрольную точку, так как она будет иметь наибольшее значение по сравнению с другими применяемыми средствами и методами контроля.

Таким образом, необходимо продолжать работать с деревом принятия решений для анализа всех опасностей на каждом этапе процесса до тех пор, пока не будут определены все критические контрольные точки.

Ответы на вопросы дерева принятия решений необходимо фиксировать, причем повторное рассмотрение вопросов для тех же самых опасностей в случае, если не были внесены какие-либо изменения, должны давать аналогичные ответы. Удобно строить для этих целей матрицу «вопрос и ответ», где будут отражены итоги работы с деревом принятия решений по каждому этапу процесса, на котором была идентифицирована опасность. Лучше всего эту таблицу разместить в специальной форме в качестве дополнительной к блок-схеме анализа опасностей.

Далее необходимо выявить из числа ККТ те, которые уже обеспечены достаточным контролем, чтобы исключить дублирование работ. В соответствии с ГОСТ 51705.1-2001 (п. 4.4.3) с целью сокращения количества критических контрольных точек без ущерба для обеспечения безопасности к ним не следует относить точки, для которых выполняются следующие условия:

– предупреждающие воздействия, которые осуществляются систематически в плановом порядке и регламентированы в Санитарных правилах и нормах, в системе технического обслуживания и ремонта оборудования, в процедурах системы качества и других системах менеджмента предприятия;

– выполнение предупреждающих воздействий, не относящихся к контрольным точкам, оценивается группой ХАССП и периодически проверяется при проведении внутренних проверок. Прежде всего необходимо удостовериться, что контроль осуществляется по тем параметрам, которые обеспечивают безопасность, т.е. устраняют опасный фактор или сводят его до допустимого уровня и с достаточной периодичностью. На рис. 1.4 изображено дерево принятия решений при анализе сырья, а также приведены пояснения хода логических рассуждений.

Чтобы определить, относится ли какой-либо из видов сырья (включая ингредиенты, воду и упаковочный материал) к критической контрольной точке, рабочая группа должна ответить на **вопрос 1** (и при необходимости на **вопросы 2** и **3**) для каждого вида используемого сырья.

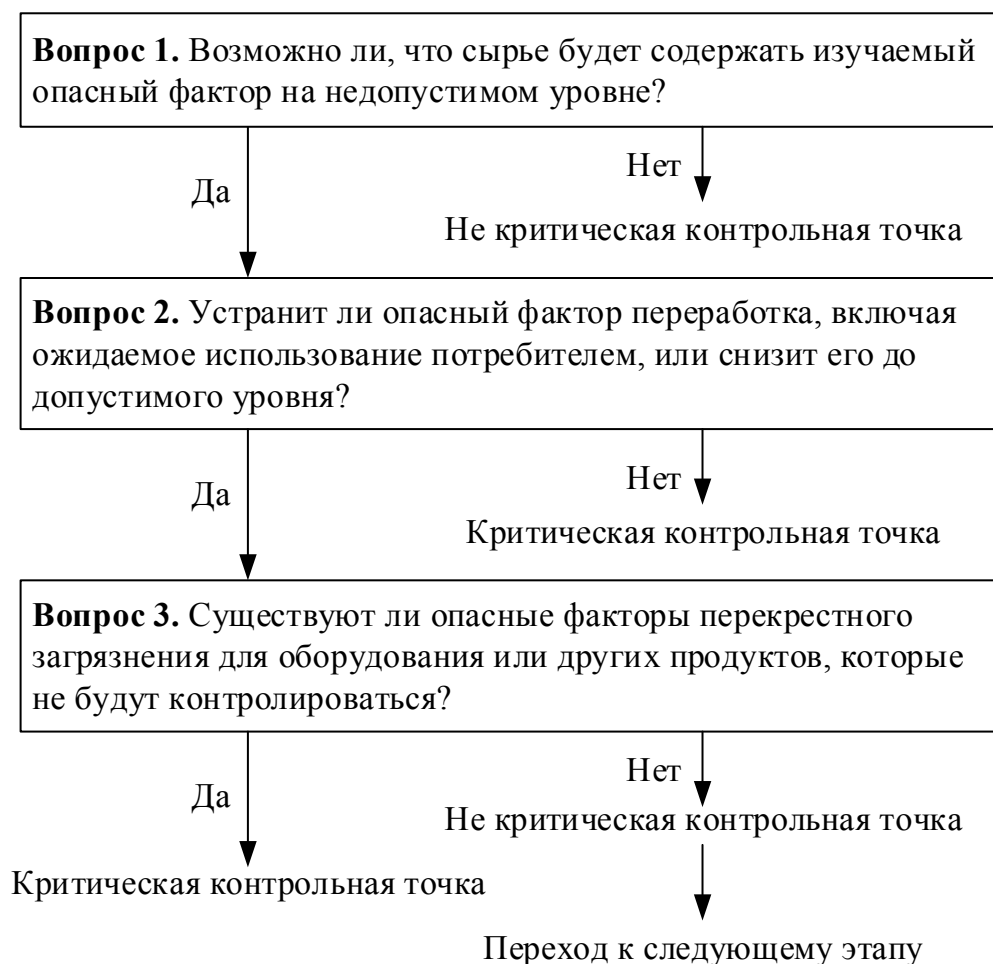


Рис. 1.4. Дерево принятия решения по критическим контрольным точкам сырья

**Вопрос 1.** Возможно ли, что сырье будет содержать изучаемый опасный фактор на недопустимом уровне?

Рабочая группа должна дать ответ на этот вопрос с учетом, например, эпидемиологической информации, прежних показателей деятельности поставщика или информации, связанной с вопросами безопасности продукта. Если рабочая группа уверена, что ответ будет отрицательным, то сырье не следует рассматривать в качестве критической контрольной точки. Если же члены рабочей группы не уверены в ответе, то они должны принять положительный ответ и перейти к **вопросу 2**.

**Вопрос 2.** Устранит ли опасный фактор переработка, включая ожидаемое использование потребителем, или снизит его до допустимого уровня?

Рабочая группа, предполагая, что опасный фактор присутствует в сырье, должна последовательно изучить производственный процесс

с использованием технологической схемы и обследовать производственную линию для того, чтобы установить, устранят ли данный опасный фактор какие-либо стадии (включая использование потребителем) или снизят его до безопасного уровня. Если ответ на этот вопрос будет положительным, то рабочая группа должна перейти к **вопросу 3**. Если ответ отрицательный, то качество сырья является критическим.

**Вопрос 3.** Существуют ли опасные факторы перекрестного загрязнения для оборудования или других продуктов, которые не будут контролироваться?

Если рабочая группа дает положительный ответ на **вопрос 3**, то качество сырья является критическим. Если ответ отрицательный, то качество сырья не является критическим, и эксперты переходят к анализу следующего вида сырья.

Результаты исследований также необходимо зафиксировать в специальной таблице анализа опасных факторов. И также надлежит исключить из числа выявленных критических контрольных точек те, которые уже подлежат контролю в других системах предприятия, таких как система теххимического контроля сырья, система менеджмента качества.

Таким образом, осуществление процедур контроля в критических контрольных точках обеспечивает безопасность производимой продукции. Если же в техническом задании установлено, что система ХАССП охватывает не только вопросы безопасности, но и вопросы качества, то определяются не только критические контрольные точки, но и так называемые критические точки для качества.

Критическая точка для качества – это этап или процедура, где проявление качественного опасного фактора можно устранить или уменьшить применением контроля.

Для однозначного определения, является ли данный этап критической контрольной точкой или контрольной точкой для качества, можно использовать схему, приведенную на рис. 1.5.



Рис. 1.5. Схема принятия решения о критической контрольной точке

Когда настоящий этап разработки плана ХАССП завершен, т.е. определены все критические контрольные точки, необходимо отметить их на блок-схеме производственного процесса и перейти к следующему этапу – определению критических пределов.

Установление критических пределов для критических контрольных точек

Рабочая группа должна установить для каждой идентифицированной критической контрольной точки критические пределы.

**Критические пределы** – это максимальные или минимальные значения биологического, химического или физического параметра, требующего контроля в критической контрольной точке в целях предотвращения, уничтожения присутствующего загрязнения или уменьшения его величины до приемлемого уровня. Критические пределы используются, чтобы показать различия между безопасными и небезопасными производственными условиями в критической контрольной точке. Они показывают момент, когда допустимая (контролируемая) ситуация переходит в недопустимую (неконтролируемую) в смысле безопасности конечного продукта.

По каждой критической контрольной точке критические пределы должны устанавливаться по одному или нескольким параметрам, т.е. в каждой критической контрольной точке будет

проводиться одно или более контрольных измерений, для того чтобы гарантировать, что опасность предотвращена или сведена до приемлемого уровня.

Так как критические пределы определяют границы между безопасными и опасными производственными условиями, важно, чтобы они были правильно установлены. Для того чтобы установить соответствующие критические пределы, рабочая группа должна изучить все критерии, влияющие на безопасность в каждой критической контрольной точке. Другими словами, необходимо детализировать опасность наряду с факторами, влияющими на предотвращение опасности или контроль. Причем критические пределы необязательно должны быть идентичны параметрам обработки.

Каждая критическая контрольная точка может иметь различные факторы, требующие контроля для обеспечения безопасности изделия, и каждый из этих факторов будет иметь соответствующий критический предел. Таким образом, все факторы, связанные с безопасностью в критической контрольной точке, должны быть идентифицированы. А уровень, при котором каждый фактор становится границей между опасным и безопасным, является критическим пределом. Факторами, которые обычно используются как критические пределы, являются: температура, время, *pH*, влажность, активность воды, концентрация соли, кислотность, содержание хлора, вязкость, наличие консервантов. При этом предпочтительнее использовать измеримый фактор, который может быть проверен испытанием или наблюдением, но могут быть использованы и органолептические показатели, такие как запах или внешний вид.

Многочисленные типы критических пределов связаны с контрольными мерами. Они могут иметь минимальное значение (например, минимальное значение устанавливается для времени и температуры обработки) или могут быть максимальной величиной (например, длительность хранения на складе). Для других критических контрольных точек необходимо, чтобы параметр находился в диапазоне между минимальным и максимальным значением, например, нитрит в беконе, где минимальный уровень – показатель микробиологической безопасности, а максимальный – гарантия химической безопасности.

Разработчики систем безопасности на различных предприятиях могут прийти к различным решениям при определении влияющих факторов и критических пределов. Например, тепловая обработка всегда принимается за критическую контрольную точку для уничтожения микроорганизмов. Предположим, что нагревание изделия до  $70^{\circ}\text{C}$  в течение 15с достаточно для обеспечения безопасности. Чтобы быть уверенными, что необходимые температура и время соблюдаются, рабочая группа на одном предприятии принимает решение выработать критические пределы для температуры в печи, для влажности, скорости конвейера в печи, толщины изделия. Контроль за этими факторами обеспечит эффективную тепловую обработку каждого изделия в течение 15 с при температуре  $70^{\circ}\text{C}$ . На другом предприятии разработчики системы ХАССП отслеживают температуру в печи и время нахождения изделия в печи.

Выделяют следующие виды критических пределов: химические, физические, процедурные и микробиологические.

**Химические пределы** могут быть связаны с возникновением химических заражений в изделии или его компонентах, или с контролем за микробиологическим заражением через определенные факторы. Можно привести следующие примеры химических пределов: приемлемые уровни для микотоксинов, *pH*, концентрация соли, активности воды *A<sub>w</sub>*, или маркировка, отсутствие аллергенов.

**Физические пределы** связаны с приемлемостью к физическим или инородным материалам. Однако они также могут быть при контроле микробиологических заражений, когда выживание или уничтожение микроорганизмов управляются физическими параметрами. Примерами факторов, связанных с физическими пределами, являются: отсутствие металла, неповрежденное решето (сито), размер ячеек сита, задержание, температура и время.

**Процедурные пределы** связаны с процедурными мерами контроля. Например, некая процедура, которая обязательно должна быть осуществлена, может стать критическим пределом. Точно так же пределом может быть подтверждение статуса, где мерой контроля является гарантия поставщика относительно специфических опасностей.

**Микробиологических пределов** следует избегать в системе безопасности, за исключением контроля нескорпортующегося сырья,



поскольку микробиологические испытания могут осуществляться только в лаборатории и в течение нескольких дней. Такой контроль не позволяет быстро определять отклонение процесса и своевременно принимать корректирующие меры. То есть возможна ситуация, когда производство продолжает выпускать продукцию в то время, когда опасность существует и о ней никто не подозревает. Нежелательность установления микробиологических пределов также обусловлена тем, что микроорганизмы редко гомогенно распределяются по всей партии и поэтому не могут быть обнаружены в отобранных образцах для анализа и в то время не могут присутствовать в других образцах продукции из той же партии. Использовать микробиологические пределы можно, только если материал гомогенен.

Микробиологические показатели лучше использовать в целях проверки, т.е. когда проводятся дополнительные испытания для подтверждения эффективности системы ХАССП. Единственным исключением из этого правила является возможность проведения микробиологических испытаний за короткий промежуток времени.

Критические пределы, основанные на субъективных данных, например, на визуальном контроле, должны сопровождаться точными требованиями в отношении допустимого уровня.

Критические пределы должен устанавливать персонал, знающий процесс производства и регламентированные требования к данной продукции. Иногда для установления критических пределов требуется проведение экспериментов, математическое моделирование, например, моделирование на компьютере выживания и характеристик роста определенных микроорганизмов в пищевых продуктах. В других случаях можно руководствоваться авторитетной технической информацией или нормативными документами (ГОСТы, руководства, литературные обзоры). Тогда критические пределы могут соответствовать или быть строже регламентированных требований. Можно прибегнуть к советам экспертов-консультантов, ассоциаций, специалистов заводов – производителей оборудования, микробиологов, инженеров. В любом случае критические пределы должны быть научно обоснованы.

Вся полученная рабочей группой ХАССП информация по критическим контрольным точкам, критическим пределам, а также мониторингу, корректирующим действиям и документированию должна быть сведена в специальную форму ХАССП плана, табл. 1.2.

Форма рабочего листа ХАССП

Наименование операции	Опасный фактор	Номер критической контрольной точки	Контролируемый параметр и его предельные значения	Процедура мониторинга	Контролирующие действия	Регистрационно-учетный документ
1	2	3	4	5	6	7

Для особо ответственных операций можно устанавливать еще один уровень контроля, так называемые рабочие пределы – это пределы допуска для тех же самых критических контрольных точек, но более строгие. Они применяются для предупреждения перехода ситуации в неконтролируемую, принятия соответствующих мер прежде, чем это произойдет. В результате снижается риск отклонения, например, критические пределы для разрушения вегетативных патогенов нагревом – 65,6°С в течение 30 мин. Для недопущения отклонений в процессе могут устанавливаться рабочие пределы 68,5°С в течение 30 мин.

**Этап 9.** Установление критических пределов для каждой критической контрольной точки (согласно соответствующим требованиям). Критический предел – это величина, отделяющая допустимый уровень от недопустимого. Для каждой критической контрольной точки критические пределы устанавливаются по одному или нескольким параметрам.

**Этап 10.** Создание системы мониторинга для каждой критической контрольной точки (что? как? когда? кто?). Мониторинг – это система постоянных наблюдений или измерений, которая позволяет удостовериться, что критические контрольные точки находятся под контролем, и сделать точные регистрационные записи для будущих проверок.

**Этап 11.** Разработка плана корректирующих действий (браковка партии, проверка средств измерений, наладка оборудования, изоляция несоответствующей продукции, переработка несоответствующей продукции или утилизация).

**Этап 12. Документирование.** К основным документам программы ХАССП относятся:

- политика в области безопасности выпускаемой продукции;
- приказ по группе ХАССП;
- информация о продукции и производстве;
- отчеты и рабочие листы ХАССП;
- процедуры мониторинга и проведения корректирующих действий;
- программа внутренней проверки системы ХАССП;
- перечень регистрационно-учетной документации.

Неотъемлемой частью системы ХАССП являются внутренние проверки, которые проводятся непосредственно после внедрения системы, а затем с установленной периодичностью не реже одного раза в год.

Свидетельством того, что предприятие разработало и внедрило систему ХАССП, служит ее сертификация.

Сертификация системы ХАССП является гарантом того, что все виды деятельности на предприятии, которые влияют на качество и безопасность продукции, идентифицированы и эффективно управляются. Свидетельством успешной сертификации системы ХАССП служит маркировка продукции соответствующим знаком.

### **3. Порядок выполнения лабораторной работы**

Лабораторная работа должна выполняться в следующей последовательности:

1. Изучить материал, представленный выше.
2. Получить у преподавателя задание. Варианты заданий приведены в Приложение А.
3. Составить таблицу анализа рисков в производственном процессе.
4. Построить дерево решений и определить критические контрольные точки.
5. Оформить полученные результаты.

#### **4. Требования к отчету**

Отчет должен содержать:

1. Титульный лист.
2. Задание.
3. Цель работы.
4. Исходные данные для выполнения работы, в соответствии с заданным вариантом.
5. Основные результаты выполнения работы.
6. Выводы по работе.

#### **СПИСОК ЛИТЕРАТУРЫ**

1. ГОСТ Р 51705.1-2001. Управление качеством пищевых продуктов на основе принципов ХАССП. Общие требования
2. Австриевских А. Н. Управление качеством на предприятиях пищевой и перерабатывающей промышленности: учебник / А. Н. Австриевских, В. М. Кантере, И. В. Сурков, Е. О. Ермолаева. – 2-е изд., испр. и доп. – Новосибирск: Сиб. унив. изд-во, 2007. – 268 с.: ил.
3. Сурков И. В. Управление качеством на предприятиях пищевой, перерабатывающей промышленности, торговли и общественного питания [Электронный ресурс] / И. В. Сурков. – URL: [https://studref.com/353900/tovarovedenie/upravlenie\\_kachestvom\\_na\\_predpriyatiyah\\_pischevoy\\_pererabatyvayuschey\\_promyshlennosti\\_torgovli\\_i\\_obshchestvenn](https://studref.com/353900/tovarovedenie/upravlenie_kachestvom_na_predpriyatiyah_pischevoy_pererabatyvayuschey_promyshlennosti_torgovli_i_obshchestvenn). (01.01.2021).

# **ЛАБОРАТОРНАЯ РАБОТА № 2**

## **ИЗУЧЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА РИСКОВ: ДИАГРАММА ПАРЕТО**

### **1. Цель работы**

Целью работы является освоение методики построения и анализа диаграммы Парето, на основе заданных статистических данных.

### **2. Краткие теоретические сведения**

Проблемами сбора, обработки и анализа результатов производственной деятельности занимается математическая статистика, которая включает в себя большое количество современных методов анализа и выявления дефектов. Статистические методы управления качеством являются обязательным элементом современных систем менеджмента качества, внедряемых на российских предприятиях. К таким методам можно отнести корреляционный и регрессионный анализы, проверку статистических гипотез, факторный анализ, построение диаграммы Парето, анализ безотказности и т.д.

#### **2.1. Построение диаграммы Парето**

При контроле качества анализ Парето – это способ организации данных, показывающий, из каких основных факторов состоит анализируемый объект.

Диаграммой Парето называют такое построение, в котором строятся полосы или столбцы гистограммы в нисходящем порядке, начиная слева. Основой графика Парето является правило «80...20»: 80% проблем являются результатом 20 % причин.

Таким образом, выяснив причины появления основных дефектов, можно устранить почти все потери, сосредоточив усилия на ликвидации именно этих потерь.

Диаграмма Парето – это инструмент, позволяющий распределить усилия для разрешения возникающих проблем и выявить основные причины, которые нужно проанализировать в первую очередь.

Правило Парето – «универсальный» принцип, который применим во множестве ситуаций. Принцип Парето можно применить к любой группе причин, вызывающих то или иное последствие, когда большая часть последствий вызвана малым количеством причин.

Анализ Парето позволяет ранжировать отдельные факторы по значимости или важности, выявить и устранить в первую очередь те причины, которые вызывают наибольшее количество проблем (несоответствий).

Результаты анализа Парето, как правило, иллюстрируются диаграммой Парето (рис. 2.1), на которой по оси абсцисс отложены результаты ранжирования причин возникновения проблем качества в порядке убывания, а по оси ординат – количество ошибок как в численном, так и в накопленном процентном выражении.

На диаграмме видна область принятия первоочередных мер, очерчивающая те причины, которые вызывают наибольшее количество ошибок. Таким образом, предупредительные мероприятия в первую очередь должны быть направлены на решение именно этих проблем.

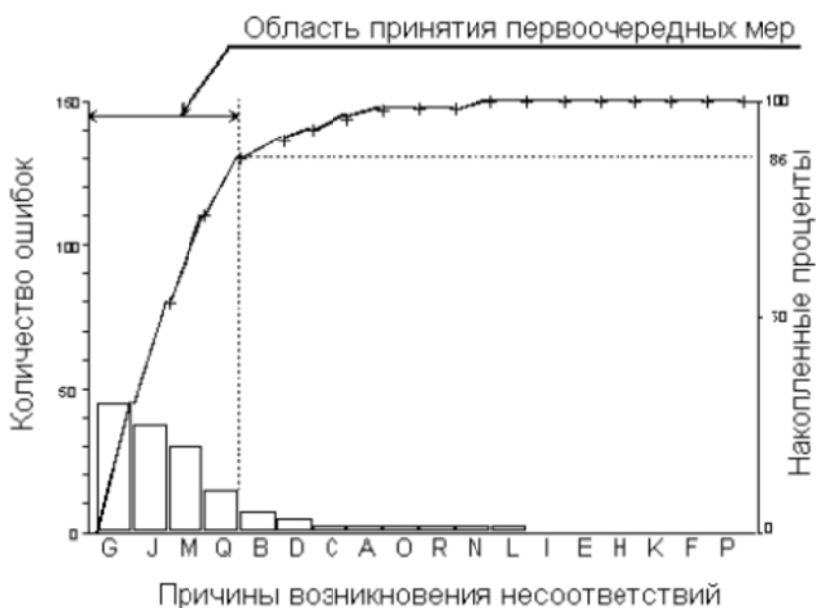


Рис. 2.1. Пример графика Парето

Различают два вида диаграмм Парето:

1. **Диаграмма Парето по результатам деятельности.** Эта диаграмма предназначена для выявления главной проблемы и отражает следующие нежелательные результаты деятельности:

- качество: дефекты, поломки, ошибки, отказы, рекламации, ремонты, возвраты продукции;
- себестоимость: объем потерь, затраты;
- сроки поставок: нехватка запасов, ошибки в составлении счетов;
- безопасность: несчастные случаи, трагические ошибки, аварии.

**2. Диаграмма Парето по причинам.** Эта диаграмма отражает причины проблем, возникающих в ходе производства, и используется для выявления главной из них:

- 1) исполнитель работы: смена, бригада, возраст, опыт работы, квалификация, индивидуальные характеристики;
- 2) оборудование: станки, агрегаты, инструменты, оснастка, организация использования, модели, штампы;
- 3) сырье: изготовитель, вид сырья, завод-поставщик, партия;
- 4) метод работы: условия производства, заказы-наряды, приемы работы, последовательность операций;
- 5) измерения: точность (указаний, чтения, приборная), верность и повторяемость (умение дать одинаковое указание в последующих измерениях одного и того же значения), стабильность (повторяемость в течение длительного периода), совместная точность, т.е. вместе с приборной точностью и тарированием прибора;
- 6) тип измерительного прибора (аналоговый или цифровой).

Построение диаграммы Парето включает следующие этапы:

1. Постановка задач и определение методов сбора данных.
2. Разработка контрольного листка для регистрации данных с перечнем видов собираемой информации.
3. Заполнение листов регистрации данных и подсчет итогов.
4. Заполнение бланка таблицы для проверки данных с целью оценки результатов расчетов по каждому проверяемому признаку в отдельности.
5. Расположение данных, полученных по каждому проверяемому признаку, в порядке значимости.
6. Построение графика, состоящего из одной горизонтальной и двух вертикальных осей. При этом на левую вертикальную ось в

соответствующем масштабе откладывается суммарное число дефектов, на правую вертикальную ось, также в соответствующем масштабе, значения накопленного процента дефектов (максимальное значение отрезка правой вертикальной оси равно 100%), а на горизонтальной оси нанесено число наименований контролируемых признаков. Максимальные значения левой и правой вертикальных осей равны между собой.

7. Построение столбчатой диаграммы. В построенных осях диаграммы откладывают число дефектов в соответствующем масштабе в виде «столбиков» и накопленный процент дефектов.

8. Построение кривой Парето. На вертикалях, соответствующих правым концам каждого интервала на горизонтальной оси, необходимо нанести точки накопленных сумм (результатов или процентов) и соединить их между собой отрезками прямых.

9. Нанесение на диаграмму всех обозначений и надписей.

10. Анализ. При построении диаграммы Парето необходимо обращать внимание на следующие моменты:

1) диаграмма Парето оказывается наиболее эффективной, если число дефектов, размещаемых на оси абсцисс, составляет 7–10;

2) в случае, когда признак «прочие» оказывается слишком большим по сравнению с другими, следует повторить анализ его содержания, а также вновь проанализировать все дефекты;

3) если дефект, стоящим первым по порядку, технически труден для анализа, следует начать с анализа дефекта, следующего за ним;

4) если обнаружится дефект, в отношении которого легко провести улучшение, то этим следует воспользоваться, не обращая внимание на порядок расположения признаков;

5) для анализа причин появления дефектов можно рекомендовать использовать причинно-следственные диаграммы (диаграммы Исикавы).

### **Пример построения диаграммы Парето**

При использовании диаграммы Парето наиболее распространенным методом анализа является *ABC*-анализ, сущность которого покажем на примере.



На складе имеется несколько видов готовой продукции разной стоимости, вся продукция подвергается сплошному выходному контролю. Из-за длительного контроля реализация задерживается, что ведет к убыткам.

Необходимо понять, какие причины приводят к данным последствиям, выявить эти причины и найти решения для их устранения.

Для правильного анализа разделим всю готовую продукцию, хранящуюся на складе, по группам в зависимости от стоимости каждого продукта, табл. 2.1.

Таблица 2.1

Информация о присутствии на складе товаров  
различных ценовых групп

Стоимость продукта, руб.	Число образцов, тыс. шт.
90–100	0,2
80–90	0,3
70–80	0,5
60–70	0,5
50–60	0,8
40–50	1,2
30–40	1,5
20–30	2,5
10–20	5,0
до 10	12,5
Итого:	25,0

Для дальнейшего анализа, построения диаграммы Парето и проведения ABC-анализа составляется табл. 2.2 с накоплением до 100%.

Построение табл. 2.2 накопленных частот осуществляется следующим образом. Сначала находят стоимость изделий конкретной партии, затем – общую стоимость как сумму произведений столбцов 1 и 2.

Таблица 2.2

## Анализ соотношения количества и стоимости образцов

Стоимость продукта, руб.	Число образцов, тыс. шт.	Стоимость продукции		Число образцов на складе	
		Накопленная стоимость, тыс. руб.	Относительная стоимость, %	Накопленная продукция, тыс. шт.	Удельный вес вида продукта, %
1	2	3	4	5	6
95	0,2	19,0	4,1	0,2	0,8
85	0,3	44,5	9,6	0,5	2,0
75	0,5	82,0	17,6	1,0	4,0
65	0,5	114,5	24,6	1,5	6,0
55	0,8	158,5	34,1	2,3	9,2
45	1,2	212,5	45,7	3,5	14,0
35	1,5	265,0	57,0	5,0	20,0
25	2,5	327,5	70,4	7,5	30,0
15	5,0	402,5	86,6	12,5	50,0
5	12,5	465,0	100,0	25,0	100,0

Затем находят значение столбца 4, который показывает, сколько процентов от общей стоимости составляют данные каждой строки, а столбец 5 таблицы формируется нарастающим итогом построчных данных столбца 2.

Данные столбца 6 формируются следующим образом: значение 0,8 из первой строки представляет собой число процентов, приходящихся на накопленный запас продукции (200) от всего количества образцов (25000).

Значение 2,0 из второй строки представляет собой число процентов, приходящихся на накопленный запас продукции (200+300), от всего ее количества. После проведения этой подготовительной работы несложно построить диаграмму Парето.

В двухмерной системе координат по оси абсцисс отложим удельный вес продукта  $n_i/N, \%$  (данные столбца 6), а по оси ординат – относительную стоимость этой продукции  $С_{Ti}/С_{T}, \%$ .

Соединив полученные точки прямыми, получим кривую Парето (или диаграмму Парето), как это показано на рис. 2.2.

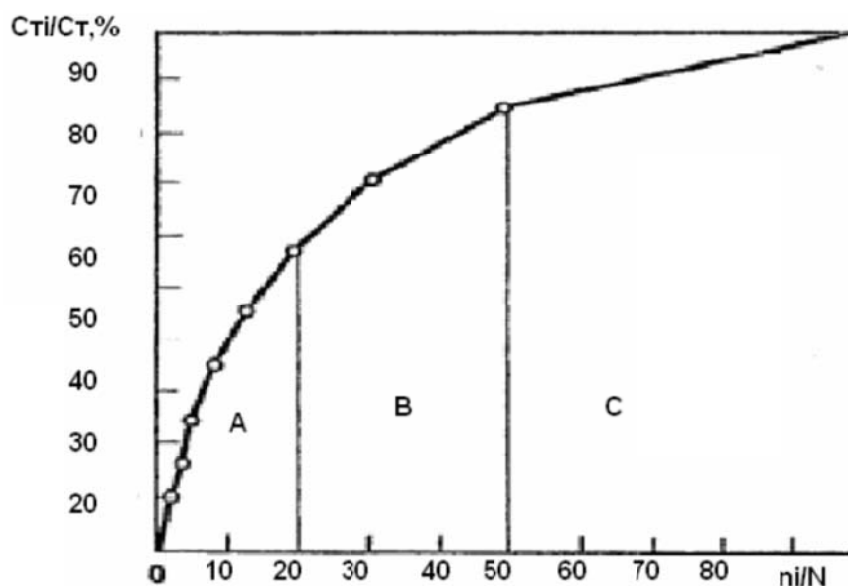


Рис. 2.2. Диаграмма Парето, построенная по данным табл. 2.1

Кривая Парето получилась сравнительно плавной в результате использования большого объема выборки.

Из анализа диаграммы Парето видно, что на долю наиболее дорогой продукции (первые 7 строк таблицы), которая составляет 20% от общего числа хранящихся на складе образцов, приходится более 50% общей стоимости всей готовой продукции, а на долю самой дешевой продукции, расположенной в последней строке таблицы и составляющей 50% от общего количества продукции на складе, приходится всего 13,3% от общей стоимости.

Назовем группу дорогой продукции группой *A*, группу дешевой продукции – группой *C* и промежуточную группу – группой *B*. Построим табл. 2.3 *ABC*-анализа полученных результатов.

Таблица 2.3

*ABC*-анализ результатов, полученных из табл. 2.1

Группа	Относительная частота количества образцов в группе, %	Относительная стоимость образцов в группе, %
<i>A</i>	20	56,7
<i>B</i>	30	30,0
<i>C</i>	50	13,3

В связи с вышесказанным можно сделать вывод, что контроль продукции на складе будет эффективнее в том случае, если контроль образцов группы *A* будет жестким (сплошным), а контроль образцов группы *C* – выборочным.

### **3. Порядок выполнения лабораторной работы**

Лабораторная работа должна выполняться в следующей последовательности:

1. Изучить материал, представленный выше.
2. Получить у преподавателя задание. Варианты заданий приведены в Приложение Б.
4. Построить диаграмму Парето и кумулятивную кривую.
5. Проанализировать полученные результаты и сделать выводы.
6. Предложить мероприятия по устранению наиболее весомых дефектов.
7. Оформить полученные результаты.

### **4. Требования к отчету**

Отчет должен содержать:

1. Титульный лист.
2. Задание.
3. Цель работы.
4. Исходные данные для выполнения работы, в соответствии с заданным вариантом.
5. Основные результаты выполнения работы в виде диаграммы Парето и кумулятивной кривой.
6. Выводы по работе.

## **СПИСОК ЛИТЕРАТУРЫ**

1. Зорин В. А., Павлов А.П., Пегачков А. А. Контроль качества продукции и услуг: учебное пособие. – М.: МАДИ (ГТУ), 2007. – 82с.
2. Орлов В. Н., Марфицын В. В. Диаграмма Парето и ее использование для анализа качества продукции: Методические указания. – Курган: КГУ, 2012. – 21 с

# Лабораторная работа № 3

## ИЗУЧЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА РИСКОВ: РЕГРЕССИОННЫЙ АНАЛИЗ

### 1. Цель работы

Целью работы является получение практических навыков, необходимых для построения регрессионных моделей, а также закрепление теоретического материала по линейному множественному регрессионному анализу.

### 2. Краткие теоретические сведения

#### 2.1. Постановка задачи регрессионного анализа

Регрессионный анализ служит для определения вида связи между переменными и предоставляет возможность для прогнозирования значения одной (зависимой) переменной, отталкиваясь от значения другой (независимой) переменной. Простейшей регрессией является линейная регрессия с одной независимой переменной:

$$\hat{y} = b \cdot x + a,$$

где  $b$  – регрессионный коэффициент;

$a$  – смещение по оси ординат.

В процессе регрессионного анализа данные исследуются методом наименьших квадратов, позволяющий сделать правдоподобные заключения о взаимосвязях, которые могут существовать между переменными. То есть, при проведении простой линейной регрессии основной задачей является такая прямая, для которой сумма квадратов вертикальных расстояний до отдельных точек данных является минимальной.

Множественная регрессия – это уравнение связи с несколькими независимыми переменными:

$$\hat{y} = f(x_1, x_2, \dots, x_p),$$

где  $\hat{y}$  – зависимая переменная (результативный признак);

$x_1, x_2, \dots, x_p$  – независимые переменные (признаки-факторы).

Для построения уравнения множественной регрессии чаще

используются следующие функции:

– линейная  $\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p$ ;

– степенная  $\hat{y} = b_0 \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot \dots \cdot x_p^{b_p}$ ;

– экспонента  $\hat{y} = e^{b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p}$ ;

– гипербола  $\hat{y} = \frac{1}{b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p}$ .

Можно использовать и другие функции, приводимые к линейному виду.

Задача линейного множественного регрессионного анализа состоит в вычислении значений коэффициентов  $b_0, b_j$  в линейном полиноме (модели):

$$\hat{y} = b_0 + \sum_{j=1}^p b_j \cdot x_j. \quad (3.1)$$

При этом результатом анализа данных является матрица вида:

$$\begin{pmatrix} y_1 & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ y_2 & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ y_n & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix},$$

где  $n$  – количество опытов;

$p$  – число факторов;

$x_{ij}$  – значение  $j$ -го фактора для  $i$ -го опыта;

$y_i$  – значение функции отклика для  $i$ -го опыта.

Вычисление значений коэффициентов  $b_0, b_j$  в линейном полиноме (3.1) в матричной форме запишется следующим образом:

$$\mathbf{B} = (\mathbf{X}^T \cdot \mathbf{X})^{-1} \cdot (\mathbf{X}^T \cdot \mathbf{Y}).$$

Здесь  $\mathbf{B}$  – вектор-столбец коэффициентов модели:

$$\mathbf{B} = \begin{pmatrix} b_0 \\ b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_p \end{pmatrix};$$

$\mathbf{X}$  – матрица значений всех рассматриваемых факторов, полученных при проведении измерений или наблюдений (регрессионная матрица):

$$\mathbf{X} = \begin{pmatrix} x_{10} & x_{11} & x_{12} & \dots & x_{1j} & \dots & x_{1p} \\ x_{20} & x_{21} & x_{22} & \dots & x_{2j} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ x_{n0} & x_{n1} & x_{n2} & \dots & x_{nj} & \dots & x_{np} \end{pmatrix};$$

$X_{i0}$  – вектор-столбец, определяющий свободный член уравнения регрессии (фиктивная переменная). В матрице исходных данных этот столбец состоит из единиц;

$Y$  – вектор-столбец опытных значений изучаемой характеристики:

$$\mathbf{Y} = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_n \end{pmatrix};$$

$\mathbf{X}^T$  – матрица, транспонированная к матрице  $\mathbf{X}$ .

Полученное уравнение регрессии отвечает только частичной выборке, т.е. данным, которые были использованы при статистической обработке. Для распространения этой зависимости на генеральную совокупность необходимо оценить значимость коэффициентов регрессии, так как может оказаться, что при условии неравенства коэффициента  $b_j$  нулю истинный коэффициент регрессии, отражающий генеральную совокупность,  $\beta_j=0$ . В этом случае пользоваться полученной моделью нельзя.

Проверка значимости полученных коэффициентов регрессии сводится к последовательной проверке гипотез  $H_0: \beta_j=0$  и  $H_1: \beta_j \neq 0$ ,  $j=0, 1, 2, \dots, p$ . Для этого вычисляются наблюдаемые значения критерия Стьюдента:

$$\hat{t}_j = \frac{|b_j|}{S_{b_j}}.$$

Здесь  $S_{b_j}^2$  – дисперсия оценок  $j$ -го коэффициента регрессии, вычисляемая по формуле:

$$S_{b_j}^2 = S_{\text{ост}}^2 \cdot C_{jj}.$$

где  $C_{jj}$  – диагональный элемент матрицы, обратной информационной

матрице  $(X^T \cdot X)$ , т.е. диагональный элемент ковариационной матрицы (или матрицы ошибок  $(X^T \cdot X)^{-1}$ ).

Остаточная дисперсия вычисляется по формуле:

$$S_{\text{ост}}^2 = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n-p-1} = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{f_1}.$$

Вычисленное значение  $\hat{t}_j$  сравнивают с табличным  $t_{(n-p-1); \alpha}^{\text{табл}}$  при числе степеней свободы  $f=n-p-1$  и выбранном уровне значимости  $\alpha$ .

Уровень значимости  $\alpha$  для проверки можно рассматривать как вероятность ошибки первого рода, т.е. ошибки, ведущей к отбрасыванию нулевой гипотезы  $H_0: \beta_j = 0$ , когда на самом деле она верна. Следовательно, эта ошибка приводит к сохранению регрессоров, чьи коэффициенты незначимы.

Ошибкой второго рода называется ошибка, приводящая к принятию нулевой гипотезы, когда она неверна. При этом из модели выбрасываются члены, которые надо было бы оставить. Ошибка второго рода для экспериментатора более значительна, поскольку приводит к смещению оценок и неправильному предсказанию отклика, а ошибка первого рода дает лишь некоторое снижение эффективности.

Если при выбранном уровне значимости:

$$\hat{t}_i \geq t_{(n-p-1); \alpha}^{\text{табл}},$$

то нулевая гипотеза отвергается и коэффициент  $b_j$  считается статистически значимым, т.е. имеет ненулевое значение и в генеральной совокупности.

Так как коэффициенты регрессии бывают связаны между собой, после отбрасывания незначимого коэффициента  $b_q$  необходимо все вычисления, связанные с регрессионным анализом, произвести заново.

Полученные оценки коэффициентов регрессии называются точечными, поскольку на числовой оси возможное значение данного коэффициента представляет одна точка. Доверительным интервалом данной случайной величины  $b_j$  называют интервал с границами  $(b_j - \varepsilon)$  и  $(b_j + \varepsilon)$ , в котором с наперед заданной вероятностью  $(1-\alpha)$  заключено истинное значение  $\beta_j$ . Вероятность  $(1-\alpha)$  называется доверительной вероятностью. При построении доверительных интервалов надо иметь в виду, что оценки коэффициентов  $b_j$  имеют нормальное



распределение и независимы от  $S_{\text{ост}}^2$ . В таком случае доверительные интервалы для значимых коэффициентов регрессии определяются неравенством:

$$b_j - t_{(n-p-1);\alpha}^{\text{табл}} \cdot S_{b_j} \leq \beta_j \leq b_j + t_{(n-p-1);\alpha}^{\text{табл}} \cdot S_{b_j}.$$

Проверка гипотезы об адекватности (значимости) уравнения регрессии – это поиск ответа на вопрос, можно ли использовать полученное уравнение или необходима более сложная модель.

Значимость уравнения регрессии определяется его предсказательной силой, т.е. возможностью надежно прогнозировать средние значения зависимой переменной  $y$  по заданным значениям независимых переменных  $x_i$ . Уравнение  $\hat{y} = f(x_1, x_2, \dots, x_p)$ , согласно которому должно проводиться прогнозирование, получено на основании статистической обработки частичной совокупности. Это уравнение может существенно отличаться от гипотетического уравнения, соответствующего генеральной совокупности. Для оценки надежности уравнения регрессии применяют  $F$ -критерий Фишера, который определяется по формуле:

$$\hat{F} = \frac{S_y^2}{S_{\text{ост}}^2}, \quad (3.2)$$

где  $S_y^2$  – дисперсия фактических значений зависимого переменного:

$$S_y^2 = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n-1} = \frac{\sum_{i=1}^n (y_i - \bar{y})^2}{f_2},$$

а среднее значение зависимого переменного рассчитывается по формуле:

$$\bar{y} = \frac{1}{n} \cdot \sum_{i=1}^n y_i.$$

Полученное по (3.2) значение  $F$ -критерия сравнивают с  $F_{(f_1; f_2; \alpha)}^{\text{табл}}$  – табличным значением при выбранном уровне значимости и числе степеней свободы  $f_1$  и  $f_2$ . Уравнение признается адекватным (значимым), если:

$$\hat{F} \geq F_{(f_1; f_2; \alpha)}^{\text{табл}}.$$

В этом случае предсказательная сила уравнения регрессии больше, чем предсказательная сила среднего значения.

Если априори степень полинома неизвестна, то расчеты необходимо произвести несколько раз, последовательно увеличивая

степень полинома до тех пор, пока полученное уравнение регрессии не станет адекватным.

Рассмотрим более общий случай, когда зависимость математического ожидания  $M[y]$  случайной величины  $y$  от независимых переменных  $x_0, x_1, \dots, x_p$ , выражается полиномом степени  $k$ , т.е.:

$$M[y] = \eta = \beta_0 + \sum_{j=1}^p \beta_j \cdot x_j + \sum_{j=1}^p \beta_{jj} \cdot x_j^2 + \sum_{j=1}^{p-1} \sum_{i=j+1}^p \beta_{ji} \cdot x_j \cdot x_i + K. \quad (3.3)$$

Число коэффициентов регрессии в этом уравнении равно

$$C_{p+k}^k = \prod_{i=1}^p \frac{k+i}{i},$$

т.е. числу сочетаний из  $(p+k)$  по  $k$ . Путем несложных преобразований уравнение (3.3) можно свести к уравнению вида:

$$\eta = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \dots + \beta_p \cdot x_p.$$

Например, примем степень  $k$  полинома равной 2. Тогда оценка  $\hat{y}$  величины  $\eta$  может быть записана в виде уравнения:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_p \cdot x_p + b_{12} \cdot x_1 \cdot x_2 + \dots + b_{p-1} \cdot x_{p-1} \cdot x_p + b_{11} \cdot x_1^2 + \dots + b_{pp} \cdot x_p^2. \quad (3.4)$$

Введем фиктивную переменную  $x_0=1$  и заменим члены второго порядка линейными, приравняв квадраты факторов и парные взаимодействия к новым переменным:

$$x_1^2 = x_{p+1}; x_2^2 = x_{p+2}; \dots; x_p^2 = x_{p+p} = x_{2p}; x_1 \cdot x_2 = x_{2p+1}; x_{p-1} \cdot x_p = x_{p1},$$

где  $p1 = C_{p+2}^2 - 1$ .

В новой системе обозначений полином (3.4) можно записать как линейное однородное уравнение:

$$\hat{y} = b_0 \cdot x_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_{p1} \cdot x_{p1}.$$

Аналогично можно нелинейные члены полинома любого порядка заменить линейными членами. Изложенная процедура возможна, если все коэффициенты входят в уравнение линейно. Такое уравнение называется линейным по параметрам. Таким образом, после линеаризации полином любого порядка можно рассматривать как линейное однородное уравнение и применять для регрессионного анализа формулы, выведенные для линейного уравнения регрессии.

### **3. Порядок выполнения работы**

Для выполнения работы необходимо:

1. Получить у преподавателя вариант задания и вид модели регрессии (Приложение В).
2. Провести регрессионный анализ с помощью пакета *MatLab*.
3. Оценить коэффициенты модели, их значимость и проверить адекватность модели.
4. Построить графики по исходным данным и по полученной модели. Построить график ошибки.
5. Оформить отчет. Сохранить в отчете числовые и графические результаты анализа.

### **4. Требования к отчету**

Отчет должен содержать:

1. Титульный лист.
2. Задание.
3. Цель работы.
4. Исходные данные для выполнения работы, в соответствии с заданным вариантом.
5. Основные результаты выполнения работы.
6. Выводы по работе.

### **5. Контрольные вопросы**

1. В чем состоит задача линейного множественного регрессионного анализа?
2. Какую ошибку называют ошибкой первого рода?
3. Какую ошибку называют ошибкой второго рода?
4. Как проводится оценка надежности уравнения регрессии?
5. Для чего вводится фиктивная переменная  $x_0=1$ ?

## **СПИСОК ЛИТЕРАТУРЫ**

1. Львовский Е. Н. Статистические методы построения эмпирических формул: учеб. пособие. 2-е изд., перераб. и доп. М.: Высш. шк., 1988. 239 с.

## **Лабораторная работа № 4**

# **ИЗУЧЕНИЕ СТАТИСТИЧЕСКИХ МЕТОДОВ АНАЛИЗА РИСКОВ: КЛАСТЕРНЫЙ АНАЛИЗ**

### **1. Цель работы**

Целью работы является получение практических навыков, необходимых для проведения кластерного анализа.

### **2. Краткие теоретические сведения**

Классификация является основой умозрительной человеческой деятельности и фундаментальным процессом научной практики, поскольку системы классификации содержат понятия, необходимые для разработки теорий в науке и приносят порядок в исследования. Наибольшее использование имеют семь основных инструментов управления качеством, предназначенных для обнаружения, идентификации, оценки и анализа проблем качества, а также для выработки управляющих (корректирующих или предупреждающих) мероприятий по устранению или недопущению несоответствующего качества процессов, продукции или услуг. Дополнительным инструментарием специалистов в области управления качеством являются, например, специальные методы обработки экспериментальных данных, методы прогнозных оценок, кластерный анализ, многофакторный анализ, функционально-стоимостной анализ, современные методы искусственного интеллекта и многое другое.

«Кластерный анализ» – это общее название множества вычислительных процедур, используемых при создании классификации. Задачу классификации можно сформулировать следующим образом. Имеется некоторое конечное множество объектов произвольной природы, представленных совокупностью соответствующих векторов. Необходимо классифицировать эти объекты, т.е. разбить их множество на заданное или произвольное количество групп (кластеров, классов, таксонов) таким образом, чтобы в каждую группу оказались включенными объекты, близкие между собой в том или ином смысле; причем число этих групп может быть неизвестно заранее. Априорная информация о классификации

объектов при этом отсутствует. В результате работы этих процедур исследуемые объекты подразделяются на кластеры – группы очень похожих в некотором смысле объектов.

**Рассмотрим пример процедуры кластерного анализа.**

Допустим, мы имеем набор данных  $A$ , состоящий из 14-ти примеров, у которых имеется по два признака  $X$  и  $Y$ . Данные по ним приведены в табл. 4.1.

*Таблица 4.1*

Набор данных  $A$

№ примера	признак $X$	признак $Y$
1	27	19
2	11	46
3	25	15
4	36	27
5	35	25
6	10	43
7	11	44
8	36	24
9	26	14
10	26	14
11	9	45
12	33	23
13	27	16
14	19	47

Данные в табличной форме не носят информативный характер. Представим переменные  $X$  и  $Y$  в виде диаграммы рассеивания, изображенной на рис. 4.1.

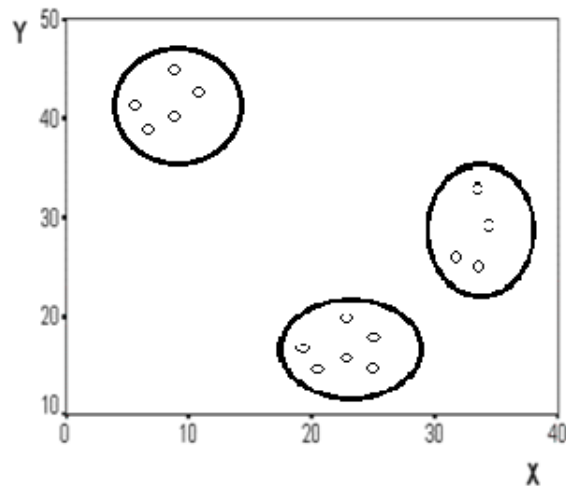


Рис. 4.1. Диаграмма рассеивания переменных  $X$  и  $Y$

На рис. 4.1 мы видим несколько групп «похожих» примеров. Примеры (объекты), которые по значениям  $X$  и  $Y$  «похожи» друг на друга, принадлежат к одной группе (кластеру); объекты из разных кластеров не похожи друг на друга.

Критерием для определения **схожести и различия кластеров** является расстояние между точками на диаграмме рассеивания. Это сходство можно «измерить», оно равно расстоянию между точками на графике. Способов определения меры расстояния между кластерами, называемой еще мерой близости, существует несколько. Наиболее распространенный способ – **вычисление евклидова расстояния** между двумя точками  $i$  и  $j$  на плоскости, когда известны их координаты  $X$  и  $Y$ :

$$D_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (4.1)$$

*Примечание.* Чтобы узнать расстояние между двумя точками, надо взять разницу их координат по каждой оси, возвести ее в квадрат, сложить полученные значения для всех осей и извлечь квадратный корень из суммы.

Когда осей больше, чем две, расстояние рассчитывается таким образом: сумма квадратов разницы координат состоит из столько же слагаемых, сколько осей (измерений) присутствует в нашем пространстве. Например, если нам нужно найти расстояние между двумя точками в пространстве трех измерений (такая ситуация представлена на рис. 4.2), формула (4.1) приобретает вид:

$$D = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}.$$

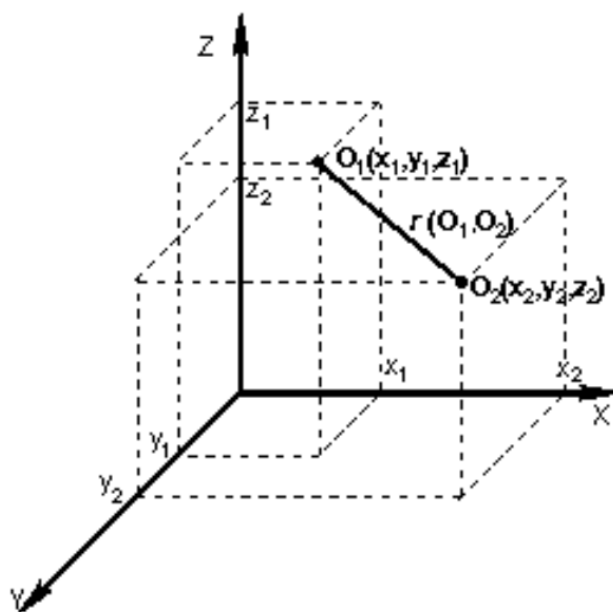


Рис. 4.2. Расстояние между двумя точками в трехмерном пространстве

Кластер имеет следующие математические характеристики: центр, радиус, среднеквадратическое отклонение, размер кластера.

**Центр кластера** – это среднее геометрическое место точек в пространстве переменных.

**Радиус кластера** – это максимальное расстояние точек от центра кластера.

Кластеры могут быть перекрывающимися. В этом случае невозможно при помощи математических процедур однозначно отнести объект к одному из двух кластеров. Такие объекты называют спорными.

**Спорный объект** – это объект, который по мере сходства может быть отнесен к нескольким кластерам.

Размер кластера может быть определен либо по радиусу кластера, либо по среднеквадратичному отклонению объектов для этого кластера. Объект относится к кластеру, если расстояние от объекта до центра кластера меньше радиуса кластера. Если это условие выполняется для двух и более кластеров, объект является спорным. Неоднозначность данной задачи может быть устранена экспертом или аналитиком.

Работа кластерного анализа опирается на два предположения. Первое предположение – рассматриваемые признаки объекта в принципе допускают желательное **разбиение пула** (совокупности)

объектов на кластеры. Второе предположение – сравнимость шкал, правильность выбора масштаба или единиц измерения признаков.

**Выбор масштаба** в кластерном анализе имеет большое значение. Рассмотрим пример. Представьте себе, что данные признака  $x$  в наборе данных  $A$  на два порядка больше данных признака  $y$ : значения переменной  $x$  находятся в диапазоне от 100 до 700, а значения переменной  $y$  в диапазоне от 0 до 1.

Тогда при расчете величины расстояния между точками, отражающими положение объектов в пространстве их свойств, переменная, имеющая большие значения, т.е. переменная  $x$ , будет практически полностью доминировать над переменной с малыми значениями, т.е. переменной  $y$ . Таким образом, из-за неоднородности единиц измерения признаков становится невозможно корректно рассчитать расстояния между точками.

Эта проблема решается при помощи предварительной стандартизации переменных. **Стандартизация** (*standardization*) или **нормирование** (*normalization*) приводит значения всех преобразованных переменных к единому диапазону значений путем выражения через отношение этих значений к некоей величине, отражающей определенные свойства конкретного признака. Существуют различные способы нормирования исходных данных.

Два наиболее распространенных способа:

– деление исходных данных на среднее квадратичное отклонение соответствующих переменных;

– вычисление  $Z$ -вклада или стандартизованного вклада.

Наряду со стандартизацией переменных, существует вариант придания каждой из них определенного коэффициента важности, или веса, который бы отражал значимость соответствующей переменной. В качестве весов могут выступать экспертные оценки, полученные в ходе опроса экспертов – специалистов предметной области. Полученные произведения нормированных переменных на соответствующие веса позволяют получать расстояния между точками в многомерном пространстве с учетом неодинакового веса переменных.

В ходе экспериментов возможно сравнение результатов, полученных с учетом экспертных оценок и без них, и выбор лучшего из них.



## **Методы кластерного анализа**

Методы кластерного анализа можно разделить на две группы:

- иерархические;
- неиерархические.

Каждая из групп включает множество подходов и алгоритмов. Используя различные методы кластерного анализа, аналитик может получить различные решения для одних и тех же данных. Это считается нормальным явлением.

Рассмотрим иерархические и неиерархические методы подробно.

### **Иерархические методы кластерного анализа**

Суть иерархической кластеризации состоит в последовательном объединении меньших кластеров в большие или разделении больших кластеров на меньшие.

#### **Иерархические агломеративные методы (*Agglomerative Nesting, AGNES*)**

Эта группа методов характеризуется последовательным объединением исходных элементов и соответствующим уменьшением числа кластеров.

В начале работы алгоритма все объекты являются отдельными кластерами. На первом шаге наиболее похожие объекты объединяются в кластер. На последующих шагах объединение продолжается до тех пор, пока все объекты не будут составлять один кластер.

#### **Иерархические дивизимные (делимые) методы (*DIVisive ANALysis, DIANA*)**

Эти методы являются логической противоположностью агломеративным методам. В начале работы алгоритма все объекты принадлежат одному кластеру, который на последующих шагах делится на меньшие кластеры, в результате образуется последовательность расщепляющих групп.

Принцип работы описанных выше групп методов в виде дендрограммы показан на рис. 4.3.

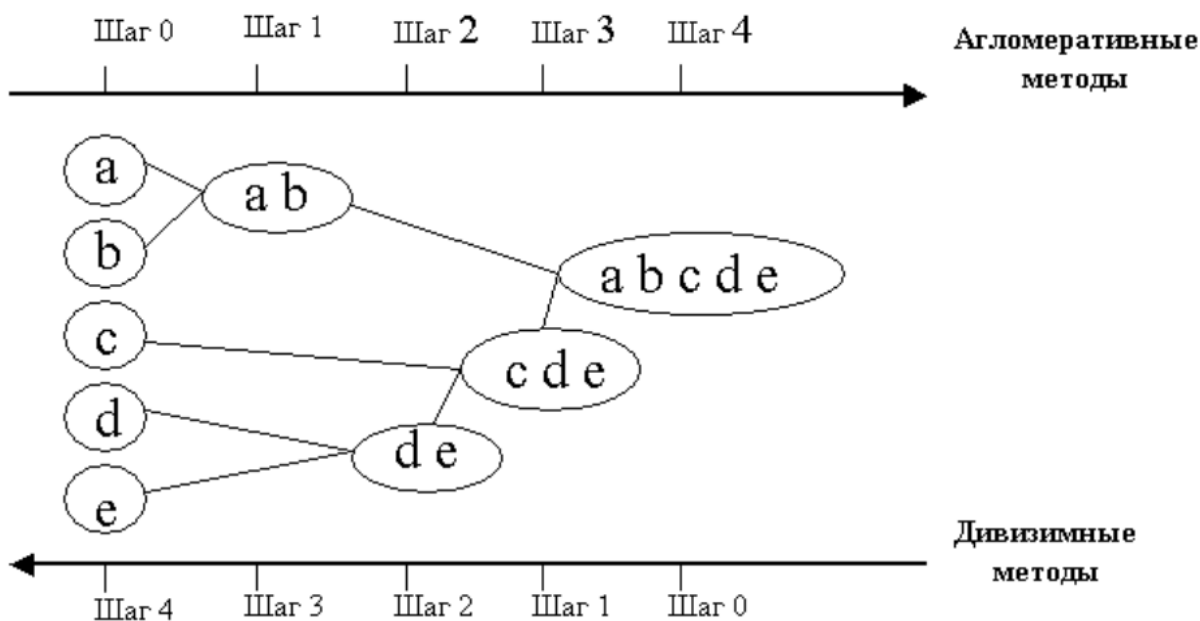


Рис. 4.3. Дендрограмма агломеративных и дивизимных методов

Иерархические методы кластеризации различаются правилами построения кластеров. В качестве правил выступают критерии, которые используются при решении вопроса о «схожести» объектов при их объединении в группу (агломеративные методы), либо разделения на группы (дивизимные методы).

Иерархические методы кластерного анализа используются при небольших объемах наборов данных. Преимуществом иерархических методов кластеризации является их наглядность.

Иерархические алгоритмы связаны с построением **дендрограмм** (от греческого *dendron* – «дерево»), которые являются результатом иерархического кластерного анализа. Дендрограмма (*dendrogram*) – древовидная диаграмма, описывает близость отдельных точек и кластеров друг к другу, представляет в графическом виде последовательность объединения (разделения) кластеров. Дендрограмма содержит  $n$  уровней, каждый из которых соответствует одному из шагов процесса последовательного укрупнения (разделения) кластеров. Дендрограмму также называют древовидной схемой, деревом объединения кластеров, деревом иерархической структуры. Существует много способов построения дендрограмм. В дендрограмме объекты могут располагаться вертикально или горизонтально. Пример вертикальной дендрограммы приведен на рис. 4.4.

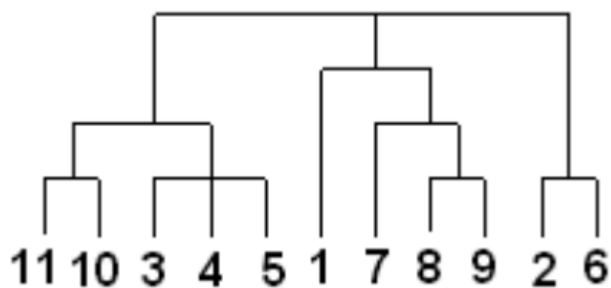


Рис. 4.4. Пример дендрограммы

Числа 11, 10, 3 и т.д. соответствуют номерам объектов или наблюдений исходной выборки. Мы видим, что на первом шаге каждое наблюдение представляет один кластер (вертикальная линия), на втором шаге наблюдаем объединение таких наблюдений: 11 и 10; 3, 4 и 5; 8 и 9; 2 и 6. На втором шаге продолжается объединение в кластеры: наблюдения 11, 10, 3, 4, 5 и 7, 8, 9. Данный процесс продолжается до тех пор, пока все наблюдения не объединятся в один кластер.

### Меры расстояния

Объединение или метод древовидной кластеризации используется при формировании кластеров несходства или расстояния между объектами. Эти расстояния могут определяться в одномерном или многомерном пространстве. Наиболее прямой путь вычисления расстояний между объектами в многомерном пространстве состоит в вычислении евклидовых расстояний. Если вы имеете двух- или трёхмерное пространство, то эта мера является реальным геометрическим расстоянием между объектами в пространстве (как будто расстояния между объектами измерены рулеткой). Однако алгоритм объединения не «заботится» о том, являются ли «предоставленные» для этого расстояния настоящими или некоторыми другими производными мерами расстояния, что более значимо для исследователя; и задачей исследователей является подобрать правильный метод для специфических применений.

**Евклидово расстояние.** Это, по-видимому, наиболее общий тип расстояния. Оно попросту является геометрическим расстоянием в многомерном пространстве и вычисляется следующим образом:

$$D_{x,y} = \sqrt{\sum_i (x_i - y_i)^2},$$

где  $x, y$  – объекты, которые требуется сравнить;  
 $i$  – размерность измерения.

**Квадрат евклидова расстояния.** Иногда может возникнуть желание возвести в квадрат стандартное евклидово расстояние, чтобы придать большие веса более отдаленным друг от друга объектам. Это расстояние вычисляется следующим образом:

$$D_{x,y} = \sum_i (x_i - y_i)^2 .$$

**Расстояние городских кварталов (манхэттенское расстояние).** Это расстояние является просто средним разностей по координатам. В большинстве случаев эта мера расстояния приводит к таким же результатам, как и для обычного расстояния Евклида. Однако отметим, что для этой меры влияние отдельных больших разностей (выбросов) уменьшается (так как они не возводятся в квадрат). Манхэттенское расстояние вычисляется по формуле:

$$D_{x,y} = \sum_i |x_i - y_i| .$$

**Расстояние Чебышева.** Это расстояние может оказаться полезным, когда желают определить два объекта как «различные», если они различаются по какой-либо одной координате (каким-либо одним измерением). Расстояние Чебышева вычисляется по формуле:

$$D_{x,y} = \max\{|x_i - y_i|\} .$$

**Степенное расстояние.** Иногда желают прогрессивно увеличить или уменьшить вес, относящийся к размерности, для которой соответствующие объекты сильно отличаются. Это может быть достигнуто с использованием степенного расстояния. Степенное расстояние вычисляется по формуле:

$$D_{x,y} = \left\{ \left( \sum_i |x_i - y_i|^p \right)^{\frac{1}{r}} \right\} ,$$

где  $r$  и  $p$  – параметры, определяемые пользователем.

Несколько примеров вычислений могут показать, как «работает» эта мера. Параметр  $p$  ответственен за постепенное взвешивание разностей по отдельным координатам, параметр  $r$  ответственен за прогрессивное взвешивание больших расстояний между объектами.

Если оба параметра  $r$  и  $p$ , равны двум, то это расстояние совпадает с расстоянием Евклида.

**Процент несогласия.** Эта мера используется в тех случаях, когда данные являются категориальными. Это расстояние вычисляется по формуле:

$$D_{x,y} = \frac{(\text{количество } x_i \neq y_i)}{i}.$$

### **Методы объединения или связи**

Когда каждый объект представляет собой отдельный кластер, расстояния между этими объектами определяются выбранной мерой. Существуют различные правила, называемые методами объединения или связи для двух кластеров. Рассмотрим некоторые из них.

**Метод ближнего соседа или одиночная связь.** Здесь расстояние между двумя кластерами определяется расстоянием между двумя наиболее близкими объектами (ближайшими соседями) в различных кластерах. Этот метод позволяет выделять кластеры сколь угодно сложной формы при условии, что различные части таких кластеров соединены цепочками близких друг к другу элементов. В результате работы этого метода кластеры представляются длинными «цепочками» или «волоконистыми» кластерами, «сцепленными вместе» только отдельными элементами, которые случайно оказались ближе остальных друг к другу.

**Метод наиболее удаленных соседей или полная связь.** Здесь расстояния между кластерами определяются наибольшим расстоянием между любыми двумя объектами в различных кластерах (т.е. «наиболее удаленными соседями»). Метод хорошо использовать, когда объекты действительно происходят из различных «рощ». Если же кластеры имеют в некотором роде удлиненную форму или их естественный тип является «цепочечным», то этот метод не следует использовать.

**Метод Варда (*Ward's method*).** В качестве расстояния между кластерами берется прирост суммы квадратов расстояний объектов до центров кластеров, получаемый в результате их объединения. В отличие от других методов кластерного анализа для оценки расстояний между кластерами, здесь используются методы дисперсионного анализа. На каждом шаге алгоритма объединяются такие два кластера, которые приводят к минимальному увеличению

целевой функции, т.е. внутригрупповой суммы квадратов. Этот метод направлен на объединение близко расположенных кластеров и «стремится» создавать кластеры малого размера.

**Метод невзвешенного попарного среднего** (метод невзвешенного попарного арифметического среднего – *unweighted pair-group method using arithmetic averages, UPGMA*). В качестве расстояния между двумя кластерами берется среднее расстояние между всеми парами объектов в них. Этот метод следует использовать, если объекты действительно происходят из различных «рощ», в случаях присутствия кластеров «цепочного» типа, при предположении неравных размеров кластеров.

**Метод взвешенного попарного среднего** (метод взвешенного попарного арифметического среднего – *weighted pair-group method using arithmetic averages, WPGMA*). Этот метод похож на метод невзвешенного попарного среднего, разница состоит лишь в том, что здесь в качестве весового коэффициента используется размер кластера (число объектов, содержащихся в кластере). Этот метод рекомендуется использовать именно при наличии предположения о кластерах разных размеров.

**Невзвешенный центроидный метод** (метод невзвешенного попарного центроидного усреднения – *unweighted pair-group method using the centroid average*). В качестве расстояния между двумя кластерами в этом методе берется расстояние между их центрами тяжести.

**Взвешенный центроидный метод** (метод взвешенного попарного центроидного усреднения – *weighted pair-group method using the centroid average, WPGMC*). Этот метод похож на предыдущий, разница состоит в том, что для учета разницы между размерами кластеров (числе объектов в них), используются веса. Этот метод предпочтительно использовать в случаях, если имеются предположения относительно существенных отличий в размерах кластеров.

### **Алгоритм кластеризации *k-means***

К наиболее простым и эффективным алгоритмам кластеризации относится *k-means*, или в русскоязычном варианте *k-средних*. Он состоит из четырех шагов:

– первый шаг: задается число кластеров  $k$ , которое должно быть сформировано из объектов исходной выборки;

– второй шаг: случайным образом выбирается  $k$  записей, которые будут служить начальными центрами кластеров. Начальные точки, из которых потом вырастает кластер, часто называют «семенами». Каждая такая запись представляет собой своего рода «эмбрион» кластера, состоящий только из одного элемента;

– третий шаг: для каждой записи исходной выборки определяется ближайший к ней центр кластера;

– четвертый шаг: производится вычисление **центроидов** – центров тяжести кластеров. Это делается путем определения среднего для значений каждого признака всех записей в кластере. Затем старый центр кластера смещается в его центроид. Таким образом, центроиды становятся новыми центрами кластеров для следующей итерации алгоритма.

Шаги 3 и 4 повторяются до тех пор, пока выполнение алгоритма не будет прервано, либо пока не будет выполнено условие в соответствии с некоторым критерием сходимости.

Остановка алгоритма производится, когда границы кластеров и расположение центроидов перестают изменяться, т.е. на каждой итерации в каждом кластере остается один и тот же набор записей. Алгоритм *k-means* обычно находит набор стабильных кластеров за несколько десятков итераций. В качестве критерия сходимости чаще всего используется сумма квадратов расстояний между центроидом кластера и всеми вошедшими в него записями. Алгоритм остановится тогда, когда ошибка достигнет достаточно малого значения.

### **3. Порядок выполнения лабораторной работы**

Лабораторная работа должна выполняться в следующей последовательности:

1. Получить у преподавателя задание. Варианты заданий приведены в Приложение Г.

2. Провести кластерный анализ данных с использованием **всех** инструментов (алгоритмов) кластерного анализа пакета *SPSS (Superior Performance Software System)*.

3. Проанализировать полученные результаты и сделать выводы на основании общей оценки результатов работы алгоритмов.

Кластерный анализ можно считать успешным, если он выполнен разными способами, проведено сравнение результатов и найдены

общие закономерности, а также найдены стабильные кластеры независимо от способа кластеризации.

4. Оформить полученные результаты.

#### **4. Требования к отчету**

Отчет должен содержать:

1. Титульный лист.
2. Задание.
3. Цель работы.
4. Исходные данные для выполнения работы, в соответствии с заданным вариантом.
5. Основные результаты выполнения работы.
6. Выводы по работе.

#### **5. Контрольные вопросы и задания**

1. Какие методы кластерного анализа вы знаете?
2. Опишите различия между агломеративными и дивизимными методами кластерного анализа.
3. Как выбрать меру сходства для проведения кластерного анализа?
4. Как провести оценку результатов кластерного анализа?
5. Опишите алгоритм кластеризации *k-means*.

### **СПИСОК ЛИТЕРАТУРЫ**

1. Чубукова И. А. *Data Mining* [Электронный ресурс] / И. А. Чубукова. – URL: [https://intuit.ru/goods\\_store/ebooks](https://intuit.ru/goods_store/ebooks). (01.01.2021).



**Варианты заданий для лабораторной работы № 1**

Вариант № 1.

Задание. Разработать систему управления безопасностью пищевых продуктов ХАССП на примере производства замороженных говяжьих котлет, рис. А1.



*Рис. А1.* Блок-схема технологических операций производства замороженных говяжьих котлет

## Вариант 2.

Задание. Разработать систему управления безопасностью пищевых продуктов ХАССП на примере производства полуфабриката – готового к употреблению после разогрева панированного мяса, рис. А2.

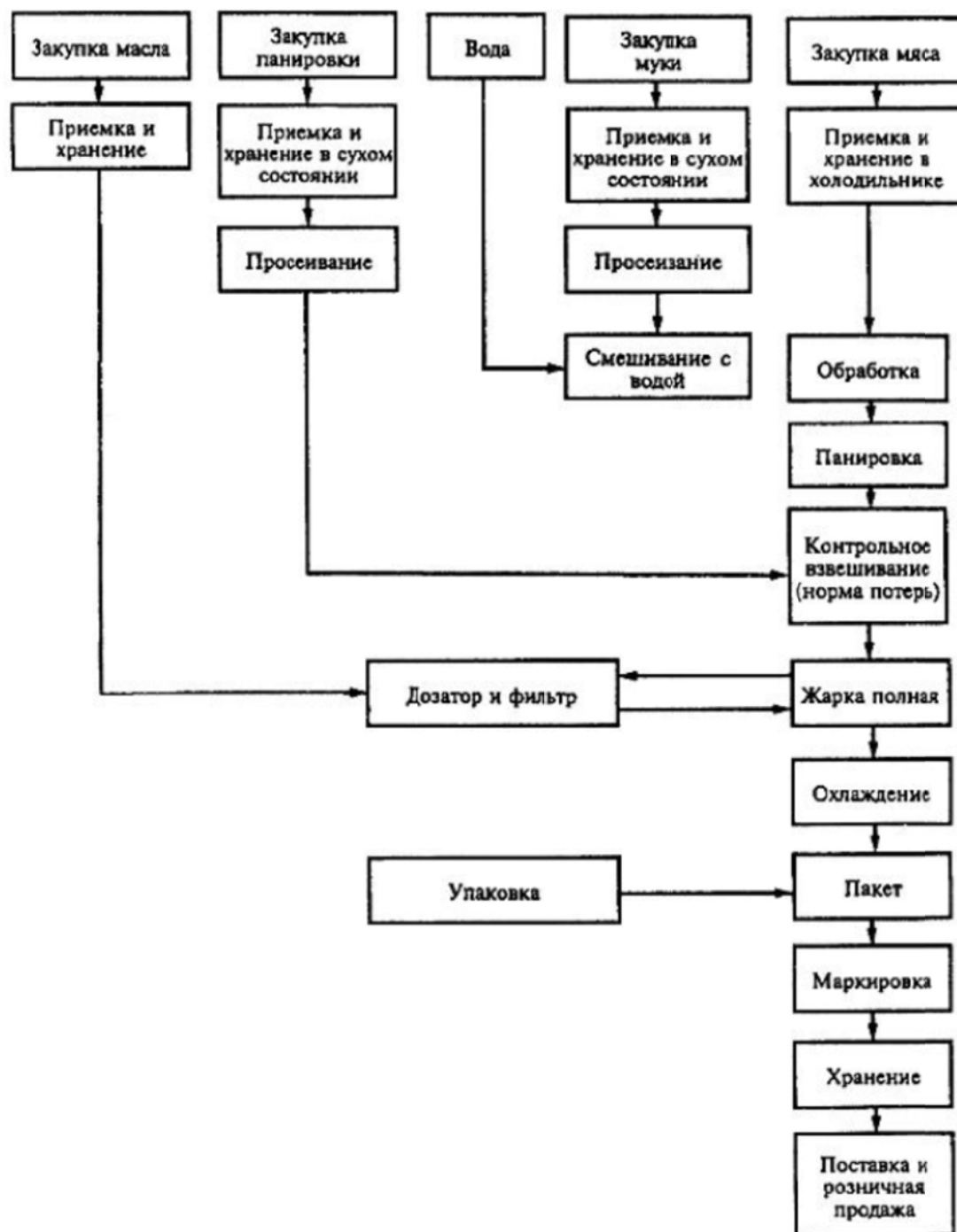


Рис. А2. Блок-схема технологических операций производства полуфабриката – готового к употреблению после разогрева панированного мяса

### Вариант 3.

Задание. Разработать систему управления безопасностью пищевых продуктов ХАССП на примере производства мороженого с шоколадной крошкой, рис. А3.

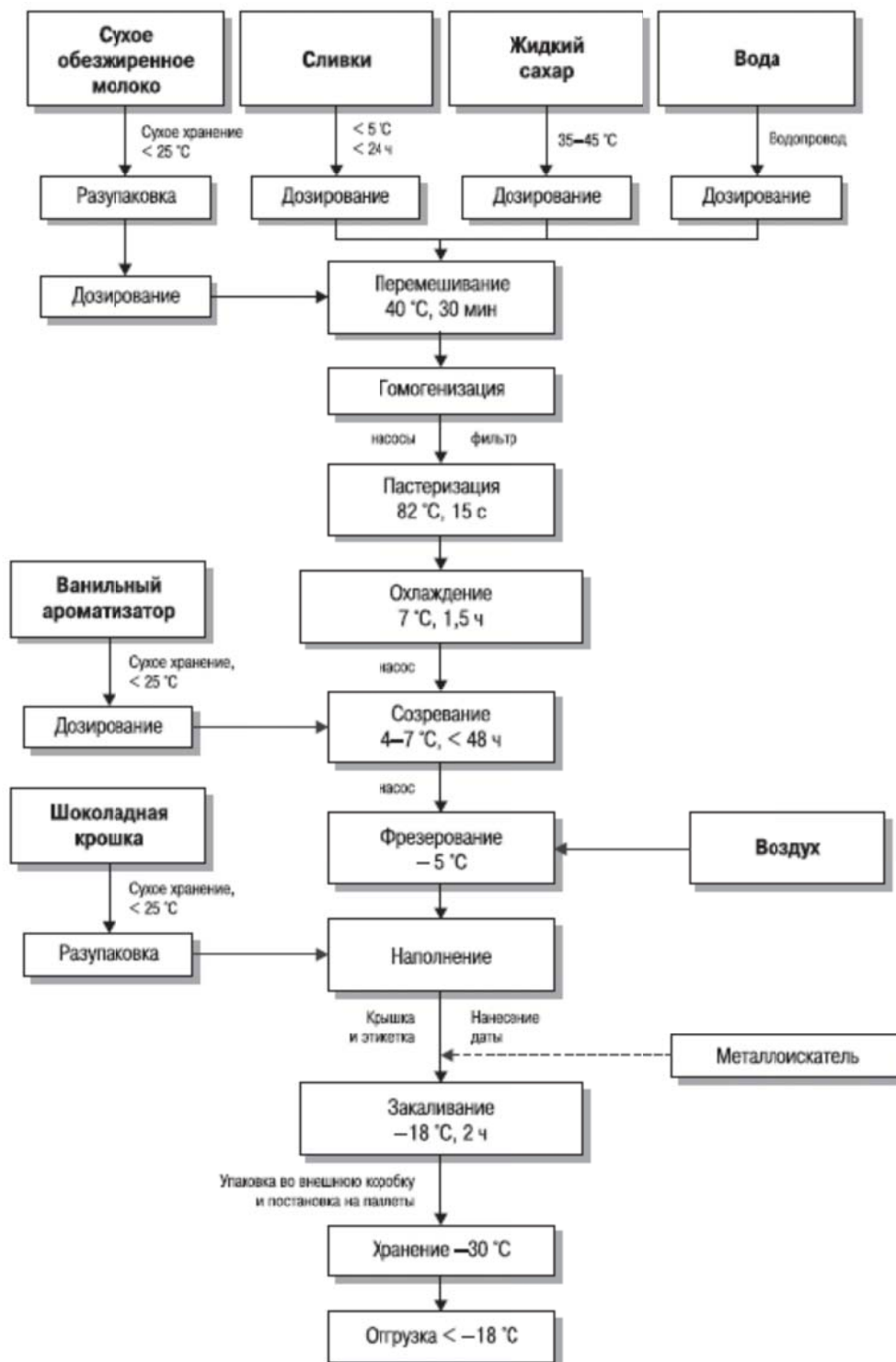


Рис. А3. Блок-схема технологических операций производства мороженого с шоколадной крошкой

## Приложение Б

### Варианты заданий для лабораторной работы № 2

Варианты заданий приведены в табл. Б1.

*Таблица Б1*

#### Варианты заданий

№ №	Вид брака	Количество некачественных изделий																
		Вар. 1	Вар. 2	Вар. 3	Вар. 4	Вар. 5	Вар. 6	Вар. 7	Вар. 8	Вар. 9	Вар. 10	Вар. 11	Вар. 12	Вар. 13	Вар. 14	Вар. 15	Вар. 16	Вар. 17
1	X1	12	20	260	5	18	20	40	24	101	95	8	8	230	410	40	640	60
2	X2	32	18	54	30	10	18	26	124	3	5	2	4	2	160	480	96	18
3	X3	8	5	120	210	24	9	84	6	80	12	5	3	7	2	12	10	3
4	X4	300	27	34	38	120	8	13	10	7	4	28	23	1060	120	36	50	218
5	X5	24	14	18	90	50	84	3	146	5	20	77	20	160	16	6	120	50
6	X6	18	58	70	20	15	169	61	8	74	34	10	70	80	24	200	70	40
7	X7	210	85	32	150	38	2	10	3	10	12	6	84	35	3	4	3	85
8	X8	30	10	16	14	4	198	6	5	22	24	43	15	42	1214	3	84	400
9	X9	14	3	12	45	75	50	8	20	60	74	98	140	120	100	149	30	8
10	X10	12	6	14	8	8	10	9	6	4	140	58	14	45	2	70	8	6
11	X11								100	14	8	3	6	16	8	8	5	10
12	X12								4	12	94	6	75	3	11	1	200	100
13	X13										4		2			27		

**Варианты заданий для лабораторной работы № 3**

Задание № 1. Оценить коэффициенты уравнения регрессии:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_3 \cdot x_3.$$

Результаты и условия эксперимента приведены в табл. В1.

*Таблица В1*

Результаты и условия эксперимента

№ эксперимента	$y$	$x_1$	$x_2$	$x_3$
1	140,28	252,36	96,67	8,37
2	142,02	262,54	100,07	9,07
3	149,90	285,70	96,78	9,35
4	147,12	277,52	101,30	9,67
5	163,62	307,95	100,35	9,45
6	173,40	322,44	104,8	10,12
7	178,86	334,88	106,17	10,35
8	186,26	350,11	109,20	11,03
9	183,53	346,10	104,48	10,38
10	198,76	374,91	106,88	12,15
11	205,30	378,49	113,14	12,98
12	206,77	397,48	112,98	11,34
13	198,42	378,39	109,07	10,95
14	216,48	393,44	114,45	12,89
15	221,45	403,84	115,23	13,71

Задание № 2. Оценить коэффициенты уравнения регрессии:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_1 \cdot x_2 + b_{11} \cdot x_1^2 + b_{22} \cdot x_2^2.$$

Результаты и условия эксперимента приведены в табл. В2.

Таблица В2

Результаты и условия эксперимента

№ эксперимента	$y$	$x_1$	$x_2$
1	72	6	145
2	63	6	70
3	57	1	145
4	49	1	70
5	61	3,5	107,5
6	67	3,5	145
7	64	6	107,5
8	56	3,5	70
9	52	1	107,5

Задание № 3. Оценить коэффициенты уравнения регрессии:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_1 \cdot x_2 + b_{11} \cdot x_1^2 + b_{22} \cdot x_2^2.$$

Результаты и условия эксперимента приведены в табл. В3.

Таблица В3

Результаты и условия эксперимента

№ эксперимента	$y$	$x_1$	$x_2$
1	58,7	1	0
2	49,2	-1	0
3	50,5	0,5	0,866
4	61,0	0,5	-0,866
5	43,8	-0,5	0,866
6	57,7	-0,5	-0,866
7	50,1	0	0

Задание № 4. Оценить коэффициенты уравнения регрессии:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2.$$

Результаты и условия эксперимента приведены в табл. В4.

Результаты и условия эксперимента

№ эксперимента	$y$	$x_1$	$x_2$
1	0,91	85	0,71
2	0,83	91	0,86
3	0,93	78	0,45
4	0,85	90	0,81
5	0,83	93	0,83
6	0,92	84	0,60
7	0,91	82	0,53
8	0,81	92	0,88
9	0,80	94	0,87
10	0,80	93	0,82
11	0,82	90	0,75
12	0,89	89	0,80
13	0,90	88	0,78
14	0,91	90	0,83
15	0,86	92	0,91
16	0,91	88	0,80
17	0,91	91	0,85
18	0,89	92	0,86
19	0,90	90	0,87
20	0,89	93	0,85

Задание № 5. Оценить коэффициенты уравнения регрессии:

$$\hat{y} = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2.$$

Результаты и условия эксперимента приведены в табл. В5.

## Результаты и условия эксперимента

№ эксперимента	$y$	$x_1$	$x_2$
1	7,00	3,90	10,00
2	7,00	3,90	14,00
3	7,00	3,70	15,00
4	7,00	4,00	16,00
5	7,00	3,80	17,00
6	7,00	4,80	19,00
7	8,00	5,40	19,00
8	8,00	4,40	20,00
9	8,00	5,30	20,00
10	10,00	6,80	20,00
11	9,00	6,00	21,00
12	11,00	6,40	22,00
13	9,00	6,80	22,00
14	11,00	7,20	25,00
15	12,00	8,00	28,00
16	12,00	8,20	29,00
17	12,00	8,10	30,00
18	12,00	8,50	31,00



**Приложение Г**  
(обязательное)

**Варианты заданий для лабораторной работы № 4**

Набор данных для вариантов задания лабораторной работы № 4 представлены в табл. Г1 – табл. Г4.

*Таблица Г1*

Набор данных  $A$  для варианта № 1

№ примера	Признак $k_1$	Признак $k_2$	Признак $k_3$	Признак $k_4$
1	144	15	4,7	0,43
2	151	19	4,9	0,43
3	157	15	4,9	0,48
4	170	7	5,2	0,73
5	152	11	5	0,77
6	145	23	4,6	0,28
7	175	24	5,5	0,4
8	149	27	4,7	0,42
9	99	10	4,3	0,43
10	113	8	3,7	0,44
11	140	18	4,6	0,44
12	102	15	4,1	0,46
13	135	11	4,2	0,5
14	150	19	4,7	0,76
15	149	6	5	0,79
16	68	15	2,3	0,38
17	136	19	4,4	0,43
18	144	24	4,9	0,43
19	72	6	2,9	0,46
20	97	7	4,2	0,47

Набор данных  $A$  для варианта № 2

№ прим ера	Приз нак $k_1$	Приз нак $k_2$	Приз нак $k_3$	Приз нак $k_4$	Приз нак $k_5$	Приз нак $k_6$	Приз нак $k_7$	Приз нак $k_8$	Приз нак $k_9$	Приз нак $k_{10}$
1	10	10	9	10	10	10	9	10	10	9
2	10	10	4	10	5	5	4	5	4	3
3	5	4	10	5	10	4	10	5	3	10
4	10	10	9	10	10	10	9	10	10	9
5	4	3	5	4	3	10	4	10	10	5
6	10	10	4	10	5	4	3	4	5	5
7	4	4	5	5	4	10	5	10	10	6
8	4	5	3	4	5	10	4	10	10	4
9	4	5	10	4	10	5	10	4	3	10
10	10	10	4	10	5	4	4	5	4	4
11	4	5	10	5	10	4	10	4	5	10
12	10	10	9	10	10	9	9	10	10	10
13	6	5	4	3	5	10	5	10	10	5
14	4	5	10	4	10	5	10	3	4	10
15	10	10	9	10	10	9	10	9	10	10
16	6	5	3	4	4	10	4	10	10	5
17	10	10	5	10	4	5	4	3	4	5
18	4	5	10	4	10	4	10	4	4	10

Набор данных  $A$  для варианта № 3

№ примера	Признак $k_1$	Признак $k_2$
1	27	19
2	11	46
3	25	15
4	36	27
5	35	25
6	10	43
7	11	44
8	36	24
9	26	14
10	26	14
11	9	45
12	33	23
13	27	16
14	10	47

Набор данных  $A$  для варианта № 4

№ примера	Признак $k_1$	Признак $k_2$	Признак $k_3$	Признак $k_4$	Признак $k_5$	Признак $k_6$
1	0	4	0	0	2	4
2	0	4	2	2	4	3
3	0	0	5	0	3	0
4	2	0	0	2	3	3
5	4	0	0	0	2	0
6	2	0	2	2	3	3
7	4	0	0	0	2	0
8	3	0	0	3	4	4
9	0	2	0	0	0	0
10	0	3	3	0	0	4
11	0	0	0	0	0	4
12	0	3	0	0	0	4
13	0	0	4	0	4	4
14	0	0	3	0	4	3
15	0	0	4	0	4	0
16	0	0	0	2	2	3
17	0	0	3	0	3	3