

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ
РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Пензенский государственный университет
архитектуры и строительства»
(ПГУАС)

Р.В. Тарасов, Л.В. Макарова

ОСНОВЫ ТЕОРИИ ЭКСПЕРИМЕНТА

Учебно-методическое пособие
по выполнению курсовой работы

Пенза 2015

УДК 005.6:001.12(075.8)

ББК 65.290-80я73

T19

Рекомендовано Редсоветом университета

Рецензенты: доктор технических наук, профессор
В.И. Логанина (ПГУАС);
заместитель директора по качеству
ООО «Строительные материалы»,
кандидат технических наук, доцент
В.Ю. Нестеров

Тарасов Р.В.

T19 Основы теории эксперимента: учебно-методическое пособие по
выполнению курсовой работы / Тарасов Р.В., Макарова Л.В. –
Пенза: ПГУАС. – 44 с.

Изложены последовательность выполнения курсовой работы и содержания расчетно-пояснительной записки. Приведены необходимые сведения о статистических методах обработки экспериментальных данных и о методах, применяемых при проведении экспериментальных исследований. Даны примеры решения типовых задач.

Учебно-методическое пособие подготовлено на кафедре «Управление качеством и технология строительного производства» и предназначено для студентов, обучающихся по направлению подготовки 27.04.02 «Управление качеством» при выполнении курсовой работы по дисциплине «Основы теории эксперимента».

© Пензенский государственный университет
архитектуры и строительства, 2015

© Тарасов Р.В., Макарова Л.В., 2015

ПРЕДИСЛОВИЕ

Учебно-методическое пособие по выполнению курсовой работы позволит овладеть следующими компетенциями:

– способностью формулировать цели и задачи исследования, выявлять приоритеты решения задач, выбирать и создавать критерии оценки (ОПК-1);

– способностью к самостоятельному обучению новым методам исследования, к изменению научного и научно-производственного профиля своей профессиональной деятельности (ОПК-2);

– способностью применять современные методы исследования, оценивать и представлять результаты выполненной работы (ОПК-6);

– способностью осуществлять постановку задачи исследования, формирование плана его реализации (ПК-6);

– способностью выбирать существующие или разрабатывать новые методы исследования (ПК-7).

ВВЕДЕНИЕ

Цель курсовой работы – закрепить теоретический материал изучаемой дисциплины, привить обучающимся знания и навыки выполнения инженерных исследований, в том числе и практических навыков экспериментальных исследований в условиях организации исследовательской работы.

В курсовой работе студенты должны овладеть практическими навыками использования методов, применяемых в экспериментальных исследованиях.

Основное внимание уделяется:

- выбору объекта исследования, параметра и факторов;
- практическому применению регрессионного анализа.

1. ЗАДАНИЕ НА КУРСОВУЮ РАБОТУ

В задании на курсовой проект, которое выдается индивидуально для каждого студента, содержится следующая информация (прил. 1):

- последовательность выполнения работы;
- сроки выполнения курсового проекта.

2. СОСТАВ И СОДЕРЖАНИЕ КУРСОВОЙ РАБОТЫ

Курсовая работа состоит из расчетно-пояснительной записки объемом 25-30 машинописных страниц в комплекте представляет собой принятое студентом решение поставленной задачи.

Расчетно-пояснительная записка должна быть написана от руки с одной стороны листа бумаги формата А4 или машинописным способом через 1,5 интервала. На каждый лист пояснительной записки наносится карандашом рамка рабочего поля, отстоящая от кромки листа слева на 20 мм, а справа, снизу и сверху – на 5 мм. Расстояние от рамки до границы текста в начале строк – не менее 5 мм, в конце строк не менее –3 мм;

от верхней и нижней строк - не менее 10 мм.

Пояснительная записка должна содержать:

- титульный лист;
- задание на курсовую работу;
- содержание;
- введение;
- основную часть;
- список использованных источников;
- приложение (при необходимости).

Титульный лист выполняется по форме, указанной в прил. 2 стандартным шрифтом.

Пояснительная записка должна излагаться грамотным литературным языком, со сжатыми и четкими формулировками, без лишних подробностей и повторений. Не допускается сокращения слов, кроме общепринятых. Страницы записки должны быть пронумерованы и, если есть таблицы, графики или рисунки, иметь название.

В расчетно-пояснительной записке предусматриваются разделы:

- введение-1...2 стр.;
- основная часть - 10...20 стр.;
- заключение -1...2 стр.
- библиографический список – 1...2 стр.

3. ПОСЛЕДОВАТЕЛЬНОСТЬ ВЫПОЛНЕНИЯ КУРСОВОЙ РАБОТЫ

Рекомендуется следующий порядок разработки проекта:

1. Ознакомиться с индивидуальным заданием, настоящими методическими указаниями и графиком выполнения курсовой работы.
2. Изучить соответствующие разделы рекомендуемой литературы.
3. Произвести необходимые описания и расчеты, в соответствии с заданием.
4. Оформить требуемые разделы расчетно-пояснительной записки согласно методическим указаниям по выполнению и оформлению курсовой работы.
5. Подготовить доклад и защитить курсовую работу.

4. КОНСУЛЬТАЦИИ И ЗАЩИТА КУРСОВОЙ РАБОТЫ

Основная цель консультаций – привить студентам навыки работы над справочной и нормативной литературой, монографиями, статьями в журналах, учебниками и т.п. На консультациях студенты должны обращаться к преподавателю со своими решениями. Задача преподавателя – оценить решенные вопросы и дать ответы на вопросы частного или принципиального характера.

Обучающийся обязан выполнить отдельные разделы работы в сроки, установленные преподавателем, и явиться в дни обязательных консультаций для контроля выполнения ими индивидуального задания в соответствующие сроки.

Студент защищает свой проект перед преподавателем в присутствии других студентов.

К защите студент предоставляет пояснительную записку. До защиты курсовая работа хранится у студента.

Оценка за проект ставится по пятибалльной системе. При этом учитывается: глубина проработки курсовой работы; качество оформления; умение докладывать и отвечать на вопросы.

В случае неудовлетворительной оценки студент дорабатывает работу или получает новое задание по усмотрению преподавателя.

Защищенная курсовая работа хранится на кафедре.

5. РАСЧЕТНО-ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Расчетно-пояснительная записка должна включать в себя следующие основные разделы:

Введение.

1. Выбор объекта исследования, параметра оптимизации и факторов.
 - 1.1 Описание объекта исследования.
 - 1.2 Выбор параметра оптимизации.

1.3 Выбор факторов.

2. Линейная парная регрессия.

На основании статистических данных, полученных при анализе деятельности предприятия, получить зависимость и проверить основные гипотезы.

2.1 Расчет коэффициентов уравнения регрессии.

2.2 Проверка адекватности модели.

2.3 Проверка значимости коэффициентов уравнения регрессии.

3. Нелинейная парная регрессия.

На основании статистических данных, полученных при анализе деятельности предприятия, получить зависимость и проверить основные гипотезы.

3.1 Расчет коэффициентов уравнения регрессии.

3.2 Проверка адекватности модели.

3.3 Проверка значимости коэффициентов уравнения регрессии.

Заключение.

Список использованных источников.

5.1. Введение

Введение расчетно-пояснительной записки должно содержать краткий обзор состояния, перспективы и пути решения поставленных задач. В обзоре необходимо отразить практическую значимость экспериментальных исследований с использованием регрессионного анализа.

5.2. Выбор объекта исследования, параметра оптимизации и факторов

Экспериментальные исследования ведутся во всех областях науки и техники. Экспериментальные исследования проводятся в лабораториях, на производстве, на опытных полях и участках, в клиниках и т.д.

Эксперимент – система операций, воздействий и наблюдений, направленных на получение информации об объекте при исследовательских операциях.

Опыт – воспроизведение исследуемого явления в определенных условиях проведения эксперимента при возможной регистрации его результатов.

Эксперимент может быть физическим, психологическим или модельным. Он может непосредственно проводиться на объекте или на его модели. Модель обычно отличается от объекта масштабом, а иногда природой.

Если модель достаточно точно описывает объект, то эксперимент на объекте может быть заменен экспериментом на модели. В последнее время наряду с физическими моделями все большее распространение получают абстрактные математические модели. Можно получать новые сведения об

объекте, экспериментируя на модели, если она достаточно точно описывает объект.

Для проведения эксперимента с наибольшей эффективностью необходим научный подход к его планированию, что позволит собрать необходимые данные, использовать для их анализа статистические методы и сделать правильные и объективные выводы.

Таким образом, планирование эксперимента – это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью. При этом существенно следующее:

- стремление к минимизации общего числа опытов;
- одновременное варьирование всеми переменными, определяющими процесс, по специальным алгоритмам;
- использование математического аппарата, формализующего многие действия экспериментатора;
- выбор четкой стратегии, позволяющей получать адекватные данные и принимать обоснованные решения после каждой серии экспериментов.

Задачи, для решения которых может использоваться планирование эксперимента, чрезвычайно разнообразны. К ним относят поиск оптимальных условий, построение интерполяционных формул, выбор существенных факторов, оценка и уточнение констант теоретических моделей, выбор наиболее приемлемых из некоторого множества гипотез о механизме явлений и т.д.

Для описания и анализа объекта исследования можно воспользоваться представлением о «черном ящике», который схематически изображен на рис. 1.

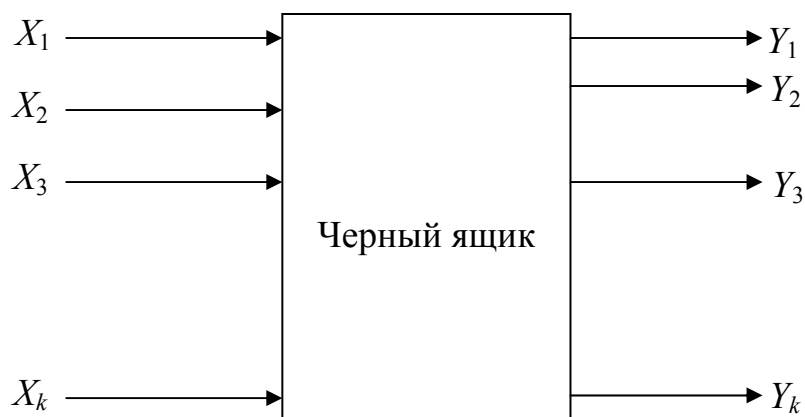


Рис. 1. Схема «черного ящика»

Стрелки справа изображают численные характеристики целей исследования, которые называются параметрами оптимизации.

Для проведения эксперимента необходимо иметь возможность воздействовать на поведение «черного ящика». Стрелки слева характеризуют способы воздействия на объект и называются факторами

При решении задачи используются математические модели объекта исследования, т.е. уравнение, связывающее параметр оптимизации с факторами. Это уравнение в общем виде выглядит следующим образом:
 $y = \varphi(x_1, x_2, \dots, x_k),$

Каждый фактор может принимать в опыте одно из нескольких значений, т.е. уровней. Если используется непрерывный ряд, то фактор способен принимать бесконечное множество значений. Однако на практике точность, с которой устанавливается некоторое значение, не безгранична. Поэтому можно считать, что любой фактор имеет определенное число дискретных уровней.

Фиксированный набор уровней факторов (т.е. установление каждого фактора на некоторый уровень) определяет одно из возможных состояний «черного ящика». Одновременно это есть условия проведения одного из возможных опытов. Если перебрать все возможные наборы состояний, то получится полное множество различных состояний данного «ящика». Одновременно это будет число возможных различных опытов.

Таким образом, целью эксперимента является установление степени влияния каждого фактора на отклик (параметр оптимизации) или получение функции, связывающей факторы и отклик. Полученную зависимость между факторами и откликом называют *поверхностью отклика*, уравнение, связывающее факторы и отклик – *регрессионным уравнением*, а определение коэффициентов этого уравнения – *оценкой коэффициентов*.

Фактором называют параметр, значение которого в эксперименте задают. Факторов может быть несколько. *Отклик* – параметр, который измеряется при различных значениях факторов, он обязательно должен быть один. Исследование объекта заключается в задании ряда значений факторов и получении для каждой совокупности факторов отклика. Отдельный акт по получению отклика по значениям факторов называется *экспериментом* или *экспериментальной точкой*. Совокупность значений факторов всех элементов, проводимых при одном исследовании, называется *планом эксперимента*. Диапазон изменений факторов, при котором проводится эксперимент, называется *факторным полем*.

При сборе статистического материала погрешность получаемых параметров часто сопоставима с величиной их изменения и зависимость теряется в случайном шуме. В этом случае поле факторов разбивается на равные интервалы и определяется среднее значение фактора и отклика для точек, попадающих в каждый интервал. Полученные точки образуют эмпирическую линию регрессии, по которой может быть получена функция зависимости отклика от факторов.

При использовании статистического подхода к планированию экспериментов и анализу данных необходимо, чтобы все участники эксперимента еще до его начала ясно понимали, что именно предстоит исследовать и каким образом нужно собирать данные или хотя бы имели представление о том, как эти данные нужно анализировать. Можно рекомендовать следующую схему:

1. Признание факта существования задачи и ее формулировка. На данном этапе необходимо выработать ясную и общепринятую формулировку проблемы исследования. Необходимо уточнить все представления о целях эксперимента. Ясная формулировка задачи во многих случаях оказывается существенной для лучшего понимания явлений и решения задачи.

2. Выбор факторов и уровней. Экспериментатор должен отобрать независимые переменные, или *факторы*, которые будут исследоваться в эксперименте. Факторы в эксперименте могут быть количественными или качественными. При исследовании количественных факторов нужно продумать то, как поддерживать желаемые значения этих факторов и как их измерять. Также необходимо выбрать значения, или *уровни*, факторов, которые будут использованы в эксперименте. Эти уровни могут быть заданы или выбраны случайным образом из множества всех возможных уровней фактора.

3. Выбор переменной отклика. При выборе отклика, или зависимой переменной, экспериментатор должен быть уверен в том, что отклик, который предстоит измерять, действительно содержит информацию об исследуемой проблеме. Нужно подумать также и о том, как будет измеряться отклик и какова вероятная точность этих измерений.

4. Выбор плана эксперимента. Этот этап является наиболее важным в процессе экспериментирования. Исследователь должен задать величину отличия от истинного отклика, которое он хочет обнаружить, и величину риска, на который он может пойти, с тем, чтобы выбрать соответствующий объем выборки (число реплик). Он должен также определить, в каком порядке будут собираться данные, и какой метод рандомизации будет применен. Необходимо всегда согласовывать между собой статистическую точность и стоимость эксперимента. Большинство рекомендуемых планов экспериментов и статистически эффективны, и экономичны, поэтому усилия экспериментатора по обеспечению статистической точности обычно приводят и к экономической эффективности.

Должна быть предложена математическая модель эксперимента, что позволит провести статистический анализ данных.

5. Проведение эксперимента. В ходе эксперимента, т.е. реального процесса сбора данных, исследователь должен внимательно следить за тем, чтобы все проходило в соответствии с планом. Особое внимание нужно обращать на рандомизацию, точность измерений и поддержание как можно большей однородности внешних условий эксперимента.

6. Анализ данных. Для анализа данных эксперимента должны применяться статистические методы. При этом важно не забывать и о точности вычислений. Современные средства вычислительной техники во многом облегчили экспериментатору эту задачу, одновременно сократив трудоемкость вычислений. В процессе анализа данных часто оказываются полезными и графические методы.

7. Выводы и рекомендации. По завершении анализа данных экспериментатор может сделать выводы относительно своих результатов. Нужно дать физическую интерпретацию статистических выводов и оценить их практическое значение, а затем вывести рекомендации об использовании результатов. Эти рекомендации могут включать и проведение дальнейших экспериментов, поскольку исследование обычно является *итеративным* процессом, когда отдельный эксперимент, отвечая на некоторые опросы, одновременно ставит новые. Представляя свои результаты и выводы, экспериментатор должен стараться использовать как можно меньше специальной статистической терминологии и формулировать свое сообщение по возможности проще. Использование рисунков и графиков – очень эффективный способ представления важных результатов эксперимента.

5.2.1. Описание объекта исследования

На предварительном этапе подготовки к процедуре планирования эксперимента целесообразно составление анкеты для сбора априорной информации.

Постановка задачи, выбор параметров оптимизации

1. Краткое описание процесса, объекта.
2. Формулировка цели исследования (если задач несколько – проранжировать их по степени важности).
3. Выбор параметров оптимизации (откликов). Заполните следующую таблицу, включив в нее все возможные отклики (табл. 1).

Т а б л и ц а 1

Параметры оптимизации

Номер отклика	Название	Размерность	Область определения	Точность	Примечание

4. Желаемый результат. Число и точность.
5. Какой результат будет считаться отличным, хорошим, удовлетворительным, неудовлетворительным.

Выбор факторов

1. Список всех «подозреваемых»: факторов, которые могут влиять на процесс.
2. Список факторов, включаемых в реальный эксперимент (табл. 2).

Факторы

Номер фактора	Название	Размерность	Область определения	Область интереса	Точность	Примечание

3. Существуют ли возможности установления значения фактора на любом заданном уровне?

4. Сохраняются ли заданные значения уровней в течение опыта?

5. Могут ли некоторые комбинации уровней факторов привести к остановке процесса (например, взрыв, не технологичность и т.д.)?

Число опытов

1. Желаемое число опытов, ограничения на число опытов.

2. Желаемый срок проведения исследования.

3. Примерная длительность одного опыта.

4. Стоимость и затраты труда при проведении одного опыта серии.

5. Желаемое число уровней для одного фактора.

6. Возможность выполнения параллельных опытов и их желаемое число.

7. Возможность проведения параллельных измерений.

8. Желаемая стратегия проведения опытов (например, по одному в день и т.д.).

Учет априорной информации

1. Условия и результаты, достигнутые при изучении аналогичных процессов.

2. Результаты предварительного эксперимента и данные (литературные и собственные) о величине ошибки эксперимента.

3. Взаимодействия факторов.

5.2.2. Выбор параметра оптимизации

При планировании эксперимента определяющим является выбор параметра оптимизации. Цель исследования должна быть сформулирована очень четко и предполагает количественную оценку. Параметр оптимизации является реакцией (откликом) на воздействие факторов, которые определяют поведение выбранной вами системы. Реакция объекта многогранна, многоаспектна. Выбор того аспекта, который представляет наибольший интерес, как раз и задается целью исследования.

При выборе параметра оптимизации необходимо учитывать ряд требований:

1. Параметр оптимизации должен быть *количественным*, задаваться числом.

Множество значений, которые может принимать параметр оптимизации, называется областью его определения. Области определения могут быть непрерывными и дискретными, ограниченными, и неограниченными.

2. Параметр оптимизации должен выражаться *одним числом*.

В большинстве случаев это получается естественно, как регистрация показания прибора. Чаще приходится производить некоторые вычисления. Так бывает при расчете выхода реакции. В химии часто требуется получать продукт с заданным отношением компонентов, например, $A:B=3:2$. Один из возможных вариантов решения подобных задач состоит в том, чтобы выразить отношение одним числом (1,5) и в качестве параметра оптимизации пользоваться значениями отклонений (или квадратов отклонений) от этого числа.

3. *Однозначность* в статистическом смысле. Заданному набору значений факторов должно соответствовать одно с точностью до ошибки эксперимента значение параметра оптимизации.

4. *Эффективность* оценивания системы.

Для успешного достижения цели исследования необходимо, чтобы параметр оптимизации оценивал эффективность функционирования системы в заранее выбранном смысле. Это требование является главным, определяющим корректность постановки задачи. Представление об эффективности не остается постоянным в ходе исследования. Оно меняется по мере накопления информации и в зависимости от достигнутых результатов. Это приводит к последовательному подходу при выборе параметра оптимизации. Так, например, на первых стадиях исследования технологических процессов в качестве параметра оптимизации часто используется выход продукта. Однако в дальнейшем, когда возможность повышения выхода исчерпана, нас начинают интересовать такие параметры, как себестоимость, чистота продукта и т.д.

5. *Эффективность параметра оптимизации в статистическом смысле*. Фактически это требование сводится к выбору параметра оптимизации, который определяется с наибольшей возможной точностью. (Если и эта точность недостаточна, тогда приходится обращаться к увеличению числа повторных опытов.)

6. Следующее требование к параметру оптимизации – требование *универсальности* или *полноты*. Под универсальностью параметра оптимизации понимается его способность всесторонне характеризовать объект. В частности, технологические параметры оптимизации недостаточно универсальны: они не учитывают экономику. Универсальностью обладают, например, обобщенные параметры оптимизации, которые строятся как функции от нескольких частных параметров.

7. Параметр оптимизации должен иметь *физический смысл, быть простым и легко вычисляемым*.

Требование физического смысла связано с последующей интерпретацией результатов эксперимента.

5.2.3. Выбор факторов

При планировании эксперимента необходим учет всех факторов которые могут влиять на процесс. Так же, как и параметр оптимизации, каждый фактор имеет область определения и принимает различные значения, т.е. варьируется на разных уровнях.

При поддержании фактора на некотором фиксированном уровне может быть ложное представление об оптимуме, так как нет гарантий, что фиксированный уровень является оптимальным. Естественно, что чем меньше уровень варьирования, тем проще организовать эксперимент. Однако это не всегда соответствует лучшим свойствам матрицы планирования с точки зрения математической обработки, поэтому фактор считается заданным, если вместе с его названием указана область его определения.

Под областью определения понимается совокупность всех значений, которые может принимать данный фактор.

Область определения может быть непрерывной и дискретной. В подавляющем большинстве случаев используются дискретные области определения. Так, для факторов с непрерывной областью определения, таких, как температура время, количество вещества и т.п., всегда выбираются дискретные множества уровней. В практических задачах области определения факторов, как правило, ограничены. Ограничения могут носить принципиальный либо технический характер.

Необходимо знать, является ли фактор переменной величиной, которую можно оценивать количественно: измерять, взвешивать, титровать и т.п., или же он – некоторая переменная, характеризующаяся качественными свойствами, т.е. факторы разделяются на количественные и качественные. Качественные факторы – это разные вещества, разные технологические способы, аппараты, исполнители и т.д.

Выбор качественных или количественных факторов зависит от степени изученности процесса. Если приходится планировать эксперимент на первой стадии, когда еще неясно какие вещества или материалы необходимо использовать для достижения оптимального результата, то в этом случае при постановке эксперимента совершенно необходимо вводить некоторое количество качественных факторов, не исключая и количественные факторы.

При планировании эксперимента факторы должны быть *управляемыми*. Это значит, что экспериментатор, выбрав нужное значение фактора, может его поддерживать постоянным в течение всего опыта, т. е. может управлять фактором. В этом состоит особенность «активного» эксперимента. Планировать эксперимент можно только в том случае, если уровни факторов подчиняются воле экспериментатора.

Чтобы точно определить фактор, нужно указать последовательность действий (операций), с помощью которых устанавливаются его конкретные значения (уровни). Такое определение фактора называется *операцио-*

нальным. Введение операционального определения обеспечивает однозначное понимание фактора.

С операциональным определением связаны выбор размерности фактора и точность его фиксирования. В основном выбор размерности фактора не представляет особой трудности. Это действительно так в тех случаях, когда существуют соответствующие методы измерений, построены измерительные шкалы, приборы, созданы эталоны и т.д. Так обстоит дело при измерении температуры, времени, давления и т.д. В некоторых случаях выбор размерности превращается в весьма трудную проблему выбора измерительной шкалы. Замена одной измерительной шкалы другой называется преобразованием шкал. Оно может быть использовано для упрощения модели объекта.

Точность замера факторов должна быть возможно более высокой. Степень точности определяется диапазоном изменения факторов.

Факторы должны быть непосредственными воздействиями на объект. Факторы должны быть *однозначны*. Невозможно управлять фактором, который является функцией других факторов. Не редко при планировании используют сложные факторы, т.е. взаимодействие отдельных факторов.

При планировании эксперимента обычно одновременно изменяется несколько факторов. Поэтому очень важно сформулировать требования, которые предъявляются к совокупности факторов. Прежде всего, выдвигается требование *совместимости*. Совместимость факторов означает, что все их комбинации осуществимы и безопасны.

При планировании эксперимента важна *независимость* факторов, т.е. возможность установления фактора на любом уровне вне зависимости от уровней других факторов. Если это условие невыполнимо, то планировать эксперимент невозможно.

Для выбора наиболее значимых факторов (или параметров оптимизации на стадии предварительного изучения объекта исследования при формализации априорных сведений можно провести психологический эксперимент, заключающийся в объективной обработке данных, полученных в результате опроса специалистов или в результате исследований, опубликованных в литературе. Такого рода эксперимент позволяет более правильно спроектировать объект исследования, принять или отвергнуть некоторые предварительные гипотезы, дать сравнительную оценку влияния различных факторов на параметры оптимизации и тем самым правильно отобрать факторы для последующего активного эксперимента, обоснованно исключив некоторые из них из дальнейшего рассмотрения. Особенность метода априорного ранжирования факторов заключается в том, что факторы ранжируются в порядке убывания вносимого им вклада. Вклад каждого фактора оценивается по величине ранга – места, которое отведено исследователем (специалистом при опросе, экспертом) данному фактору при ранжи-

ровании всех факторов с учетом их предполагаемого (количественно неизвестного) влияния на параметры оптимизации.

Результаты опроса специалистов обрабатываются следующим образом. Сначала определяют сумму рангов по факторам $\left(\sum_1^m a_{ij}\right)$, а затем разность (Δi) между суммой каждого фактора и средней суммой рангов и сумму квадратов отклонений (S) :

$$\Delta i = \sum_1^m a_{ij} - \frac{\sum_1^k \sum_1^m a_{ij}}{k} = \sum_1^m a_{ij} - T$$

$$S = \sum_1^m (\Delta a)^2$$

где a_{ij} – ранг каждого i -го фактора j -го исследователя (специалиста);
 m – число исследователей;
 k – число факторов;
 T – средняя сумма рангов.

Полученные значения позволяют построить среднюю априорную диаграмму рангов, но предварительно необходимо оценить степень согласованности мнений всех исследователей с помощью коэффициента конкордации ω :

$$\omega = \frac{12S}{m^2(k^3 - k) - m \sum_1^m T_j},$$

где $T_j = \sum(t_j^3 - t_j)$; t_j – число одинаковых рангов в j -м ранжировании.

После расчета коэффициента необходимо оценить его значимость с помощью χ^2 -критерия, значение которого определяют по формуле:

$$\chi^2 = \frac{12S}{mk(k+1) - \frac{1}{k-1} \sum_1^m T_j}$$

Гипотеза о наличии согласия может быть принимается, если при заданном числе степеней свободы табличное значение χ^2 меньше расчетного для 5%-ного уровня значимости.

Оценив согласованность мнений всех исследователей, строят среднюю диаграмму рангов, откладывая по одной оси факторы, а по другой – соответствующие суммы рангов. Чем меньше сумма рангов данного фактора, тем выше его место в диаграмме. С помощью последней оценивается значимость факторов.

В случае неравномерного экспоненциального убывания распределения часть факторов можно исключить из дальнейшего рассмотрения, отнеся их

влияние к шумовому полю. Если же их распределение равномерное, то в эксперимент рекомендуется включать все факторы.

Построение средней априорной диаграммы рангов по известным литературным источникам полезно с той точки зрения, что она по существу является сокращенным литературным обзором по объему исследования.

Пример. В процессе предварительного изучения нового вида бетона были опрошены четыре специалиста ($m=4$). Данные опросы были использованы для априорного ранжирования факторов с целью выделения наиболее существенных из них. Проводился опрос с помощью анкеты, содержащей 8 факторов ($k=8$), которые нужно было проранжировать с учетом степени их влияния на прочность бетона. Матрица рангов, полученная из анкет, приведена в табл. 3

Используя данные табл. 3 рассчитаем коэффициент конкордации.

Т а б л и ц а 3

Матрица рангов

Исследователи (m)	Факторы ($k=8$)								T_j
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	
1	7	4	4	6	2	6	2	4	36
2	6	3	6	7	2	7	3	4	18
3	7	3	4	6	1	6	3	4	18
4	8	1	5	6	1	6	4	5	18
$\sum_1^m a_{ij}$	28	11	19	25	6	25	12	17	$\sum T_j = 90$
Δi	15,15	-9,40	1,40	4,60	-14,40	4,60	-8,40	-3,40	
$(\Delta i)^2$	229,52	88,36	1,96	21,16	207,36	21,16	70,56	11,56	$S = 651,64$

$$T_1 = (3^3 - 3) + (2^3 - 2) + (2^3 - 2) = 36$$

$$T_2 = (2^3 - 2) + (2^3 - 2) + (2^3 - 2) = 18;$$

$$T_3 = (2^3 - 2) + (2^3 - 2) + (2^3 - 2) = 18;$$

$$T_4 = (2^3 - 2) + (2^3 - 2) + (2^3 - 2) = 18;$$

$$\sum_1^4 T_j = 36 + 18 + 18 + 18 = 90;$$

$$T = \frac{\sum_1^4 T_j}{J_{\max}} = \frac{90}{7} = 12,85;$$

$$\omega = \frac{12S}{m^2(k^3 - k) - m \sum_1^m T_j} = \frac{12 \cdot 651,64}{4^2(8^3 - 8) - 4 \cdot 90} = 0,99 > 0.$$

Так как величина коэффициента конкордации существенно отличается от нуля, можно считать, что между мнениями исследователей имеется существенная связь. Практически все исследователи одинаково ранжируют факторы (найденное значение близко к единице).

Значимость коэффициента конкордации проверяли по χ^2 -критерию:

$$\chi^2 = \frac{12 \cdot 651,64}{4 \cdot 8 \cdot (8+1) - \frac{1}{8-1} \cdot 90} = 28,41.$$

Для 5%-ного уровня значимости при числе степеней свободы $f = 8-1=7$ $\chi^2=14,067$. В связи с тем, что табличное значение χ^2 -критерия меньше расчетного, можно с 95%-ной доверительной вероятностью утверждать, что мнение исследователей относительно степени влияния факторов согласуется в соответствии с коэффициентом конкордации $\omega = 0,99$. Это позволяет построить среднюю диаграмму рангов для рассматриваемых факторов (рис. 2).

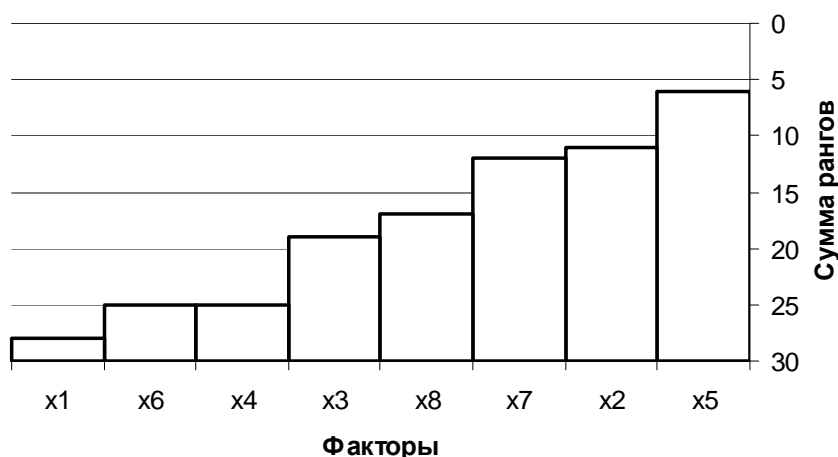


Рис. 2. Средняя диаграмма рангов

Анализ диаграммы показывает, что распределение – равномерное, убывание – немонотонное.

По результатам проведенного психологического эксперимента было решено оставить для дальнейшего рассмотрения пять факторов, занимающие первые пять мест.

5.3. Линейная парная регрессия

К так называемым «парным» зависимостям типа $y = f(x)$ относится подавляющее большинство всех формул, используемых в естественнонаучных и технических дисциплинах. С помощью метода наименьших квадратов можно построить парные зависимости оптимальной формы.

Когда непрерывным изменениям измеряемой величины x в некоторых характеристиках сопутствуют непрерывные изменения другой величины y , то утверждают, что между x и y имеется корреляция.

Метод, анализирующий корреляционную зависимость между несколькими переменными величинами, называют корреляционным анализом (correlation analysis). В частности, когда переменных величин только две, анализ называется простым корреляционным анализом, когда же одновременно подвергают анализу более трех переменных величин, то анализ называют сложным корреляционным анализом.

Когда между измеренной величиной x какой-либо характеристики и измеренной величиной y другой характеристики имеется корреляция, то такую взаимозависимость можно выразить количественно, создав формулу зависимости. Такую формулу зависимости называют регрессией (regression).

Для графической интерпретации регрессии и корреляции можно использовать корреляционную диаграмму.

Обычно корреляционной диаграммой (scatter diagram) называют такую диаграмму, на которую нанесены точки значений двух видов характеристик x и y , относящиеся к выборке объемом более 30 образцов, причем в качестве шкалы принимаются соответственно x и y по оси абсцисс и по оси ординат (рис. 3).

При составлении корреляционной диаграммы следует учитывать следующие моменты:

- 1) нет ли точек, которые стоят необычно далеко на отлете?
- 2) нельзя ли разграничить их на группы?

Что касается точек, стоящих на отлете, то поскольку это в большинстве случаев происходит из-за причин, вызванных погрешностями в измерениях или другими специфическими факторами, необходимо уточнить эти причины. Если же причина выяснена и сделано все возможное для ее устранения, то такие точки исключаются.

Благодаря разделению на группы можно получить много различной информации, поэтому очень важно рассмотреть ее по отдельным слоям: по сырью, по оборудованию, по операторам, по дням и часам осуществления производственных операций. В частности, в ситуации, когда точки после нанесения на диаграмму сосредоточиваются в виде отдельных групп, совершенно необходимо разделять такие данные послойно.

Если провести прямые линии параллельно оси абсцисс и оси ординат через точки (\bar{x}, \bar{y}) средних значений x и y , то плоская поверхность, по которой разбросаны точки, окажется разделенной на четыре части, которые секторами корреляционного поля.

Корреляционная диаграмма иллюстрирует, что точки, расположенные в I секторе, будут превышать средние значения, а точки, расположенные в III секторе, окажутся меньше средних значений, следовательно, при увеличении x в обоих секторах увеличивается и y , или при увеличении y увеличивается и x . С другой стороны, когда на этой диаграмме, во II и IV секто-

рах будет нанесено много точек, тогда при увеличении x соответственно уменьшается y .

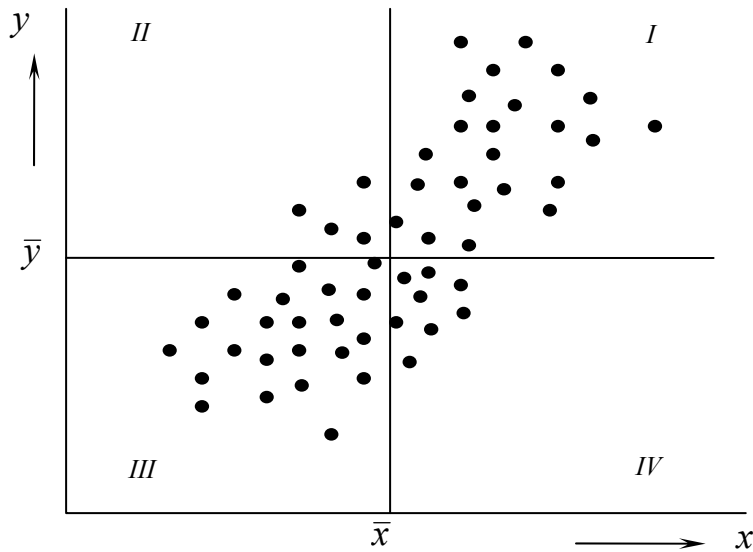


Рис. 3. Корреляционная диаграмма

Что касается нанесенных точек x, y , то считают: величина (сумма произведений отклонений), полученная умножением отклонений от соответствующих средних значений, представляет собой:

в I и III секторах $\sum (x - \bar{x})(y - \bar{y}) > 0$ - положительное корреляционное соотношение;

во II и IV секторах $\sum (x - \bar{x})(y - \bar{y}) < 0$ отрицательное корреляционное соотношение.

После нахождения суммы произведений, можно, исходя из нее, приблизительно понять степень корреляции, ибо если эта сумма составит значительную положительную величину, то это будет положительной корреляцией, если же она составит значительную отрицательную величину, то это будет отрицательной корреляцией. Вместе с тем, если рассматривать степень корреляции только по сумме произведений, то нужно учесть, что она изменяется в зависимости от рассеивания значений x и y . Поэтому в качестве критерия корреляции принимают сумму произведений, деленную на произведение корней квадратных из суммы квадратов каждого из отклонений x и y , что и называют коэффициентом корреляции (correlation coefficient).

Коэффициент корреляции данных выборки обычно выражают символом r , а формула имеет следующий вид:

$$r = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sqrt{\sum (x - \bar{x})^2} \sqrt{\sum (y - \bar{y})^2}}$$

или

$$r = \frac{\sum xy - \frac{(\sum x)(\sum y)}{n}}{\sqrt{\left[\sum x^2 - \frac{(\sum x)^2}{n}\right] \cdot \left[\sum y^2 - \frac{(\sum y)^2}{n}\right]}}$$

Кроме того, если обозначить сумму квадратов x , сумму квадратов y , а также сумму произведений x и y соответственно через $S(xx)$, $S(yy)$, $S(xy)$, то r можно выразить следующей формулой:

$$r = \frac{S(xy)}{\sqrt{S(xx) \cdot S(yy)}}$$

где $S(xx) = \sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n}$.

$$S(yy) = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n}$$

$$S(xy) = \sum (x - \bar{x})(y - \bar{y}) = \sum xy - \frac{(\sum x)(\sum y)}{n}$$

При вычислении коэффициента корреляции возможны две ситуации:

1. Ситуация, когда вычисления делаются непосредственно на основании измеренных значений. Данная ситуация возникает в том случае, когда, число данных сравнительно невелико.

2. Ситуация, когда вычисления производят по измеренным значениям, классифицированным в корреляционной таблице. Когда число измеренных величин велико, то производить непосредственные вычисления по измеренным значениям очень трудно. Тогда как при наличии одной характеристики очень удобно классифицировать измеренные величины и построить гистограмму. Для этого устанавливают несколько групп как для x , так и для y . Затем определяют частоту спаренных значений x и y , входящих в группы. В результате из созданной корреляционной таблицы удобно определить коэффициент корреляции r .

Коэффициент корреляции занимает промежуточное значение между -1 и $+1$. Причем, если вслед за увеличением x увеличивается и y , то коэффициент корреляции становится положительным, а если вслед за увеличением x уменьшается y , то он становится отрицательным. Поэтому при приближении $|r|$ к единице корреляция вполне вероятна, тогда как при приближении $|r|$ к нулю она маловероятна. Поскольку r представляет собой статистическую величину, вычисленную на основании опытных данных, то необходимо проверить значимость коэффициента корреляции.

Пусть двумерная генеральная совокупность (x, y) распределена нормально. Из этой совокупности извлечена выборка объема n и по ней най-

ден выборочный коэффициент корреляции $r \neq 0$. Требуется проверить нулевую гипотезу $H_0: r_r = 0$ о равенстве нулю генерального коэффициента корреляции.

Если нулевая гипотеза принимается, то это означает, что x и y некоррелированы; в противном случае – коррелированы.

Для того, чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе $H_1: r_r \neq 0$, надо вычислить наблюдаемое значение критерия $T_{набл}$

$$T_{набл} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$$

и по таблице критических точек распределения Стьюдента (табл. 2, прил.), по заданному уровню значимости α и числу степеней свободы $k=n-2$ найти критическую точку $t_{кр}(\alpha; k)$ двусторонней критической области. Если $|T_{набл}| < t_{кр}$ – нет оснований отвергнуть нулевую гипотезу. Если $|T_{набл}| > t_{кр}$ – нулевую гипотезу отвергают.

5.3.1. Расчет коэффициентов уравнения регрессии

Если между двумя характеристиками существует корреляция, то контролируя один параметр, можно контролировать и другой. В подобной ситуации необходимо знать, насколько увеличивается или уменьшается y в соответствии с увеличением или уменьшением x .

Что касается определенной области изменения x , то поскольку она связана линейной формулой

$$y = b_0 + b_1x$$

становится возможным строить оценку y , исходя из значений x .

В данной формуле, поскольку x является независимой переменной, а y – зависимой переменной, то b_1 называется коэффициентом регрессии. Кроме того, b_0 является постоянным членом. Следовательно, формула складывается из двух частей: из y – члена, который находится под влиянием изменений x , и из x – устойчивого члена, который является независимой переменной. Такая формула зависимости определяет линию регрессии.

Для определения линии регрессии необходимо статистически оценить коэффициент регрессии b_1 и постоянное число b_0 .

Для этого должны быть удовлетворены два следующих условия:

1. Линия регрессии должна проходить сквозь точки (\bar{x}, \bar{y}) средних значений x и y .

2. Сумма квадратов отклонений от линии регрессии значений y по всем точкам должна быть наименьшей. Для того чтобы удовлетворить этим двум условиям, целесообразно определить b_1, b_0 так, чтобы выражение $\sum (\bar{y} - \bar{Y}_i)^2 = \min$.

Таким образом, подставляя вместо Y_i подставить $b_0 + b_1x$, получим

$$U = \sum_{i=1}^n [y_i - (b_0 + b_1x_i)]_{\min}^2$$

или

$$U = \sum_{i=1}^n \Delta_i^2, \quad i = 1, n.$$

Для решения поставленной задачи необходимо в каждом конкретном случае вычислить значения коэффициентов b_0 и b_1 , минимизирующие сумму отклонений U . для этого необходимо вычислить частные производные функции U по коэффициентам b_0 и b_1 и приравнять их нулю:

$$\begin{cases} \frac{\partial U}{\partial b_0} = 0 \\ \frac{\partial U}{\partial b_1} = 0 \end{cases}$$

Решая эту систему уравнений, находим искомые значения b_0 и b_1 . Решая систему нормальных уравнений получим:

$$\begin{cases} \frac{\partial U}{\partial b_0} = \sum_{i=1}^n [y_i - (b_0 + b_1x_i)] = 0; \\ \frac{\partial U}{\partial b_1} = \sum_{i=1}^n [y_i - (b_0 + b_1x_i)]x_i = 0. \end{cases}$$

Преобразуем полученную систему нормальных уравнений:

$$b_0n + b_1 \sum x_i = \sum y_i, \quad b_0 \sum x_i + b_1 \sum x_i^2 = \sum (y_i x_i)$$

Такую систему можно решить с помощью определителей

$$b_0 = \frac{\Theta_1}{\Theta}, \quad b_1 = \frac{\Theta_2}{\Theta}$$

где Θ – главный определитель.

После нахождения определителей получим

$$b_0 = \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - (\sum x)^2}; \quad b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}.$$

В результате вычислений получим уравнение регрессии, отображающее с некоторой вероятностью зависимость y от x , построенную по экспериментальным точкам.

Пример. Рассчитать коэффициент корреляции и определить уравнение регрессии при изучении зависимости между плотностью и пористостью образцов каменных строительных материалов.

Отберем выборку объемом 20-30 образцов. Прежде чем, определим водопоглощение образцов, которое характеризует открытую пористость материала, измерим объем образца и массу (табл. 4).

Т а б л и ц а 4

Данные для вычисления коэффициента корреляции
и коэффициентов регрессии

Номер образца	x	y	x^2	y^2	xy
1	2	3	4	5	6
1	2,32	7,7	5,38	59,29	17,86
2	2,41	6,9	5,81	47,61	16,63
3	2,50	6,3	6,25	39,69	15,75
4	2,42	6,5	5,86	42,25	15,73
5	2,51	6,0	6,30	36,00	15,06
6	2,34	7,7	5,47	59,29	18,02
7	2,42	7,0	5,86	49,00	16,94
8	2,39	8,0	5,71	64,00	19,12
9	2,38	8,2	5,66	67,24	19,52
10	2,43	7,5	5,90	56,25	18,23
11	2,30	9,1	5,29	82,81	20,93
12	2,35	8,8	5,52	77,44	20,68
13	2,43	6,5	5,90	42,25	15,80
14	2,34	7,6	5,48	57,76	17,78
15	2,50	6,4	6,25	40,96	16,00
16	2,44	7,0	5,95	49,00	17,08
17	2,33	8,4	5,43	70,56	19,57
18	2,40	7,3	5,76	53,29	17,52
19	2,47	6,2	6,10	38,44	15,31
20	2,39	8,1	5,71	65,61	19,36
Σ	48,07	147,2	115,59	1098,74	352,89

Открытая пористость строительных материалов равна отношению объема всех пор насыщающихся водой, к объему материала, т.е. объемному водопоглощению

$$P_o = \frac{m_2 - m_1}{V}$$

где m_1 и m_2 – масса образцов в сухом и насыщенном водой состоянии, соответственно.

Как и другие статистические расчеты, вычисление коэффициента корреляции между плотностью материала x (г/см³) пористостью y (%), а также коэффициентов уравнения регрессии удобно представить в табличной форме (табл. 4).

Некоторое представление о возможной корреляции можно получить, построив корреляционную диаграмму (рис. 4).

Рассматривая эту диаграмму, мы не видим значительно рассеянных точек, поэтому можно полагать, что имеется корреляция.

Рассчитаем коэффициент корреляции.

$$S(xx) = \sum (x - \bar{x})^2 = \sum x^2 - \frac{(\sum x)^2}{n} = 115,59 - \frac{2310,72}{20} = 0,06$$

$$S(yy) = \sum (y - \bar{y})^2 = \sum y^2 - \frac{(\sum y)^2}{n} = 1098,74 - \frac{21667,84}{20} = 15,35$$

$$S(xy) = \sum (x - \bar{x})(y - \bar{y}) = \sum xy - \frac{(\sum x)(\sum y)}{n} = 352,89 - \frac{48,07 \cdot 147,2}{20} = -0,9$$

$$r = \frac{S(xy)}{\sqrt{S(xx) \cdot S(yy)}} = \frac{-0,9}{\sqrt{0,06 \cdot 15,35}} = -0,94$$

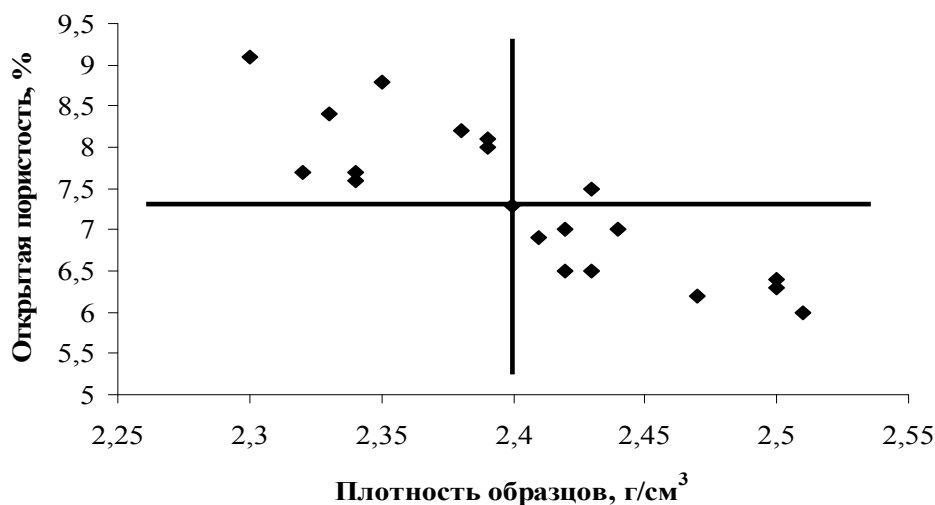


Рис. 4. Корреляционная диаграмма

Так как значение коэффициента корреляции отрицательное, следовательно, существует обратная линейная связь (с ростом значений x значения y уменьшаются).

Проверим гипотезу о значимости выборочного коэффициента корреляции. Найдем наблюдаемое (эмпирическое значение критерия) по формуле:

$$T_{набл} = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} = \frac{0,94\sqrt{20-2}}{\sqrt{1-0,94^2}} = \frac{3,98}{0,35} = 11,37$$

По таблице критических точек распределения Стьюдента (табл. 2. прил.), по уровню значимости $\alpha=0,05$, помещенному в верхней строке таблицы, и числу степеней свободы $k = n - 2 = 20 - 2 = 18$ находим критическую точку двусторонней критической области $t_{кр}(0,05;18)=2,1$.

Так как $|T_{набл}| > t_{кр}$ – отвергаем нулевую гипотезу о равенстве нулю генерального коэффициента корреляции, т.е. коэффициент корреляции значимо отличается от нуля; следовательно, x и y коррелированы.

Рассчитаем коэффициенты уравнения регрессии.

$$b_0 = \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - (\sum x)^2} = \frac{147,2 \cdot 115,59 - 352,89 \cdot 48,07}{20 \cdot 115,59 - 48,07^2} = 47,62$$

$$b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} = \frac{20 \cdot 352,89 - 48,07 \cdot 147,2}{20 \cdot 115,59 - 48,07^2} = -16,76$$

Уравнение регрессии или формула, которая отображает с некоторой вероятностью зависимость y от x , построенная по экспериментальным точкам, изображенным на рисунке 4, имеет вид

$$y = 47,62 - 16,76x.$$

5.3.2. Проверка адекватности модели

После вычисления коэффициентов модели необходимо оценить ее пригодность. Такая проверка называется проверкой адекватности модели

Если расставить экспериментальные точки относительно линии регрессии, то они будут характеризоваться некоторым разбросом, причем для каждой точки разброс будет различным (разная дисперсия воспроизводимости).

Для характеристики среднего разброса относительно линии регрессии вполне подходит остаточная сумма квадратов. Неудобство состоит в том, что она зависит от числа коэффициентов в уравнении: введите столько коэффициентов, сколько вы провели независимых опытов, и получите остаточную сумму, равную нулю. Поэтому предпочитают относить ее на один «свободный» опыт. Число таких опытов называется числом степеней свободы (f).

Числом степеней свободы в статистике называется разность между числом опытов и числом коэффициентов (констант), которые уже вычислены по результатам этих опытов независимо друг от друга.

Существует следующее правило: в планировании эксперимента число степеней свободы для дисперсии адекватности равно числу различных опытов, результаты которых используются при подсчете коэффициентов регрессии, минус число определяемых коэффициентов.

Остаточная сумма квадратов, деленная на число степеней свободы, называется *остаточной дисперсией*, или *дисперсией адекватности* ($s_{ад}^2$).

$$s_{ад}^2 = \frac{\sum_{i=1}^N \Delta y_i^2}{f}$$

Остаточная сумма квадратов (сумма квадратов невязок) представляет собой разность между предсказанным значением, рассчитанным по полученному уравнению регрессии, и полученным в результате эксперимента: $\Delta y = \hat{y} - \bar{y}$.

Для проверки гипотезы об адекватности можно использовать F-критерий (критерий Фишера).

$$F = \frac{s_{\text{ад}}^2}{s_{\{y\}}^2}$$

где $s_{\{y\}}^2$ – дисперсия параметра оптимизации.

При подсчете дисперсии параметра оптимизации квадрат разности между значением y_q в каждом опыте и средним значением из n повторных наблюдений \bar{y} нужно просуммировать по числу опытов в матрице N , а затем разделить на $N(n-1)$.

$$s_{\{y\}}^2 = \frac{\sum_1^N \sum_1^n (y_{iq} - \bar{y}_i)^2}{N(n-1)}$$

Дисперсию воспроизводимости проще всего рассчитывать, когда соблюдается равенство числа повторных опытов во всех экспериментальных точках. На практике чаще всего число повторных опытов различно.

При усреднении дисперсий можно пользоваться взвешенным значением дисперсий, взятым с учетом числа степеней свободы

$$s_{\{y\}}^2 = \frac{s_1^2 f_1 + s_2^2 f_2 + \dots + s_N^2 f_N}{f_1 + f_2 + \dots + f_N} = \frac{\sum_{i=1}^N s_i^2 f_i}{\sum_{i=1}^N f_i},$$

где s_1^2 – дисперсия первого опыта;

s_2^2 – дисперсия второго опыта и т.д.;

f_1 – число степеней свободы в первом опыте, равное числу параллельных опытов n_1 минус 1, т.е. $f_1 = n_1 - 1$;

f_2 – число степеней свободы во втором опыте и т.д.

Число степеней свободы средней дисперсии принимается равным сумме чисел степеней свободы дисперсий, из которых она вычислена.

Если взять среднее значение дисперсий без учета числа степеней свободы, то это будет ошибкой, также как если взять среднее значение стандартных отклонений. Стандартные отклонения нужно возвести в квадрат и затем взять взвешенное среднее, как указано выше.

Прежде чем проверять адекватность модели необходимо провести проверку однородности дисперсий.

Если сравниваемое количество дисперсий больше двух и одна дисперсия превышает остальные, то можно воспользоваться критерием Кохрена. Это критерий пригоден для случаев, когда во всех точках имеется одина-

ковое число повторных опытов. При этом подсчитывается дисперсия в каждой горизонтальной строке матрицы.

$$s^2 = \frac{\sum_1^n (y_q - \bar{y})^2}{n-1},$$

а затем из всех дисперсий находится наибольшая s_{\max}^2 , которая делится на сумму всех дисперсий. Критерий Кохрена – это отношение максимальной дисперсии к сумме всех дисперсий.

$$G = \frac{s_{\max}^2}{\sum_1^N s_i^2},$$

С этим критерием связаны числа степеней свободы $f_1 = n - 1$ и $f_2 = N$. Гипотеза об однородности дисперсий подтверждается, если экспериментальное значение критерия Кохрена не превышает табличного значения. Тогда можно усреднять дисперсии и пользоваться формулой

$$S_{\{y\}}^2 = \frac{\sum_1^N \sum_1^n (y_{iq} - \bar{y}_i)^2}{N(n-1)}.$$

Удобство использования критерия Фишера состоит в том, что проверку гипотезы можно свести к сравнению с табличным значением (прил. 2).

Таблица построена следующим образом. Столбцы связаны с определенным числом степеней свободы для числителя f_1 строки – для знаменателя f_2 . На пересечении соответствующей строки и столбца стоят критические значения F-критерия. Как правило, в технических задачах используется уровень значимости 0,05.

Если рассчитанное значение F-критерия не превышает табличного, то с соответствующей доверительной вероятностью модель можно считать адекватной. При превышении табличного значения эту приятную гипотезу приходится отвергать.

Если модель адекватна, то можно перейти к крутому восхождению. Если нет – приходится преодолевать дополнительные трудности. Это мы обсудим ниже. Но во всех случаях интересно проверять еще значимость отдельных коэффициентов регрессии.

5.3.3. Проверка значимости коэффициентов

Проверка значимости каждого коэффициента проводится независимо.

Ее можно осуществлять двумя равноценными способами: проверкой по t -критерию Стьюдента или построением доверительного интервала. При использовании полного факторного эксперимента или регулярных дробных реплик доверительные интервалы для всех коэффициентов (в том чис-

ле и эффектов взаимодействия) равны друг другу. Прежде всего надо, конечно, найти дисперсию коэффициента регрессии $s_{\{b_j\}}^2$

$$s_{\{b_j\}}^2 = \frac{s_{\{\bar{y}\}}^2}{N}.$$

Из формулы видно, что дисперсии всех коэффициентов равны друг другу, так как они зависят только от ошибки опыта и числа опытов.

Теперь легко построить доверительный интервал (Δb_j)

$$\Delta b_j = \pm t s_{\{b_j\}}.$$

Здесь t – табличное значение критерия Стьюдента при числе степеней свободы, с которыми определялась $s_{\{y\}}^2$, и выбранном уровне значимости (обычно 0,05); $s_{\{b_j\}}$ – квадратичная ошибка коэффициента регрессии

$$s_{\{b_j\}} = + \sqrt{s_{\{b_j\}}^2}.$$

Формулу для доверительного интервала можно записать в следующей эквивалентной форме:

$$\Delta b_j = \pm \frac{t s_{\{y\}}}{\sqrt{N}}$$

Коэффициент значим если его абсолютная величина больше доверительного интервала. Доверительный интервал задается верхней и нижней границами $b_j + \Delta b_j$ и $b_j - \Delta b_j$.

Для отыскания значений t -критерия можно воспользоваться таблицей, фрагмент из которой приведен в прил. 2.

Таблица построена следующим образом. Столбцы соответствуют различным степеням свободы и значениям критерия.

Пусть в двух разных задачах случайно оказались два численно равных коэффициента регрессии. Доверительные интервалы для них оказались различными. Из них значим только второй

Задача	b_j	Δb_j
1	5,3	$\pm 5,5$
2	5,3	$\pm 2,6$

В действительности, чем уже доверительный интервал (при заданном α), тем с большей уверенностью можно говорить о значимости коэффициента.

Если абсолютная величина коэффициента больше, чем доверительный интервал, то коэффициент значим.

Можно проверять значимость коэффициентов по критерию, то воспользоваться формулой

$$t = \frac{|b_j|}{s_{\{b_j\}}}.$$

Вычисленное значение t -критерия сравнивается с табличным при заданном α и соответствующем числе степеней свободы. Полученные выводы о значимости коэффициентов, конечно, должны совпадать с предыдущими.

Так производится проверка значимости коэффициентов.

Пример: Проверим адекватность модели и значимость коэффициентов регрессии на примере.

Прежде чем делать выводы о адекватности линейной модели необходимо проверить однородность дисперсий, для чего можно воспользоваться критерием Кохрена. Сначала подсчитаем дисперсии в каждой горизонтальной строке матрицы (табл. 5).

В нашем случае эта формула для расчета дисперсии приобретает вид:

$$s^2 = \frac{\sum_1^n (y_q - \bar{y})^2}{n-1} = \frac{\sum_1^2 (y_q - \bar{y})^2}{2-1} = 2(\Delta y)^2$$

Максимальная дисперсия оказалась в третьем опыте. Экспериментальный критерий Кохрена рассчитаем по формуле:

$$G = \frac{s_{\max}^2}{\sum_1^N s_i^2} = \frac{2}{6,32} = 0,32.$$

Т а б л и ц а 5

Расчет дисперсии воспроизводимости

Номер опыта	y'	y''	\bar{y}	Δy	$(\Delta y)^2$	S_i^2
1	84,8	83,2	84,0	0,8	0,64	1,28
2	77,2	76,8	77,0	0,2	0,04	0,08
3	83,0	81,0	82,0	1	1	2
4	83,5	82,5	83,0	0,5	0,25	0,5
5	80,1	79,9	80,0	0,1	0,01	0,02
6	81,9	80,1	81,0	0,9	0,81	1,62
7	78,5	77,5	78,0	0,5	0,25	0,5
8	78,6	79,4	79,0	0,4	0,16	0,32
Σ					3,16	6,32

Табличный критерий Кохрена равен: $G = 0,68$ (прил. 2). Экспериментальный критерий не превышает значения табличного, значит гипотеза об однородности дисперсий подтверждается.

Найдем дисперсию воспроизводимости. Для двух повторных опытов она приобретает следующий вид:

$$s_{\{y\}}^2 = \frac{2 \sum_{i=1}^N (y_{iq} - \bar{y}_i)^2}{N} = \frac{2 \cdot 3,16}{8} = 0,79.$$

Остаточная дисперсия (дисперсия адекватности) составила

$$s_{\text{ад}}^2 = \frac{\sum_{i=1}^N \Delta y_i^2}{f} = \frac{\sum_{i=1}^8 \Delta y_i^2}{f} = \frac{64}{8-3-1} = 16,0.$$

Табличное значение критерия Фишера составило (прил. 2) $F=6,04$. $F_{\text{расч}} = 16,0 / 0,79 = 20,25$. Так как расчетное значение коэффициента Фишера превышает табличное, то модель нельзя считать адекватной.

Проверим значимость коэффициентов.

Найдем доверительный интервал по формуле:

$$\Delta b_j = \pm \frac{ts_{\{y\}}}{\sqrt{N}} = \frac{2,306 \cdot 0,89}{\sqrt{8}} = \pm 0,727.$$

Сравним полученную величину интервала с абсолютными значениями коэффициентов. Сред полученных коэффициентов незначимым оказывается коэффициент b_1 . Остальные коэффициенты значимы.

Можно проверить значимость коэффициентов по t -критерию. Для отыскания значений t -критерия можно воспользоваться табл. 3 приложения. Сначала необходимо найти квадратичную ошибку коэффициента регрессии по формуле.

$$s_{\{b_j\}} = + \sqrt{s_{\{b_j\}}^2} = \sqrt{\frac{0,79}{8}} = 0,31.$$

Тогда

$$t_1 = \frac{|b_j|}{s_{\{b_j\}}} = \frac{0,5}{0,31} = 1,61 < 2,306;$$

$$t_2 = \frac{|b_i|}{s_{\{b_j\}}} = \frac{1}{0,31} = 3,22 > 2,306 \text{ и т. д.}$$

Незначимым является коэффициент b_1 .

Таким образом, окончательное уравнение регрессии будет иметь следующий вид:

$$y = 80,5 + x_3 + x_1x_2 + x_1x_3 + x_2x_3 - x_1x_2x_3.$$

5.4. Нелинейная парная регрессия

В том случае, когда гипотеза линейности может быть отброшена или когда при графическом изображении точек нелинейность явно просматривается, есть смысл получить по экспериментальным данным нелинейную

(квадратичную или высших порядков) формулу парной зависимости. При этом можно рассчитывать, что нелинейная формула даст меньшую остаточную дисперсию, т.е. лучше предскажет результат.

Чтобы отличать квадратичную парную регрессию от множественной и с целью упрощения расчетов целесообразно изменить систему обозначений коэффициентов регрессии. С учетом этой оговорки формулу квадратичной парной регрессии можно представить в виде

$$y = a + bx + cx^2$$

5.4.1. Расчет коэффициентов уравнения регрессии

Коэффициенты квадратичного уравнения a , b и c можно найти, решая систему трех нормальных уравнений с тремя неизвестными:

$$\begin{aligned} an + b \sum x + c \sum x^2 &= \sum y \\ a \sum x + b \sum x^2 + c \sum x^3 &= \sum xy \\ a \sum x^2 + b \sum x^3 + c \sum x^4 &= \sum x^2 y \end{aligned}$$

Пример: В результате проведенных наблюдений были получены экспериментальные данные, представленные в табл. 6.

Т а б л и ц а 6

Экспериментальные данные

№	x	y	xy	x^2	x^2y	x^3	x^4
1	1,70	25,00	42,50	2,89	72,25	4,91	8,35
2	3,40	34,00	115,60	11,56	393,04	39,30	133,63
3	4,00	57,00	228,00	16,00	912,00	64,00	256,00
4	4,10	82,00	336,20	16,81	1378,42	68,92	282,58
5	5,30	98,00	519,40	28,09	2752,82	148,88	789,05
Σ	18,50	296,00	1241,70	75,35	5508,53	326,02	1469,61

Взятые из таблицы значения подставляем в нормальные уравнения

$$\begin{aligned} 5,00a + 18,50b + 75,35c &= 296,00 \\ 18,50a + 75,35b + 326,01c &= 1241,70 \\ 75,35a + 326,01b + 1469,60c &= 5508,53 \end{aligned}$$

Для решения системы уравнений воспользуемся матричным методом.

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}$$

$$\Delta_1 = \begin{vmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 296,00 & 18,50 & 75,35 \\ 1241,70 & 75,35 & 326,01 \\ 5508,53 & 326,01 & 1469,60 \end{vmatrix} = 8689,93$$

$$\Delta_2 = \begin{vmatrix} a_{11} & b_1 & a_{13} \\ a_{21} & b_2 & a_{23} \\ a_{31} & b_3 & a_{33} \end{vmatrix} = \begin{vmatrix} 5,00 & 296,00 & 75,35 \\ 18,50 & 1241,70 & 326,01 \\ 75,35 & 5508,53 & 1469,60 \end{vmatrix} = -2651,44$$

$$\Delta_3 = \begin{vmatrix} a_{11} & a_{12} & b_1 \\ a_{21} & a_{22} & b_2 \\ a_{31} & a_{32} & b_3 \end{vmatrix} = \begin{vmatrix} 5,00 & 18,50 & 296,00 \\ 18,50 & 75,35 & 1241,70 \\ 75,35 & 326,01 & 5508,53 \end{vmatrix} = 1564,46$$

$$\Delta = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = \begin{vmatrix} 5,00 & 18,50 & 75,35 \\ 18,50 & 75,35 & 326,01 \\ 75,35 & 326,01 & 1469,60 \end{vmatrix} = 379,30$$

$$a = \frac{\Delta_1}{\Delta} = 22,91; \quad b = \frac{\Delta_2}{\Delta} = -6,99; \quad c = \frac{\Delta_3}{\Delta} = 4,12$$

Таким образом, уравнение можно представить в следующем виде:

$$y = 22,91 - 6,99x + 4,12x^2$$

5.4.2. Проверка адекватности модели

Проверка адекватности модели осуществляется по критерию Фишера (см п. 5.3.2).

5.4.3. Проверка значимости коэффициентов

Проверка значимости коэффициентов уравнения регрессии проводится по методике, указанной в п. 5.3.3.

5.5 Заключение

В заключении необходимо привести основные выводы об используемых методах регрессионного анализа, описать их преимущества и недостатки с указанием сфер их практического применения.

ПРИЛОЖЕНИЕ

Приложение 1

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РФ

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО
ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ АРХИТЕКТУРЫ И
СТРОИТЕЛЬСТВА»

КАФЕДРА УПРАВЛЕНИЕ КАЧЕСТВОМ И ТСП

ЗАДАНИЕ НА КУРСОВУЮ РАБОТУ по дисциплине «Основы теории эксперимента»

Студент _____.

Группа УК-11м

1. Тема: «**Линейная и нелинейная парная регрессия**»

2. Срок представления к защите: « 2 » июня 2015 г

3. Содержание пояснительной записки:

Введение.

1. Выбор объекта исследования, параметра оптимизации и факторов

1.1 Описание объекта исследования.

1.2 Выбор параметра оптимизации.

1.3 Выбор факторов

2. Линейная парная регрессия

На основании статистических данных, полученных при анализе деятельности предприятия, получить зависимость и проверить основные гипотезы.

2.1 Расчет коэффициентов уравнения регрессии.

2.2 Проверка адекватности модели.

2.3 Проверка значимости коэффициентов уравнения регрессии.

3. Нелинейная парная регрессия

На основании статистических данных, полученных при анализе деятельности предприятия, получить зависимость и проверить основные гипотезы.

3.1 Расчет коэффициентов уравнения регрессии.

3.2 Проверка адекватности модели.

3.3 Проверка значимости коэффициентов уравнения регрессии.

Заключение.

Список использованных источников.

Руководитель работы _____ /*к.т.н., доцент Р.В Тарасов*/

подпись, дата,

инициалы, фамилия

Задание принял: _____

подпись, дата

Сроки выполнения работы:

25 % –

75% -

Сдача работы – со _____ по

50 % –

100 % –

Приложение 2

Критические значения статистических критериев

Критерий Фишера $F_{кр}$ при $\alpha = 0,05$

f_2	Число степеней свободы f_1 (числитель)											
	1	2	3	4	5	6	7	8	9	10	20	∞
1	161	199	216	225	230	234	237	239	241	242	248	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,66	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,80	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,56	4,37
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	3,87	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,44	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,15	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	2,94	2,71
10	4,96	4,1	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,77	2,54
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,12	1,84
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,57	1,00

Критерий Фишера $F_{кр}$ при $\alpha=0,1$

f_2	Число степеней свободы f_1 (числитель)											
	1	2	3	4	5	6	7	8	9	10	20	∞
1	39,9	49,5	53,6	55,8	57,2	58,2	58,9	59,4	59,9	60,2	61,7	63,3
2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38	9,39	9,44	9,49
3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24	5,23	5,18	5,13
4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94	3,92	3,84	3,76
5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32	3,30	3,21	3,10
6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96	2,94	2,84	2,72
7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72	2,70	2,59	2,47
8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56	2,54	2,42	2,29
9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44	2,42	2,30	2,16
10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35	2,32	2,20	2,06
20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96	1,94	1,79	1,61
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	1,60	1,42	1,00

Критерий Кохрена $G_{кр}$ при $\alpha=0,05$

№	Число степеней свободы f											
	1	2	3	4	5	6	7	8	9	10	16	∞
2	0,99	0,97	0,93	0,90	0,87	0,85	0,83	0,81	0,80	0,78	0,73	0,50
3	8	5	9	6	7	3	3	6	1	8	4	0
4	0,96	0,87	0,79	0,74	0,70	0,67	0,65	0,63	0,61	0,60	0,54	0,33
5	7	1	8	6	7	7	3	3	7	2	7	3
6	0,90	0,76	0,68	0,62	0,59	0,56	0,53	0,51	0,50	0,48	0,43	0,25
7	6	8	4	9	0	0	6	8	2	8	7	0
8	0,84	0,68	0,59	0,54	0,50	0,47	0,45	0,43	0,42	0,41	0,36	0,20
9	1	4	8	4	6	8	6	9	4	2	4	0
10	0,78	0,61	0,53	0,48	0,44	0,41	0,39	0,38	0,36	0,35	0,31	0,16
15	1	6	2	0	5	8	8	2	8	7	4	7
20	0,72	0,56	0,48	0,43	0,39	0,37	0,35	0,33	0,32	0,31	0,27	0,14
∞	7	1	0	1	7	3	4	8	6	5	6	3
	0,68	0,52	0,43	0,39	0,36	0,33	0,31	0,30	0,29	0,28	0,24	0,12
	0	0	8	1	0	6	8	4	3	3	6	5
	0,63	0,47	0,40	0,35	0,32	0,30	0,29	0,27	0,26	0,25	0,22	0,11
	8	8	3	8	9	7	0	7	6	7	3	1
	0,60	0,44	0,37	0,33	0,30	0,28	0,26	0,25	0,24	0,23	0,20	0,10
	2	5	4	1	3	2	7	4	4	5	3	0
	0,47	0,33	0,27	0,24	0,22	0,20	0,19	0,18	0,17	0,16	0,14	0,06
	1	5	5	2	0	3	1	2	4	7	3	7
	0,38	0,27	0,22	0,19	0,17	0,16	0,15	0,14	0,13	0,13	0,11	0,05
	9	0,00	0	2	3	0	0	2	6	0	1	0
	0,00	0	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00
	0		0	0	0	0	0	0	0	0	0	0

Распределение Пирсона χ^2

α	Число степеней свободы f											
	1	2	3	4	5	6	7	8	9	10	15	20
0,0	3,8	6,0	7,8	9,5	11,1	12,6	14,1	15,5	16,9	18,3	25,0	31,4
5	2,7	4,6	6,3	7,8	9,2	10,6	12,0	13,4	14,7	16,0	22,3	28,4
0,1	1,64	3,22	4,64	6,0	7,3	8,6	9,8	11,0	12,2	13,4	19,3	25,0
0,2												

Критерий Стьюдента $t_{кр}$

α	Число степеней свободы f											
	1	2	3	4	5	6	7	8	9	10	20	∞
0,0	12,71	4,30	3,18	2,78	2,57	2,45	2,36	2,31	2,26	2,23	2,09	1,96
5	6,31	2,92	2,35	2,13	2,01	1,94	1,89	1,86	1,83	1,81	1,72	1,65
0,1	3,08	1,89	1,64	1,53	1,48	1,44	1,41	1,40	1,38	1,37	1,33	1,28
0,2												

Приложение 3
Министерство образования и науки Российской Федерации

ФГБОУ ВПО «ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИ-
ТЕТ АРХИТЕКТУРЫ И СТРОИТЕЛЬСТВА»
КАФЕДРА «УПРАВЛЕНИЕ КАЧЕСТВОМ И ТСП»

КУРСОВАЯ РАБОТА

На тему:

«ЛИНЕЙНАЯ И НЕЛИНЕЙНАЯ ПАРНАЯ РЕГРЕССИЯ»

Автор работы: _____
Направление: 27.04.02 «Управление качеством»
Обозначение: _____ Группа УК - 11м
Руководитель: _____
Работа защищена: _____ Оценка

Пенза 20____

Пример расчета линейной и нелинейной парной регрессии

Большинство научных исследований связано с экспериментом. Он может непосредственно проводиться на объекте или на его модели.

Планирование эксперимента – это процедура выбора числа и условий проведения опытов, необходимых и достаточных для решения поставленной задачи с требуемой точностью.

При решении задачи используются математические модели объекта исследования, т.е. уравнение, связывающее параметр оптимизации с факторами. Это уравнение в общем виде выглядит следующим образом:

$$y = f(x_1, x_2, x_3, \dots, x_n)$$

Для определения связи между двумя необходимыми параметрами можно воспользоваться парной регрессией. Парная регрессия представляет собой уравнение, описывающее связь между двумя переменными: зависимой переменной y и независимой переменной x .

Рассмотрим примеры использования линейной и нелинейной парной регрессии для определения зависимостей между оцениваемыми показателями при производстве сахара-песка.

При производстве сахара используются диффузионные аппараты, которые имеют определенные недостатки, что не позволяет обеспечить оптимальные условия проведения процесса экстрагирования для получения максимального выхода сахара. Основными факторами, влияющими на ход процесса экстрагирования и полноту извлечения сахара из стружки, являются температурный режим, особенно в начальной стадии процесса, направление движения фаз, гидродинамика процесса, соотношение расхода масс экстрагента и стружки. Таким образом, целесообразно определить зависимость выхода сахара (y , т/смену) от температуры в диффузионном аппарате (x , °C), что позволит найти оптимальный температурный режим для максимального выхода сахара. Исходные данные представлены в табл. 1.

Т а б л и ц а 1П4

Результаты эксперимента

№	x	y	x^2	y^2	xy	\hat{y}	$(y - \hat{y})$	$(y - \hat{y})^2$
1	68	110	4624	12100	7480	112,6	-2,6	6,76
2	69	112	4761	12544	7728	119,8	-7,8	60,84
3	70	130	4900	16900	9100	127	3	9
4	71	135	5041	18225	9585	134,2	0,8	0,64
5	72	150	5184	22500	10800	141,4	8,6	73,96
6	73	150	5329	22500	10950	148,6	1,4	1,96
7	74	150	5476	22500	11100	155,8	-5,8	33,64
Σ	497	937	35317	127269	66743			186,8

Для установления связи между изучаемыми переменными рассчитаем коэффициент парной корреляции:

$$\tau = \frac{\Sigma xy - \left(\frac{1}{n}\right)(\Sigma x)(\Sigma y)}{\sqrt{\left[\Sigma x^2 - \left(\frac{1}{n}\right)(\Sigma x)^2\right] \left[\Sigma y^2 - \left(\frac{1}{n}\right)(\Sigma y)^2\right]}}$$

Коэффициент корреляции – это статистический показатель зависимости двух случайных величин. Коэффициент корреляции может принимать значения от -1 до +1. При этом значение -1 будет говорить об отсутствии корреляции между величинами, 0 - о нулевой корреляции, а +1 – о полной корреляции величин. Т.е., чем ближе значение коэффициента корреляции к +1, тем сильнее связь между двумя случайными величинами.

$$\tau = \frac{66743 - \frac{1}{7} \cdot 497 \cdot 937}{\sqrt{(35317 - \frac{1}{7} \cdot 497^2)(127269 - \frac{1}{7} \cdot 937^2)}} = 0,92$$

Проведем оценку значимости коэффициента парной корреляции с помощью критерия Стьюдента.

$$\hat{t} = \frac{\tau \sqrt{n-2}}{\sqrt{1-\tau^2}}$$

$$\hat{t} = \frac{0,92 \sqrt{7-2}}{\sqrt{1-0,92^2}} = 13,4$$

Сравним полученное значение с критическим значением t-критерия, который определяется по таблице распределения Стьюдента:

$$t_{кр} = 2,36, t_{кр} < t$$

Таким образом, коэффициент корреляции значим.

Найдем выборочные коэффициенты регрессии по следующей формуле:

$$b_0 = \frac{\Sigma y \Sigma x^2 - \Sigma xy \Sigma x}{n \Sigma x^2 - (\Sigma x)^2}$$

$$b_0 = \frac{937 \cdot 35317 - 66743 \cdot 497}{7 \cdot 35317 - 497^2} = -377$$

$$b_1 = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2}$$

$$b_1 = \frac{7 \cdot 66743 - 497 \cdot 937}{7 \cdot 35317 - 497^2} = 7,2$$

Уравнение регрессии имеет следующий вид (рис. 1):

$$y = -377 + 7,2x$$

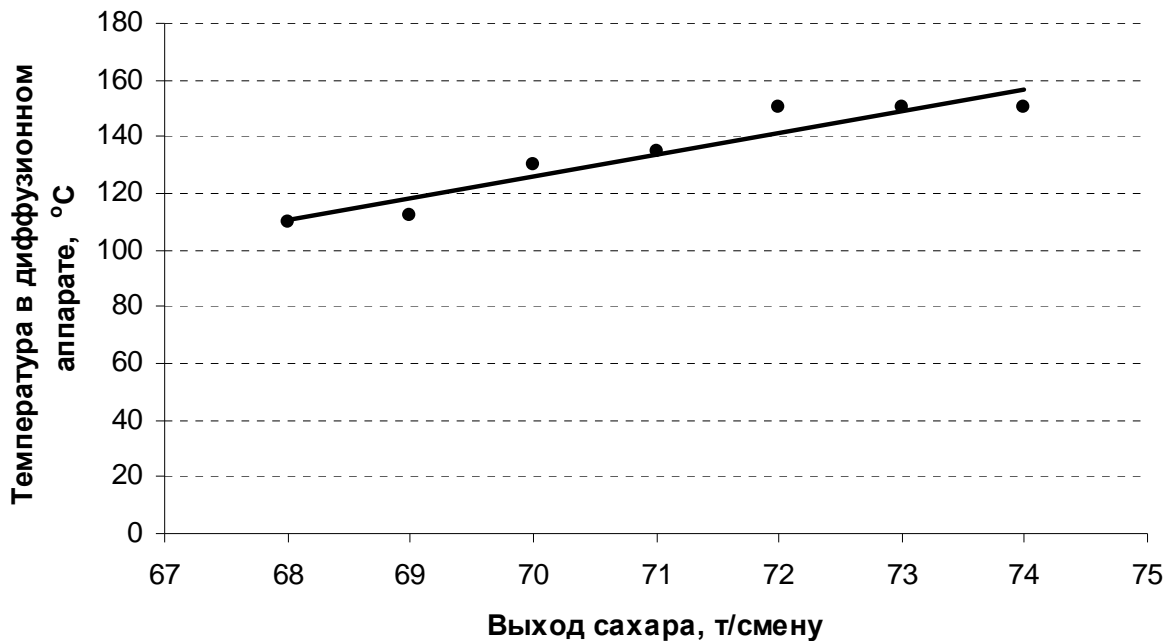


Рис.1 Зависимость выхода сахара от температуры в диффузионном аппарате

С помощью коэффициентов можно выяснить силу влияния факторов. Чем больше величина коэффициента, тем большее влияние оказывает фактор. В данном случае коэффициент имеет знак плюс. Это означает, что с увеличением значения фактора значение параметра оптимизации увеличивается.

Для практического использования моделей регрессии большое значение имеет их адекватность, т.е. соответствие фактическим статистическим данным.

Значение дисперсии адекватности модели можно вычислить по следующей формуле:

$$S_{ад}^2 = \frac{\sum (y - \hat{y})^2}{n - 2}$$

$$S_{ад}^2 = \frac{186,8}{7 - 2} = 37,36$$

Далее найдем дисперсию воспроизводимости $S_{воспр}^2$:

$$S_{воспр}^2 = \frac{\sum y^2 - \left(\frac{1}{n}\right)(\sum y)^2}{n - 1}$$

$$S_{воспр}^2 = \frac{127269 - \frac{1}{7} \cdot 937^2}{7 - 1} = 307,5$$

Для того, чтобы проверить гипотезу об адекватности модели можно воспользоваться критерием Фишера:

$$F = \frac{S_{ад}^2}{S_{воспр}^2}$$

$$F = \frac{37,36}{307,5} = 0,12$$

$$F_{таб} = 3,79$$

$$F_{таб} > F_{расч}$$

На основании полученных данных можно сделать вывод, что модель адекватна.

Для выполнения нормы – 150 т/смену – температура в диффузионном аппарате должна достигать 73°C:

$$x = \frac{y + 377}{7,2} = \frac{150 + 377}{7,2} = 73C$$

На практике для получения 150 т сахара в смену температура в аппарате должна быть в пределах 72-74°C.

В том случае, если предполагаемый характер процесса носит нелинейный характер, имеет смысл получить квадратичную зависимость.

Проведем анализ того же процесса в вакуум-аппарате. Принцип действия данного оборудования и диффузионного аппарата похожи, так как они направлены на одну цель – достичь наибольшего выхода сахара-песка.

Продолжение прил. 4

В качестве переменных примем:

x – температура в вакуум - аппарате, °C;

y – выход сахара, т/смену.

Исходные данные представлены в табл. 2.

Т а б л и ц а 2П4

Результаты эксперимента

№	x	y	xy	x^2	x^2y	x^3	x^4
1	80	120	9600	6400	768000	512000	40960000
2	81	130	10530	6561	852930	531441	43046721
3	82	140	11480	6724	941360	551368	45212176
4	83	145	12035	6889	998905	571787	47458321
5	84	150	12600	7056	105840	592704	49787136
Σ	330	685	56245	33630	3667035	2759300	226464354

Составим и решим систему уравнений:

$$an + b\Sigma x + c\Sigma x^2 = \Sigma y$$

$$a\Sigma x + b\Sigma x^2 + c\Sigma x^3 = \Sigma xy$$

$$a\Sigma x^2 + b\Sigma x^3 + c\Sigma x^4 = \Sigma x^2y$$

$$5a + 330b + 33630c = 685$$

$$330a+33630b+2759300c=56245$$

$$33630a+2759300b+226464354c=3667035$$

Уравнение регрессии будет иметь следующий вид (рис. 2):

$$y=-145-9,07x+0,06x^2$$

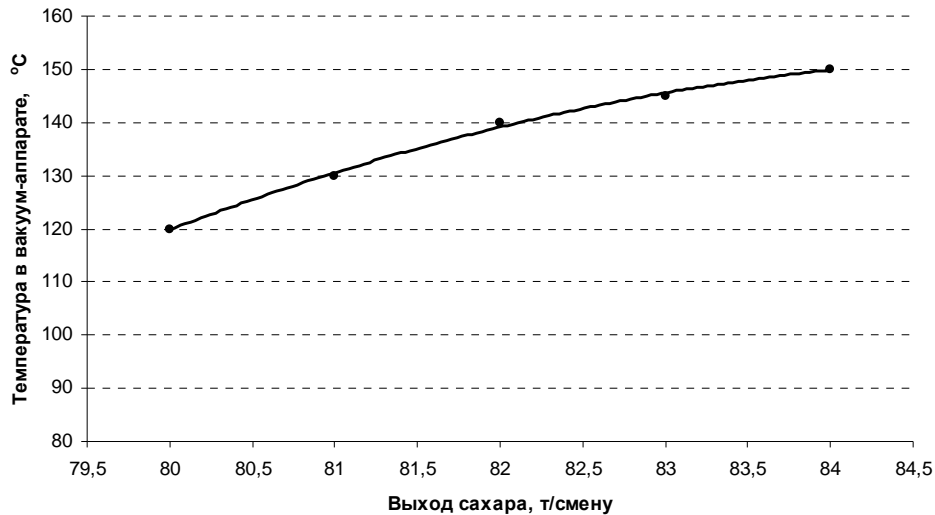


Рис. 2. Зависимость выхода сахара от температуры в вакуум-аппарате

Проверим адекватность модели

$$S_{ад}^2 = \frac{584467}{5-2} = 194822$$

$$S_{воспр}^2 = \frac{94425 - \frac{1}{5} \cdot 685^2}{5-1} = 145$$

$$F = \frac{194822}{145} = 1344$$

$$F_{таб} = 5,05$$

Так как $F_{табл} < F_{расч}$ можно сделать вывод, что модель адекватна.

Таким образом, на основании статистических данных, полученных при анализе деятельности предприятия, были получены зависимости между технологическими режимами работы оборудования и конечным выходом сахара-песка. На основании полученных данных можно сделать вывод, что на ОАО «Атмис-сахар» правильно подобран температурный режим. Выход сахара соответствует нормативным показателям.

РЕКОМЕНДУЕМАЯ ЛИТЕРАТУРА

1. Гмурман, В.Е. Теория вероятностей и математическая статистика [Текст] / В.Е. Гмурман – М.: Высшая школа. – 2003. – 480 с.
2. Данилов, А.М. Теория вероятностей и математическая статистика учебное пособие / А.М. Данилов, А.А. Данилов – Пенза: Пензенский гос. архит.-строит. ин-т, 1996. – 167 с.
3. Зейдель, А.Н. Элементарные оценки ошибок измерений [Текст] / А.Н. Зейдель – М.: Наука, 1967. – 88 с.
4. Ивченко, Г.И. Сборник задач по математической статистике учебное пособие / Г.И. Ивченко, Ю.И. Медведев, А.В. Чистяков – М.: Высшая школа, 1989. – 256 с.
5. Линник, Ю.В. Метод наименьших квадратов и основы теории обработки наблюдений [Текст] / Ю.В. Линник – М.: Физматгиз, 1962. – 114 с.
6. Налимов, В.В. Применение математической статистики при анализе вещества [Текст] / В.В. Налимов – М.: Физматгиз, 1960. – 430 с.
7. Супрун, А.Н. Вычислительная математика для инженеров-экологов методическое пособие / А.Н. Супрун, В.В. Найдено. – М.: Изд-во АСВ, 1996. – 391 с.
8. Тейлор, Дж. Введение в теорию ошибок [Текст] / Дж. Тейлор – М.: Мир, 1985. – 272 с.
9. Щиголев Б.М. Математическая обработка наблюдений [Текст] / Б.М. Щиголев. – М.: Физматгиз, 1962. – 344 с.
10. Адлер, Ю.П. Введение в планирование эксперимента [Текст] / Ю.П. Адлер – М.: Metallургия, 1969. – 157 с.
11. Львовский, Е.Н. Статистические методы построения эмпирических формул учебное пособие /Е.Н. Львовский – М.: Высшая школа, 1988. – 239 с.
12. Монтгомери, Д.К. Планирование и анализ данных [Текст] / Д.К. Монтгомери – Л.: Судостроение, 1980. – 384 с.
13. Кальгин, А.А. Лабораторный практикум по технологии бетонных и железобетонных изделий [Текст] / А.А. Кальгин, Ф.Гю Сулейманов – М.: Высш. шк., 1994. – 272 с.
14. Королев, Е.В. организация и проведение научно-исследовательской работы студентов технических специальностей [Текст] / Е.В. Королев, В.И. Логанина, В.С. Демьянова, Р.В. Тарасов – Пенза: ПГУАС, 2012. – 172 с.

ОГЛАВЛЕНИЕ

ПРЕДИСЛОВИЕ	3
ВВЕДЕНИЕ.....	4
1. ЗАДАНИЕ НА КУРСОВУЮ РАБОТУ	5
2. Состав и содержание курсовой работы.....	5
3. Последовательность выполнения курсовой работы	6
4. Консультации и защита курсовой работы	6
5. Расчетно-пояснительная записка.....	6
5.1. Введение.....	7
5.2. Выбор объекта исследования, параметра оптимизации и факторов.....	7
5.2.1. Описание объекта исследования	11
5.2.2. Выбор параметра оптимизации	12
5.2.3. Выбор факторов.....	14
5.3. Линейная парная регрессия.....	18
5.3.1. Расчет коэффициентов уравнения регрессии	22
5.3.2. Проверка адекватности модели	26
5.3.3. Проверка значимости коэффициентов.....	28
5.4. Нелинейная парная регрессия.....	31
5.4.1. Расчет коэффициентов уравнения регрессии	32
5.4.2. Проверка адекватности модели	33
5.4.3. Проверка значимости коэффициентов.....	33
5.5. Заключение	33
ПРИЛОЖЕНИЕ	34

Учебное издание

Тарасов Роман Викторович
Макарова Людмила Викторовна

ОСНОВЫ ТЕОРИИ ЭКСПЕРИМЕНТА

Учебно-методическое пособие
по выполнению курсовой работы

В авторской редакции
Верстка Т.Ю. Симутина

Подписано в печать 18.04.15. Формат 60×84/16.
Бумага офисная «Снегурочка». Печать на ризографе.
Усл.печ.л. 2,55. Уч.-изд.л. 2,75. Тираж 80 экз.
Заказ № 118.

Издательство ПГУАС.
440028, г.Пенза, ул. Германа Титова, 28