О. Л. Любимцева

Моделирование процессов

Учебное пособие

Нижний Новгород
2025

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Нижегородский государственный архитектурно-строительный университет»

О. Л. Любимцева

Моделирование процессов

Утверждено редакционно-издательским советом университета
в качестве учебного пособия

Нижний Новгород
ННГАСУ
2025

ББК 15.12
Л 93
УДК 519.876.2

Рецензенты:

А. В. Баландин – канд. физ.-мат. наук, доцент кафедры алгебры, геометрии и дискретной математики ФГАОУ ВО «Национального исследовательского Нижегородского государственного университета им. Н.И. Лобачевского»

Д. К. Отдельнова – начальник метрологической службы ООО «АВАТАР»

Любимцева, О. Л. Моделирование процессов : учебное пособие / О. Л. Любимцева ; Министерство образования и науки Российской Федерации, Нижегородский государственный архитектурно-строительный университет. – Нижний Новгород : ННГАСУ, 2025. – 75 с.: ил. – ISBN 978-5-528-00634-5. – Текст : непосредственный.

Учебное пособие «Моделирование процессов» содержит основные понятия теории моделирования, классификации моделей и моделирования, основы планирования эксперимента и основы по строения регрессионных моделей для исследования технологических процессов производства.

Учебное пособие предназначено для студентов, обучающихся по направлению подготовки 27.03.01 «Стандартизация и метрология», профиль «Стандартизация и сертификация». Пособие будет интересно студентам разных форм обучения, занимающихся описанием и моделированием различных процессов производства.

Учебное пособие разработано на кафедре Инженерной графики и информационного моделирования ННГАСУ.

ISBN 978-5-528-00634-5

СОДЕРЖАНИЕ

Математическое моделирование – третий путь познания.
Академик РАН, док.ф-м наук, А.Н.Тихонов

ВВЕДЕНИЕ

В начале прошлого столетия слова «модель», «моделирование» были известны очень узкому кругу высокопрофессиональных специалистов, связанных или с исследованием сложных физических и природных процессов и явлений, или с созданием сложных технических объектов. В наше время слова «модель» и «моделирование» не вызывают особого трепета и используются в обычной жизни. Компьютерные информационные технологии расширили возможности моделирования, и сегодня трудно представить научно-исследовательскую и серьезную проектную деятельность без использования методологии и современных средств построения и использования моделей. В последние десятилетия моделирование оформилось в самостоятельную междисциплинарную область знаний со своими объектами, закономерностями, подходами и методами исследования.

Учебное пособие предназначено для студентов направления подготовки 24.03.01 Стандартизация и метрология. Любой технологический процесс должен быть управляем, а для этого следует понимать модель, по принципу которой он функционирует. Данное учебное пособие содержит теоретический материал по основным разделам дисциплины «Моделирование процессов» и позволяет студентам получить знания и представления об основах и методологии моделирования, о построении и применении моделей в технологических процессах.

<u>Обозначения</u>

 – Замечание

 – Историческая справка

 – Пример

ОСНОВНЫЕ ПОНЯТИЯ И ОПРЕДЕЛЕНИЯ

Под моделью понимают такой материальный или мысленно представляемый объект, который в процессе познания (изучения) замещает объект–оригинал, сохраняя некоторые важные для данного исследования типичные его черты. Модель – реально существующая или мысленно представляемая система, которая, замещая и отображая оригинал с определенной целью, находится с ним в отношениях подобия (сходства). Таким образом, можно сказать, что *модель* – это объект любой природы, который при исследовании способен замещать реально существующий объект с целью получения новой информации о последнем.

Процесс – определенная совокупность действий, направленных на достижение поставленной цели.

Система – это совокупность взаимосвязанных элементов и компонентов, имеющая вполне конкретную структуру и вполне конкретное целевое назначение.

Элемент системы – часть системы, не подвергаемая дальнейшему делению.

Моделирование – замещение исследуемого объекта (оригинала) его условным образом, описанием или другим объектом (моделью) и познание свойств оригинала путем исследования свойств модели. Моделирование – это построение (или выбор из уже существующих) модели, ее изучение и использование с целью получения новых знаний об исследуемом объекте. Реальная польза от моделирования может быть получена при выполнении следующих условий:

- модель должна быть адекватной оригиналу в том смысле, что должна с достаточной точностью отображать интересующие исследователя характеристики оригинала;
- модель должна устранять проблемы, связанные с физическими измерениями каких-то сигналов или характеристик оригинала.

Моделирование базируется на нескольких основополагающих принципах :

1. Принцип информационной достаточности – при полном отсутствии информации об объекте построение его модели невозможно. Существует некоторый уровень априорной информации об объекте, только при достижении которого может быть построена адекватная модель. При наличии полной информации об объекте построение его модели не имеет смысла.
2. Принцип осуществимости – создаваемая модель должна обеспечивать достижение поставленной цели исследования с вероятностью, существенно отличающейся от нуля.
3. Принцип множественности моделей – создаваемая модель должна отражать в первую очередь те свойства реального объекта (системы), которые интересуют исследователя. Для полного исследования объекта необходимо достаточно большое количество моделей, отражающих исследуемый объект с разных сторон и с разной степенью детализации.
4. Принцип агрегатирования – в большинстве исследований систему целесообразно представить как совокупность подсистем, для описания которых оказываются пригодными стандартные схемы.
5. Принцип параметризации – модель строится в виде известной системы, параметры которой неизвестны.

Виды моделей и моделирования

1. По способу реализации модели подразделяются на материальные и идеальные.

Материальное моделирование – это моделирование, при котором исследование объекта выполняется с использованием его материального аналога, воспроизводящего основные физические, геометрические, динамические, функциональные характеристики объекта.

Идеальное моделирование отличается от материального тем, что основано не на материальной аналогии объекта и модели, а на аналогии идеальной, мыслеобразной и всегда носит теоретический характер.

Идеальное моделирование является первичным по отношению к материальному.

Мысленный образ реального объекта, сложившийся в голове исследователя, в научной литературе называется *когнитивной моделью*. Создавая такую модель, исследователь часто упрощает объект, чтобы получить более лаконичное и компактное описание. Представление когнитивной модели на естественном языке называется *содержательной моделью*. В естественно-научных дисциплинах и технике содержательную модель часто называют *технической постановкой проблемы*.

2. По функциональному признаку.

Моделирование может быть *статическим* и *динамическим*. Статическим называется моделирование, при котором среди параметров объекта и модели отсутствует время и сами параметры объекта со временем не изменяются. При динамическом моделировании объект исследования и его параметры во времени существенно изменяются.

Моделирование может быть *детерминированным* и *стохастическим*. Детерминированное моделирование отображает детерминированные процессы, то есть процессы, в которых предполагается отсутствие всяких случайных воздействий. Стохастическое моделирование отображает вероятностные процессы и события.

Моделирование может быть *дискретным* и *непрерывным*. Модель дискретная, если она описывает поведение системы только в дискретные моменты времени. Модель непрерывная, если она описывает поведение системы для всех моментов времени из некоторого промежутка времени.

Моделирование может быть *полным*, *неполным* и *приближенным*. Полным называется моделирование, при котором достигается полное подобие исследуемого объекта и модели во времени и в пространстве. Неполным называется моделирование, при котором реализуется неполное подобие исследуемого объекта и модели во времени и в пространстве. Приближенным называется моделирование, при котором некоторые проявления исследуемого объекта не моделируются совсем.

Математическая модель – совокупность математических объектов (уравнений, систем уравнений и неравенств, алгебраических выражений и т. д.), описывающих языком математических символов исследуемый объект и его отношения с окружающим миром.

Математическое моделирование – это построение математической модели (или выбор имеющейся «модели-заготовки»), ее исследование с целью получения новой информации об объекте и использование для описания свойств и предсказания поведения объекта.

Классификация математических моделей

1. Сложность объекта

Все объекты моделирования можно разделить на две группы: простые объекты и системы. При моделировании *простых объектов* не рассматривается внутренне строение объекта, не выделяются составляющие его элементы или подпроцессы.

Для *сложных систем* характерно наличие большого числа взаимосвязанных и взаимодействующих элементов. Их поведение многовариантно. При моделировании систем возникают большие трудности, так как модели должны учитывать свойства и поведение отдельных элементов, а также взаимосвязи между ними. Такие модели называются *структурными моделями*.

2. Оператор модели

Оператор модели определяется совокупностью уравнений. Если оператор обеспечивает линейную зависимость выходных факторов от входных, то математическая модель называется *линейной.* В противном случае модель называется *нелинейной*.

3. Параметры модели

В зависимости от вида используемых множеств параметров модели делятся на *качественные* и *количественные.*

4. Цели моделирования

В зависимости от цели моделирования выделяют:

– *дескриптивные* – установление законов изменения параметров модели;
– *оптимизационные* - для определения оптимальных (наилучших) с точки зрения некоторого критерия параметров объекта и технологических режимов;
– *управленческие модели* - для принятия эффективных управленческих решений.

5. Метод реализации модели

В зависимости от метода реализации выделяют *аналитические* и *алгоритмические* математические модели.

Метод является аналитическим, если он позволяет получить выходные факторы в виде аналитических выражений. Аналитические методы бывают *алгебраическими* и *приближенными*.

В алгоритмических моделях математические соотношения для объекта исследования заменяются алгоритмом. Алгоритмические модели бывают *численными* и *имитационными*.

6. Метод построения

Различают модели, созданные с помощью *аналитических* и *статистических методов*.

В основе аналитических моделей процессов лежат фундаментальные законы, выраженные в виде функциональных соотношений (алгебраических, интегрально–дифференциальных и т. д.). Поэтому аналитические модели описывают и раскрывают сущность процессов и явлений, протекающих в исследуемом объекте и определяющих его свойства и поведение. В качестве примера аналитических моделей можно назвать дифференциальные уравнения.

В основе статистических моделей лежат результаты экспериментального исследования объекта. Поэтому эти модели также называют эмпирическими, идентифицируемыми, вероятностно-статистическими, опытно–статистическими. Статистические модели рассматривают исследуемый объект как «черный ящик» и не раскрывают сущность процессов и явлений, протекающих в нем, они просто отражают одну из возможных зависимостей выходных переменных от входных, то есть носят частный характер, в отличие от аналитических моделей, которые имеют более общий характер. Примеры эмпирических моделей – корреляционные, регрессионные модели, модели дисперсионного анализа.

Алгоритм построения модели

Степень реализации принципов моделирования в каждой конкретной модели может быть различной, что в значительной степени зависит от соблюдения технологии моделирования исследователем.

Технологии комплексного моделирования представляют собой последовательность следующих действий:

1) определение цели моделирования;

2) разработка концептуальной модели;

3) формализация модели;

4) программная реализация модели;

5) планирование модельных экспериментов;

6) реализация плана эксперимента;

7) анализ и интерпретация результатов моделирования.

Алгоритм построения эмпирической модели существенно отличается. Перечислим основные пункты.

1. *Выявление противоречия и формулирование проблемы*. Данный этап моделирования является самым ответственным. От правильности формулирования проблемы исследования зависят результаты моделирования.

2. *Определение объекта исследования*. *Постановка задачи (задач) исследования.* На этом этапе проблему необходимо «раздробить»: выделить задачи, четко их сформулировать, определить стратегию и тактику решения каждой из них. Все задачи исследования должны быть хорошо структурируемыми – это позволит быстрее найти пути их решения. Очень важно также правильно и корректно поставить вопрос в каждой задаче. Определить ее приоритет и место в общем списке решаемых задач.

3. *Анализ априорной информации*. Формулирование гипотезы исследования. Анализ априорной информации базируется на изучении уже имеющихся результатов исследования подобных объектов и решения подобных задач другими исследователями и выявлении аналогов с целью повышения эффективности собственного исследования.

4. *Выбор входных и выходных факторов*. *Фактор* – измеряемая переменная величина, принимающая в каждый момент времени некоторое определенное значение из своей области определения. При моделировании технологических процессов можно сформулировать следующие требования к входным факторам:

– они должны быть взаимно независимыми;
– количественными и сравнительно легко измеряемыми;
– простыми и иметь физический смысл;
– они должны быть универсальными и полными с точки зрения описания свойств и структуры исследуемого объекта.

5. *Формализация задачи*. Современный математический аппарат требует, чтобы для решения задача была поставлена формально, то есть в виде математической формулы.

6. *Планирование и проведение эксперимента*. Основой планирования эксперимента является теория планирования факторного эксперимента. В алгоритме построения эмпирической модели этот этап предшествует построению модели. В алгоритме построения аналитической модели планирование и проведение эксперимента осуществляется после построения модели и выполняется для оценки точности аналитической модели. При планировании эксперимента определяются окончательное количество самих входных факторов и количество их уровней. *Уровень фактора* – конкретное значение фактора из его области определения при экспериментальном исследовании объекта. Совокупность уровней входных факторов объекта (по одному уровню от фактора) определяет одно состояние объекта

7. *Обработка результатов эксперимента.*

8. *Построение модели.* Результаты экспериментальных исследований обрабатываются математическим аппаратом статистического анализа (регрессионного, дисперсионного, корреляционного и т. д.)

9. *Проверка адекватности модели*. Адекватность модели характеризует ее соответствие экспериментальным данным. Проверка осуществляется по специальным критериям.

10. *Интерпретация результатов моделирования*. На этапе интерпретации оценивается, насколько результаты моделирования (в частности, модель) соответствуют здравому смыслу и существующей информации о поведении и свойствах объекта.

11. *Оценка пригодности модели.* Применяют при работе с регрессионной моделью.

12. *Решение задачи оптимизации.* Применяется при работе с моделями математического программирования.

13. *Использование модели. Документирование результатов.* После получения и проверки модели для ее дальнейшего использования необходимо оформить результаты моделирования.

КРИТЕРИИ СОГЛАСИЯ

Общая постановка задачи

В математической статистике часто выделяют особый раздел, в котором рассматривается проверка гипотез. Статистическая проверка гипотез применяется для того, чтобы использовать полученную по выборке информацию для суждения о законе распределения генеральной совокупности. Обычно статистическая гипотеза проверяется с помощью критериев согласия, которые позволяют оценить соответствие того или иного теоретического закона распределения некоторому эмпирическому ряду распределения. Критерии согласия должны дать ответ на вопрос, можно ли принять для данного эмпирического распределения модель, выражаемую некоторым теоретическим законам распределения. Одними из таких критериев являются критерии согласия Пирсона и Колмогорова – Смирнова.

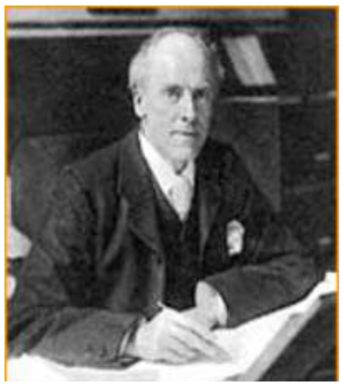
Критерий согласия Пирсона

Критерий χ^2 Пирсона – непараметрический метод, применяют для проверки гипотезы о соответствии эмпирического распределению предполагаемому теоретическому распределению F(x) при большом объеме выборки ($n \geq 100$). Основная и альтернативные гипотезы о соответствии выборочного распределения ($F_n(x)$) теоретическому закону ($F(x;Q)$, где Q – параметр гипотетического распределения) можно записать как:

$$H_0: F_n(x) = F(x;Q)$$

$$H_1: F_n(x) \neq F(x;Q).$$

***Карл Пирсон** - английский математик, статистик, биолог и философ, основатель математической статистики (опубликовано более 400 основополагающих работ). Разработал теорию корреляции, критерии согласия, алгоритмы принятия решений и оценки параметров. С его именем связаны такие широко используемые термины и методы, как:*



Карл Пирсон (1857 – 1936)

- *коэффициент корреляции Пирсона и корреляционный анализ;*
- *критерий согласия Пирсона (критерий хи-квадрат);*
- *нормальное распределение;*
- *распределение Пирсона.*

Пирсон ввёл наглядное представление распределения случайной величины с помощью гистограммы, ввёл и исследовал понятия стандартного отклонения, коэффициента асимметрии распределения. Для распределений, не соответствующих нормальному закону, Пирсон предложил «метод моментов», позволяющий найти теоретический закон, наилучшим образом соответствующий эмпирической выборке.

Критерий применим для любых видов функций $F(x)$, даже при неизвестных значениях их параметров, что обычно имеет место при анализе результатов механических испытаний. В этом заключается его универсальность.

Использование критерия χ^2 предусматривает разбиение размаха варьирования выборки на интервалы и определение числа наблюдений n_j (частоты) для каждого из интервалов, иными словами, необходимо построить интервальный статистический ряд. Для удобства оценок параметров распределения интервалы выбирают одинаковой длины (h):

$$h = \frac{X_{max} - X_{min}}{m}, \tag{1}$$

где m – число интервалов, значение зависит от объема выборки и определяется по формуле:

$$m = 1 + log_2 N \tag{2}$$

Интервалы, содержащие менее пяти наблюдений, объединяются с соседними. Однако, если число таких интервалов составляет менее двадцати процентов от их общего количества, допускаются интервалы с частотой n_j не менее двух ($n_j \geq 2$).

Статистика для проверки критерия Пирсона:

$$\chi^2 = \sum_{j=1}^{m} \frac{\left(n_j - Np_j\right)^2}{Np_j}, \tag{3}$$

где p_j – вероятность попадания случайной величины в j-ый интервал, вычисляемая в соответствии с гипотетическим законом распределения $F(x; Q)$.

Основную гипотезу проверяют путем сравнения значения выборочной статистики с критическим значением $\chi^2_{(\alpha;\nu)}$, где $\nu = m - r - 1$ (m – число интервалов, r – число параметров, оцениваемых по рассматриваемой выборке). Если $\chi^2_{\text{в}} < \chi^2_{(\alpha;\nu)}$, то основная гипотеза не отвергается и выборка соответствует заданному закону распределения.

Недостатком критерия согласия Пирсона является потеря части первоначальной информации, связанная с необходимостью группировки результатов наблюдений в интервалы и объединения отдельных интервалов с малым числом наблюдений. В связи с этим рекомендуется дополнять проверку соответствия распределений по критерию χ^2 другими критериями. Особенно это необходимо при сравнительно малом объеме выборки ($n \approx 100$).

1. Для управления качеством выплавляемой стали необходимо знание законов распределения ее механических свойств. Для исследования взято 374 наблюдения (N =374) над сталью марки 35ГС. Необходимо проверить гипотезу о том, что механическое свойство (предел текучести) стали подчиняется нормальному закону распределения при уровне значимости α = 0,05. Необходимые данные представлены в таблице.

№ интервала (i)	$x_i - x_{i+1}$	n_i
1	40-41	20
2	41-42	112
3	42-43	154
4	43-44	73
5	44-45	13
6	45-46	2

Решение

Необходимо проверить гипотезу:

$$H_0: F_n(x) = N(a; \sigma^2)$$
$$H_1: F_n(x) \neq N(a; \sigma^2)$$

Так как последний интервал содержит меньше пяти элементов, то необходимо определить какой процент от всех интервалов он составляет и принять решение для дальнейшего исследования. Всего шесть интервалов, что составляет сто процентов, тогда один интервал – приближенно семнадцать процентов. Это менее двадцати процентов от общего объема и интервал содержит два элемента, следовательно, согласно Замечанию, его можно рассматривать как отдельный элемент ряда, и он не даст существенных искажений результата.

Определим оценки основных числовых характеристик:

Выборочное среднее: $\bar{x} = \frac{1}{N}\sum_{i=1}^{k} \text{x}_{\text{ср}_i} n_i$, *где* $N = \sum_{i=1}^{k} n_i$ (4)

$$\bar{x} = \frac{40{,}5 \cdot 20 + 41{,}5 \cdot 112 + 42{,}5 \cdot 154 + 43{,}5 \cdot 73 + 44{,}5 \cdot 13 + 45{,}5 \cdot 2}{374} = 42{,}37$$

Выборочное среднее квадратическое отклонение: S = 0,94.

Вероятность попадания случайной величины в интервал (a;b) равна разности значений функции распределения на концах этого интервала. Так как рассматривается возможность принадлежности нормальному распределению, то , $p\,(a < x < b) = \Phi\left(\frac{b-\bar{x}}{S}\right) - \Phi\left(\frac{a-\bar{x}}{S}\right)$. (5)

$P\,(40 < x < 41) = \Phi\left(\frac{41-42{,}37}{0{,}94}\right) - \Phi\left(\frac{40-42{,}37}{0{,}94}\right) = \Phi(-1{,}46) - \Phi(-2{,}52) = 0{,}4941 - 0{,}4279 = 0{,}0662$

$P\,(41 < x < 42) = \Phi\left(\frac{42-42{,}37}{0{,}94}\right) - \Phi\left(\frac{41-42{,}37}{0{,}94}\right) = \Phi\,(-0{,}39) - \Phi\,(-1{,}46) = 0{,}4279 - 0{,}1517 = 0{,}2762$

$P\,(42 < x < 43) = 0{,}4003$

$P\,(43 < x < 44) = 0{,}2096$

$P\,(44 < x < 45) = 0{,}0392$

$P\,(45 < x < 46) = 0{,}0025$

Проверяем:

$\sum_{i=1}^{6} p_i = 0{,}0662 + 0{,}2762 + 0{,}4003 + 0{,}2096 + 0{,}0392 + 0{,}0025 = 0{,}994 \approx 1$

Для дальнейших расчетов построим таблицу:

$x_i - x_{i+1}$	n_i	$N \cdot p_i$	$n_i - N \cdot p_i$	$(n_i - N \cdot p_i)^2$	$\frac{(n_i - N \cdot p_i)^2}{N \cdot p_i}$
40-41	20	24,76	-4,76	22,66	0,92
41-42	112	103,30	8,70	75,69	0,73
42-43	154	149,71	4,29	18,40	0,12
43-44	73	78,39	-5,39	29,05	0,37
44-45	13	14,66	1,66	2,76	0,19
45-46	2	0,94	1,09	1,12	1,20

По формуле (3) определяем выборочное значение статистики Пирсона:

$\chi^2_{\text{в}} = 3{,}53.$

$\chi^2_{\text{кр}(0{,}05;6-2-1)} = 7{,}815.$

Так как выборочное значение статистики меньше критической точки, то основная гипотеза не отвергается, следовательно, механическое свойство (предел текучести) стали подчиняется нормальному закону распределения при уровне значимости α = 0,05.

Критерий Колмогорова – Смирнова

Критерий Колмогорова-Смирнова – непараметрический критерий согласия, в классическом понимании предназначен для проверки простых гипотез о принадлежности анализируемой выборки некоторому известному закону распределения. Критерий позволяет найти точку, в которой сумма накопленных расхождений между двумя распределениями является наибольшей, и оценить достоверность этого расхождения. Наиболее известно применение данного критерия для проверки исследуемых совокупностей на нормальность распределения. В качестве меры расхождения между теоретическим и эмпирическим распределениями рассматривают максимальное значение абсолютной величины разности эмпирической функции распределения $F_n(x)$ и соответствующей теоретической функцией распределения $F(x)$

$$D_{\text{в}} = \sup_{-\infty \le x \le +\infty} |F_n(x) - F(x)|, \tag{6}$$

называемой статистикой критерия Колмогорова – Смирнова.

Предельное распределение этой статистики для случая проверки простой гипотезы было получено А.Н. Колмогоровым. При $n \longrightarrow \infty$ функция распределения статистики $D\sqrt{n}$ (λ) сходится равномерно к функции распределения Колмогорова $K(\lambda) = \sum_{k=-\infty}^{+\infty} (-1)^k e^{-2\lambda^2 k^2}$.

Приведем таблицу соответствия уровня значимости и значения критической точки по распределению Колмогорова.

α	0,005	0,01	0,025	0,05	0,1	0,15	0,2	0,25
$\lambda_{1-\alpha}$	1,73	1,63	1,48	1,36	1,22	1,14	1,07	1,02

План процедуры проверки гипотезы:

1. Строят эмпирическую ($F_n(x)$) и предполагаемую теоретическую $F(x)$ функции распределения.

2. Определяется мера расхождения (D) между теоретической и эмпирическими функциями по формуле (6) и вычисляется величина $\lambda = D\sqrt{n}$.

3. Если $\lambda > \lambda_{\text{кр}}$, то основная гипотеза отвергается.

О критериях Колмогорова и Смирнова

В литературе, особенно переводной, иногда используют термин «критерий Колмогорова – Смирнова» по отношению к процедурам проверки непараметрических статистических гипотез. Однако анализ публикаций академика А.Н. Колмогорова (1903 – 1987) и члена- корреспондента АН СССР Н.В. Смирнова (1900 – 1966) свидетельствует о том, что такого критерия не существует. У А.Н. Колмогорова и Н.В. Смирнова нет совместных работ, они никогда не изучали одновременно один и тот же статистический критерий. Право на существование имеет лишь термин «критерий типа Колмогорова – Смирнова».

В 1933 г. была опубликована на итальянском языке знаменитая статья А.Н. Колмогорова «Об эмпирическом определении закона распределения», где доказывалась теорема о предельном распределение статистики критерия Колмогорова для проверки согласия эмпирического распределения с заданным непрерывным теоретическим распределением. Больше к этой тематике он не возвращался.

Колмогоров
Андрей Николаевич

Смирнов
Николай Васильевич

В те же 1930-е годы Н.В. Смирнов изучал различные непараметрические статистики. В статье 1939 г. «Об уклонениях эмпирической кривой распределения» он рассматривал число пересечений эмпирической функции распределения $F_n(x)$с кривой $F(x) + \frac{\lambda}{\sqrt{n}}$ и число выходов $F_n(x)$ за пределы полосы $F(x) \pm\frac{\lambda}{\sqrt{n}}$ при произвольном фиксированном $\lambda > 0$. Из предельной теоремы для числа пересечений эмпирической функции распределения с указанной выше кривой было получено утверждение об асимптотическом поведении статистики (в настоящее время ее называют «односторонней одновыборочной статистикой Смирнова»).

Частным случаем теоремы о предельном поведении числа выходов эмпирической функции распределения за пределы полосы является указанная выше теорема А.Н. Колмогорова. Подчеркнем, что методы доказательства в работах А.Н. Колмогорова и Н.В. Смирнова принципиально различны.

Все рассматриваемые критерии «типа Колмогорова-Смирнова» обычно называют критерии, построенные на основе эмпирических функций *распределения (для одной или нескольких выборок), в которых используются операция взятия супремума.*

Преимущество критерия Колмогорова–Смирнова заключается в том, что для его реализации не требуется группирование данных (с неизбежной потерей информации), а дает возможность рассматривать индивидуальные наблюдаемые значения. Этот критерий можно успешно применять для малых выборок. Считается, что его мощность выше, чем у критерия Пирсона.

ЭМПИРИЧЕСКАЯ МОДЕЛЬ ПРОЦЕССА В РАМКАХ ДИСПЕРСИОННОГО АНАЛИЗА

Подавляющее большинство изучаемых объектов относится к классу сложных систем, характеризующихся значительным числом взаимосвязанных параметров. Задача исследования таких систем заключается в установлении зависимости между входными параметрами–факторами и выходными параметрами–показателями качества функционирования системы. Для решения любой задачи необходимо иметь математическую модель исследуемого объекта. В свою очередь, модель объекта получают, используя результаты опытов. Эксперименты проводятся во всех областях экономики. Цель этих экспериментов – либо установить новые факты об исследуемом объекте, либо сравнить влияния различных условий на рассматриваемый процесс. Для того чтобы провести эксперимент наиболее эффективно, необходим научный подход. Под статистическим планированием эксперимента понимают такую организацию экспериментального исследования, которая позволит собрать необходимые данные, применить для их анализа статистические методы и сделать правильные и объективные выводы.

Возникновение современных статистических методов планирования эксперимента связано с первой работой Рональда А.Фишера. Фишер разработал и впервые применил дисперсионный анализ в качестве важнейшего метода статистического анализа в планировании эксперимента. В 1935 г. вышла монография Р.Фишера по планированию эксперимента “The Design of Experiments”, давшая название всему направлению. Этой работой были заложены основы методов планирования эксперимента, которые успешно развиваясь объединялись под общим названием дисперсионного анализа.

Другим известным автором работ по планированию экспериментов был Фрэнк Йейтс (1902 – 1994). Фрэнк Йейтс – британский статистик, продолжатель работ Карла Пирсона и Рональда Фишера по созданию и развитию методов

Фрэнк Йейтс (слева), Роналд Фишер (в центре) и Уильям Кокрэн. 1952 г.

математической статистики. Труды в области планирования эксперимента, дисперсионного анализа, исследования операций и др.

Планирование эксперимента в дисперсионном анализе состоит в выборе в соответствии с условиями проведения наблюдений такого способа группирования наблюдений, который позволил бы найти оценки и проверить гипотезы относительно исследуемого процесса.

Дисперсионный анализ – это статистический метод анализа результатов наблюдений, зависящих от различных, одновременно действующих факторов, выбор наиболее важных факторов и оценка их влияния.

Факторами обычно называют внешние условия, влияющие на эксперимент. Например, температура и атмосферное давление, тип оборудования и т.п. Нас будут интересовать те факторы, действие которых значительно и поддается проверке. В условиях эксперимента каждый фактор может принимать одно или несколько значений, которые будем называть уровнями, в результате чего можно исследовать влияние контролируемого фактора на эксперимент. В зависимости от количества контролируемых факторов, включенных в анализ, различают однофакторный анализ (по одному признаку), двухфакторный анализ (по двум признакам) и многофакторный анализ.

Для проведения дисперсионного анализа необходимо соблюдение следующих условий: результаты наблюдений должны быть независимыми случайными величинами, имеющими нормальное распределение и одинаковую дисперсию.

Постановка задачи

Пусть на автоматической линии несколько станков параллельно выполняют некоторую операцию. Для правильного планирования последующей обработки важно знать, насколько однотипны средние размеры деталей. В данной ситуации только один фактор влияет на размер детали – станки, на которых они изготавливаются. Исследователя интересует, насколько существенно влияние фактора. Предположим, что совокупности размеров деталей, изготавливаемых на каждом станке, имеют нормальное распределение и равные дисперсии.

Формализуем задачу. Необходимо сравнить a различных уровней одного фактора (a станков). Разные уровни одного фактора в рамках дисперсионного анализа еще называются *обработками*. Наблюдаемый отклик (j наблюдений (экспериментов), $j = \overline{1, n}$) на каждую из a обработок есть случайная величина. Данные эксперимента принято представлять в виде таблицы (матрицы наблюдений).

Фактор А (обработка/уровень)	**Наблюдение**			
	1	2	…	n
1	y_{11}	y_{12}	…	y_{1n}
2	y_{21}	y_{22}	…	y_{21}
. . .	. . .	. . .		. . .
a	y_{a1}	y_{a2}	…	y_{an}

y_{ij} ($i = \overline{1, a}$; $j = \overline{1, n}$) – размер детали, изготовленной на станке i и проверенный во время эксперимента под номером j, или y_{ij} – есть j*-ое* наблюдение при *i-ой* обработке.

Модели дисперсионного анализа

Каждое наблюдение y_{ij} можно представить в виде линейной статистической модели

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij}, \qquad i = \overline{1; a};\ j = \overline{1; n} \tag{7}$$

где μ – параметр, общий для всех обработок, представляющий собой общее среднее (истинное значение);

τ_i – параметр, характеризующий *i-ую* обработку, называемый эффектом *i-ой* обработки;

ε_{ij} – вариация результатов внутри отдельного уровня (характеризует влияние всех неучтенных моделью факторов), $\varepsilon_{ij} \sim N(0, \sigma^2)$.

Модель (7) соответствует однофакторному дисперсионному анализу (ОДА), так как исследуется только один фактор. Предположим, что измерения проводятся в случайном порядке, чтобы обеспечить как можно большую однородность внешних условий, при которых применяются обработки. Такой план эксперимента называется *полностью рандомизированным.*

Рандомизация означает, что распределение экспериментального материала и порядок, в котором должны проводиться отдельные опыты или прогоны эксперимента, устанавливаются случайным образом.

Статистическая модель (7) описывает две различные ситуации относительно эффектов обработок. С одной стороны, эти *a* обработок могут быть заданы экспериментом, с другой – задаются случайным образом из большой совокупности обработок. В первом случае модель будет носить название *модель постоянных эффектов*, во втором – *модель случайных эффектов*.

Модель постоянных эффектов

В данном случае эффект обработки определяется как отклонение среднего по обработке от общего среднего, например, $\tau_1 = \overline{y_1} - \mu$; $\tau_2 = \overline{y_2} - \mu$ ... Тогда, если фактор не оказывает влияние, то эффект по каждой обработке должен быть равен нулю. Именно это предположение и будет основной гипотезой. В качестве альтернативной рассмотрим предположение, что найдется хотя бы одна обработка, эффект которой будет отличен от нуля.

Таким образом: $H_0: \tau_1 = \tau_2 = \ ... = \tau_a = 0$

$H_1: \tau_i \neq 0$

Отличительной особенностью модели является то, что вывод, сформулированный в результате анализа, будет применим только к тем уровням, что рассматривались в эксперименте. Вывод нельзя распространять на похожие обработки, не исследованные в эксперименте. Иными словами, вывод носит локальный характер и может понадобиться оценка эффектов обработки.

Модель случайных эффектов

Как сказано выше, экспериментаторам приходится исследовать факторы с большим числом возможных уровней. Если при этом из совокупности уровней фактора выбираются случайным образом a уровней, то говорят, что такой *фактор случаен*. Сам вид модели не изменится, однако суть компонент будет иная, а именно, τ_i как и ε_{ij} станет случайной переменной. Если дисперсия τ_i равна σ_τ^2 и случайная величина $\tau_i \sim N(0, \sigma_\tau^2)$, при этом τ_i и ε_{ij} независимы, то проверяем следующую гипотезу:

$$H_0: \sigma_\tau^2 = 0$$

$$H_1: \sigma_\tau^2 > 0$$

Поясним, если σ_τ^2 равна нулю, то эффекты всех обработок одинаковы, в противном случае между ними существуют различия.

Так как уровни фактора выбраны случайным образом, то результаты, полученные при исследовании, должны относиться ко всей совокупности уровней фактора. Таким образом, вывод является глобальным.

Однофакторный дисперсионный анализ для модели постоянных эффектов

Введем следующие обозначения:

$y_{i.}$ – сумма наблюдений i-ой обработки;

$\overline{y_{i.}}$ – среднее наблюдений для i-ой обработки;

$y_{..}$ – общая сумма всех наблюдений;

$\overline{y_{..}}$ - общее среднее всех наблюдений.

Точка вместо нижнего индекса обозначает суммирование по индексу, который она заменяет

$$y_{i.} = \sum_{j=1}^{n} y_{ij}; \qquad \overline{y_{i.}} = \frac{y_{i.}}{n} \tag{8}$$

$$y_{..} = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij}; \qquad \overline{y_{..}} = \frac{y_{..}}{N}, \ N=an \tag{9}$$

Так как рассматривается модель постоянных эффектов, то

$$H_0: \tau_1 = \tau_2 = \ldots = \tau_a = 0$$

$$H_1: \tau_i \neq 0$$

Таким образом, если основная гипотеза не отвергается, то каждое наблюдение складывается из общего среднего (μ) и реализации случайной ошибки (ε). Соответствующая процедура проверки и есть дисперсионный анализ. Общая скорректированная сумма (отклонение от общего среднего) может быть записана в виде:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{..}}\right)^2=\sum_{i=1}^{a}\sum_{j=1}^{n}\left[\left(\overline{y_{i.}}-\overline{y_{..}}\right)+\left(y_{ij}-\overline{y_{i.}}\right)\right]^2=$$

$$=\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\overline{y_{i.}}-\overline{y_{..}}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{i.}}\right)^2+2\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\overline{y_{i.}}-\overline{y_{..}}\right)\left(y_{ij}-\overline{y_{i.}}\right)=$$

$$=n\sum_{i=1}^{a}\left(\overline{y_{i.}}-\overline{y_{..}}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{i.}}\right)^2+2\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\overline{y_{i.}}-\overline{y_{..}}\right)\left(y_{ij}-\overline{y_{i.}}\right).$$

Рассмотрим последнее слагаемое

$$S=\sum_{i=1}^{a}\sum_{j=1}^{n}\left(\overline{y_{i.}}-\overline{y_{..}}\right)\left(y_{ij}-\overline{y_{i.}}\right)=\sum_{i=1}^{a}\left(\overline{y_{i.}}-\overline{y_{..}}\right)\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{i.}}\right),$$

но $\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{i.}}\right)=\sum_{j=1}^{n}y_{ij}-n\overline{y_{i.}}=y_{i.}-n\frac{y_{i.}}{n}=0$. Следовательно , S = 0.

Таким образом, общую изменчивость данных, характеризуемую общей скорректированной суммой квадратов, можно разбить на сумму квадратов отклонений средних по обработкам от общего среднего и сумму квадратов отклонений при каждом наблюдении от своего среднего:

$$\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{..}}\right)^2=n\sum_{i=1}^{a}\left(\overline{y_{i.}}-\overline{y_{..}}\right)^2+\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij}-\overline{y_{i.}}\right)^2 \tag{10}$$

При этом отклонение средних по обработкам от общего среднего характеризуют различия между обработками, в то время как отклонения наблюдений внутри обработки от своего среднего могут быть обусловлены только случайной ошибкой. Запись (10) чаще всего представляют в виде:

$$SS_{\text{общ}}=SS_{\text{обр}}+SS_{\text{ош}}\ , \tag{11}$$

где $SS_{\text{общ}}$ – общая сумма квадратов,

$SS_{\text{обр}}$ – сумма квадратов, обусловленной обработками , то есть различия между обработками,

$SS_{\text{ош}}$ – сумма квадратов, обусловленной ошибкой, то есть различиями внутри обработок.

Совместив выражения (10) и (11), получим упрощенные формулы для расчетов:

$$SS_{\text{общ}}=\sum_{i=1}^{a}{y_{ij}}^2-\frac{1}{N}{y_{..}}^2 \tag{12}$$

$$SS_{\text{обр}} = \frac{1}{n}\sum_{i=1}^{n} y_{i.}^{\ 2} - \frac{1}{N} y_{..}^{\ 2} \quad (13)$$

$$SS_{\text{ош}} = SS_{\text{общ}} - SS_{\text{обр}} \quad (14)$$

Величину $SS_{\text{обр}}$ также называют *рассеиванием по факторам*, то есть рассеиванием за счет исследуемого фактора.
$SS_{\text{ош}}$ – *остаточное рассеивание*, то есть рассеивание за счет неучтенных факторов.

Зная суммы квадратов, можно оценить соответствующие дисперсии (средние суммы квадратов) – межгрупповую и внутригрупповую:

$$MS_{\text{обр}} = \frac{SS_{\text{обр}}}{a-1}\ ;\ MS_{\text{ош}} = \frac{SS_{\text{ош}}}{a(n-1)} \quad (15)$$

Если влияние всех уровней фактора одинаково, то $MS_{\text{обр}}$ и $MS_{\text{ош}}$ – оценки общей дисперсии. Следовательно, для проверки влияния фактора достаточно проверить гипотезу о равенстве дисперсий:

$$H_0\colon MS_{\text{обр}} = MS_{\text{ош}}$$

$$H_1\colon MS_{\text{обр}} > MS_{\text{ош}}$$

Для проверки гипотезы используют статистику Фишера $F_{\text{в}} = \frac{MS_{\text{обр}}}{MS_{\text{ош}}}$.

$F_{\text{кр}(\alpha;a-1;a(n-1))}$ – критическая точка при уровне значимости α и степенями свободы: *(a-1)* и *a(n-1)*.

Если $F_{\text{в}} > F_{\text{кр}(\alpha;a-1;a(n-1))}$, то основная гипотеза модели постоянных эффектов отвергается, то есть фактор оказывает влияние на исследуемый процесс.

Если $F_{\text{в}} < F_{\text{кр}(\alpha;a-1;a(n-1))}$, то основная гипотеза модели постоянных эффектов не отвергается, то есть фактор не оказывает влияние на исследуемый процесс.

При обработке данных, как правило, результаты вносят в так называемую таблицу Дисперсионного анализа.

Источник изменчивости	Сумма квадратов (SS)	Степень свободы (df)	Средняя сумма квадратов (MS)	$\boldsymbol{F}_{\text{в}}$
Между обработками	$SS_{\text{обр}}$	a-1	$MS_{\text{обр}}$	$\frac{MS_{\text{обр}}}{MS_{\text{ош}}}$
Внутри обработки (ошибка)	$SS_{\text{ош}}$	$a(n$-$1)$	$MS_{\text{ош}}$	
Сумма	$SS_{\text{общ}}$	N-1		

2. Для изготовителя представляет интерес предел прочности на растяжение синтетического волокна, идущего на ткань, предназначенную для мужских рубашек. По мнению изготовителя, предел прочности зависит от процентного содержания хлопка в волокне. Рассмотрены пять уровней содержания хлопка: 15%, 20%, 25%, 30%, 35%. Для каждого из уровней были проведены по пять наблюдений. Результаты эксперимента представлены в таблице.

Процентное содержание хлопка в волокне	**Наблюдения**				
	1	**2**	**3**	**4**	**5**
15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

Решение

В формулировке задачи указаны уровни фактора, следовательно, имеет место модель постостоянных эффектов. Тогда необходимо проверить гипотезу:

$$H_0: \tau_1 = \tau_2 = \tau_3 = \tau_4 = \tau_5 = 0$$

$$H_1: \tau_i \neq 0.$$

Определим основные параметры модели: a = 5; n = 5; N = an = 25.

$y_{1.} = 7 + 7 + 15 + 11 + 9 = 49$

$y_{2.} = 12 + 17 + 12 + 18 + 18 = 77$

Процентное содержание хлопка в волокне	**Наблюдения**					$\boldsymbol{y_{i.}}$
	1	**2**	**3**	**4**	**5**	
15	7	7	15	11	9	49
20	12	17	12	18	18	77
25	14	18	18	19	19	88
30	19	25	22	19	23	108
35	7	10	11	15	11	54

$y_{..} = 7 + 7 + 15 + \cdots + 11 + 15 + 11 = 376$

или $y_{..} = 49 + 77 + 88 + 108 + 54 = 376$

Определим скорректированные суммы квадратов:

$$SS_{\text{общ}} = (7^2 + 7^2 + 15^2 + 11^2 + 9^2 + 12^2 + \ldots + 15^2 + 11^2) - \frac{1}{25}376^2 = 636{,}96$$

$SS_{обр} = \frac{1}{5}(49^2 + 77^2 + 88^2 + 108^2 + 54^2) - \frac{1}{25}376^2 = 475{,}76$

$SS_{ош} = 636{,}96 - 475{,}76 = 161{,}20$

Соберем все данные в таблице Дисперсионного анализа

Источник изменчивости	$\boldsymbol{SS}$	**df**	**MS**	$\boldsymbol{F}_{в}$
Содержание хлопка в волокне	475,76	4	$\frac{475{,}76}{4} = 118{,}94$	$\frac{118{,}94}{8{,}06} = 14{,}76$
Ошибка	161,20	20	$\frac{161{,}20}{20} = 8{,}06$	
Сумма	636,96	24		

$F_{кр(0{,}01;4;20)} = 4{,}43$ (Приложение Б)

Сравним выборочное значение статистики с критической точкой при уровне значимости в 0,01.

14,76 > 4,43.

Так как выборочное значение статистики больше критической точки, то основная гипотеза отвергается, следовательно, содержание хлопка в волокде действительно оказывает влияние на его прочность.

Оценивание параметров модели

В зависимости от полученного ответа на вопрос, сформулированный гипотезой, экспериментатор либо продолжит исследования, либо начнет поиск нового фактора, возможно, оказывающего влияние. Эти исследования необходимы для управления процессом производства, качества продукции. Итак, предположим, что гипотеза отвергается и фактор оказывает влияние на исследуемый процесс. Для проведения дальнейших исследований необходимо найти оценки параметров модели. Суть проблемы состоит в минимизации случайной ошибки, то есть минимизации влияния факторов, неучтенных при проведении эксперимента. Для решения поставленной задачи воспользуемся методом наименьших квадратов (МНК). Рассмотрим сумму квадратов ошибок, и с учетом (7) имеем:

$$L = \sum_{i=1}^{a} \sum_{j=1}^{n} \varepsilon_{ij}^2 = \sum_{i=1}^{a} \sum_{j=1}^{n} \left(y_{ij} - \mu - \tau_i\right)^2 \to min \qquad (16)$$

Таким образом, необходимо определить такие оценки параметров $\mu(\hat{\mu})$ и $\tau_i(\widehat{\tau_i})$, чтобы сумма (16) была минимальной. Для этого определим частные производные по этим параметрам и решим систему из (a+1) уравнения:

$$\begin{cases} \left.\dfrac{\partial L}{\partial \mu}\right|_{\hat{\mu},\hat{\tau_i}} = 0 \\ \left.\dfrac{\partial L}{\partial \tau_i}\right|_{\hat{\mu},\hat{\tau_i}} = 0 \end{cases} \quad \Rightarrow \quad \begin{cases} -2\sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \hat{\mu} - \hat{\tau_i}\right) = 0 \\ -2\sum_{j=1}^{n}\left(y_{ij} - \hat{\mu} - \hat{\tau_i}\right) = 0 \end{cases}$$

Распишем систему после упрощения

$$\begin{cases} N\hat{\mu} + n\widehat{\tau_1} + n\widehat{\tau_2} + \cdots + n\widehat{\tau_a} = y_{..} \\ n\hat{\mu} + n\widehat{\tau_1} \qquad\qquad\qquad\qquad = y_{1.} \\ n\hat{\mu} + \qquad n\widehat{\tau_2} \qquad\qquad\quad = y_{2.} \\ -\,-\,-\,-\,-\,-\,-\,-\,-\,-\,-\,-\,-\,- \\ n\hat{\mu} + \qquad\qquad\qquad n\widehat{\tau_a} = y_{a.} \end{cases} \qquad (17)$$

Заметим, что сумма последних a уравнений дает первое, следовательно, уравнения не являются независимыми и решение системы не единственно. Так как эффекты обработок были определны как отклонение от математического ожидания общего среднего, то логично наложить ограничесние вида:

$$\sum_{i=1}^{a}\hat{\tau_i} = 0.$$

При таком ограничении система уравнений (17) примет вид:

$$\begin{cases} N\hat{\mu} = y_{..} \\ n(\hat{\mu} + \widehat{\tau_1}) = y_{1.} \\ n(\hat{\mu} + \widehat{\tau_2}) = y_{2.} \\ -\,-\,-\,-\,-\,-\,-\,- \\ n(\hat{\mu} + \widehat{\tau_a}) = y_{a.} \end{cases} \quad \Rightarrow \quad \begin{cases} \hat{\mu} = \overline{y_{..}} \\ \widehat{\tau_1} = \overline{y_{1.}} - \overline{y_{..}} \\ \widehat{\tau_2} = \overline{y_{2.}} - \overline{y_{..}} \\ -\,-\,-\,-\,-\,-\,-\,- \\ \widehat{\tau_a} = \overline{y_{a.}} - \overline{y_{..}} \end{cases} \qquad (18)$$

Данное решение не является единственно возможным. Если условия изменятся, то будут получены иные решения.

2.1 Оценить параметры модели, представленной в предыдущем примере.

Решение

$$\hat{\mu} = \overline{y}_{..} = \frac{y_{..}}{N} = \frac{365}{25} = 15{,}04$$

$$\widehat{\tau_1} = \overline{y_{1.}} - \overline{y}_{..} = \frac{y_{1.}}{n} - \overline{y}_{..} = \frac{49}{5} - 15{,}04 = 9{,}8 - 15{,}04 = -\,5{,}24$$

$$\widehat{\tau_2} = \frac{77}{5} - 15{,}04 = 15{,}4 - 15{,}04 = 0{,}36$$

$\widehat{\tau_3} = \frac{88}{5} - 15{,}04 = 17{,}6 - 15{,}04 = 2{,}56$

$\widehat{\tau_4} = \frac{108}{5} - 15{,}04 = 21{,}6 - 15{,}04 = 6{,}56$

$\widehat{\tau_5} = \frac{54}{5} - 15{,}04 = 10{,}8 - 15{,}04 = -4{,}24$

Определим минимальное значение эффекта обработок по абсолютноной величине. Именно эта обработка будет наиболее предпочтительна в плане влияния посторонних факторов на процесс производства. $min|\widehat{\tau_i}| = 0.36$, следовательно, наилучшая прочность волокна будет получена при содержании в нем 20% хлопка.

Сравнение отдельных средних по обработкам

Предположим, фактор оказывает влияние на исследуемый процесс, то есть установлено, что между обработками есть различия. Однако нет ответа на вопрос, какие именно обработки различаются, следовательно, будут полезны дальнейшие исследования. Напомним, что среднее по i-ой обработке определяется как $\mu_i = \mu + \tau_i$, при этом μ_i оценивается величиной $\overline{y_{i.}}$

Метод контрастов

Особенностью метода контрастов является то, что до начала эксперимента необходимо сделать предположения (выдвинуть гипотезы) относительно сравнения уровней фактора и их комбинаций. Гипотез должно быть не более *(a-1)* и альтернативная гипотеза должна представлять обобщение всех обработок. Каждая из гипотез может представлять собой линейную комбинацию средних значений по обработкам. Линейная комбинация называется *контрастом* (C_i) и должна удовлетворять условию: *сумма коэффициентов контрастов равна нулю.* Например, $c_1\mu_1 + c_2\mu_3 = c_3\mu_4$, при этом, $\sum_{i=1}^{3} c_i = c_1 + c_2 + c_3 = 0$. Коэффициенты контрастов (c_i) должны подбираться до начала эксперимента. Для проверки статистических гипотез используется скорректированная сумма квадратов контрастов

$$SS_c = \frac{\left(\sum_{i=1}^{a} c_i y_i\right)^2}{n\sum_{i=1}^{a} {c_i}^2} \qquad (19)$$

Сумма обладает одной степенью свободы. Сумму квадратов контрастов сравнивают со средним квадратом ошибки. Статистика, которая при этом получается, подчиняется F – распределению с одной и *(N – a)* степенями свободы.

2.2 Провести сравнение обработок с помощью метода контрастов.

Решение

Сформулируем гипотезы для сравнения средних в условиях примера этой главы.

$H_0: \mu_1 = \mu_2$

$H_0: \mu_4 = \mu_5$

$H_0: \mu_1 + \mu_2 = 2\mu_3$

$H_1: 2\mu_1 + \mu_2 + \mu_3 = 2\mu_4 + 2\mu_5$

Определим контрасты в соответствии с гипотезами:

$C_1: 1\overline{y_{1.}} - 1\overline{y_{2.}} = 0 \mid 1 + (-1) = 0$

$C_2: 1\overline{y_{4.}} - 1\overline{y_{5.}} = 0 \mid 1 + (-1) = 0$

$C_3: 1\overline{y_{1.}} + 1\overline{y_{2.}} - 2\overline{y_{3.}} = 0 \mid 1 + 1 + (-2) = 0$

$C_4: 2\,\overline{y_{1.}} + 1\,\overline{y_{2.}} + 1\overline{y_{3.}} - 2\overline{y_{4.}} - 2\overline{y_{5.}} = 0 \mid 2 + 1 + 1 + (-2) + (-2) = 0$

Рассчитаем значение контрастов:

$C_1 = 9{,}8 - 15{,}4 = -5{,}6$

$C_2 = 21{,}6 - 10{,}8 = 10{,}8$

$C_3 = 9{,}8 + 15{,}4 - 2 \cdot 17{,}6 = -10$

$C_4 = 2 \cdot 9{,}8 + 15{,}4 + 17{,}6 - 2 \cdot 21{,}6 - 2 \cdot 10{,}8 = -12{,}2$

Определяем скорректированные суммы контрастов:

$$SS_{C_1} = \frac{(-5{,}6)^2}{5 \cdot (1^2 + (-1)^2)} = 3{,}136$$

$$SS_{C_2} = \frac{10{,}8^2}{5 \cdot (1^2 + (-1)^2)} = 11{,}664$$

$$SS_{C_3} = \frac{(-10)^2}{5 \cdot (1^2 + 1^2 + (-2)^2)} = 3{,}333$$

$$SS_{C_4} = \frac{(-12{,}2)^2}{5 \cdot (2^2 + 1^2 + 1^2 + (-2)^2 + (-2)^2)} = 2{,}126$$

Составим таблицу дисперсионного анализа:

Источник изменчивости	$\boldsymbol{SS}$	**df**	**MS**	$\boldsymbol{F_{\text{в}}}$
$C_1: \mu_1 = \mu_2$	3,14	1	3,14	0,39
$C_2: \mu_4 = \mu_5$	11,66	1	11,66	1,45
$C_3: \mu_1 + \mu_2 = 2\mu_3$	3,33	1	3,33	0,41
$C_4: 2\mu_1 + \mu_2 + \mu_3 = 2\mu_4 + 2\mu_5$	2,13	1	2,13	0,26
Ошибка	161,20	20	8,06	
Сумма	636,96	24		

$F_{\text{кр}(0{,}01;1;20)} = 8{,}1$

Все выборочные значения статистики меньше значения критической точки, следовательно, гипотезы не отвергаются.

Вывод:

- в среднем прочность сиснтетических волокон, содержащих 15% хлопка и 20%, одинакова (для первой гипотезы);
- в среднем прочность сиснтетических волокон, содержащих 30% хлопка и 35%, одинакова (для второй гипотезы);
- если скрутить нить из волокон, содержащих 15% хлопка и 20%, то по прочности она будет сопоставима с нитью, скрученной из удвоенного волокна, содержащего 25% хлопка.

Рассмотрев метод контрастов, может сформироваться мнение, что вполне достаточно сравнить поправно все обработки. Данное мнение ошибочно, так как в результате проверки будет накапливаться ошибка I рода. Существует целый ряд критериев, лишенных данного недостатка.

Множественный критерий размахов Дункана

Критерий позволяет выполнять парные сравнения обработок с использованием шагового порядка сравнения. При этом устанавливается защитный уровень доли ошибок для набора проверок, а не для доли ошибок отдельных проверок. Для его применения средние по обработкам распределяются в поряке убывания и для каждого среднего определяется стандарт ошибки

$$S_{\bar{y}_i} = \sqrt{\frac{1}{n} MS_{\text{ош}}}\,, \tag{20}$$

где $MS_{\text{ош}}$ значение, определенное в результате проведения дисперсионного анализа.

Рассмотрим алгоритм исследования с помощью множественного критерия размахов Дункана.

1. Расположить все средние в порядке убывания.
2. Определить стандарт ошибки.
3. По таблице значимых размахов Дункана (Приложение В) определяются величины $r_\alpha(p,f)$, где α – уровень значимости; $p = \overline{2,a}$; f – число степеней свободы ошибки.
4. Значимые размахи Дункана преобразуются в набор *(a-1)* наименьших значимых размахов по формуле:

$$R_p = r_\alpha(p,f) \cdot S_{\bar{y}_i} \tag{21}$$

5. Последовательно определяют размахи между всевозможными парами средних, начиная с наибольшего, и сравнивают полученные результаты с наименьшими значимыми размахами. Заметим, что самый большой размах будет между $max\ \bar{y}_i$ и $min\ \bar{y}_i$.
6. Сформулировать вывод. Если значение разности больше соответствующего наименьшего значимого размаха, то различия между данными средними по обработкам значима (существенна). В противном случае – разность незначима и средние сопоставимы.

2.3 Сравнить средние по обработкам в условиях примера этой главы

Решение

$\overline{y_1} = 9{,}8;\ \overline{y_2} = 15{,}4;\ \overline{y_3} = 17{,}6;\ \overline{y_4} = 21{,}6; \overline{y_5} = 10{,}8$

1. Располагаем в порядке убывания $\overline{y_4};\ \overline{y_3};\ \overline{y_2};\ \overline{y_5}; \overline{y_1}$

2. Так как объем выборок для обработок совпадает, то значение стандарта ошибки будет единым для всех $S_{\bar{y}} = \sqrt{\frac{1}{5} \cdot 8{,}06} \approx 1{,}27.$

3. Из таблицы Приложения В находим значимые размахи Дункана
(α = 0,05; f = 20).

$r_{0{,}05}(2; 20) = 2{,}95$

$r_{0{,}05}(3; 20) = 3{,}10$

$r_{0{,}05}(4; 20) = 3{,}18$

$r_{0{,}05}(5; 20) = 3{,}25$

4. Определяем наименьшие значимые размахи:

$$R_2 = r_{0,05}(2{,}20) \cdot 1{,}27 = 2{,}95 \cdot 1{,}27 \approx 3{,}747$$
$$R_3 = r_{0,05}(3{,}20) \cdot 1{,}27 = 3{,}10 \cdot 1{,}27 \approx 3{,}937$$
$$R_4 = 3{,}18 \cdot 1{,}27 \approx 4{,}039$$
$$R_5 = 3{,}25 \cdot 1{,}27 \approx 4{,}128$$

5. Сравнение всевозможных пар начинаем с наибольшего среднего

$$\overline{y_4}:\quad \overline{y_4} - \overline{y_1} \vee R_5$$
$$\overline{y_4} - \overline{y_5} \vee R_4$$
$$\overline{y_4} - \overline{y_2} \vee R_3$$
$$\overline{y_4} - \overline{y_3} \vee R_2$$

Полученные значения являются размахами между указанными обработками. Размахи сравним с соответствующим наименьшим значимым размахом.

$$\overline{y_4}:\quad \overline{y_4} - \overline{y_1} = 21{,}6 - 9{,}8 = 11{,}8 > 4{,}128(R_5)$$
$$\overline{y_4} - \overline{y_5} = 21{,}6 - 10{,}8 = 10{,}8 > 4{,}039(R_4)$$
$$\overline{y_4} - \overline{y_2} = 21{,}6 - 15{,}4 = 6{,}2 > 3{,}937(R_3)$$
$$\overline{y_4} - \overline{y_3} = 21{,}6 - 17{,}6 = 4 > 3{,}747(R_2)$$

Вывод: все разности значимы, то есть прочность синтетического волокна, содержащего 30% хлопка, существенно отличаются от прочности других волокон эксперимента.

6. Отбрасываем наибольшее среднее в списке и продолжаем сравнивать обработки. В данном случае максимальной становится $\overline{y_3}$**,** а наибольшим из наименьших размахов – R_4.

$$\overline{y_3}:\quad \overline{y_3} - \overline{y_1} = 17{,}6 - 9{,}8 = 7{,}8 > 4{,}039(R_4)$$
$$\overline{y_3} - \overline{y_5} = 17{,}6 - 10{,}8 = 6{,}8 > 3{,}937(R_3)$$
$$\overline{y_3} - \overline{y_2} = 17{,}6 - 15{,}4 = 2{,}2 < 3{,}747(R_2)$$

Вывод: не значимой является только последняя разность, следовательно, прочность синтетического волокна, содержащего 25% хлопка и 20%, существенного различия не имее.

7. Повторим алгоритм до последнего сравнения

$$\overline{y_2}:\quad \overline{y_2} - \overline{y_1} = 15{,}4 - 9{,}8 = 5{,}6 > 3{,}937(R_3)$$
$$\overline{y_2} - \overline{y_5} = 15{,}4 - 10{,}8 = 4{,}6 > 3{,}747(R_2)$$

Вывод: разности значимы, то есть прочность синтетического волокна, содержащего 20% хлопка, существенно отличается от прочности волокон 15% и 35% содержания хлопка.

8. $\overline{y_5}:\quad \overline{y_5} - \overline{y_1} = 10{,}8 - 9{,}8 = 1 < 3{,}747(R_2)$

Вывод: разность незначима, следовательно, прочность синтетического волокна, содержащего 35% хлопка и 15%, существенного различия не имее.

1. Множественный критерий размахов Дункана проводят после исследования значимости фактора.
2. Если первое сравнение имеет знак меньше, то дальнейшие сравнения не проводят. Они не имеют смысла и делают вывод о том, что все обработки несущественно отличаются друг от друга.

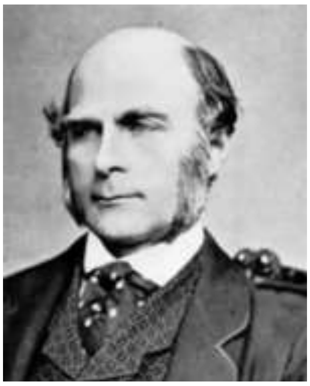
РЕГРЕССИОННЫЕ МОДЕЛИ

При изучении процессов функционирования сложных систем приходится иметь дело с целым рядом одновременно действующих случайных величин. Для уяснения механизма явлений, причинно-следственных связей между элементами системы, по полученным наблюдениям пытаются установить взаимоотношения этих величин. До сих пор отсутствуют принятые всеми аналитические модели многих процессов. Достаточно часто для построения моделей по результатам экспериментальных исследований используют математический аппарат регрессионного и корреляционного анализа.

Регрессия – термин от латинского корня [gressio], означающего «движение», и приставки [re], характеризующей его обратное направление.

Общеупотребительное значение – возвращение к ранее пройденным периодам, формам существования, стадиям, состояниям.

Слово «корреляция» происходит от латинского [correlatio] — «соотношение».



Фрэнсис Гальтон — британский исследователь, географ, антрополог, психолог, статистик.

Карл Пирсон — британский математик, статистик, биолог, расовый теоретик, философ

Фрэнсис Гальтон (1822-1911) - двоюродный брат Чарлза Дарвина (1809-1882) впервые (1889) ввел в употребление понятие «регрессия» и разработал основы корреляционного анализа. Он заложил основы новой науки и дал ей имя, но в стройную научную дисциплину ее превратил математик Карл Пирсон (1857-1936). В результате появился широко известный сегодня коэффициент корреляции по Пирсону.

Ф. Гальтон и К.Пирсон изучали взаимозависимости роста и массы людей разного возраста и столкнулись с необходимостью введения таких показателей указанной зависимости, которые бы отражали связь между исследуемыми характеристиками человека, но не определяли бы друг друга строго однозначно.

В настоящее время «регрессия» и «корреляция» – основные понятия статистики.

В естественных науках большей частью имеют дело со строгими (функциональными) зависимостями, при которых каждому значению одной переменной соответствует единственное значение другой. Однако в реальном мире таких зависимостей нет. Это связано с целым рядом причин и, в частности, с тем, что, во-первых, при анализе влияния одной переменной на другую не учитывается целый ряд других факторов, влияющих на нее. Во-вторых, это влияние может быть не прямым, а проявляться через цепочку других факторов; в-третьих, многие такие воздействия носят случайный характер и т.д. Поэтому в подобных случаях говорят не о функциональных, а о корреляционных либо статистических зависимостях. *Статистической* называют *зависимость*, при которой изменение одной из величин влечет изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой. Такую статистическую зависимость называют *корреляционной.* Примерами корреляционной связи являются зависимости между пределами прочности и текучести стали определенной марки, между погрешностями размера и погрешностью формы поверхности детали, между температурой испытания и прочностью материала и т. д.

Основные положения корреляционного анализа

Корреляционный анализ — это количественный метод определения тесноты и направления связи между выборочными переменными величинами.

Цель корреляционного анализа — выявление оценки силы связи между случайными величинами (признаками), которые характеризуют некоторый реальный процесс.

Основные задачи корреляционного анализа:

– измерение степени связности двух и более явлений;

– отбор факторов, оказывающих наиболее существенное влияние на результативный признак, на основании измерения степени связности между явлениями;
– обнаружение неизвестных причинных связей.

Основные практические приёмы корреляционного анализа:

– построение корреляционного поля и составление корреляционной таблицы;
– вычисление выборочных коэффициентов корреляции;
– проверка статистических гипотез значимости связи.

Предпосылки корреляционного анализа:

– переменные величины должны быть случайными;
– случайные величины должны иметь нормальное распределение.

Рассмотрим простейший случай корреляционного анализа – двумерную модель. Пусть X и Y – случайные нормально распределенные величины.

Наиболее простым и эффективным способом выявления взаимосвязей между явлениями, с которого начинается корреляционный анализ, является *графический метод*. Для этого на координатном поле наносят точки, соответствующие значениям изучаемых признаков X и Y. На оси абсцисс откладывают значения признака X, на оси ординат –признака Y. Совокупность точек образует корреляционное поле. По характеру расположения точек на корреляционном поле можно судить о направлении и силе связи.

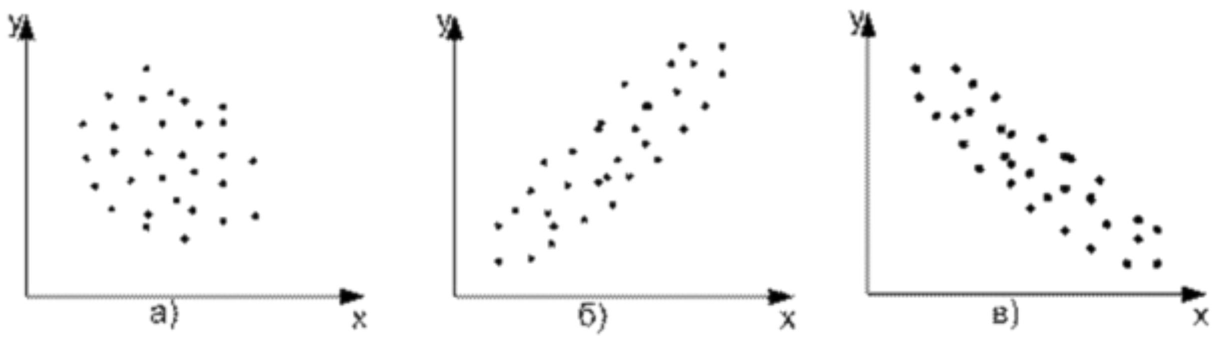


Рисунок 1 – Корреляционное поле

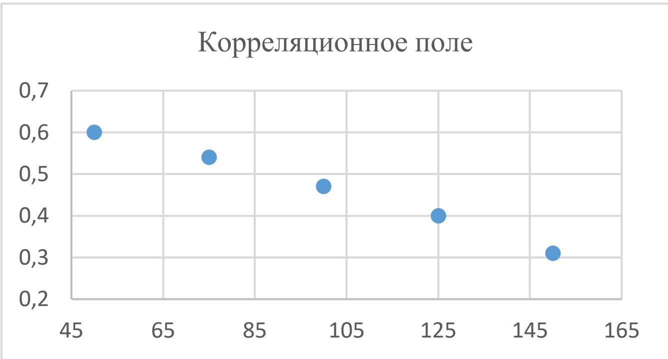
Если точки беспорядочно разбросаны по полю, то зависимость между переменными отсутствует (рис.1(*а*)); если точки образуют эллипс, то есть концентрируются вокруг оси, идущей из нижнего левого угла в верхний правый (или наоборот), то имеется прямая (или обратная) зависимость между исследуемыми признаками (рис. 1(б; в))

3. Предложены экспериментальные данные по обкатыванию поверхности шаровым инструментом и шероховатости обработанной поверхности

Сила прижима (X), кгс	50	75	100	125	150
Шероховатость (Y), мкм	0,6	0,54	0,47	0,4	0,31

Построить корреляционное поле и сделать предположения относительности связи между переменными.

Решение



Судя по расположению точек на диаграмме, можно предположить, что между силой прижима и шероховатостью существует обратная линейная связь.

Более совершенным показателем степени тесноты корреляционной связи является *линейный коэффициент корреляции*. При расчете этого показателя учитываются не только отклонения индивидуальных значений признака от средней, но и сама величина этих отклонений. Согласно общей теории Математической статистики, а именно разделу, посвященному системе случайных величин, коэффициент корреляции ρ определяется как отношение ковариации между величинами X и Y к произведению средних квадратических отклонений X и Y соответственно:

$$\rho_{xy} = \frac{cov(X;Y)}{\sigma_X \sigma_Y} \qquad (22)$$

Оценкой коэффициента корреляции является выборочный коэффициент корреляции:

$$r_{xy} = \frac{\overline{xy} - \bar{x}\bar{y}}{\sqrt{S_x^2 \cdot S_y^2}} \qquad (23)$$

где $\overline{xy}$ – выборочное среднее произведения значений случайных величин;

$\bar{x}$, $\bar{y}$ – выборочное среднее X и Y соответственно;

S_x^2, S_y^2 – выборочные дисперсии X и Y соответственно.

Коэффициента корреляции определен в интервале от минус единицы до плюс единицы ($|r_{xy}| \le 1$). Связи могут быть слабыми и сильными (тесными). Шкала Чеддока используется для оценки линейного коэффициента корреляции Пирсона:

$|r_{xy}| < 0{,}1$ – связь практически отсутствует;

$0{,}1 \le |r_{xy}| < 0{,}3$ – слабая;

$0{,}3 \le |r_{xy}| < 0{,}5$ – умеренная;

$0{,}5 \le |r_{xy}| < 0{,}7$ – заметная;

$0{,}7 \le |r_{xy}| < 0{,}9$ – высокая;

$0{,}9 \le |r_{xy}| < 1$ – весьма высокая.

Шкала разработана американским профессором социологии и статистики Робертом Эмметом Чеддоком (1879–1940). В случае если коэффициент корреляции равен 0, обе переменные полностью независимы друг от друга.

Если случайные величины X и Y независимы, то они не коррелируемы, но из некоррелируемости не следует их независимость, так как равенство нулю коэффициента корреляции указывает лишь на отсутствие линейной связи между переменными, но не отсутствие связи вообще между ними.

Для статистического вывода о наличии или отсутствии корреляционной связи между исследуемыми переменными необходимо произвести проверку значимости выборочного коэффициента корреляции. В связи с тем, что надежность статистических характеристик, в том числе и коэффициента корреляции, зависит от объема выборки, может сложиться такая ситуация, когда величина коэффициента корреляции будет целиком обусловлена случайными колебаниями в выборке, на основании которой он вычислен. Проверим значимость коэффициента корреляции.

Нулевая гипотеза состоит в том, что коэффициент корреляции равен нулю (не значим), альтернативная - не равен нулю:

$$H_0\colon\ \rho_{xy} = 0$$

$$H_1\colon\ \rho_{xy} \neq 0.$$

Выборочный коэффициент корреляции при определенных предпосылках связан со случайной величиной t, подчиняющейся распределению Стьюдента с (*n*-*2*) степенями свободы. Статистика имеет вид:

$$t_{\text{в}} = \frac{r_{xy}\sqrt{n-2}}{\sqrt{1-r_{xy}^2}} \tag{24}$$

При проверке гипотезы необходимо установить уровень значимости α, который дает практическую уверенность в том, что ошибочные заключения будут сделаны только в очень редких случаях. Уровень значимости выражает вероятность того, что нулевая гипотеза отвергается, в то время как она верна. Очевидно, что имеет смысл выбирать эту вероятность как можно меньше. При заданном уровне значимости α определяется критическое значение $t_{\text{кр}(\alpha;n-2)}$.

Если $|t| \geq t_{\text{кр}}$, то нулевая гипотеза на уровне значимости α отвергается, то есть связь между переменными значима. В противном случае ($|t| \leq t_{\text{кр}}$) – нулевая гипотеза на уровне значимости α не отвергается. Интервальная оценка для коэффициента корреляции (доверительный интервал):

$$\left(r_{xy} - t_{\text{кр}}\sqrt{\frac{1-r_{xy}{}^2}{n-2}}; r_{xy} + t_{\text{кр}}\sqrt{\frac{1-r_{xy}{}^2}{n-2}}\right) \tag{25}$$

3.1 Определить коэффициент корреляции и проверить его значимость

Решение

$\boldsymbol{Y}$	$\boldsymbol{X}$	$\boldsymbol{xy}$	$(\boldsymbol{x_i}-\boldsymbol{\bar{x}})^2$	$(\boldsymbol{y_i}-\boldsymbol{\bar{y}})^2$
0,6	50	$0{,}6\ \cdot 50 = 30$	$(50-100)^2 = 2500$	$(0{,}6-0{,}46)^2 = 0{,}018$
0,54	75	40,5	625	0,006
0,47	100	47	0	0,000036
0,4	125	50	625	0,0041
0,31	150	46,5	2500	0,0237
$\bar{y} = 0{,}46$	$\bar{x} = 100$	$\overline{xy} = 42{,}8$	$S_x^2 = 1562{,}5$	$S_y^2 \approx 0{,}013$
			$S_x \approx 39{,}53$	$S_y \approx 0{,}11$

$$r_{xy} = \frac{42{,}8 - 100 \cdot 0{,}46}{39{,}53 \cdot 0{,}11} \approx -0{,}83$$

Так как значение коэффициента корреляции по абсолютной величине принадлежит [0,7; 0,9), то наблюдается высокая (сильная) линейная связь. Судя по знаку значения r_{xy}, связь обратная. Проверим гипотезу о значимости коэффициента корреляции:

$$H_0\colon\ \rho = 0$$
$$H_1\colon\ \rho \neq 0.$$

Определим выборочное значение статистики по формуле (24).

$$t_{\text{в}} = \frac{-0{,}83\sqrt{5-2}}{\sqrt{1-(-0{,}83)^2}} = \frac{-0{,}83 \cdot 1{,}73}{\sqrt{1-0{,}6889}} \approx \frac{-1{,}44}{0{,}56} \approx -2{,}56$$

Критическая точка при уровне значимости 0,05 определяется по таблице Приложения Г: $t_{\text{кр}(0{,}05;5-2)} = 3{,}18$. Сравним выборочное значение с критической точкой.

$$2{,}56 < 3{,}18 \Rightarrow H_0 \text{ – не отвергается.}$$

Таким образом, коэффициент корреляции незначим и продолжать построение линейной модели не имеет смысла, так как переменную X можно исключить из уравнения регрессии.

Модель парной линейной регрессии

В сложных технологических процессах при рассмотрении взаимосвязей между двумя переменными X и Y выделяют одну из величин как *независимую* (объясняющую), а другую – как *зависимую* (объясняемую). В этом случае изменение первой из них может служить причиной для изменения другой.

Например, изменение свойства сырья вызывает случайные изменения свойства продукции. Однако, как было указано выше, такая зависимость не является однозначной в том смысле, что каждому конкретному значению объясняющей переменной (набору объясняющих переменных) может соответствовать не одно, а множество значений из некоторой области. Другими словами, каждому конкретному значению объясняющей переменной (набору объясняющих переменных) соответствует некоторое вероятностное распределение зависимой переменной (рассматриваемой как случайная величина). Поэтому анализируют, как объясняющая(ие) переменная(ые) влияет(ют) на зависимую переменную «в среднем». Зависимость такого типа, выражаемая соотношением $M(Y|x) = f(x)$, называется *регрессией* Y на X. При рассмотрении зависимости двух случайных величин говорят о *парной регрессии.* Зависимость нескольких переменных, выражаемая функцией $M(Y|x_1, x_2 \dots x_n) = f(x_1, x_2 \dots x_n)$, называют *множественной регрессией*.

Регрессионный анализ — набор методов, которые оценивают связь между несколькими переменными с помощью построения математических моделей. Они определяют, как изменения одного параметра влияют на другой.

Математический аппарат регрессионного анализа позволяет:

- оценить неизвестные параметры предлагаемой к исследованию регрессионной модели;
- проверить статистическую значимость параметров модели;
- проверить адекватность модели;
- оценить точность модели.

Вид регрессионной модели предлагает сам исследователь, при этом он исходит из следующего:

- физической сущности изучаемого объекта или явления;
- характера экспериментального материала;
- анализа априорной информации.

Самым простым для моделирования является объект, у которого один входной и один выходной фактор.

Для начала построения эмпирической модели необходимо иметь данные экспериментальных исследований объекта (в виде таблицы или графика), в которых каждому значению входного фактора (X) соответствует значение выходного фактора (Y), то есть известна пара чисел (x_i , y_i).

Парная линейная регрессия (теоретическое линейное уравнение регрессии) представляет собой линейную функцию между условным математическим ожиданием $M(Y|X=x)$ зависимой переменной Y и одной переменной X: $M(Y|X=x)=\beta_0+\beta_1\cdot x+\varepsilon$, где x – значение входной переменной; β_0 и β_1 – коэффициенты регрессии; ε – остаток.

По выборке ограниченного объема можно построить так называемое *эмпирическое* (*выборочное*) *уравнение регрессии* $\hat{y}=b_0+b_1\cdot x$, где $\hat{y}$ – значения выходной переменной, определенной (предсказанной) по линейной модели. Эмпирическое уравнение регрессии определяется на основе конечного числа статистических данных. Поэтому коэффициенты эмпирического уравнения регрессии являются случайными величинами, изменяющимися от выборки к выборке. Коэффициенты β_0 и β_1 будем оценивать по выборке с помощью *метода наименьших квадратов (МНК)*.

Оценка коэффициентов регрессии с помощью МНК

Суть метода в том, что подбираются такие величины b_0 и b_1 , при которых сумма квадратов отклонений измеренных величин y от предсказанных $\hat{y}$ была бы минимальной.

Пусть $y_i=\beta_0+\beta_1\cdot x_i+\varepsilon_i$ и $\hat{y}_i=b_0+b_1\cdot x_i$, тогда

$\varepsilon_i=y_i-\hat{y}_i=y_i-(b_0+b_1\cdot x_i)$

Сумма квадратов отклонений представлена в виде:

$$S^2=\sum_{i=1}^{m}\varepsilon_i^2=\sum_{i=1}^{m}\big(y_i-(b_0+b_1\cdot x_i)\big)^2\rightarrow\ min$$

Решение поставленной задачи сводится к определению экстремума указанной функции двух переменных. С этой целью находим частные производные функции по коэффициентам b_0 и b_1 и приравниваем их к нулю.

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum_{i=1}^{m}(y_i - (b_0 + b_1 \cdot x_i)) = 0 \\ \dfrac{\partial S}{\partial b_1} = -2\sum_{i=1}^{m}(y_i - (b_0 + b_1 \cdot x_i)) \cdot x_i = 0 \end{cases}$$

Решим полученную систему уравнений:

$$\begin{cases} \sum_{i=1}^{m} y_i - mb_0 - b_1 \sum_{i=1}^{m} x_i = 0 \\ \sum_{i=1}^{m} y_i x_i - b_0 \sum_{i=1}^{m} x_i - b_1 \sum_{i=1}^{m} x_i^2 = 0 \end{cases}$$

Из первого уравнения системы следует:

$$-mb_0 = b_1 \sum_{i=1}^{m} x_i - \sum_{i=1}^{m} y_i$$

$$b_0 = \frac{\sum_{i=1}^{m} y_i - b_1 \sum_{i=1}^{m} x_i}{m}$$

$$b_0 = \frac{1}{m}\sum_{i=1}^{m} y_i - b_1 \frac{1}{m}\sum_{i=1}^{m} x_i$$

$$b_0 = \bar{y} - b_1\bar{x} \qquad (26)$$

Из второго уравнения системы следует:

$$b_1 \sum_{i=1}^{m} x_i^2 = \sum_{i=1}^{m} y_i\, x_i - b_0 \sum_{i=1}^{m} x_i$$

$$b_1\overline{x^2} = \overline{yx} - b_0\bar{x}$$

$$b_1 = \frac{\overline{yx} - b_0\bar{x}}{\overline{x^2}}$$

$$b_1 = \frac{\overline{yx} - (\bar{y} - b_1\bar{x}\,)\bar{x}}{\overline{x^2}}$$

$$b_1 = \frac{\overline{yx} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} \qquad (27)$$

1. Нетрудно доказать, что имеет место соотношение $b_1 = r_{xy} \cdot \frac{\sqrt{S_y^2}}{\sqrt{S_x^2}}$.

Таким образом, коэффициент регрессии пропорционален коэффициенту корреляции.

2. Смысл коэффициентов регрессии: b_1- угловой коэффициент регрессии показывает, на сколько единиц в среднем изменяется переменная Y при изменении переменной X на единицу; b_0 – свободный коэффициент дает прогнозируемое значение зависимой переменной Y при $x = 0$.

4. Определить регрессионную модель опираясь на данные таблицы.

x	53	57	60	63	64	66	63	62	62	66	69	67
y	51	54	57	59	63	58	60	59	58	63	65	62

Решение

Y	***X***	***xy***	$(x_i)^2$
51	53	2703	2809
54	57	3078	3249
57	60	3420	3600
59	63	3717	3969
63	64	4032	4096
58	66	3828	4356
60	63	3780	3969
59	62	3658	3844
58	62	3596	3844
63	66	4158	4356
65	69	4485	4761
62	67	4154	4489
$\bar{y} = 59{,}08$	$\bar{x} = 62{,}67$	$\overline{xy} = 3717{,}42$	$\overline{x^2} = 3945$

$$b_1 = \frac{3717{,}42 - 59{,}08 \cdot 62{,}67}{3945 - 62{,}67^2} = \frac{14{,}86}{18} \approx 0{,}82$$

$$b_0 = 59{,}08 - 0{,}82 \cdot 62{,}67 = 7{,}50$$

Регрессионная модель имеет вид: $\hat{y} = 7{,}50 + 0{,}82x$.

Анализ адекватности регрессионной модели

Для практического использования моделей регрессии большое значение имеет их адекватность, то есть соответствие фактическим статестическим данным. Другими словами, насколько широко рассеяны точки наблюдений относительно линии регрессии. Следует напомнить, что цель регрессионного анализа – объяснить поведение зависимой переменной *Y*.

Пусть на основе выборочных наблюдений построено уравнение регрессии $\hat{y} = b_0 + b_1 \cdot x$, тогда значение зависимой переменной Y в каждом наблюдении можно разложить на две составляющие: $y_i = \hat{y_i} + e_i$, где остаток e_i - та часть зависимой переменной *Y*, которую невозможно объяснить с помощью уравнения регрессии. Следовательно,

$$S_y^2 = S_{\hat{y}}^2 + S_e^2 \qquad (28)$$

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

$S_{\hat{y}}^2 = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2$;

$S_e^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y_i})^2$.

Отношение объясненной части дисперсии переменной к общей дисперсии называется *коэффициентом детерминации* (R^2) и определяется как

$$R^2 = \frac{S_{\hat{y}}^2}{S_y^2} = 1 - \frac{S_e^2}{S_y^2} \qquad (29)$$

Коэффициент детерминации характеризует долю вариации (разброса) зависимой переменной, объясненную с помощью уравнения регрессии.

Значение R^2 должно находиться в диапазоне от нуля до единицы: $0 \leq R^2 \leq 1$. Модель считается более качественной, если значение коэффициента детерминации близко к 1. Если $R^2 = 1$, то эмпирические точки (x_i; y_i) лежат точно на линии регрессии и между переменными Y и X существует линейная функциональная зависимость. Если $R^2 = 0$, то вариация зависимой переменной полностью обусловлена неучтенными в модели факторами. Достаточно качественной можно признать модель с коэффициентом детерминации выше 0,8.

В анализе важнейшей оценкой является проверка качества уравнения регрессии. Принята следующая схема проверки:

– проверка статистической значимости коэффициентов уравнения регрессии;
– проверка общего качества уравнения регрессии.

На начальном этапе статистического анализа построенной модели наиболее важной является задача установления наличия линейной зависимости между переменными Y и X. Эта проблема может быть решена с помощью проверки *гипотезы о статистической значимости коэффициента регрессии*

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

При этом, если H_0 не отвергается, то есть основание считать, что величина Y не зависит от X (точнее связь между этими двумя переменными далека от линейной зависимости). В этом случае говорят, что коэффициент b_1 статистически незначим (он слишком близок к нулю). При отклонении H_0 коэффициент b_1 считается статистически значимым, что указывает на наличие определенной линейной зависимости между Y и X.

Значимость коэффициента регрессии b_1 проверяется с помощью анализа отношения

$$t_{\text{в}} = \frac{b_1}{S_{b_1}} \text{ , где } S_{b_1} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2}{(n-2)\sum(x_i - \bar{x})^2}} \tag{30}$$

b_1 – оценка коэффициента регрессии, полученная по наблюдаемым данным; S_{b_1} – стандартная ошибка коэффициента регрессии.

Статистика имеет распределение Стьюдента с числом степеней свободы $df = n - 2$.

Проверим гипотезу о значимости свободного коэффициента:

$$H_0 : \beta_0 = 0;$$

$$H_1 : \beta_0 \neq 0.$$

Для параметра b_0 критерий проверки гипотезы о незначимом отличии его от нуля имеет вид:

$$t_{\text{в}} = \frac{b_0}{S_{b_0}}\text{, где } S_{b_0} = \sqrt{\frac{\sum(y_i - \hat{y}_i)^2 \sum x_i^2}{n(n-2)\sum(x_i - \bar{x})^2}} \tag{31}$$

После проверки значимости каждого из коэффициентов регрессии обычно проверяется общее качество уравнения, которое оценивается по тому, как хорошо эмпирическое уравнение регрессии согласуется со статистическими данными. Коэффициент детерминации рассматривают, как правило, в качестве основного показателя, отражающего меру качества регрессионной модели, описывающей связь между зависимой и независимыми переменными модели. На практике для этого проверяют гипотезу о статистической значимости коэффициента детерминации R^2:

$$H_0: R^2 = 0;$$

$$H_1: R^2 > 0.$$

Для проверки данной гипотезы используется следующая F – статистика:

$$F_{\text{в}} = \frac{R^2(n-2)}{1-R^2} \tag{32}$$

со степенями свободы $df_1 = 1$; $df_2 = n-2$.

Если выборочное значение статистики больше значения критической точки, то основная гипотеза отвергается и коэффициент детерминации значим.

При анализе адекватности модели возможны следующие варианты:

1. Коэффициент детерминации значим и все коэффициенты регрессии значимы. Тогда модель может быть использована для принятия решений и прогнозирования.
2. Коэффициент детерминации значим и угловой коэффициент регрессии значим. Тогда модель может быть использована для принятия некоторых решений, но не для прогнозирования.
3. Коэффициент детерминации значим, все коэффициенты регрессии не значимы. Тогда модель считают полностью неадекватной.

4.1 Определить регрессионную модель, опираясь на данные таблицы.

x	53	57	60	63	64	66	63	62	62	66	69	67
y	51	54	57	59	63	58	60	59	58	63	65	62

Проверить адекватность модели.

Решение

Ранее была построена регрессионная модель вида: : $\hat{y} = 7{,}50 + 0{,}82x$.

1. Вычислим коэффициент детерминации (29).

$S_y^2 = 2{,}18$; $S_e^2 = 14{,}41$

$$R^2 = 1 - \frac{2{,}18}{14{,}41} = 0{,}85$$

Значение коэффициента детерминации расположено в интервале от 0,7 до 0,99, следовательно, модель достаточно качественная (на 85 процентов переменная X объясняет значения переменной Y).

$\boldsymbol{Y}$	$\boldsymbol{X}$	$\boldsymbol{\hat{y}}$	$\boldsymbol{y_i - \hat{y}_i}$	$\boldsymbol{y_i - \bar{y}}$
51	53	51,13	-0,13	-8,08
54	57	54,42	-0,42	-5,08
57	60	56,89	0,11	-2,08
59	63	59,36	-0,36	-0,08
63	64	60,18	2,82	3,92
58	66	61,83	-3,83	-1,08
60	63	59,36	0,64	0,92
59	62	58,53	0,47	-0,08
58	62	58,53	-0,53	-1,08
63	66	61,83	1,17	3,92
65	69	64,30	0,70	5,92
62	67	62,65	-0,65	2,92
			$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 26{,}13$	$\sum_{i=1}^{n}(y_i - \bar{y})^2 = 172{,}92$

2. Проверяем гипотезы относительно коэффициентов регрессии.

$H_0 : \beta_1 = 0$

$H_1 : \beta_1 \neq 0$

Для проверки воспользуемся формулой (30).

$$S_{b_1} = \sqrt{\frac{26{,}13}{(12-2)\cdot 216{,}67}} \approx 0{,}11;\ t_{\text{в}} = \frac{0{,}82}{0{,}11} \approx 7{,}49;\ t_{\text{кр}\,(0{,}05;\,12-2)} = 2{,}23$$

Так как выборочное значение статистики больше значения критической точки, то основная гипотеза отвергается, то есть угловой коэффициент регрессии значим.

Y	X	$\hat{y}$	$y_i - \hat{y}_i$	$x_i - \bar{x}$
51	53	51,13	-0,13	-9,67
54	57	54,42	-0,42	-5,67
57	60	56,89	0,11	-2,67
59	63	59,36	-0,36	0,33
63	64	60,18	2,82	1,33
58	66	61,83	-3,83	3,33
60	63	59,36	0,64	0,33
59	62	58,53	0,47	-0,67
58	62	58,53	-0,53	-0,67
63	66	61,83	1,17	3,33
65	69	64,30	0,70	6,33
62	67	62,65	-0,65	4,33
			$\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 = 26,13$	$\sum_{i=1}^{n}(x_i - \bar{x})^2 = 216,67$

Проверим гипотезу о значимости свободного коэффициента:

$H_0: \beta_0 = 0;$

$H_1: \beta_0 \neq 0.$

Для проверки воспользуемся формулой (31).

$$S_{b_0} = \sqrt{\frac{26,13 \cdot 47342}{12 \cdot (12-2) \cdot 216,67}} \approx 6,90\ ;\ t_{в} = \frac{7,50}{6,90} \approx 1,09;\ t_{кр\ (0,05;\ 12-2)} = 2,23$$

Так как выборочное значение статистики меньше значения критической точки, то основная гипотеза не отвергается, то есть свободный коэффициент регрессии не значим.

3. Проверим гипотезу о значимости коэффициента детерминации

H_0: $R^2 = 0$;

H_1: $R^2 > 0$.

Для проверки воспользуемся формулой (32).

$$F_{в} = \frac{0,85 \cdot (12-2)}{1-0,85} = 56,16;\ F_{кр\ (0,05;\ 1;\ 12-2)} = 4,96.$$

Так как выборочное значение статистики больше значения критической точки, то основная гипотеза отвергается, следовательно, коэффициент детерминации значим.

4. Принятие решения относительно регрессионной модели.

Коэффициент детерминации значим, и угловой коэффициент регрессии значим. Следовательно, регрессионная модель вида $\hat{y} = 7,50 + 0,82x$ может быть использована для принятия некоторых решений, но не для прогнозирования.

ЛИНЕЙНОЕ ПРОГРАММИРОВАНИЕ

Многие задачи планирования народного хозяйства, управления промышленностью тесно связаны с определением оптимального плана использования ограниченных ресурсов. Необходимость решения таких задач привела к широкому применению математических методов и появлению нового раздела математики – математического программирования. Предметом математического программирования является решение задач о нахождении точек экстремумов функций нескольких переменных на множествах, определяемых линейными или нелинейными ограничениями (равенствами и неравенствами).

Оптимизация – это выбор наилучшего решения с учетом ограничений на этот выбор. Математическая формулировка задач оптимизации в качестве обязательных элементов содержит

<u>*Целевую функцию*</u> $(F(x_1, x_2, ..., x_n))$ <u>*при условиях*</u> $G_i\ (x_1, x_2, ..., x_n) < b_i\ (i = \overline{1, m})$, где $F(x)$ и G_i − заданные функции, а b_i − некоторые действительные числа. Оптимизация заключается в определении значений переменных $x_1, x_2, ...x_n$, при которых целевая функция минимальна (или максимальна).

В зависимости от свойств функций $F(x)$ и G_i математическое программирование можно рассматривать как ряд самостоятельных дисциплин, занимающихся изучением и разработкой методов решения определенных классов задач. Прежде всего, задачи математического программирования делятся на задачи линейного и нелинейного программирования. При этом если все функции $F(x)$ и G_i линейные, то соответствующая задача является задачей линейного программирования. Если же хотя бы одна из указанных функций нелинейная, то соответствующая задача является задачей нелинейного программирования. Наиболее изученным разделом математического программирования является линейное программирование. Для решения задач линейного программирования разработан целый ряд эффективных методов, алгоритмов и программ.

Формализация проблемы как задачи линейного программирования (ЗЛП) подразумевает следующие этапы:

– понять проблему, составить описательную модель задачи;
– идентифицировать основные переменные задачи;
– выбрать некоторую количественную меру эффективности для целевой функции;
– представить эту меру эффективности как линейную функцию относительно основных переменных;
– идентифицировать и представить все ограничения как линейные уравнения или неравенства относительно основных переменных;
– собрать количественные данные или сделать соответствующие оценки для всех параметров модели.

Общая форма записи модели ЗЛП

$$F(x) = c_1x_1 + c_2x_2 + \ldots + c_nx_n \rightarrow max \qquad (33)$$

$$\begin{cases} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n \le b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n \le b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n \le b_m \\ x_i \ge 0, i = 1,2,\ldots,n \end{cases} \qquad (34)$$

где a_{ij}, b_i, c_i - заданные постоянные величины

<u>*Допустимое решение*</u> – это совокупность чисел (план) *(x_1, x_2, …, x_n)*, удовлетворяющих ограничениям задачи (принадлежащих допустимому множеству).

<u>*Оптимальное решение*</u> – план, при котором целевая функция принимает максимальное (минимальное) значение.

Задачу линейного программирования достаточно часто записывают в векторной форме

$$F(X) = CX \rightarrow max\ (min) \qquad (35)$$

$$AX \le B,\ X \ge 0, \qquad (36)$$

где

$$A = \begin{pmatrix} a_{11} & a_{12} \dots & a_{1n} \\ a_{21} & a_{22} \dots & a_{2n} \\ \dots & \dots & \dots \\ a_{m1} & a_{m2} \dots & a_{mn} \end{pmatrix}; \quad X = (x_1, x_2, \dots x_n); \; C = (c_1, c_2, \dots c_n); \; B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{pmatrix}$$

Сформулируем задачу планирования производства в общем виде.

Для изготовления двух видов продукции P_1 и P_2 используют три вида сырья: S_1, S_2, S_3. Запасы сырья, количество единиц сырья, затрачиваемых на изготовление единицы продукции, а также величина прибыли, получаемая от реализации единицы продукции, приведены в таблице. Необходимо составить такой план выпуска продукции, чтобы при ее реализации получить максимальную прибыль.

Вид сырья	**Запасы сырья**	**Количество единиц сырья, затрачиваемых на изготовление единицы продукции**	
		P_1	P_2
S_1	b_1	a_{11}	a_{12}
S_2	b_2	a_{21}	a_{22}
S_3	b_3	a_{31}	a_{32}
Прибыль от единицы продукции		c_1	c_2

Целевая функция:

$$F(x_1, x_2) = c_1x_1 + c_2x_2 \rightarrow max, \tag{37}$$

где x_1 и x_2 – количество изделий продукции P_1 и P_2 соответственно;

$F(x_1, x_2)$ – прибыль (прибыль от единицы одной продукции умноженной на ее количество плюс прибыль от другой продукции умноженной на количество) .

Ограничения:

На изготовление продукции P_1 необходимо затратить a_{11} количество единиц сырья S_1, a_{21} – количество единиц сырья S_2 и a_{31} – сырья S_3. Продукции P_1 выпускается предположительно x_1 штук, следовательно, на это количество продукции будет затрачено:

$(a_{11} \cdot x_1)$ сырья S_1; $(a_{21} \cdot x_1)$ сырья S_2 и $(a_{31} \cdot x_1)$ сырья S_3.

Аналогично можно рассуждать относительно продукции P_2:

$(a_{12} \cdot x_2)$ сырья S_1; $(a_{22} \cdot x_2)$ сырья S_2 и $(a_{32} \cdot x_2)$ сырья S_3.

Однако запасы сырья ограничены. Отсюда следует:

$(a_{11} \cdot x_1 + a_{12} \cdot x_2)$ – это количество сырья S_1 необходимо, но запас данного сырья составляет b_1. Следовательно, необходимое количество не должно превышать имеющееся, но при этом максимально приближаться к нему:

$$a_{11} \cdot x_1 + a_{12} \cdot x_2 \le b_1.$$

Аналогичные рассуждения проводят при формировании всех ограничений. Кроме того, количество изготавливаемой продукции есть величина неотрицательная. Таким образом, система ограничений имеет вид:

$$\begin{cases} a_{11} \cdot x_1 + a_{12} \cdot x_2 \le b_1 \\ a_{21} \cdot x_1 + a_{22} \cdot x_2 \le b_2 \\ a_{31} \cdot x_1 + a_{32} \cdot x_2 \le b_3 \\ x_1, x_2 \ge 0. \end{cases}$$

Итак, общий вид задачи, которую предстоит решить:

$$F(x_1, x_2) = c_1x_1 + c_2x_2 \rightarrow max,$$

$$\begin{cases} a_{11} \cdot x_1 + a_{12} \cdot x_2 \le b_1 \\ a_{21} \cdot x_1 + a_{22} \cdot x_2 \le b_2 \\ a_{31} \cdot x_1 + a_{32} \cdot x_2 \le b_3 \\ x_1, x_2 \ge 0. \end{cases} \qquad (38)$$

Решением задачи будет пара (x_1,x_2).

Задачи линейного программирования можно решать различными способами. Рассмотрю лишь два из них: графический метод и симплекс – метод.

Графический метод

Графический (геометрический) метод применим только для задач малой размерности (количество переменных в задаче равно 2 или 3), поскольку он основан на геометрическом построении множества допустимых решений ЗЛП.

Графический метод реализуется в два этапа:

– построение допустимого множества решений ЗЛП;
– нахождение оптимального решения среди всех допустимых.

Рассмотрим применение графического метода на основе задачи (38). Введем на плоскости прямоугольную систему координат. Тогда допустимую область задачи (38) можно изобразить графически, как множество точек плоскости, координаты которых удовлетворяют сразу всем неравенствам задачи.

Рассмотрим неравенство $a_{11} \cdot x_1 + a_{12} \cdot x_2 \leq b_1$. Каждое такое неравенство определяет полуплоскость, лежащую по одну сторону прямой $a_{11} \cdot x_1 + a_{12} \cdot x_2 = b_1$. Координаты точек другой полуплоскости удовлетворяют противоположному неравенству $a_{11} \cdot x_1 + a_{12} \cdot x_2 > b_1$ (рис.2).

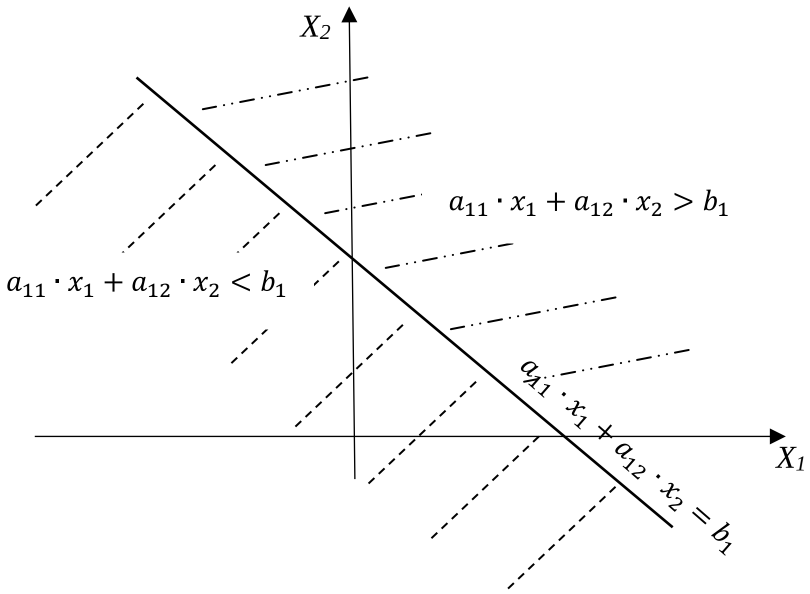


Рисунок 2 – Расположение полуплоскостей относительно прямой $a_{11} \cdot x_1 + a_{12} \cdot x_2 = b_1$

Чтобы определить, какую именно полуплоскость определяет данное неравенство, достаточно взять произвольную точку плоскости (x_1, x_2). (например, начало координат) и подставить в неравенство числа x_1, x_2 . Допустимую область задачи (38) составляют точки пересечения полуплоскостей, определяемых каждым из ограничений.

5. Определить оптимальное решение задачи с помощью графического метода.

$F(x_1, x_2) = 2x_1 + 2x_2 \rightarrow \max$

$$\begin{cases} x_1 + 2x_2 \leq 6 \quad (1) \\ 2x_1 + x_2 \leq 8 \quad (2) \\ x_2 \leq 2 \quad (3) \\ x_1, x_2 \geq 0. \quad (4) \end{cases}$$

Решение

Рассмотрим первое ограничение задачи $x_1 + 2x_2 \le 6$.
Построим прямую $x_1 + 2x_2 = 6$. Для этого определим точки пересечения прямой с осями координат.
X_2 : ($x_1 = 0$), тогда $2x_2 = 6$; $x_2 = 3$. Следовательно, (0; 3)
X_1 : ($x_2 = 0$); $x_1 = 6$. Следовательно, (6; 0).
Построим данную прямую на плоскости X_1OX_2
Аналогично построим прямые (2) и (3).
(2) $2x_1 + x_2 = 8$;
X_1: $2x_1 = 8 \rightarrow x_1 = 4 \rightarrow (4; 0)$
X_2: $x_2 = 8 \rightarrow (0; 8)$.
(3) $x_2 = 2$.

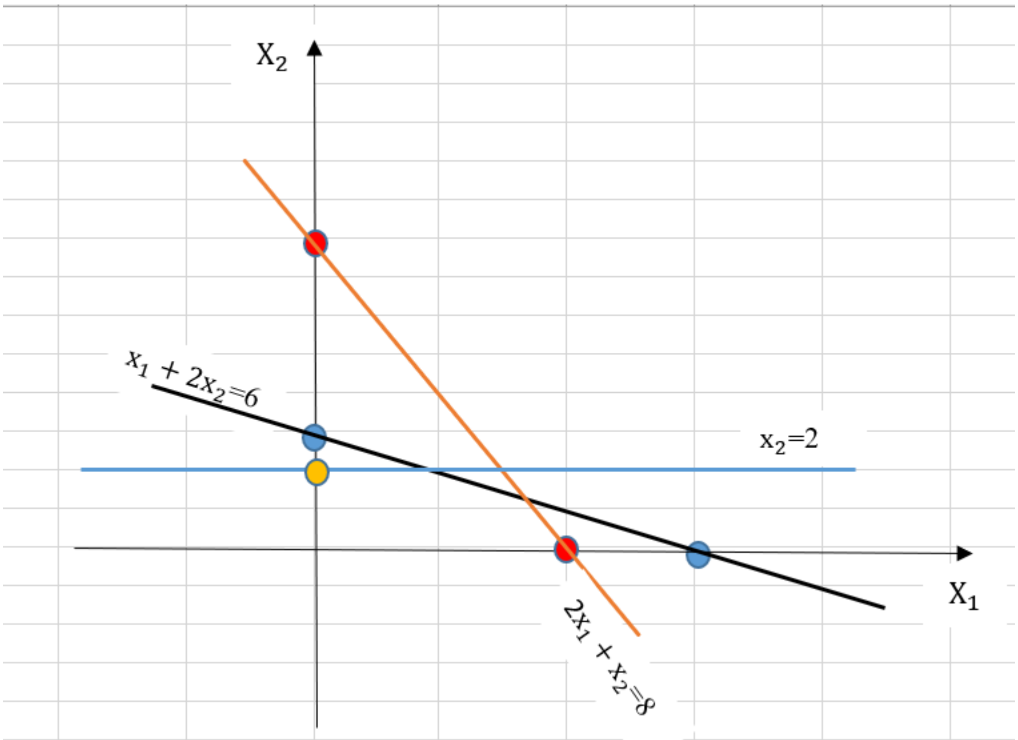


Рисунок 3 – Расположение прямых-ограничений

Так как ограничения представлены неравенствами, то рассматривать в декартовой системе координат надо не прямые, а полуплоскости. Так для первого ограничения $x_1 + 2x_2 \le 6$ мы должны выделить ту полуплоскость, точки которой при подстановке в неравенство будут удовлетворять ему. Эта полуплоскость расположена «ПОД» прямой (1). Если рассуждать аналогично и в отношении других полуплоскостей, то можно выделить многоугольник OABCD (штриховка), который будет определять множество допустимых решений.

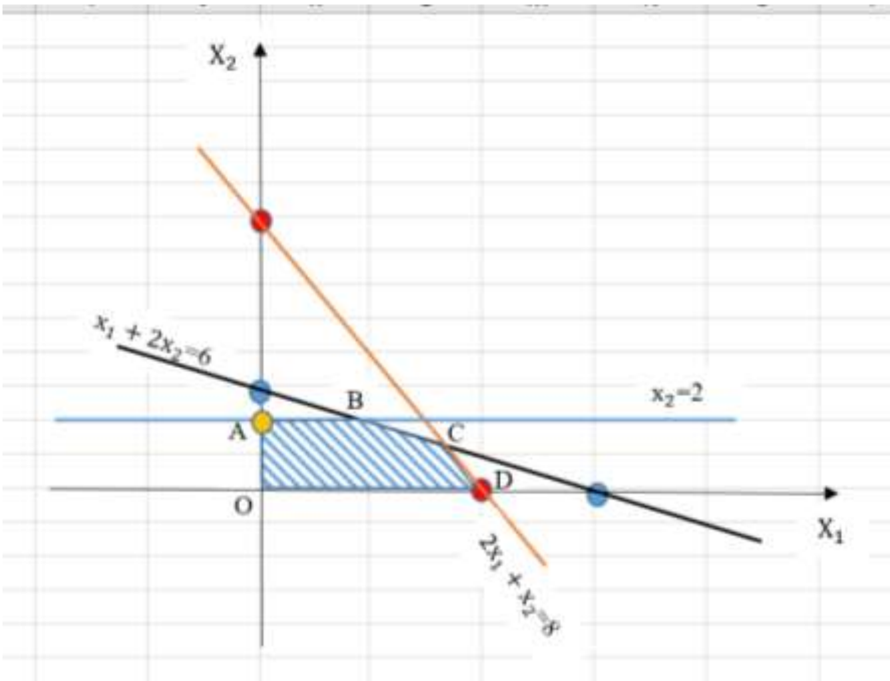


Рисунок 4 – Допустимое множество решений

Нахождение оптимального решения требует определения направления наискорейшего роста целевой функции $F(x_1, x_2)$. Такое направление задается градиентом целевой функции: grad $F(x_1,x_2) = \left(\frac{\partial F}{\partial x_1}, \frac{\partial F}{\partial x_2}\right)$.

Так как $F(x_1, x_2) = 2x_1 + 2x_2$, то направление наискорейшего роста функции F определяется вектором $\vec{n}$= (2; 2) с началом в точке O(0;0).

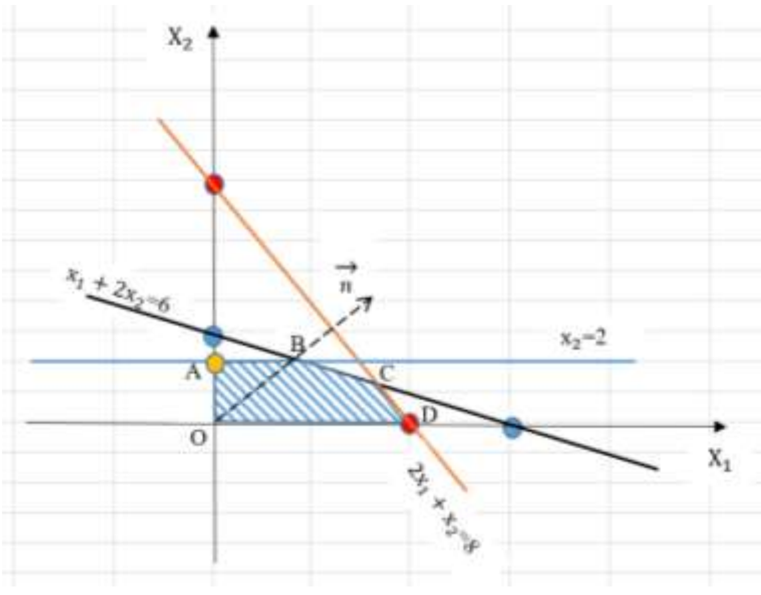


Рисунок 5 – Направление наискорейшего роста функции F

Построим *линию нулевого уровня* $F = 0$, то есть $F(x_1, x_2) = 2x_1 + 2x_2 = 0$. Эта прямая проходит через точку O(0;0) и перпендикулярна вектору $\vec{n}$. Заметив, что каждая из линий уровня является прямой, параллельной линии нулевого уровня $F = 0$, будем передвигать линию нулевого уровня $F = 0$ параллельно самой себе в направлении вектора до тех пор, пока она пересекается с точками пятиугольника *OABCD*. Очевидно, что последними точками, в которых передвигаемая линия пересекается с пятиугольником *OABCD*, могут быть только вершины пятиугольника. В нашем случае – это точка С.

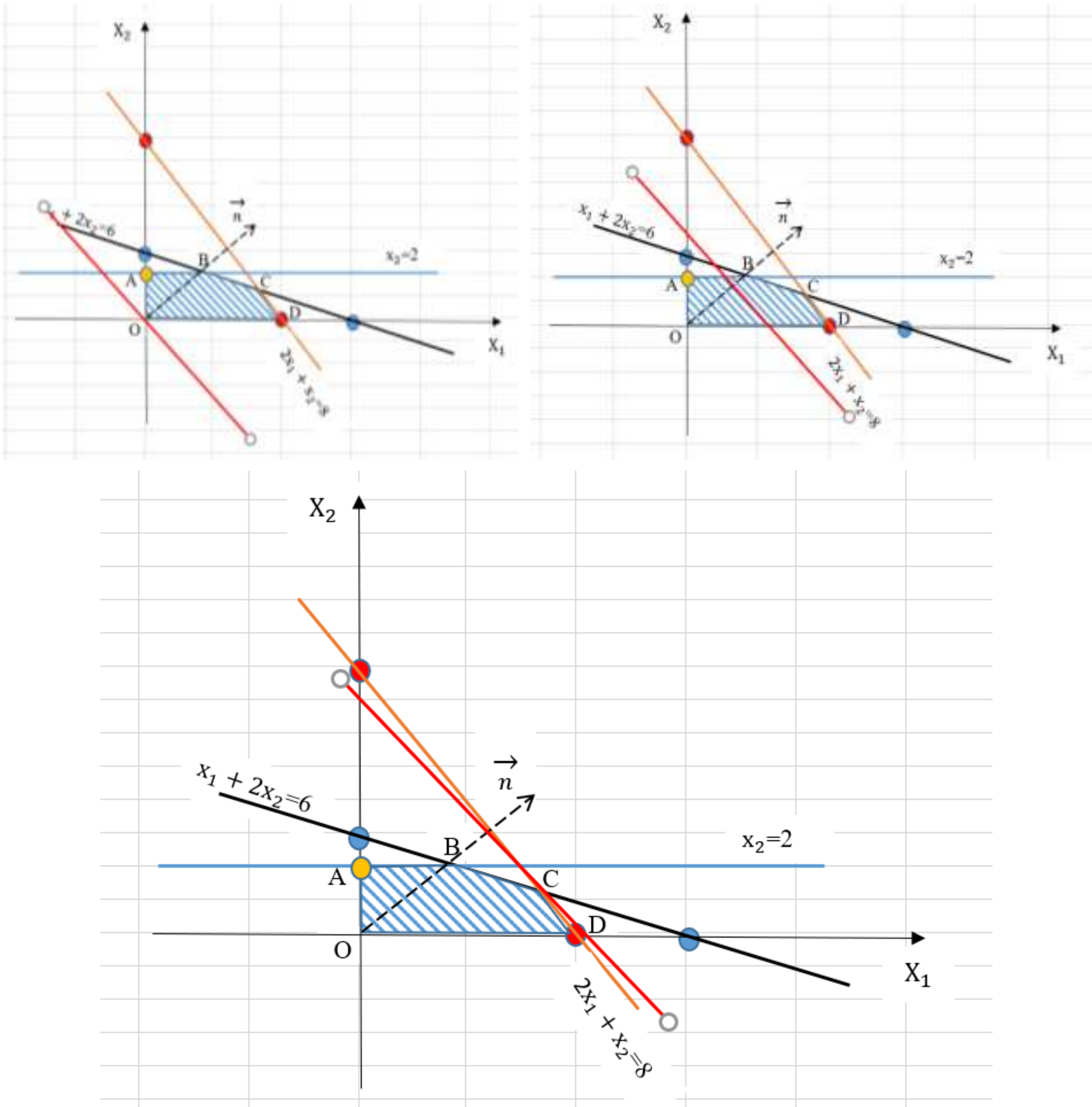


Рисунок 6 – Определение наиболее вероятного решения

Точка С есть точка пересечения прямых (1) и (2). Построим систему уравнений (1) и (2) и найдем ее решение.

$$\begin{cases} x_1 + 2x_2 = 6, \\ 2x_1 + x_2 = 8. \end{cases}$$

Решением будет пара $\left(\frac{10}{3}; \frac{4}{3}\right)$.

Подсчитаем теперь значение, которые принимает функция

$F\left(\frac{10}{3}; \frac{4}{3}\right) = \frac{28}{3} = 9{,}3(3)$.

Для проверки возьмем любую точку из множества, предположим (3;1).

$F(3;1) = 8$, $8 < 9{,}3$.

Ответ: функция принимает максимальное значение равное 9,3 при $x_1 = \frac{10}{3}$ и $x_2 = \frac{4}{3}$.

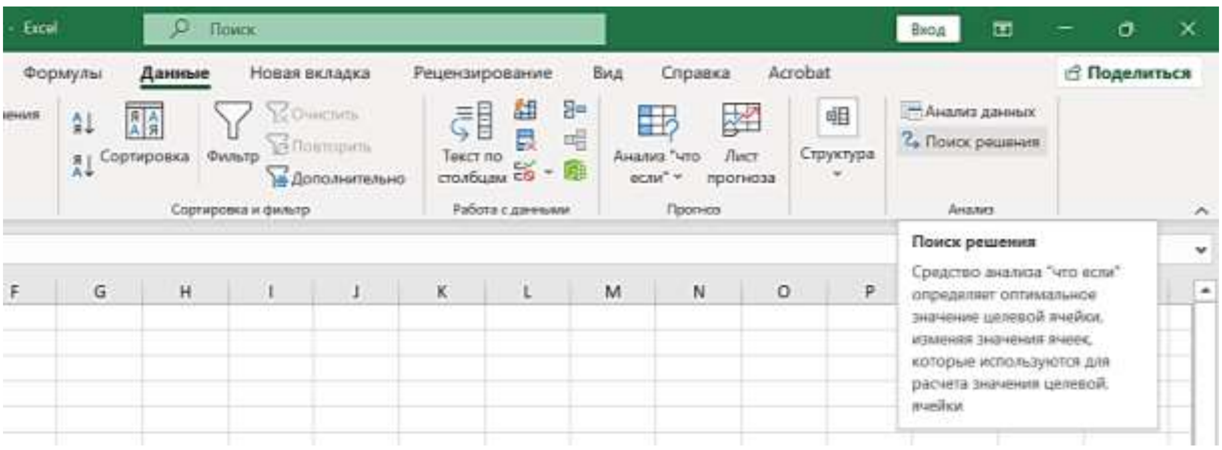
В рассмотренной задаче не было указано, что определяется количество продукции. Не следует забывать, что если речь идет о количестве, то должны быть ограничения на целочисленность значений.

Симплекс – метод

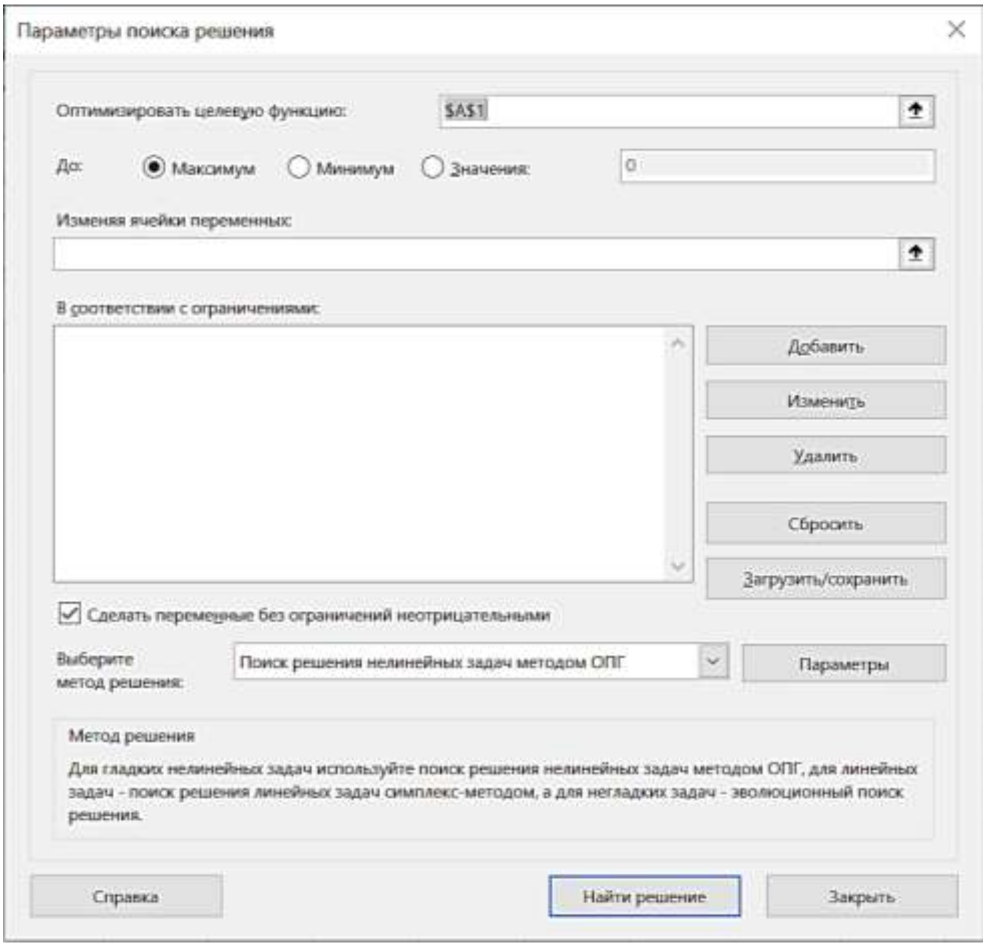
Симплекс-метод является универсальным методом решения задач линейного программирования с любым числом переменных и с любым числом ограничений. Тем не менее исходная форма задачи, к которой непосредственно применим симплекс-метод, должна иметь специальный вид.

В рамках дисциплины опустим математическую сторону вопроса и обратимся к обработке данных задачи с помощью программных средств, в частности MS Excel.

Для решения задач оптимизации в MS Excel необходима надстройка *«Поиск решения»* (Данные → Анализ → Поиск решения).



Инструмент «Поиск решения» предполагает работу с диалоговым окном



Поясним запрос в диалоговом окне.

1. *Оптимизировать целевую функцию* : запрос на адрес ячейки, в которой помещается Целевая функция;
2. *До*: необходимо указать направление оптимизации;
3. *Изменяя ячейки переменных*: указывается диапазон ячеек, содержащих переменные (значения переменных);
4. *В соответствии с ограничениями* : в указанное поле вводят диапазоны (ячейки) левой и правой частей множества ограничений.
5. *Выберите метод решения*: выбираем из раскрывающегося контекстного меню симплекс-метод.
6. После заполнения всех полей выбираем «Найти решение».

Рассмотрим, как работает схема на конкретном примере.

6. Для изготовления трех видов изделий А, В, С используется токарное, фрезерное, сварочное и шлифовальное оборудование. Затраты времени на обработку одного изделия для каждого из типов оборудования указаны в таблице. В ней же указан общий фонд рабочего времени каждого из типов используемого оборудования, а также прибыль от реализации одного изделия каждого вида. Требуется определить, сколько изделий и какого вида следует изготовить предприятию, чтобы прибыль от их реализации была максимальной. Составить математическую модель задачи.

Тип оборудования	***Затраты времени (станко – часы) на обработку одного изделия каждого вида***			***Общий фонд рабочего времени оборудования (часы)***
	A	***B***	***C***	
Фрезерное	2	4	5	120
Токарное	1	8	6	280
Сварочное	7	4	5	240
Шлифовальное	4	6	7	360
Прибыль (руб)	10	14	12	
Количество изделий	x_1	x_2	x_3	

Решение

Математическая модель задачи имеет следующий вид:

$$F(x_1, x_2, x_3) = 10x_1 + 14x_2 + 12x_3 \to max$$

$$\begin{cases} 2x_1 + 4x_2 + 5x_3 \le 120 \\ x_1 + 8x_2 + 6x_3 \le 280 \\ 7x_1 + 4x_2 + 5x_3 \le 240 \\ 4x_1 + 6x_2 + 7x_3 \le 360 \\ x_1 \ge 0 \\ x_2 \ge 0 \\ x_3 \ge 0 \\ x_1, x_2, x_3 \in Z. \end{cases}$$

Перенесем математическую модель в MS Excel. Воспользуемся инструментом «Поиск решения». Ввод данных в диалоговое окно разбивается на два этапа:

1. Указание целевой функции и переменных;
2. Добавление ограничений.

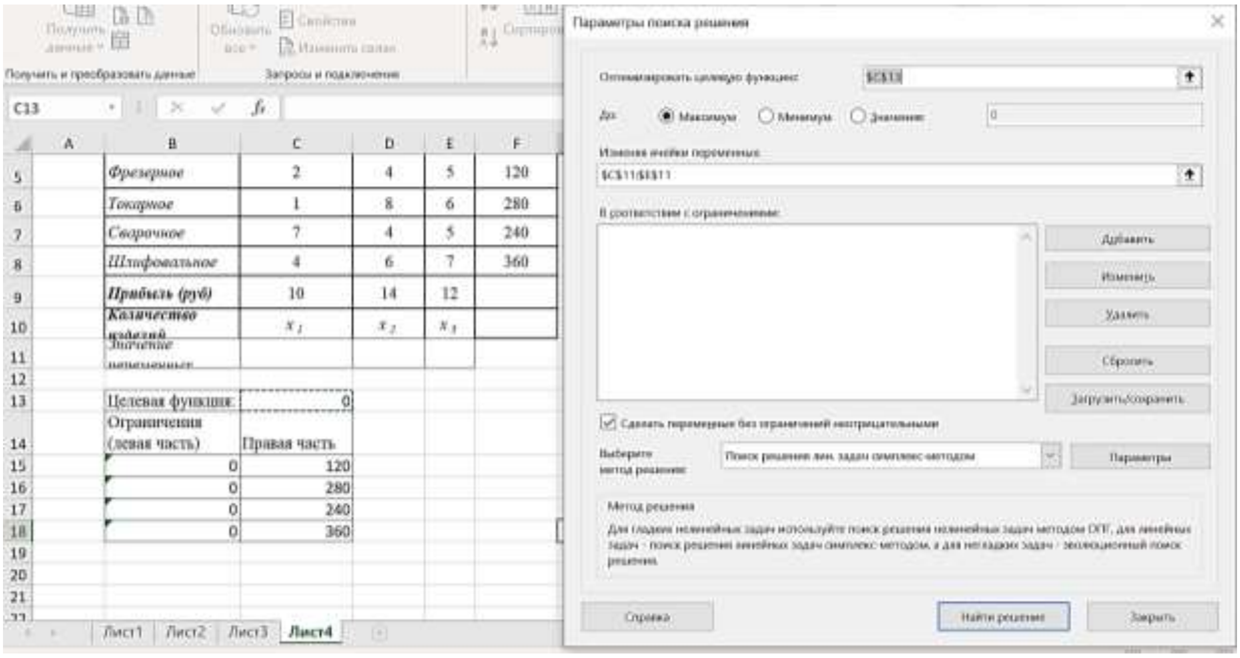


Рисунок 7 – Первый этап заполнения диалогового окна

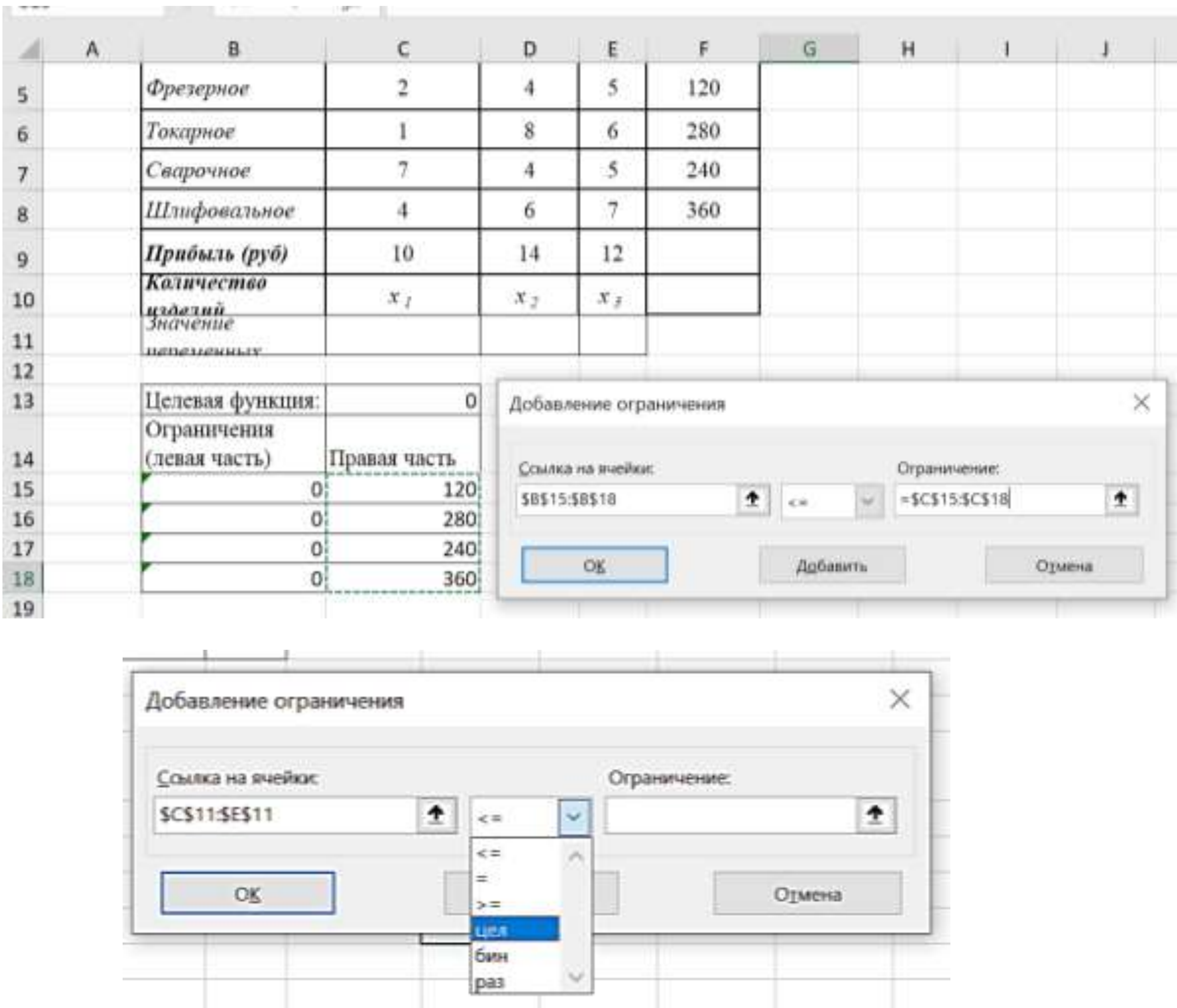


Рисунок 8 – Добавление ограничений

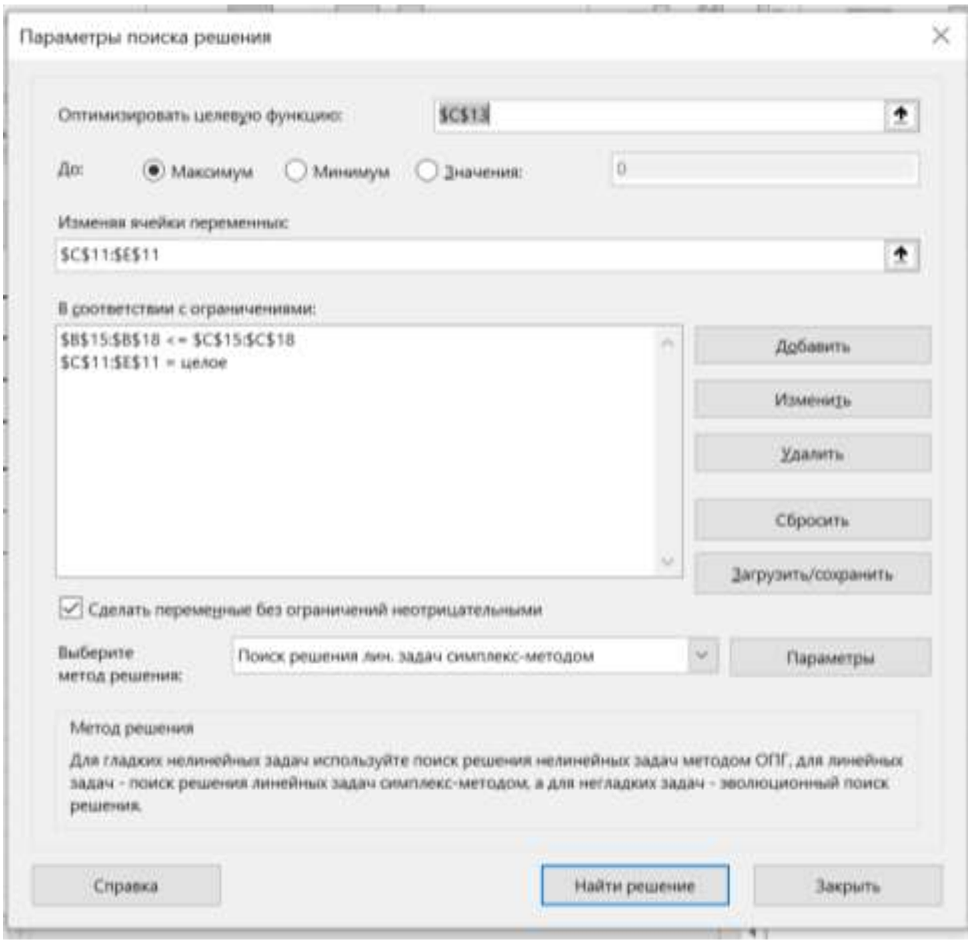


Рисунок 9 – Результат заполнения Диалогового окна

Запустив «Поиск решения» на выполнение, получим оптимальное решение задачи:

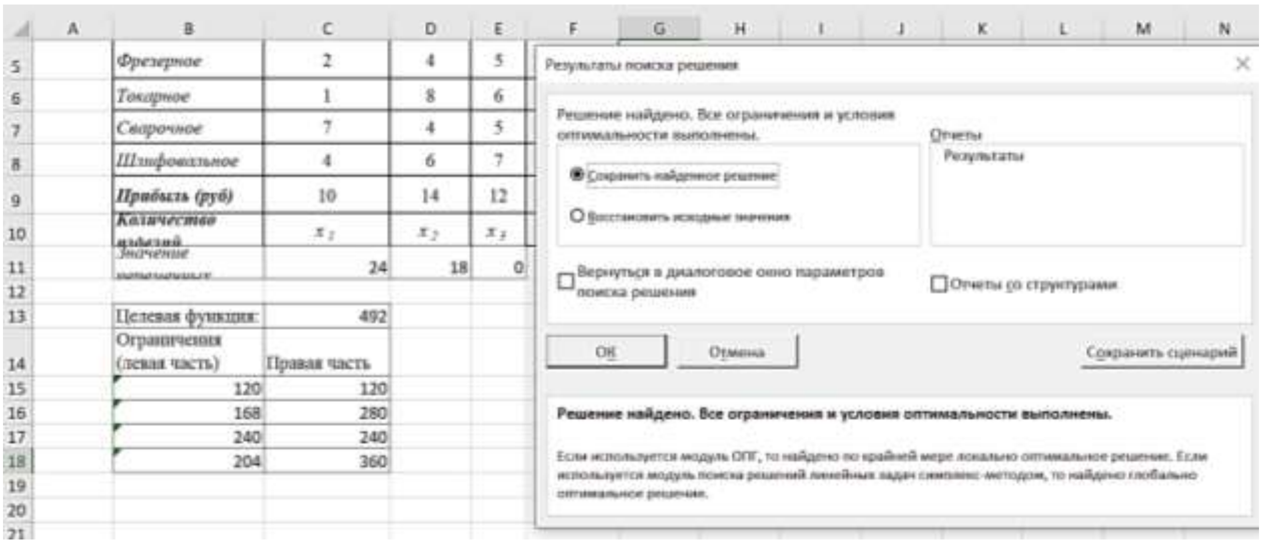


Рисунок 10 – Результаты поиска решений

Ответ: для получения максимальной прибыли (492 руб) необходимо изготовить 24 изделия вида А, 18 изделий вида В и 0 изделий вида С.

Проверим решение. Для этого рассмотрим иные целочисленные значения, удовлетворяющие ограничениям. Так как изделия вида В наиболее дорогие, то их количество оставим неизменным, но добавим изделий вида С.

Пусть $x_2 = 18, x_1 = 16, x_3 = 3$. В этом случае значение целевой функции составит 448 рублей. Следует заметить, что достаточно сложно подобрать целочисленные значения, которые удовлетворяли бы первому ограничению.

Транспортная задача

Под названием «Транспортная задача» объединяется широкий круг задач с единой математической моделью. Подобные задачи относятся к задачам динейного программирования и широко применяются как в теоретических разработках, так и в практике планирования раздичных бизнес – процессов. Кроме того, к задачам транспортного типа сводятся многие другие задачи линейного программирования - задачи о назначениях, сетевые, календарного планирования.

Постановка транспортной задачи

Некоторый однородный продукт, сосредоточенный у m поставщиков в количестве a_i $(i = \overline{1;m})$ единиц, необходимо доставить n потребителям в количестве b_j $(j = \overline{1;n})$ единиц. Известна стоимость c_{ij} перевозки единицы груза от поставщика i к потребителю j. Требуется составить такой план перевозок, при котором запросы всех потребителей полностью удовлетворены и суммарные затраты на перевозку всех грузов минимальны.

Исходные данные транспортной задачи записываются в виде таблицы:

	Потребители				
Поставщики		$\boldsymbol{b_1}$	$\boldsymbol{b_2}$	**…**	$\boldsymbol{b_n}$
	$\boldsymbol{a_1}$	c_{11}	c_{12}	**…**	c_{1n}
	$\boldsymbol{a_2}$	c_{21}	c_{22}	**…**	c_{2n}
	…	**…**	**…**	**…**	**…**
	$\boldsymbol{a_m}$	c_{m1}	c_{m2}	…	c_{mn}

Исходные данные могут быть представлены в виде вектора запасов поставщиков А = $(a_1; a_2; …; a_m)$, вектора запросов потребителей

B = $(b_1; b_2 ; \dots ; b_n)$, матрицы стоимостей C:

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1n} \\ c_{21} & c_{22} & \dots & c_{2n} \\ \text{-----} & \text{-----} & \text{-----} & \text{-----} \\ c_{m1} & c_{m2} & \dots & c_{mn} \end{pmatrix}$$

В транспортных задачах под поставщиками и потребителями понимаются различные промышленные предприятия, склады, магазины и т.д. Однородными считаются грузы, которые могут быть перевезены одним видом транспорта. Под стоимостью перевозок понимаются тарифы, расстояния, время, расход топлива и т.п.

Математическая модель транспортной задачи

Транспортная задача относится к двух индексным задачам линейного программирования. В результате решения задачи необходимо найти матрицу

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \text{-----} & \text{-----} & \text{-----} & \text{-----} \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix}$$

где x_{ij} количество единиц груза, запланированных к перевозке от поставщика i к потребителю j.

Так как $c_{ij} \cdot x_{ij}$ определяют затраты на перевозку груза, то суммарные затраты на перевозку всех грузов: $\sum_{i=1}^{m} \sum_{j=1}^{n} (c_{ij} \cdot x_{ij})$. По условию задачи требуется обеспечетить минимум суммарных затрат, следовательно, целевая функция задачи: $F(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} (c_{ij} \cdot x_{ij}) \to min$. Система ограничений задачи состоит из двух групп уравнений:

1. Состоит из m уравнений и описывает тот факт, что запасы всех m поставщиков вывозятся полностью и имеют вид: $\sum_{j=1}^{n} x_{ij} = a_i, \quad i = \overline{1, m}$.

2. Состоит из n уравнений и выражает требование удовлетворить запросы всех n потребителей полностью: $\sum_{i=1}^{m} x_{ij} = b_j,\ j = \overline{1,n}$.

Учитывая условия неотрицательности объемов перевозок, математическая модель имеет вид:

$$F(X) = \sum_{i=1}^{m} \sum_{j=1}^{n} c_{ij} x_{ij} \to min.$$

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i, \ \ i = \overline{1,m}, \\ \sum_{i=1}^{m} x_{ij} = b_j, \ \ j = \overline{1,n}, \\ x_{ij} \geq 0 \end{cases}$$

Матрицу перевозок *X*, удовлетворяющую ограничениям, называют *планом перевозок* транспортной задачи. План *X**, при котором целевая функция *F(X)* достигает минимума, называется *оптимальным*.

В рассматриваемой модели транспортной задачи предполагается, что суммарные запасы поставщиков равны суммарным запросам потребителей, $\sum_{i=1}^{m} a_i = \sum_{j=1}^{n} b_j$. Такая задача называется *задачей с правильным балансом*, а ее модель – *закрытой*. Если же это равенство не выполняется, то задача называется *задачей с неправильным балансом*, а ее модель – *открытой* .

7. Построить математическую модель транспортной задачи заданной таблицей

a_i \ b_j	20	30	40
40	3	5	7
50	4	6	10

Решение

Проверим соотношение суммарных запасов и суммарных запросов.
$\sum_{i=1}^{2} a_i = 40 + 50 = 90$; $\sum_{j=1}^{3} b_j = 20 + 30 + 40 = 90$; $90 = 90$, следовательно, рассматриваемая модель – закрытая. Введем переменные:

$$X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

Целевая функция имеет вид:

$$F(X) = 3x_{11} + 5x_{12} + 7x_{13} + 4x_{21} + 6x_{22} + 10x_{23} \to min$$

Система ограничений:

$$\begin{cases} x_{11} + x_{12} + x_{13} = 40, \\ x_{21} + x_{22} + x_{23} = 50, \\ x_{11} + x_{21} = 20, \\ x_{12} + x_{22} = 30, \\ x_{13} + x_{23} = 40, \ \ x_{ij} \geq 0, i = 1,2;\ j = 1,2,3. \end{cases}$$

Рассмотрим случай $\sum_{i=1}^{m} a_i \neq \sum_{j=1}^{n} b_j$, то есть модель открытая. Возможны две ситуации:

1. Суммарные запасы больше суммарного потребления $\sum_{i=1}^{m} a_i > \sum_{j=1}^{n} b_j$. Тогда система ограничений имеет вид:

$$\begin{cases} \sum_{j=1}^{n} x_{ij} \leq a_i, \ \ i = \overline{1,m}, \\ \sum_{i=1}^{m} x_{ij} = b_j, \ \ j = \overline{1,n}, \\ x_{ij} \geq 0 \end{cases}$$

Для перевода открытой модели к закрытой необходимо ввести фиктивного (n + *1*) потребителя, потребность которого определяется как:

$$b_{n+1} = \sum_{i=1}^{m} a_i - \sum_{j=1}^{n} b_j.$$

2. Суммарные запасы меньше суммарного потребления $\sum_{i=1}^{m} a_i < \sum_{j=1}^{n} b_j$. Тогда система ограничений имеет вид:

$$\begin{cases} \sum_{j=1}^{n} x_{ij} = a_i, \ \ i = \overline{1,m}, \\ \sum_{i=1}^{m} x_{ij} \leq b_j, \ \ j = \overline{1,n}, \\ x_{ij} \geq 0. \end{cases}$$

Для перевода открытой модели к закрытой необходимо ввести фиктивного (m + *1*) поставщика, запасы которого определяется как:

$$a_{m+1} = \sum_{j=1}^{n} b_j - \sum_{i=1}^{m} a_i.$$

Стоимость единицы перевозки груза (c_{ij}) для фиктивного потребителя (поставщика) полагают равной нулю, так как груз в обоих случаях не перевозится.

8. Построить математическую модель транспортной задачи заданной таблицей

a_i \ b_j	30	40	60
40	7	8	6
60	6	5	10
50	4	3	9

Решение

Проверяем модель: 40 + 60 + 50 = 150; 30 + 40 + 60 = 130; 150 > 130, таким образом модель – открытая. Рассматривается первая ситуация, следовательно, необходимо ввести фиктивного потребителя и его потребность составит: 150 – 130 = 20. Тогда имеем,

a_i \ b_j	30	40	60	20
40	7	8	6	0
60	6	5	10	0
50	4	3	9	0

Введем переменные: $X = \begin{pmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{pmatrix}$

Целевая функция имеет вид:

$F(X) = 7x_{11} + 8x_{12} + 6x_{13} + 6x_{21} + 5x_{22} + 10x_{23} + 4x_{31} + 3x_{32} + 9x_{33} \to min$

Система ограничений:

$$\begin{cases} x_{11} + x_{12} + x_{13} + x_{14} = 40, \\ x_{21} + x_{22} + x_{23} + x_{24} = 60, \\ x_{31} + x_{32} + x_{33} + x_{34} = 50, \\ x_{11} + x_{21} + x_{31} = 30, \\ x_{12} + x_{22} + x_{32} = 40, \\ x_{13} + x_{23} + x_{33} = 60, \\ x_{14} + x_{24} + x_{34} = 20, \\ x_{ij} \geq 0, i = \overline{1,3};\ j = \overline{1,4}. \end{cases}$$

Дальнейшее решение связано с MS Excel инструментом «Поиск решения».

СПИСОК ЛИТЕРАТУРЫ

1. Ю.П. Адлер, Е.В. Маркова, Ю.В. Грановский «Планирование эксперимента при поиске оптимальных условий» М., Наука, 1976 – 280 стр.

2. Зедгинидзе И.Г. Планирование эксперимента для исследования многокомпонентных систем. М., Наука, 1976 – 390 стр.

3. Иванова В. М., Калинина В. Н., Нешумова Л. А. и др. Математическая статистика, изд. 2, перераб. и доп. , М. Высш. Школа, 1981-371стр.

4. Монтгомери Д.К. «Планирование эксперимента и анализ данных» Пер с Англ. – Л.: Судостроение. 1980 -384стр.

5. А.И.Орлов «Непараметрические критерии согласия Колмогорова, Смирнова, омега-квадрат и ошибки при их применении». Научный журнал КубГАУ, №97(03), 2014 года Электронный адрес : http://ej.kubagro.ru/2014/03/pdf/47.pdf

6. Спиридонов А.А. Планирование эксперимента при исследовании технологических процессов. М. Машиностроение, 1981 – 184 стр.

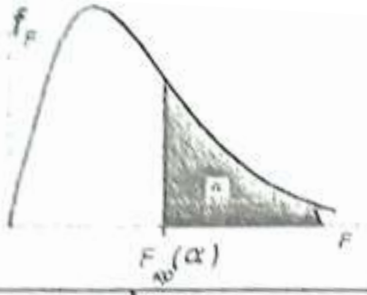
7. В. А. Штерензон «Моделирование технологических процессов»: конспект лекций. Екатеринбург: Изд-во Рос. гос. проф.-пед. ун-та, 2010. 66 с.

Приложение А

Критические точки распределения χ^2

Число степеней свободы, k	Уровень значимости α					
	0,01	0,025	0,05	0,95	0,975	0,89
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17,5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30,2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Приложение Б



Обратное F-распределение Фишера

$$F_{кр} = F_{кр}(\alpha,\ m_1,\ m_2)$$

Число степ.	Уровень значимости для правосторонней критической области $\alpha = 0{,}05$									
m2 / m1	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35

Число степ.	Уровень значимости для правосторонней критической области $\alpha = 0{,}01$									
m2 / m1	1	2	3	4	5	6	7	8	9	10
1	4052,2	4999,5	5403,4	5624,6	5763,6	5859,0	5928,4	5981,1	6022,5	6055,8
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37

Приложение В

Таблица значимых размахов Дункана

$r_{0,05}(p, f)$

f	p: 2	3	4	5	6	7	8	9	10	20	50	100
1	18,0	18,0	18,01	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0	18,0
2	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09	6,09
3	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50	4,50
4	3,93	4,01	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02	4,02
5	3,64	3,74	3,79	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83	3,83
6	3,46	3,58	3,64	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68	3,68
7	3,35	3,47	3,54	3,58	3,60	3,61	3,61	3,61	3,61	3,61	3,61	3,61
8	3,26	3,39	3,47	3,52	3,55	3,56	3,56	3,56	3,56	3,56	3,56	3,56
9	3,20	3,34	3,41	3,47	3,50	3,52	3,52	3,52	3,52	3,52	3,52	3,52
10	3,15	3,30	3,37	3,43	3,46	3,47	3,47	3,47	3,47	3,48	3,48	3,48
11	3,11	3,27	3,35	3,39	3,43	3,44	3,45	3,46	3,46	3,48	3,48	3,48
12	3,08	3,23	3,33	3,36	3,40	3,42	3,44	3,44	3,46	3,48	3,48	3,48
13	3,06	3,21	3,30	3,35	3,38	3,41	3,42	3,44	3,45	3,47	3,47	3,47
14	3,03	3,18	3,27	3,33	3,37	3,39	3,41	3,42	3,44	3,47	3,47	3,47
15	3,01	3,16	3,25	3,31	3,36	3,38	3,40	3,42	3,43	3,47	3,47	3,47
16	3,00	3,15	3,23	3,30	3,34	3,37	3,39	3,41	3,43	3,47	3,47	3,47
17	2,98	3,13	3,22	3,28	3,33	3,36	3,38	3,40	3,42	3,47	3,47	3,47
18	2,97	3,12	3,21	3,27	3,32	3,35	3,37	3,39	3,41	3,47	3,47	3,47
19	2,96	3,11	3,19	3,26	3,31	3,35	3,37	3,39	3,41	3,47	3,47	3,47
20	2,95	3,10	3,18	3,25	3,30	3,34	3,36	3,38	3,40	3,47	3,47	3,47
30	2,89	3,04	3,12	3,20	3,25	3,29	3,32	3,35	3,37	3,47	3,47	3,47
40	2,86	3,01	3,10	3,17	3,22	3,27	3,30	3,33	3,35	3,47	3,47	3,47
60	2,83	2,98	3,08	3,14	3,20	3,24	3,28	3,31	3,33	3,47	3,48	3,48
100	2,80	2,95	3,05	3,12	3,18	3,22	3,26	3,29	3,32	3,47	3,53	3,53
∞	2,77	2,92	3,02	3,09	3,15	3,19	3,23	3,26	3,29	3,47	3,61	3,67

$r_{0,01}(p, f)$

f*	p: 2	3	4	5	6	7	8	9	10	20	50	100
1	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0	90,0
2	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0	14,0
3	8,26	8,5	8,6	8,7	8,8	8,9	8,9	9,0	9,0	9,3	9,3	9,3
4	6,51	6,8	6,9	7,0	7,1	7,1	7,2	7,2	7,3	7,5	7,5	7,5
5	5,70	5,96	6,11	6,18	6,26	6,33	6,40	6,44	6,5	6,8	6,8	6,8
6	5,24	5,51	5,65	5,73	5,81	5,88	5,95	6,00	6,0	6,3	6,3	6,3
7	4,95	5,22	5,37	5,45	5,53	5,61	5,69	5,73	5,8	6,0	6,0	6,0
8	4,74	5,00	5,14	5,23	5,32	5,40	5,47	5,51	5,5	5,8	5,8	5,8
9	4,60	4,86	4,99	5,08	5,17	5,25	5,32	5,36	5,4	5,7	5,7	5,7
10	4,48	4,73	4,88	4,96	5,06	5,13	5,20	5,24	5,28	5,55	5,55	5,55
11	4,39	4,63	4,77	4,86	4,94	5,01	5,06	5,12	5,15	5,39	5,39	5,39
12	4,32	4,55	4,68	4,76	4,84	4,92	4,96	5,02	5,07	5,26	5,26	5,26
13	4,26	4,48	4,62	4,69	4,74	4,84	4,88	4,94	4,98	5,15	5,15	5;15
14	4,21	4,42	4,55	4,63	4,70	4,78	4,83	4,87	4,91	5,07	5,07	5,07
15	4,17	4,37	4,50	4,58	4,64	4,72	4,77	4,81	4,84	5,00	5,00	5,00
16	4,13	4,34	4,45	4,54	4,60	4,67	4,72	4,76	4,79	4,94	4,94	4,94
17	4,10	4,30	4,41	4,50	4,56	4,63	4,68	4,73	4,75	4,89	4,89	4,89
18	4,07	4,27	4,38	4,46	4,53	4,59	4,64	4,68	4,71	4,85	4,85	4,85
19	4,05	4,24	4,35	4,43	4,50	4,56	4,61	4,64	4,67	4,82	4,82	4,82
20	4,02	4,22	4,33	4,40	4,47	4,53	4,58	4,61	4,65	4,79	4,79	4,79
30	3,89	4,06	4,16	4,22	4,32	4,36	4,41	4,45	4,48	4,65	4,71	4,71
40	3,82	3,99	4,10	4,17	4,24	4,30	4,34	4,37	4,41	4,59	4,69	4,69
60	3,76	3,92	4,03	4,12	4,17	4,23	4,27	4,31	4,34	4,53	4,66	4,66
100	3,71	3,86	3,98	4,06	4,11	4,17	4,21	4,25	4,29	4,48	4,64	4,65
∞	3,64	3,80	3,90	3,98	4,04	4,09	4,14	4,17	4,20	4,41	4,60	4,68

* f — степени свободы.

Приложение Г

Таблица критических значений t-критерия Стьюдента

Число степеней свободы	Доверительные вероятности				
	0,9	0,95	0,98	0,99	0,999
1	6,31375	12,7062	31,821	63,6559	636,578
2	2,91999	4,30266	6,96455	9,924988	31,5998
3	2,35336	3,18245	4,54071	5,840848	12,9244
4	2,13185	2,77645	3,74694	4,60408	8,61008
5	2,01505	2,57058	3,36493	4,032117	6,8685
6	1,94318	2,44691	3,14267	3,707428	5,95872
7	1,89458	2,36462	2,99795	3,499481	5,40807
8	1,85955	2,30601	2,89647	3,355381	5,04137
9	1,83311	2,26216	2,82143	3,249843	4,78089
10	1,81246	2,22814	2,76377	3,169262	4,58676
11	1,79588	2,20099	2,71808	3,105815	4,43688
12	1,78229	2,17881	2,68099	3,054538	4,31784
13	1,77093	2,16037	2,6503	3,012283	4,22093
14	1,76131	2,14479	2,62449	2,976849	4,14031
15	1,75305	2,13145	2,60248	2,946726	4,07279
16	1,74588	2,1199	2,58349	2,920788	4,01487
17	1,73961	2,10982	2,56694	2,898232	3,96511
18	1,73406	2,10092	2,55238	2,878442	3,92174
19	1,72913	2,09302	2,53948	2,860943	3,88332
20	1,72472	2,08596	2,52798	2,845336	3,84956
21	1,72074	2,07961	2,51765	2,831366	3,8193
22	1,71714	2,07388	2,50832	2,818761	3,79223
23	1,71387	2,06865	2,49987	2,807337	3,76764
24	1,71088	2,0639	2,49216	2,796951	3,74537
25	1,70814	2,05954	2,4851	2,787438	3,72514
26	1,70562	2,05553	2,47863	2,778725	3,70666
27	1,70329	2,05183	2,47266	2,770685	3,68949
28	1,70113	2,04841	2,46714	2,763263	3,67392
29	1,69913	2,04523	2,46202	2,756387	3,65952
30	1,69726	2,04227	2,45726	2,749985	3,64598
60	1,67065	2,0003	2,39012	2,660272	3,46015
120	1,65765	1,97993	2,35783	2,617417	3,37342

Любимцева Ольга Львовна

Моделирование процессов

Учебное пособие

Редактор:
Н. В. Викулова

Подписано в печать формат 60х90 1/16. Бумага газетная. Печать трафаретная.
Уч. изд. л. 4,5. Усл. печ. л. 4,7. Тираж 60 экз. Заказ №

Федеральное государственное образовательное учреждение высшего образования
«Нижегородский государственный архитектурно-строительный университет»
603000, Нижний Новгород, ул. Ильинская, 65.
Полиграфический центр ННГАСУ, 603000, Нижний Новгород, ул. Ильинская, 65
http://www.nngasu.ru, rector@nngasu.ru