

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ  
федеральное государственное автономное образовательное учреждение  
высшего образования  
«Национальный исследовательский технологический университет «МИСИС»  
НОВОТРОИЦКИЙ ФИЛИАЛ

Кафедра металлургических технологий и оборудования

**Шаповалов А.Н.**

**Бирюкова О.Д.**

**Куницина Н.Г.**

## **ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА**

Методические указания  
по выполнению домашнего задания / контрольной работы  
для студентов направления подготовки 22.03.02 Металлургия  
Профиль – Обработка металлов давлением,  
всех форм обучения

Новотроицк, 2023

УДК 378  
ББК 74.58  
Ш 24

**Рецензенты:**

Доцент кафедры металлургических технологий и оборудования  
Новотроицкого филиала ФГАОУ ВО «Национальный исследовательский  
технологический университет «МИСИС», к.т.н., доцент  
**Е.В. Братковский**

Зав. кафедрой электроники и микроэлектроники  
ФГБОУ ВО «МГТУ им. Г.И. Носова», к.т.н., доцент  
**Д.Ю. Усатый**

Шаповалов А.Н. Планирование эксперимента: методические указания по выполнению домашнего задания / контрольной работы / А.Н. Шаповалов, О.Д. Бирюкова, Н.Г. Куница. – Новотроицк: НФ НИТУ «МИСИС», 2023. – 48 с.

Методические указания предназначены для закрепления материала при выполнении домашнего задания / контрольной работы, а также для организации самостоятельной работы студентов при изучении дисциплины «Планирование эксперимента». Изложены виды экспериментов и основные методы обработки экспериментальных данных, представлен пример выполнения задания.

Указания составлены в соответствии с требованиями образовательного стандарта высшего образования НИТУ «МИСИС» и предназначены для студентов направления подготовки 22.03.02 Metallurgy, обучающихся по профилю «Обработка металлов давлением».

*Рекомендовано Методическим советом НФ НИТУ «МИСИС»*

© ФГАОУ ВО «Национальный  
исследовательский технологический  
университет «МИСИС»,  
Новотроицкий филиал, 2023

## Содержание

Введение .....	4
1 Виды научных экспериментов и их особенности .....	5
2 Случайная величина и ее числовые характеристики.....	7
3 Элементарные методы первичной обработки экспериментальных данных....	10
4 Планирование, организация и обработка результатов активного.....	
эксперимента.....	17
4.1 Выбор параметров оптимизации и факторов .....	17
4.2 Планирование активного эксперимента .....	18
4.2.1 Методика составления матрицы планирования полного.....	
факторного эксперимента.....	19
4.2.2 Методика составления матрицы планирования полного и.....	
дробного факторного эксперимента.....	22
4.3 Проведение активного эксперимента и статистическая проверка	
результатов эксперимента .....	23
4.3.1 Проверка экспериментальных данных на наличие ошибок .....	23
4.3.2 Проверка воспроизводимости результатов экспериментов.....	24
4.4 Методика статистической обработки результатов активного.....	
эксперимента.....	25
4.4.1 Определение коэффициентов в уравнении регрессии .....	25
4.4.2 Методика определения значимости факторов по величине .....	
коэффициентов уравнения регрессии .....	27
4.4.3 Методика проверки адекватности и точности регрессионной.....	
модели.....	27
4.4.4 Использование дополнительных функций MS Excel .....	29
5 Обработка результатов пассивного эксперимента .....	31
5.1 Статистическая проверка результатов пассивного эксперимента .....	31
5.2 Статистическая обработка результатов пассивного.....	
эксперимента.....	32
5.2.1 Дисперсионный анализ .....	32
5.2.2 Корреляционный анализ.....	36
5.2.3 Регрессионный анализ .....	39
5.2.4 Использование дополнительных функций MS Excel .....	41
6 Варианты заданий и требования к оформлению работы .....	42
7 Пример выполнения работы.....	43
Библиографический список.....	47

## Введение

Как известно, эксперимент является основой познавательного процесса, а его результаты позволяют установить зависимости между изучаемыми параметрами. Получив результаты эксперимента, исследователь должен извлечь из них полезную информацию или, другими словами, провести обработку и анализ экспериментальных данных. От того, насколько корректно и правильно будут обработаны и представлены полученные экспериментальные данные, зависит их информативность, удобство использования, области их возможного применения. При неумелой обработке данных даже при очень большом их массиве не удастся добиться требуемой точности и выявить существующие закономерности.

В настоящее время процедура обработки экспериментальных данных достаточно хорошо формализована и исследователю необходимо только ее правильно использовать. Методические указания содержат в себе последовательность действий при планировании эксперимента и основы обработки экспериментальных данных с использованием методов статистического анализа (дисперсионный, корреляционный, регрессионный), и предназначено для самостоятельной работы студентов по дисциплине «Планирование эксперимента», в том числе выполнения домашнего задания / контрольной работы.

## 1 Виды научных экспериментов и их особенности

Теория предполагает, что эксперимент может быть пассивным и активным.

При пассивном эксперименте информация об исследуемом объекте накапливается путем пассивного наблюдения, то есть информацию получают в условиях обычного функционирования объекта. Активный эксперимент проводится с применением искусственного воздействия на объект по специально разработанному плану.

При пассивном эксперименте существуют только факторы в виде входных контролируемых, но неуправляемых переменных, и экспериментатор находится в положении пассивного наблюдателя. Задача планирования в этом случае сводится к оптимальной организации сбора информации и решению таких вопросов, как выбор количества и частоты измерений, выбор метода обработки результатов измерений.

Наиболее часто целью пассивного эксперимента является построение математической модели объекта. Примером пассивного эксперимента является измерение контролируемых параметров технологического процесса производства (агломерационного, доменного или сталеплавильного) с определением математической зависимости результатов процесса (производительности, параметров качества или расходных коэффициентов материалов) от изменяющихся параметров.

Активный эксперимент позволяет быстрее и эффективнее решать задачи исследования, но более сложен, требует больших материальных затрат и может помешать нормальному ходу технологического процесса. Тем не менее, учитывая преимущества активного эксперимента, тогда, когда это возможно, предпочтение следует отдавать ему.

При активном эксперименте факторы должны быть управляемыми и независимыми.

Активный эксперимент предполагает возможность воздействия на ход процесса и выбора в каждом опыте уровней факторов. При планировании активного эксперимента решается задача рационального выбора факторов, существенно влияющих на объект исследования, и определения соответствующего числа проводимых опытов. Увеличение числа включенных в рассмотрение факторов приводит к резкому возрастанию числа опытов, уменьшение – к существенному увеличению погрешности опыта. Фактор считается заданным только тогда, когда при его выборе указывается его область определения – совокупность значений, которые может принимать данный фактор. В эксперименте используется ограниченная часть области определения, задаваемая обычно в виде дискретного множества уровней. Выбранные факторы должны

быть однозначно управляемыми и операциональными, то есть поддающимися регулированию с поддержанием на заданном уровне в течение всего опыта при соблюдении последовательности необходимых для этого действий. Должна быть назначена также точность измерения факторов в выбранном диапазоне измерения.

Совокупности факторов должны отвечать требованиям совместимости и независимости. Соблюдение первого требования означает, что все комбинации факторов осуществимы и безопасны, второго - возможность установления фактора на любом уровне независимо от уровней других факторов.

При выборе между активным и пассивным экспериментов необходимо учитывать особенности проведения экспериментов на действующих агрегатах:

- недостаточная точность исходной информации и выходных показателей. Причины могут быть как субъективные, то есть плохая организация контроля, так и объективные – ограниченная точность измерительной аппаратуры – при промышленных объемах точность не удастся выдержать. В связи с недостаточной точностью данных исследователь вынужден выполнять большой объем дополнительной работы для проверки правильности данных, например, составления материального и теплового балансов плавки и др.;

- сложность и многофакторность исследуемых процессов. В связи с этим при изучении влияния одного фактора приходится фиксировать значения всех других факторов (параметров), что не всегда возможно. Так некоторые факторы просто не поддаются количественной оценке;

- возможности изменения факторов в широких пределах ограничены технологическим режимом ведения процесса, выход за рамки которого может привести к ухудшению технико-экономических показателей и аварийным ситуациям.

В связи с вышеуказанными особенностями исследователи вынужденно ограничивают свою деятельность пассивным экспериментом, который заключается в сборе данных о работе металлургического агрегата в рамках технологического режима и их статистическом анализе. Но в любом случае полученные данные необходимо правильно обработать.

## 2 Случайная величина и ее числовые характеристики

Результаты эксперимента представляют из себя случайные величины.

В теории вероятности для общей характеристики случайной величины используют параметры, называемые числовыми характеристиками. Наиболее часто применяются следующие: среднее арифметическое, математическое ожидание, дисперсия, среднеквадратическое отклонение.

Среднее арифметическое значений случайной величины определяется как частное от деления суммы всех значений случайной величины на количество измерений (опытных данных):

$$\bar{X} = \frac{1}{n} \cdot \sum_{i=1}^n X_i,$$

где  $n$  – количество опытных данных;

$X_i$  – результат  $i$ -го измерения случайной величины.

Математическое ожидание – число, равное произведению всех возможных значений случайной величины на вероятности получения этих значений

$$M_{(X)} = \sum_{i=1}^n P_i \cdot X_i,$$

где  $P_i$  – вероятность получения  $X_i$  значения случайной величины в  $i$ -ом измерении.

Математическое ожидание приближенно равно среднему арифметическому значений случайной величины, причем с увеличением количества измерений случайной величины среднее арифметическое стремится к математическому ожиданию.

Дисперсия – мера рассеяния (характеристика колеблемости) значений случайной величины относительно математического ожидания (оценивает рассеяние случайной величины около ее математического ожидания). Дисперсию определяют как среднее квадратов отклонений случайных величин от среднего арифметического

$$D_X = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2.$$

Среднеквадратическое отклонение (отдельного измерения) также является характеристикой колеблемости значений случайной величины от среднего.

Его определяют как корень квадратный из дисперсии

$$\sigma_X = \sqrt{D_X} = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

Основная задача экспериментального исследования случайных величин состоит в том, чтобы установить как распределяются экспериментальные точки на числовой оси, на плоскости или в пространстве.

Распределение случайной величины относительно его математического ожидания (среднего) описывается законом распределения случайной величины, который устанавливает связь между возможными значениями случайной величины и вероятностями их появления (P).

Уравнение кривой нормального распределения вероятностей случайной величины имеет вид

$$\varphi(X) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{1}{2\sigma^2} \cdot (X - \bar{X})^2}.$$

Для оценки симметрии распределения используют коэффициент асимметрии  $K_a$ , который должен находиться в пределах от  $-3$  до  $+3$ , при этом если  $K_a < 0$ , то асимметрия левосторонняя, если  $K_a = 0$ , то имеет место симметричное распределение, иначе – правосторонняя асимметрия:

$$K_a = (\bar{X} - M_o) / \sigma,$$

где  $M_o$  – мода случайной величины (значение  $x$ , частота появления которого наибольшая).

Крутость кривой распределения, характеризующая отклонение вершины фактического распределения от нормального оценивается величиной эксцесса:

$$e_k = \frac{\mu_4}{\sigma_x^4} - 3,$$

где  $\mu_4$  – центральный эмпирический момент четвертого порядка

$$\mu_4 = \sum (X - \bar{X})^4 \cdot m / \sum m,$$

где  $m$  – частота появления значения  $x$ ;

$\sum m$  – сумма всех частот.

Совокупность значений наблюдаемых случайных величин, полученных в



результате эксперимента, называется генеральной. Таким образом, генеральная совокупность представляет собой полный набор случайной величины  $X$ . На практике зачастую свойства генеральной совокупности оценивают по данным выборочной, которая представляет собой часть значений случайной величины  $X$  из генеральной совокупности. Выборочная совокупность может быть направленной или случайной, но в любом случае для распространения каких-то заключений, полученных по такой выборке на всю генеральную совокупность необходимо обеспечение случайности самой выборки, то есть ее представительности.

В случае если свойства генеральной совокупности оценивают по данным выборочной, для оценки рассеяния случайной величины относительно ее математического ожидания применяют исправленные характеристики дисперсии и среднеквадратического отклонения:

- «исправленная» дисперсия ( $D^*$ ) – оценивает рассеяние случайной величины из выборочной совокупности относительно ее математического ожидания:

$$D^* = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2.$$

- «исправленное» среднеквадратическое отклонение

$$\sigma_X^* = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}.$$

На практике «исправленными» характеристиками пользуются, если объем выборки менее 30.

Кроме рассмотренных, для характеристики случайной величины используются следующие параметры:

- вариационный размах  $R$  – амплитуда колебания ( $X_{\max} - X_{\min}$ );

- коэффициент вариации ( $V$ ) – характеризует колеблемость случайной величины относительно среднего и рассчитывается по среднеквадратичному отклонению:  $V = 100 \cdot \sigma / X$ .

- уровень значимости ( $p$ ) – это вероятность того, что признанное неверным и отброшенное утверждение в действительности может оказаться правильным (вероятность ошибки первого рода);

- доверительная вероятность (коэффициент надежности) ( $P$ ) – вероятность того, что признанное верным утверждение действительно неверно,  $P=1-p$  (вероятность ошибки второго рода) и др.

### 3 Элементарные методы первичной обработки экспериментальных данных

Первичная обработка экспериментальных данных заключается в их систематизации (помещение данных в таблицы или на отдельные карточки), определении диапазона изменения исследуемых параметров (минимального и максимального значений) средних значений величин и их частотных характеристик (тип распределения), построении графиков (для наглядности), подборе эмпирических формул (аппроксимация), оценки точности экспериментальных данных и исключение грубых ошибок (определение адекватности).

Методика определения стандартных числовых характеристик результатов эксперимента (среднее арифметическое, математическое ожидание, дисперсия, среднеквадратическое отклонение) изложена в п. 2 настоящего пособия.

Оценку точности результатов эксперимента (проверка на наличие ошибок) можно проводить по правилу трех сигм (среднеквадратическое отклонение), согласно которому при нормальном распределении вероятность непопадания случайной ошибки в пределы  $\pm 3\sigma$  является ничтожно малой. При использовании такого подхода считают, что наибольшая возможная ошибка не превышает некоторого числа  $\Delta$

$$\Delta = 3 \cdot \sigma^* = 3 \cdot \sqrt{\frac{\sum (X_i - \bar{X})^2}{n-1}},$$

где  $\sigma^*$  – среднеквадратическая ошибка отдельных измерений выборочной совокупности (в случае расчета по генеральной – используется не скорректированное значение ошибки).

Все значения, отклоняющиеся от среднего на величину, большую трех сигм, исключаются как содержащие грубую ошибку.

В качестве других показателей точности, кроме  $\Delta$ , используют также вероятную ошибку ( $\rho$ ), среднюю абсолютную ошибку ( $\nu$ ), меру точности ( $H$ ).

Вероятная ошибка – это величина, которая при большом числе наблюдений делит область распределения ошибок на две равные части, то есть с одинаковой вероятностью половина отклонений  $|X_i - \bar{X}|$  будет меньше вероятной ошибки, а половина больше. Вероятная ошибка характеризует точность отдельных измерений и показывает насколько тесно ошибки концентрируются возле нуля. В общем случае вероятная ошибка определяется по уравнению

$$\rho = \frac{0,477}{h} = 0,675 \cdot \sigma^*,$$

где  $\sigma$  – среднеквадратическое отклонение значений отдельных измерений;

$h$  – мера точности отдельных измерений  $h = 1/\sigma \cdot \sqrt{2}$ .

Средняя абсолютная ошибка отдельного измерения определяется из выражения

$$\vartheta = \sigma^* \cdot \frac{2}{\sqrt{2 \cdot \pi}} = 0,7979 \cdot \sigma^*.$$

Вычисления среднего значения измеряемой величины и ошибок отдельных измерений не всегда является достаточным при первичной обработке результатов эксперимента. Необходимо также оценить точность полученных результатов: найти меру точности, среднеквадратическую, вероятную и наибольшую вероятную ошибку среднего арифметического.

Мера точности ( $H$ ) среднего арифметического определяется из выражения

$$H = h \cdot \sqrt{n} = \sqrt{\frac{n \cdot (n-1)}{2 \sum (\bar{X} - X_i)}},$$

где  $h$  – мера точности отдельных измерений.

Средняя квадратичная ошибка среднего арифметического ( $\sigma_o$ ) определяется из выражения

$$\sigma_o = \sqrt{\frac{\sum (X_i - \bar{X})^2}{n \cdot (n-1)}} = \frac{\sigma^*}{\sqrt{n}},$$

где  $\sigma^*$  – среднеквадратическая ошибка отдельных измерений выборочной совокупности (в случае расчета по генеральной – используется не скорректированное значение ошибки).

Вероятную ошибку среднего арифметического ( $\rho_o$ )

$$\rho_o = 0,675 \cdot \sigma_o$$

Наибольшую вероятную ошибку среднего арифметического ( $\Delta_o$ ) определяют из равенства

$$\Delta_o = 3 \cdot \sigma_o.$$

Кроме оценки точности результатов эксперимента и отсеивания грубых ошибок, в объем первичной обработки экспериментальных данных входит ра-

бота по представлению их в более наглядном виде: составление таблиц распределения, построение частотных графиков, диаграмм или гистограмм.

Этого можно добиться следующим образом:

- опытные данные разбивают на несколько групп;
- все величины в группе представляют одной величиной – центром интервала, который называется признаком группирования;
- для сравнения групп интервалы делают одинаковыми по размеру.

Порядок распределения данных по группам следующий:

- определяется интервал группы, в который будут включаться отдельные величины;
- устанавливается расположение интервалов, то есть границы групп или их центры;
- определяется общая шкала всех интервалов и в них распределяются величины, полученные в процессе наблюдения;
- составляется таблица, которая позволяет наглядно видеть все величины, входящие в группы;

Прерывные (дискретные) и непрерывные величины делятся на группы по-разному.

Если случайная величина прерывна и размах варьирования ( $R$ ) меньше единицы измерения, то интервал группирования принимают равным единице исследуемой случайной переменной величины. В качестве примера ниже приведено исследование производительности (форм/ч) формовочной машины в течение произвольно выбранных 45 рабочих смен.

2	7	5	2	4	5	4	7	3
4	6	6	3	5	5	5	6	6
3	2	4	6	5	5	5	6	6
6	3	4	6	6	6	6	6	3
4	5	3	4	4	6	5	7	7

Максимальное значение наблюдаемой величины (количество форм в час):  $X_{\max} = 17$ , минимальное  $X_{\min} = 12$ . Разность между ними или размах варьирования  $R = X_{\max} - X_{\min} = 17 - 12 = 5$ . Эта разность меньше единицы измерения, поэтому интервал группирования можно принять равным единице, а группы составить из отдельных значений изучаемой переменной величины. После этого приступают к распределению отдельных величин по интервалам (таблица 1).

Таблица 1 – Распределение производительности формовочной машины

Признак группирования	Частота (количество случаев), $n_i$	Относительная частота, $f_i$	Кумулятивная абсолютная частота	Кумулятивная относительная частота
12	3	0,067	3	0,067
13	6	0,133	9	0,200
14	8	0,178	17	0,378
15	10	0,222	27	0,600
16	14	0,311	41	0,911
17	4	0,089	45	1,000
Сумма	45	1,000		

В столбец 1 заносят признаки группирования, то есть отдельные значения производительности формовочной машины по мере ее возрастания. В столбец 2 – число значений, соответствующих данной производительности. Это будет количество элементов в каждом интервале группирования или частота наблюдаемого явления. Все значения суммируют и получают общую сумму наблюдений – 45.

В результате такой группировки значений изучаемой величины (производительности формовочной машины) получают наглядное распределение производительности формовочной машины.

Для сравнения распределения частоты выпадения той или иной группы данных необходимо частоты каждой группы разделить на общее количество наблюдений, то есть определить относительную частоту получения  $i$ -ой производительности формовочной машины

$$f_i = \frac{n_i}{N}.$$

Сумма относительных частот должна быть всегда равна единице. Значения частоты могут быть выражены в процентах. При этом они покажут, сколько процентов от всей совокупности данных содержится в каждой группе. Кроме этой частоты, определяют обычно кумулятивную частоту (абсолютную и относительную), путем суммирования отдельных значений групповой частоты. Кумулятивная частота показывает, какое количество значений или сколько процентов от всей совокупности находится в интервале от минимального уровня до рассматриваемого в настоящее время.

Если изучаемая переменная величина является непрерывной, то необходимо выбрать интервал группирования. Интервал должен быть выбран таким, чтобы общее число групп было не менее 5 и не более 15. При выборе числа групп необходимо чтобы величина совокупности данных охватывалась всеми

группами. В качестве примера проведены результаты исследования частотного распределения содержания углерода при анализе плавков стали 35Л для производства отливок.

Для исследования выбрали 56 плавков, проведенных в течение 1 года.

0,35	0,36	0,30	0,34	0,33	0,35	0,37	0,38
0,35	0,35	0,35	0,35	0,35	0,36	0,37	0,35
0,36	0,30	0,34	0,35	0,38	0,35	0,35	0,46
0,34	0,31	0,33	0,36	0,38	0,41	0,36	0,35
0,32	0,31	0,34	0,37	0,40	0,41	0,37	0,36
0,35	0,32	0,33	0,37	0,39	0,40	0,36	0,35
0,35	0,32	0,35	0,42	0,39	0,38	0,35	0,42

При минимальном и максимальном содержании углерода в изучаемых плавках 0,3 и 0,46 %, соответственно, для определения приблизительного значения интервала группирования разделим разность между этими значениями на количество групп (например, на 8). Принимаем величину интервала группирования равной 0,02 %.

$$\frac{X_{\max} - X_{\min}}{8} = \frac{0,46 - 0,30}{8} = 0,02.$$

После этого приступаем к группированию значений содержания углерода в плавках стали (таблица 2).

Таблица 2 - Распределение содержания углерода в плавках стали 35Л

Граница групп	Середина интервала	Частота (количество случаев), $n_i$	Относительная частота, $f_i$	Кумулятивная абсолютная частота	Кумулятивная относительная частота
0,30-0,32	0,31	4	0,071	4	0,071
0,32-0,34	0,33	6	0,107	10	0,178
0,34-0,36	0,35	22	0,393	32	0,571
0,36-0,38	0,37	11	0,196	43	0,767
0,38-0,40	0,39	6	0,107	49	0,874
0,40-0,42	0,41	4	0,071	53	0,945
0,42-0,44	0,43	2	0,036	55	0,981
0,44-0,46	0,45	1	0,018	56	1,000
Сумма		56	1,000		

Из опытных данных выбираем те значения, которые входят в ту или иную группу (столбец 1) и записываем их количество в столбец 3. Это есть частота появления того или иного значения в общей сумме. Возможны затруднения при включении значений, совпадающих с границами групп. Такие величины можно

включать в группы одним из 3-х способов:

- половину спорных значений включают в низшую группу, а половину – в высшую;

- спорную величину, совпадающую с границей групп, вписывают в низшую группу;

- спорную величину, совпадающую с границей групп, вписывают в высшую группу.

Таблица 2 составлена по последнему способу. Относительная частота, а также кумулятивная относительная и абсолютная частоты определяются таким же образом, как и при составлении таблицы для дискретных величин.

Кроме составления таблиц, распределение частоты можно изобразить графически. При построении графиков по оси абсцисс откладывают интервалы группирования, а по оси ординат – абсолютные или относительные частоты. Вид графика зависит от характера изучаемой совокупности. Рассмотрим применяемые виды графиков.

Точечный график применяется для описания распределения прерывной переменной и заключается в нанесении на график в виде точек всех значений переменной. На рисунке 1 показан график распределения производительности формовочной машины (таблица 1).

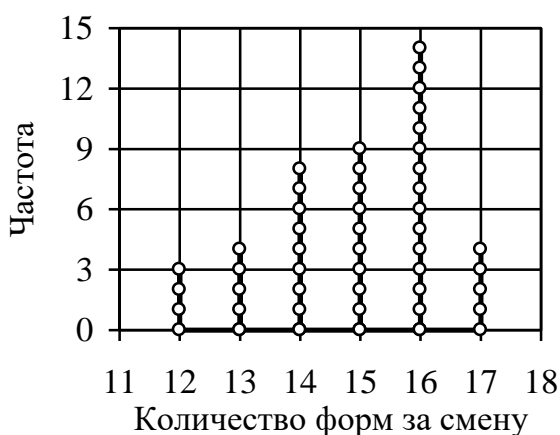


Рисунок 1 – Точечный график распределения

Гистограмма применяется для изображения распределения частоты непрерывной переменной. Здесь частота распределения изображается прямоугольником, основание которого равно ширине интервала, а высота соответствует частоте. На рисунке 2 приведена гистограмма распределения содержания углерода в плавках стали 35Л (таблица 2).

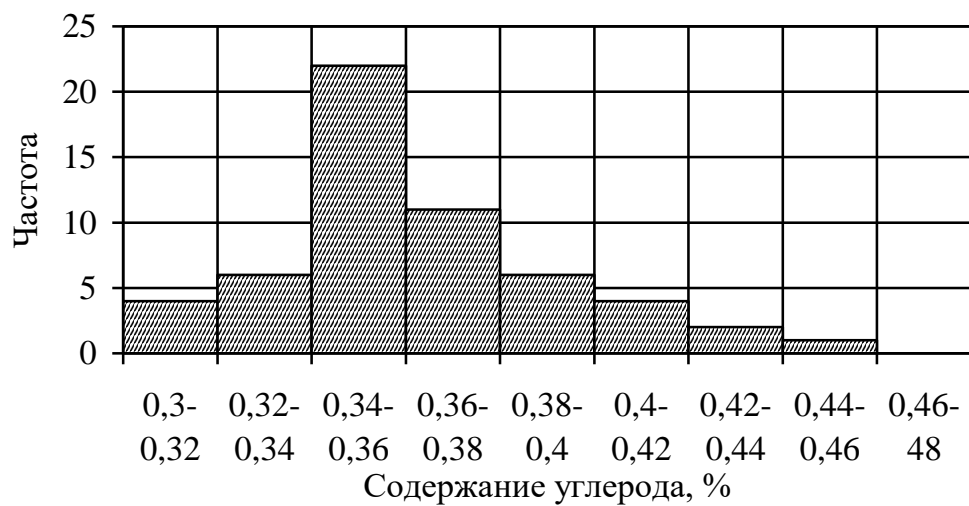


Рисунок 2 - Гистограмма распределения

Таким образом, построенные графики позволяют наглядно представить и исследовать распределение частоты исследуемой переменной.



## **4 Планирование, организация и обработка результатов активного эксперимента**

Ниже приведены основные этапы планирования, реализации и обработки результатов активного эксперимента.

### **4.1 Выбор параметров оптимизации и факторов**

Параметр оптимизации – количественная характеристика процесса, наиболее полно отражающая существо или эффективность его, обладающая однозначностью, статистической эффективностью, простотой и ясностью физического смысла, ограниченностью области определения.

В зависимости от характера металлургического процесса и производимой продукции, в качестве параметра оптимизации можно выбирать удельные расходы компонентов шихты или критерии качества продукции.

Можно выбирать несколько параметров оптимизации, однако при этом существенно увеличивается объем работ. В настоящих рыночных отношениях в качестве параметра оптимизации может выступать себестоимость продукции или уровень прибыли.

Фактор – независимая переменная величина, влияющая на процесс. Факторы бывают количественными и качественными. Первые можно выразить числом, вторые – нельзя. Качественные факторы неудобны, требуют увеличения числа опытов. Для использования качественных факторов при проведении машинной обработки результатов эксперимента необходимо заменить их количественными факторами.

К факторам предъявляются определенные требования:

- независимость друг от друга, то есть отсутствие прямой связи;
- совместимость факторов – комбинация нескольких факторов не должна вести к нарушению технологии;
- возможность измерения и поддержания на определенном уровне возможность измерения и поддержания на нужном уровне с требуемой точностью.

После выбора факторов требуется указать области их определения, которые определяются техническими возможностями эксперимента и технологической целесообразностью.

В активном эксперименте факторы могут принимать только фиксированные значения, которые называют уровнями фактора. Количество принимаемых уровней факторов зависит от выбранной структуры факторной модели и принятого плана эксперимента. Обычно активный эксперимент проводят при варьировании факторов на двух или трех уровнях. Уровни факторов представляют в кодированном виде: верхний уровень фактора обозначают «+1», нижний «-1», а

основной «0». За основной уровень фактора обычно принимается его среднее значение в исследуемом процессе.

Основной уровень можно рассматривать как исходную точку для построения плана эксперимента. Основной уровень называют также нулевым и при составлении плана эксперимента симметрично относительно него выбирают экспериментальные точки (значения факторов).

После выбора нулевого уровня необходимо установить интервалы варьирования факторов. Эта операция состоит в выборе для каждого фактора двух уровней, на которых он будет варьироваться в эксперименте.

Интервалом варьирования называют некоторое число (свое для каждого фактора), прибавление которого к основному уровню дает верхний уровень фактора, а вычитание – нижний. При этом, выбор минимального и максимального уровней (интервала варьирования фактора) фактора производится исходя из возможностей модели и технологической целесообразности.

Обычно интервал варьирования составляет до 20 % от нулевого значения фактора.

## 4.2 Планирование активного эксперимента

Цель планирования эксперимента состоит в раздельном или одновременном решении задач двух типов. В первом случае исследователю необходимо решить задачу оптимизации, то есть определить такие значения переменных факторов ( $X$ ) для которых величина параметра оптимизации ( $Y$ ) была бы минимальной или максимальной (в зависимости от вида параметра). При решении задач второго типа исследователя интересует математическое описание влияния каждого фактора на параметр оптимизации.

В реальных исследованиях решаются задачи обоих типов, то есть сначала определяют общий вид зависимости параметра оптимизации от переменных факторов, а затем выявляют оптимальные условия.

Важными этапами, при планировании эксперимента, являются выборы математического описания отклика параметра оптимизации на воздействие факторов, основного уровня для каждого фактора и интервалов его варьирования.

Математическое описание отклика параметра оптимизации на воздействие необходимо для нахождения кратчайшего пути достижения искомым значений переменных факторов. Часто вид функции параметра оптимизации неизвестен, поэтому его аппроксимируют полиномом и ограничиваются линейными членами и взаимодействиями первого порядка:

$$Y = b_0 + \sum b_i \cdot X_i + \sum b_{ij} \cdot X_{ij},$$

где  $b$  – выборочные коэффициенты регрессии;  
 $Y$  – целевая функция.

Все опыты с комбинацией факторов записываются в виде матрицы планирования эксперимента, которую заполняют путем чередования знаков (в уровнях факторов): в первом столбце знаки меняются поочередно, во втором – через 2, в третьем – через 4, в четвертом – 8 и т.д. по степеням двойки. Эксперимент, в котором реализуются все возможные сочетания уровней факторов, называется полярным факторным экспериментом (ПФЭ). При проведении эксперимента с неполным сочетанием возможных уровней факторов эксперимент называют дробным факторным экспериментом (ДФЭ).

#### 4.2.1 Методика составления матрицы планирования полного факторного эксперимента

Эксперимент, в котором реализуются все возможные сочетания уровней факторов, называется полным (полярным) факторным экспериментом (ПФЭ). Все опыты с комбинацией факторов записываются в виде матрицы планирования эксперимента, которую заполняют путем чередования знаков (в уровнях факторов): в первом столбце знаки меняются поочередно, во втором – через 2, в третьем – через 4, в четвертом – 8 и т.д. по степеням двойки.

Планирование эксперимента для получения линейной модели основано на варьировании факторов на двух уровнях – верхнем и нижнем.

Если число факторов известно, то можно определить число опытов, необходимое для реализации всех возможных сочетаний уровней факторов:

$$N=2^k,$$

где  $N$  – число опытов;  
 $k$  – число факторов;  
2 – число уровней фактора.

Матрица планирования ПФЭ типа  $2^2$  представлена в таблице 3

Таблица 3 - Матрица планирования ПФЭ типа  $2^2$

№ опыта	Факторы		Отклик $Y$
	$X_1$	$X_2$	
1	-1	-1	$Y_1$
2	+1	-1	$Y_2$
3	-1	+1	$Y_3$
4	+1	+1	$Y_4$
$\Sigma$	0	0	

В таблице 4 приведена матрица планирования полного факторного эксперимента типа  $2^3$ , количество опытов в котором составляет 8 шт.

Таблица 4 – Матрица планирования ПФЭ типа  $2^3$

№ опыта	$X_1$	$X_2$	$X_3$	Y
1	-1	-1	-1	Y1
2	+1	-1	-1	Y2
3	-1	+1	-1	Y3
4	+1	+1	-1	Y4
5	-1	-1	+1	Y5
6	+1	-1	+1	Y6
7	-1	+1	+1	Y7
8	+1	+1	+1	Y8
Сумма	0	0	0	

Результатом проведения ПФЭ  $2^2$  (таблица 3) является модель вида:  $Y=b_0+b_1 \cdot X_1+b_2 \cdot X_2$ , а для ПФЭ  $2^3$  (таблица 4) – модель вида:  $Y=b_0+b_1 \cdot X_1+b_2 \cdot X_2+b_3 \cdot X_3$ .

Приведенные модели являются линейными, но фактически в выбранных интервалах варьирования факторов процесс может изменяться и не по линейному закону. Один из часто встречающихся видов нелинейности связан с тем, что влияние одного фактора зависит от уровня, на котором находится другой фактор, то есть имеет место эффект взаимодействия двух факторов.

Для учета взаимного влияния исследуемых факторов на параметр оптимизации в матрицу планирования добавляют столбец эффекта взаимодействия интересующих факторов. Рассмотрим составление матрицы планирования ПФЭ  $2^2$  с эффектом взаимодействия (таблица 5). Для того, чтобы получить столбец эффекта взаимодействия факторов  $X_1$  и  $X_2$ , необходимо перемножить построчно столбцы, принадлежащие факторам  $X_1$  и  $X_2$ .

Таблица 5 - Матрица планирования ПФЭ  $2^2$  с эффектом взаимодействия

№ опыта	$X_1$	$X_2$	$X_1 X_2$	Y
1	-1	-1	+1	Y1
2	+1	-1	-1	Y2
3	-1	+1	-1	Y3
4	+1	+1	+1	Y4
	0	0	0	

В этом случае математическая модель будет в виде:

$$Y=b_0+b_1X_1+b_2X_2+b_{12} X_1X_2.$$

При этом следует учесть, что опыты проводятся только по столбцам  $X_1, X_2, \dots X_n$ . А столбцы взаимодействий факторов служат только для расчета. С ростом числа факторов число возможных взаимодействий быстро растет.

В таблице 6 приведена матрица планирования ПФЭ  $2^3$  с эффектом взаимодействия. Для того, чтобы получить столбец эффекта взаимодействия факторов  $X_1$  и  $X_2$ , необходимо перемножить построчно столбцы, принадлежащие факторам  $X_1$  и  $X_2$ , и т.д.

При этом следует учесть, что опыты проводятся только по столбцам  $X_1, X_2, \dots X_n$ , а столбцы взаимодействий факторов служат только для расчета. Столбец  $X_1X_2X_3$  получается перемножением всех трех столбцов. Получаемая модель имеет вид:  $Y=b_0+b_1 \cdot X_1+b_2 \cdot X_2+b_3 \cdot X_3+b_{12} \cdot X_1 \cdot X_2+b_{13} \cdot X_1 \cdot X_3+b_{23} \cdot X_2 \cdot X_3+b_{123} \cdot X_1 \cdot X_2 \cdot X_3$ .

Таблица 6 - Матрица планирования полного факторного эксперимента  $2^3$  с эффектами взаимодействия

№ опыта	$X_1$	$X_2$	$X_3$	$X_1X_2$	$X_1X_3$	$X_2X_3$	$X_1X_2X_3$	Y
1	-1	-1	-1	+1	+1	+1	-1	Y1
2	+1	-1	-1	-1	-1	+1	+1	Y2
3	-1	+1	-1	-1	+1	-1	+1	Y3
4	+1	+1	-1	+1	-1	-1	-1	Y4
5	-1	-1	+1	+1	-1	-1	+1	Y5
6	+1	-1	+1	-1	+1	-1	-1	Y6
7	-1	+1	+1	-1	-1	+1	-1	Y7
8	+1	+1	+1	+1	+1	+1	+1	Y8
Сумма	0	0	0	0	0	0	0	

Эффект взаимодействия двух факторов  $X_1X_2$  называется эффектом взаимодействия первого порядка, трех факторов – второго порядка и т.д.

Количество опытов в ПФЭ значительно превосходит число определяемых коэффициентов линейной модели и растет с увеличением числа факторов (в геометрической прогрессии) по уравнению  $N=(2-3)^k$ , где 2-3 – число уровней изменения фактора. Так, например, число опытов для ПФЭ типа  $2^3$  равно 8, для  $2^4$  равно 16, для  $2^5$  – 32 и т.д. Таким образом, по сравнению с количеством факторов число опытов несоизмеримо велико, что существенно осложняет работу исследователя. Поэтому желательно уменьшить число опытов, при этом, не потеряв смысл эксперимента. Это достигается при помощи планирования дробного факторного эксперимента.

#### 4.2.2 Методика составления матрицы планирования полного и дробного факторного эксперимента

Составление матрицы планирования дробного факторного эксперимент (ДФЭ) удобно проводить на основе матрицы планирования ПФЭ с эффектом взаимодействия. Рассмотрим планирование ДФЭ на примере матрицы планирования ПФЭ  $2^3$  с эффектом взаимодействия (таблица 6). Если есть основания считать, что в выбранных интервалах варьирования эффектами взаимодействия факторов можно пренебречь, тогда столбцы взаимодействия факторов  $X_1X_2$ ,  $X_1X_3$ ,  $X_2X_3$ ,  $X_1X_2X_3$  можно использовать для новых факторов  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ . В этом случае, при замене всех взаимодействий на дополнительные факторы вместо 128 опытов ПФЭ  $2^7$  можно провести всего 8 опытов.

Таким образом, ДФЭ реализует часть ПФЭ, которую называют дробной репликой. Дробные реплики получают путем замены некоторых эффектов взаимодействия (или всех) на новые факторы. Число опытов ДФЭ находится по формуле:

$$N=2^{n-p},$$

где  $n$  – число факторов;

$p$  – число взаимодействий в ПФЭ, замененное дополнительными факторами.

Пример матрицы планирования ДФЭ для семи факторов, полученного из матрицы ПФЭ  $2^3$  с эффектом взаимодействия путем замены взаимодействий на факторы  $X_4$ ,  $X_5$ ,  $X_6$ ,  $X_7$ , приведен в таблице 7.

Таблица 7 – Матрица планирования ДФЭ  $2^{7-4}$

№ опыта	$X_1$	$X_2$	$X_3$	$X_4$ ( $X_1X_2$ )	$X_5$ ( $X_1X_3$ )	$X_6$ ( $X_2X_3$ )	$X_7$ ( $X_1X_2X_3$ )	Y
1(7)*	-1* <sup>2</sup>	-1	-1	+1	+1	+1	-1	Y1
2(4)	+1	-1	-1	-1	-1	+1	+1	Y2
3(6)	-1	+1	-1	-1	+1	-1	+1	Y3
4(1)	+1	+1	-1	+1	-1	-1	-1	Y4
5(8)	-1	-1	+1	+1	-1	-1	+1	Y5
6(3)	+1	-1	+1	-1	+1	-1	-1	Y6
7(5)	-1	+1	+1	-1	-1	+1	-1	Y7
8(2)	+1	+1	+1	+1	+1	+1	+1	Y8
Сумма	0	0	0	0	0	0	0	

\* в скобках указаны номера рандомизованных опытов

Обычно ДФЭ используют в качестве отсеивающего эксперимента, но он может иметь и самостоятельное значение при большом количестве факторов.

Количество опытов ДФЭ  $2^{7-4} = 8$ , что совпадает с количеством опытов в ПФЭ  $2^3$ . Таким образом 8 опытов позволяют оценить влияние 7 факторов, вместо 128 для ПФЭ  $2^7$ .

После проведения опытов рассчитываются коэффициенты при независимых переменных в математической модели, которая в данном случае будет иметь вид:  $Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_4 \cdot X_4 + b_5 \cdot X_5 + b_6 \cdot X_6 + b_7 \cdot X_7$ .

### 4.3 Проведение активного эксперимента и статистическая проверка результатов эксперимента

После построения матрицы, приступают к проведению экспериментов и определяют значения параметра оптимизации в каждом опыте. Причем каждый из опытов матрицы проводят не менее 3 раз. Это будут параллельные опыты. Эксперименты следует проводить рандомизованно, то есть в случайном порядке.

Получаемые экспериментальные данные проверяются на наличие ошибок и воспроизводимость.

#### 4.3.1 Проверка экспериментальных данных на наличие ошибок

Выявление грубых ошибок параллельных опытов по критерию Стьюдента, расчетное значение которого для каждого n-го опыта ( $t_{\text{расч, im}}$ ) определяется из выражения:

$$t_{\text{расч, im}} = \frac{|Y_{im} - \bar{Y}_i^*|}{\sigma_{Yi}^*},$$

где  $Y_{im}$  – значение параметра оптимизации в n-ом параллельном опыте по i-ой строке матрицы;

$m$  – количество параллельных опытов в i-ом опыте матрицы;

$\bar{Y}_i^*$  – среднее значение параметра оптимизации в i-ой строке матрицы без учета оцениваемого результата опыта;

$\sigma_{Yi}^*$  – среднеквадратическое отклонение параметра оптимизации в i-ом опыте матрицы без учета оцениваемого результата опыта, определяемая по уравнению

$$\sigma_{Yi}^* = \sqrt{\frac{\sum_{i=1}^m (Y_{im} - \bar{Y}_i)^2}{m - 1}}.$$

Расчетное значение критерия для каждого параллельного опыта сравнивают с табличным значением критерия Стьюдента, определяемом из справочной литературы при уровне значимости (вероятность ошибочного вывода) –  $\alpha$  (обычно принимают 0,05) и числе степеней свободы  $f = m-1$  [9]. Результат опыта отсеивают если  $t_{\text{расч}} > t_{\alpha}$ . В этом случае опыт следует провести повторно.

Определение опытных значений критерия Стьюдента и его табличной величины упрощается при проведении расчетов в среде электронных таблиц MS Excel, где можно использовать встроенные функции для расчета:

- среднего арифметического      СРЗНАЧ(число1; число2; ...)
- абсолютной величины          ABS(число)
- среднеквадратического  
отклонения по выборке      СТАНДОТКЛОН(число1; число2; ...)
- табличного значения  
критерия Стьюдента          СТЬЮДРАСПОБР(вероятность  $\alpha$ ; степени свободы  $f$ )

#### 4.3.2 Проверка воспроизводимости результатов экспериментов

Проверка воспроизводимости результатов параллельных опытов проводится (при одинаковом количестве параллельных опытов) по критерию Кохрена, расчетное значение которого определяется из выражения:

$$G = \sigma_{i \text{ MAX}}^2 / \sum_{i=1}^N \sigma_i^2,$$

где  $\sigma_{i \text{ MAX}}^2$  – наибольшая из дисперсий  $Y_i$  по каждой строке матрицы планирования;

$N$  – количество опытов в матрице планирования (от 1 до  $i$ );

$\sigma_i^2$  – дисперсия  $Y$  полученная в  $i$ -ом опыте

$$\sigma_i^2 = \frac{\sum_{m=1}^m (Y_{im} - \bar{Y}_i)^2}{m-1}.$$

Расчетный критерий сравнивают с табличным, определяемым из справочных данных при уровне значимости  $\alpha$ , количестве опытов –  $N$  и числе степеней свободы  $m-1$  [9].

Если соблюдается условие  $G_{\text{расч}} \leq G_{\alpha}$ , то есть расчетный критерий Кохрена не превышает табличный, то эксперимент считается воспроизводимым, то есть проведенным в статистически равных условиях. В противном случае, опыт с максимальным значением дисперсии необходимо исключить из результатов



эксперимента, а затем повторить его для достижения статистически равных условий.

Проверка воспроизводимости экспериментов по критерию Кохрена упрощается при проведении расчетов в среде электронных таблиц MS Excel, где можно использовать встроенные функции для расчета:

- дисперсии по выборке (параллельным опытам)      ДИСП(число1;число2; ...)
- наибольшего значения      МАКС(число1;число2; ...)

#### 4.4 Методика статистической обработки результатов активного эксперимента

##### 4.4.1 Определение коэффициентов в уравнении регрессии

Простейшей формой зависимости параметра оптимизации от влияющих факторов является линейная модель вида:  $Y=b_0+b_1 \cdot X_1+b_2 \cdot X_2+\dots$

Для получения модели необходимо определить коэффициенты, показывающие влияние каждого фактора на параметр оптимизации, по формуле:

$$b_j = \frac{\sum_{i=1}^N X_{ij} \cdot \bar{Y}_i}{N},$$

где  $i$  – номер опыта (от 1 до  $N$ );

$j$  – номер фактора ( $j = 0, 1 \dots n$ );

$\bar{Y}_i$  – среднее значение параметра оптимизации в  $i$ -ом опыте;

$X_{ij}$  – кодированное значение  $j$ -го фактора в  $i$ -ом опыте;

$N$  – количество опытов.

При определении коэффициента  $b_0$  (свободного члена уравнения регрессии)  $X_{i0} = 1$ , а при расчете коэффициентов, находящихся при факторах  $X_{ij}$  – кодированное значение  $j$ -го фактора в  $i$ -ом опыте по матрице планирования.

Коэффициенты при независимых переменных (факторах) указывают на силу влияния факторов. Чем больше численная величина коэффициента, тем большее влияние оказывает фактор на величину параметра оптимизации. Если коэффициент имеет знак (+), то это влияние положительное, и наоборот.

Для реальных производственных процессов исследуемые зависимости часто отличаются от линейных. Один из часто встречающихся видов нелинейности связан с тем, что влияние одного фактора зависит от уровня, на котором

находится другой фактор, то есть имеет место эффект взаимодействия двух факторов. При учете взаимного влияния факторов получаемая модель имеет вид:

$$Y = b_0 + b_1 \cdot X_1 + b_2 \cdot X_2 + b_3 \cdot X_3 + b_{12} \cdot X_1 \cdot X_2 + b_{13} \cdot X_1 \cdot X_3 + b_{23} \cdot X_2 \cdot X_3 + b_{123} \cdot X_1 \cdot X_2 \cdot X_3.$$

Расчет коэффициентов, учитывающих взаимодействие двух факторов, ведется по уравнению

$$b_{jk} = \frac{\sum_{i=1}^N X_{ij} \cdot X_{ik} \cdot \bar{Y}_i}{N}.$$

Аналогично рассчитываются коэффициенты при взаимодействии трех и более факторов.

Полученные, таким образом, уравнения регрессии можно использовать для прогнозирования и оптимизации процесса, однако при этом следует помнить, что:

- входящие в уравнение регрессии факторы необходимо подставлять в кодированном виде, то есть в диапазоне от -1 (нижний уровень фактора) до +1 (верхний уровень фактора);
- адекватность уравнения регрессии ограничивается интервалами варьирования входящих в модель факторов.

Уравнение регрессии можно привести к явному виду, для чего следует пересчитать входящие в него коэффициенты

- для коэффициентов при факторах  $b_j^* = \frac{2 \cdot b_j}{\Delta X_j}$ ;

- для свободного члена  $b_0^* = b_0 - \sum_1^n b_j^* \cdot \bar{X}_j$ ,

где  $\bar{X}_j$  – среднее значение j-го фактора в изученном интервале варьирования;

$\Delta X_j$  – абсолютное изменение значения j-го фактора в изученном интервале варьирования;

n – число факторов в модели.

Коррекция коэффициентов, учитывающих взаимодействие факторов, ведется по аналогичным уравнениям.

#### 4.4.2 Методика определения значимости факторов по величине коэффициентов уравнения регрессии

Значимость коэффициентов в уравнении регрессии оценивается с помощью критерия Стьюдента. Для этого для каждого коэффициента уравнения регрессии (за исключением свободного члена) рассчитывается значение критерия Стьюдента по уравнению

$$t_{bj} = \frac{|b_j|}{\sigma_b},$$

где  $b_j$  – численное значение коэффициента при факторах;

$$\sigma_b = \sqrt{\frac{\sum \sigma_{Yi}^2}{N^2 \cdot m}} - \text{среднеквадратическое отклонение коэффициентов уравне-}$$

ния регрессии;

$m$  – число параллельных опытов;

$N$  – число опытов в матрице;

$$\sigma_{Yi}^2 = \frac{\sum_{m=1}^m (Y_{im} - \bar{Y}_i)^2}{m-1} - \text{дисперсия } Y \text{ полученная в } i\text{-ом опыте.}$$

Если расчетное значение критерия Стьюдента для данного коэффициента меньше табличного значения ( $t_{bj} < t_a$ ), полученного для  $t$ -распределения Стьюдента при уровне значимости –  $\alpha$  и числе степеней свободы  $f = N \cdot (m-1)$ , то данный коэффициент считается малозначащим. Факторы, соответствующие малозначащим коэффициентам, не оказывают существенного влияния на изучаемый процесс. Поэтому их отсеивают (исключают) из модели или фиксируют на постоянном уровне.

Вычисление дисперсий параметров оптимизации по опытам ( $\sigma_{Yi}^2$ ) и табличной величины  $t$ -критерия Стьюдента ( $t_a$ ) удобно проводить с помощью описанных ранее встроенных функций электронных таблиц MS Excel.

#### 4.4.3 Методика проверки адекватности и точности регрессионной модели

Проверка адекватности математической модели проводится с помощью критерия Фишера. Расчетное значение критерия Фишера определяют из выражения:

$$F_p = \frac{\sigma_{ад}^2}{\sigma_Y^2},$$

где  $\sigma_{ад}^2$  – дисперсия адекватности модели  $\sigma_{ад}^2 = \frac{m}{N-n} \cdot \sum_{i=1}^N (\bar{Y}_i - Y_{i,расч})^2$ ,

$\sigma_Y^2 = \frac{1}{N} \cdot \sum_{i=1}^N \sigma_{Y_i}^2$  - дисперсия воспроизводимости эксперимента;

$\sigma_{Y_i}^2$  – дисперсия  $Y$  полученная в  $i$ -ом опыте;

$n$  – число значимых факторов в модели;

$\bar{Y}_i$  - среднее значение параметра оптимизации в  $i$ -ом опыте;

$Y_{i,расч}$  – расчетное значение параметра оптимизации, полученное по найденному уравнению регрессии для  $i$ -го опыта.

Адекватность математической модели проверяют сравнением критериев Фишера – расчетного ( $F_p$ ) и табличного ( $F_a$ ). Математическая модель считается адекватной при условии, что расчетное значение критерия Фишера меньше табличного, полученного по таблицам  $F$ -распределения Фишера при уровне значимости (вероятность ошибочного вывода) –  $\alpha$  (обычно принимают 0,05), числе степеней свободы  $k_1 = N-n$  и  $k_2 = N \cdot (m-1)$ . Теоретическое (табличное) значение  $f$ -критерия Фишера можно определить с помощью встроенной функции MS Excel - ФРАСПОБР(вероятность  $\alpha$ ;  $k_1$ ;  $k_2$ ).

В полученной адекватной модели коэффициенты в уравнении регрессии и их знаки показывают, как изменяется значение параметра оптимизации, если соответствующий фактор изменить на величину одного интервала варьирования.

Неадекватность модели говорит о том, что не все влияющие факторы были учтены, неправильно подобраны интервалы варьирования или зависимость значительно отличается от линейной. Последнее встречается наиболее часто. В этом случае результаты эксперимента используют для отсеивания незначущих факторов.

После подтверждения адекватности модели осуществляют проверку ее точности, то есть пригодности к практическому использованию для прогнозирования и поиска оптимальных решений. Для проверки точности (работоспособности) модели используют коэффициент детерминации ( $R^2$ ). Коэффициент детерминации показывает, какая доля изменений параметра оптимизации (переменной  $Y$ ) учтена в модели и обусловлена влиянием на нее учтенных в модели факторов. Его значение вычисляют по формуле

$$R^2 = 1 - \frac{(N-n) \cdot \sigma_{ад}^2 + N \cdot (m-1) \cdot \sigma_{\bar{Y}}^2}{m \cdot \sum_1^N (\bar{Y}_i - \bar{Y})^2 + N \cdot (m-1) \cdot \sigma_{\bar{Y}}^2},$$

где  $\bar{Y}$  - среднее значение параметра за все опыты  $\bar{Y} = \frac{1}{N} \cdot \sum_1^N \bar{Y}_i$ .

Модель считается работоспособной при  $R^2 \geq 0,75$ .

#### 4.4.4 Использование дополнительных функций MS Excel

Частичное решение задач по статистической обработке результатов активного эксперимента обеспечивается дополнительной функцией MS Excel «Регрессия». Указанная функция входит в подменю «Анализ данных» (рисунок 3), которое активируется при установке надстройки «Пакет анализа».

Функция «Регрессия» позволяет определить коэффициенты уравнения регрессии, расчетные значения критерия Стьюдента для каждого коэффициента (t-статистика), коэффициент детерминации и другие статистические параметры.

После вызова функции «Регрессия» из окна диалога «Анализ данных» откроется соответствующее диалоговое окно (рисунок 3), в котором необходимо указать диапазоны с данными по зависимой функции Y (входной интервал Y) и независимым переменным X (входной интервал X), а также параметры вывода результатов анализа.

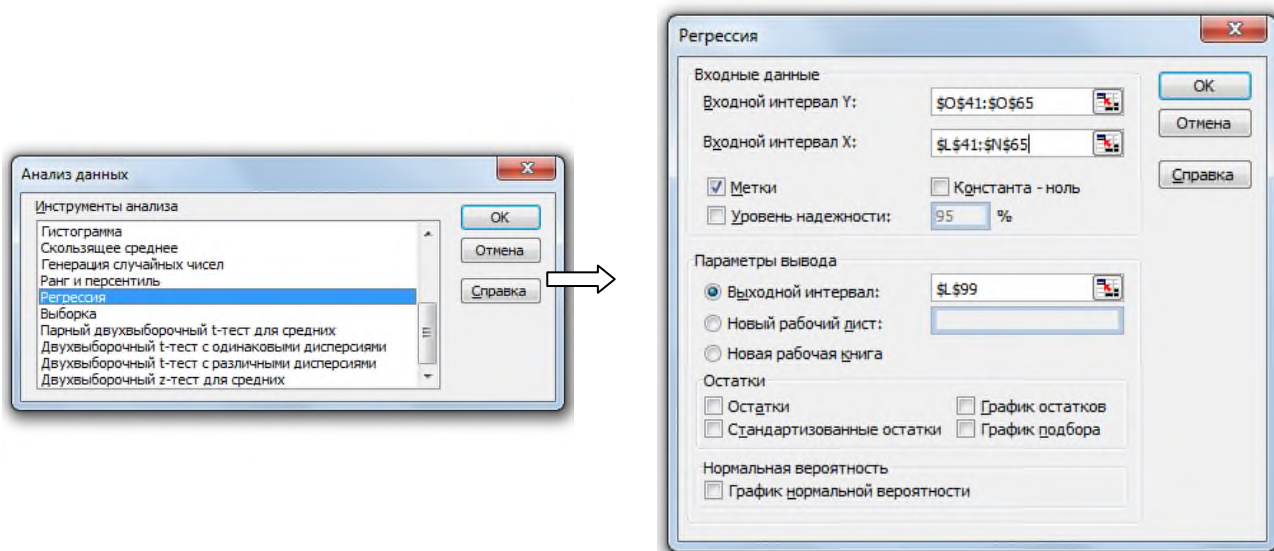


Рисунок 3 – Вызов и вид диалогового окна функции «Регрессия»

Результаты расчета с помощью функции «Регрессия» выводятся в виде таблицы, в которой указаны:

- величина коэффициента корреляции (Множественный R);
- значение коэффициента детерминации (R-квадрат);

- расчетное значение F-критерия Фишера (F), используемое для проверки значимости коэффициента корреляции (детерминации), а также значение уровня значимости, соответствующее вычисленному значению F;
- значения коэффициентов уравнения регрессии (Коэффициенты);
- расчетные значения t-критерия (t-статистика) для каждого коэффициента уравнения регрессии по которым определяется их значимость;
- др. статистические параметры исследуемых факторов.

Таким образом, использование возможностей функции «Регрессия» существенно упрощает трудоемкость статистической обработки результатов эксперимента, особенно при анализе многофакторных процессов.

Предварительный парный корреляционно-регрессионный анализ удобно проводить с использованием встроенной функции MS Excel «Мастер диаграмм» (из меню «Вставка») в режиме построения «точечной диаграммы», по которой можно оценить вид взаимосвязи между исследуемыми параметрами, а также построить «линию тренда» (различного типа) с указанием уравнения регрессии и коэффициента детерминации (достоверности аппроксимации).

## 5 Обработка результатов пассивного эксперимента

### 5.1 Статистическая проверка результатов пассивного эксперимента

Для пассивного эксперимента процесс выявления и устранения грубых ошибок основан на использовании правила трех сигм, согласно которому случайная величина (при нормальном ее распределении) не отклоняется от математического ожидания (среднего значения) по абсолютной величине больше утроенного среднеквадратического отклонения, то есть 3-х сигм.

Среднеквадратическое отклонение отдельного измерения ( $X_i$ ) генеральной совокупности является характеристикой колеблемости значений случайной величины ( $X_i$ ) от среднего ( $\bar{X}$ ) и определяется из выражения

$$\sigma_x = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^N (X_i - \bar{X})^2}.$$

Для выборочной совокупности применяется скорректированное значение среднеквадратического отклонения

$$\sigma_x^* = \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (X_i - \bar{X})^2}.$$

Экспериментальные данные считаются статистически достоверными (не содержащими грубые ошибки), если выполняется условие

$$|X_i - \bar{X}| < 3 \cdot \sigma_x^* = 3 \cdot \sqrt{\frac{1}{N-1} \cdot \sum_{i=1}^N (X_i - \bar{X})^2}.$$

Результаты наблюдений, для которых данное условие не выполняются – исключаются из выборки.

Выявление «ошибочных» результатов наблюдений по правилу «трех сигм» удобно проводить в среде электронных таблиц MS Excel, где можно использовать встроенные функции для:

- |  |   |
|--|---|
| - определения среднеквадратического отклонения по выборке                  | СТАНДОТКЛОН(число1; число2; ...)                            |
| - определения среднеквадратического отклонения по генеральной совокупности | СТАНДОТКЛОНП(число1; число2; ...)                           |
| - проверки условий (достоверности или ошибочности опытных данных)          | ЕСЛИ(лог_условие; значение_если_истина; значение_если_ложь) |

## 5.2 Статистическая обработка результатов пассивного эксперимента

Первичная обработка экспериментальных данных заключается в их систематизации (помещение данных в таблицы или на отдельные карточки), определении средних и граничных значений величин и их частотных характеристик (тип распределения), построении графиков (для наглядности), подборе эмпирических формул для математического описания исследуемого процесса.

Количественная оценка экспериментальных данных проводится методами математической статистики, с помощью которых производятся следующие работы:

- определяются основные факторы и степень их влияния на изучаемый процесс (дисперсионный анализ);
- обобщаются экспериментальные данные в математические зависимости – модели (регрессионный анализ);
- устанавливается достоверность полученных зависимостей и их адекватность изучаемым процессам (корреляционный анализ).

### 5.2.1 Дисперсионный анализ

На практике часто приходится сравнивать два периода технологического процесса, две технологии, две группы изделий и т.д., и, на основе этого сравнения, делать выводы об идентичности или различии сравниваемых явлений. Для сравнения средних арифметических двух выборок пользуются методом, который основан на сравнении дисперсий и называется дисперсионным анализом. Наиболее часто для сравнения средних двух выборок используется t-критерий Стьюдента. При сравнении средних двух выборок различают два случая: первый – при равенстве среднеквадратических отклонений двух выборок; второй – при неравенстве отклонений.

Для выбора той или иной схемы проверки гипотезы о равенстве средних двух выборок необходимо провести оценку различия дисперсий рассматриваемых выборок.

В качестве критерия сравнения дисперсий двух выборок применяют F-критерий Фишера:

$$F = \frac{N_1 \cdot (N_2 - 1) \cdot \sigma_{X,1}^2}{N_2 \cdot (N_1 - 1) \cdot \sigma_{X,2}^2},$$

где  $\sigma_X^2$  – дисперсия случайной величины  $X$  в рассматриваемой выборке (в числителе большая из двух дисперсий, условно принимаемая за первую);

$N_1$  и  $N_2$  – количество наблюдений (измерений) соответственно в 1-ой и 2-



ой выборках.

При проверке гипотезы о статистически значимом равенстве или различии двух дисперсий сравнивают расчетное значение критерия Фишера с критическим, которое определяют по таблицам F-распределения [9] при уровне значимости  $\alpha$  и с  $k_1=N_1-1$  и  $k_2=N_2-1$  степенями свободы ( $k_1$  – число степеней свободы для выборки с большей дисперсией). В случае, если  $F > F_{\alpha}$ , принимается версия о статистически значимом различии дисперсий и наоборот. Теоретическое (табличное) значение f-критерия Фишера можно определить с помощью встроенной функции MS Excel - FРАСПОБР(вероятность  $\alpha$ ;  $k_1$ ;  $k_2$ ).

Оценку статистического различия дисперсий рассматриваемых выборок можно выполнить посредством дополнительной функции MS Excel «Двухвыборочный F-тест для дисперсии». Указанная функция входит в подменю «Анализ данных», которое активируется при установке надстройки «Пакет анализа». Для ее вызова необходимо выбрать требуемое имя в окне диалога «Анализ данных». В результате появится диалоговое окно (рисунок 4), в котором необходимо указать интервалы переменных (первая – с большей из дисперсий, вторая – с меньшей) и выходной интервал.

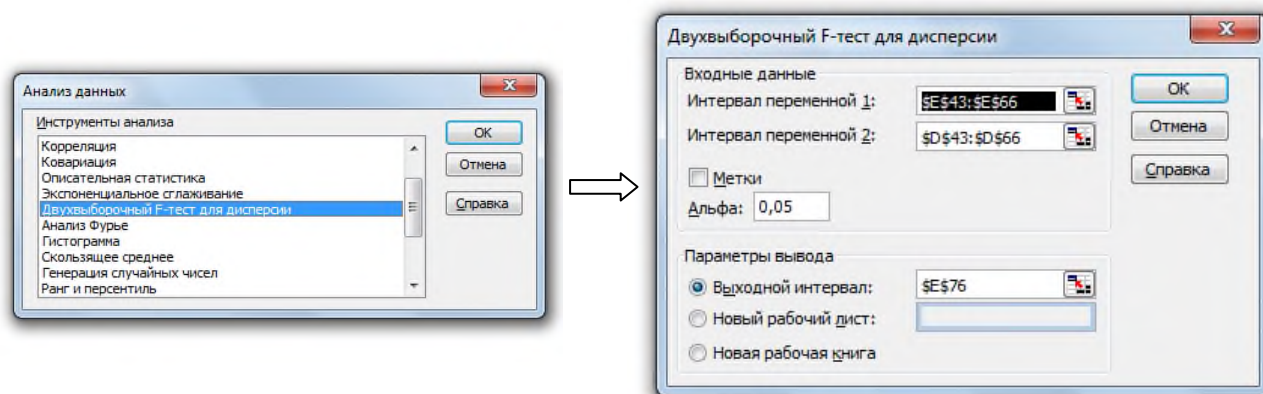


Рисунок 4 – Вызов и вид диалогового окна функции «Двухвыборочный F-тест для дисперсии»

Результаты расчета с помощью функции «Двухвыборочный F-тест для дисперсии» выводятся в виде таблицы, в которой указаны средние значения переменных, их количество и дисперсии для исследуемых выборок, а также расчетные и критические значения критерия Фишера, которые необходимо сравнить и сделать вывод о статистическом равенстве (или различии) дисперсий переменной в рассматриваемых выборках.

При проверке гипотезы о статистически значимом различии средних арифметических двух выборок при статистическом равенстве дисперсий t - критерий Стьюдента определяют из выражения:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{(N_1 - 1) \cdot \sigma_{\bar{X},1}^2 + (N_2 - 1) \cdot \sigma_{\bar{X},2}^2}} \cdot \sqrt{\frac{N_1 \cdot N_2 \cdot (N_1 + N_2 - 2)}{N_1 + N_2}},$$

где  $\bar{X}$  - среднее значение случайной величины в соответствующей выборке.

Различие между средними арифметическими двух выборок считается статистически значимым, если расчетное значение t-критерия Стьюдента превышает критическое (табличное) [9], определяемое по таблицам для t-распределения при уровне значимости  $\alpha$  и числе степеней свободы  $f = N_1 + N_2 - 2$ . Теоретическое (табличное) значение t-критерия Стьюдента ( $t_\alpha$ ) также можно определить с помощью встроенной функции MS Excel - СТЬЮДРАСПОБР(вероятность  $\alpha$ ; степени свободы  $f$ ). При выполнении условия  $t < t_\alpha$  принимается гипотеза о статистическом равенстве средних арифметических в рассматриваемых выборках.

Проверку гипотезы о статистически значимом различии средних арифметических двух выборок при статистическом равенстве дисперсий можно выполнить посредством дополнительной функции MS Excel «Двухвыборочный t-тест с одинаковыми дисперсиями». Указанная функция входит в подменю «Анализ данных», которое активируется при установке надстройки «Пакет анализа». Для ее вызова необходимо выбрать требуемое имя в окне диалога «Анализ данных». В результате появится диалоговое окно (рисунок 5), в котором необходимо указать интервалы переменных и выходной интервал.

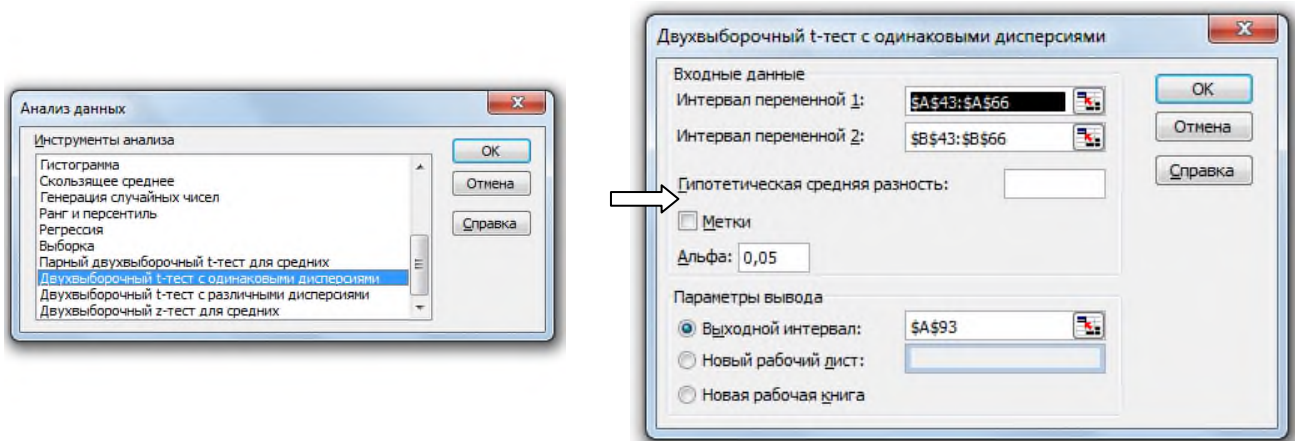


Рисунок 5 – Вызов и вид диалогового окна функции «Двухвыборочный t-тест с одинаковыми дисперсиями»

Результаты расчета с помощью функции «Двухвыборочный t-тест с одинаковыми дисперсиями» выводятся в виде таблицы, в которой указаны средние значения переменных, их количество и дисперсии для исследуемых выборок, а также расчетные и критические значения критерия Стьюдента, которые необходимо сравнить и сделать вывод о статистическом равенстве (или различии) средних в рассматриваемых выборках: различие между средними арифметиче-

скими двух выборок считается статистически значимым, если расчетное значение t-критерия Стьюдента (t-статистика) по абсолютной величине превышает критическое (t критическое двухстороннее).

В случае сравнения средних двух выборок с различными значениями дисперсий t-критерий Стьюдента определяют из выражения:

$$t = \frac{|\bar{X}_1 - \bar{X}_2|}{\sqrt{\frac{\sigma_{x,1}^2}{N_1} + \frac{\sigma_{x,2}^2}{N_2}}},$$

Версия о статистически значимом различии между средними двух выборок отвергается, если полученное значение t-критерия меньше критического ( $t_a$ ), которое для этого случая определяется по таблицам для t-распределения Стьюдента при уровне значимости  $\alpha$  и числе степеней свободы  $f$ , определяемом из выражения

$$f = \frac{\left(\frac{\sigma_{x,1}^2}{N_1} + \frac{\sigma_{x,2}^2}{N_2}\right)^2}{\frac{(\sigma_{x,1}^2/N_1)^2}{N_1 - 1} + \frac{(\sigma_{x,2}^2/N_2)^2}{N_2 - 1}}.$$

Так как результат вычисления обычно не бывает целым числом, значение  $f$  округляется до целого для получения критического значения из t-таблицы распределения Стьюдента. Табличное значение t-критерия Стьюдента ( $t_a$ ) также можно определить с помощью встроенной функции MS Excel - СТЫЮДРАСПОБР(вероятность  $\alpha$ ; степени свободы  $f$ ). Версия о значимом различии между средними двух выборок принимается, если выполняется условие:  $t > t_a$ .

Проверку гипотезы о статистически значимом различии средних арифметических двух выборок с различными значениями дисперсий можно выполнить посредством дополнительной функции MS Excel «Двухвыборочный t-тест с различными дисперсиями». Указанная функция входит в подменю «Анализ данных», которое активируется при установке надстройки «Пакет анализа». Работа с данной функцией проводится в той же последовательности, что и с рассмотренной ранее функцией «Двухвыборочный t-тест с одинаковыми дисперсиями».

## 5.2.2 Корреляционный анализ

Корреляционный анализ заключается в выявлении связи между двумя или более изучаемыми переменными посредством определения коэффициентов корреляции или корреляционных отношений.

Наиболее простой случай взаимосвязи между двумя факторами – линейная зависимость. Парный линейный коэффициент корреляции определяют методом наименьших квадратов из выражения:

$$r_{xy} = \frac{\sum_{i=1}^N (X_i - \bar{X}) \cdot (Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^N (X_i - \bar{X})^2 \cdot \sum_{i=1}^N (Y_i - \bar{Y})^2}},$$

где  $N$  – количество опытов (измерений);

$X_i$  – значение случайной величины в  $i$ -ом опыте;

$\bar{X}$  и  $\bar{Y}$  – средние значения параметров  $X$  и  $Y$  за опытный период.

Парный коэффициент корреляции может изменяться от  $-1$  до  $+1$ . Чем больше по абсолютной величине коэффициент корреляции, тем меньше рассеяние экспериментальных точек от линейной зависимости и тем теснее линейная связь между двумя переменными  $X$  и  $Y$ , которая характеризуется уравнением типа  $Y = a + b \cdot X$  (при функциональной линейной зависимости – коэффициент корреляции равен  $\pm 1$ ). Знак коэффициента корреляции указывает на характер изменения случайной величины  $Y$  в зависимости от изменения  $X$ . Так, если коэффициент корреляции положителен, то значения функции  $Y$  возрастают с увеличением переменной  $X$  (коэффициент  $b$  в уравнении линейной зависимости больше нуля). В этом случае говорят о положительной (прямой) линейной зависимости, в противном случае речь идет об отрицательной (обратной) линейной зависимости.

Для расчета линейного коэффициента корреляции можно использовать функцию MS Excel – КОРРЕЛ(массив1;массив2) из категории «статистические». Для вывода результатов в виде таблицы применяют дополнительную функцию «Корреляция» из пакета «Анализ данных». После ее вызова из окна диалога «Анализ данных» откроется соответствующее диалоговое окно (рисунок б), в котором необходимо указать диапазон ячеек с анализируемыми данными (входной интервал) и место, где будет указана таблица результатов анализа (выходной интервал).

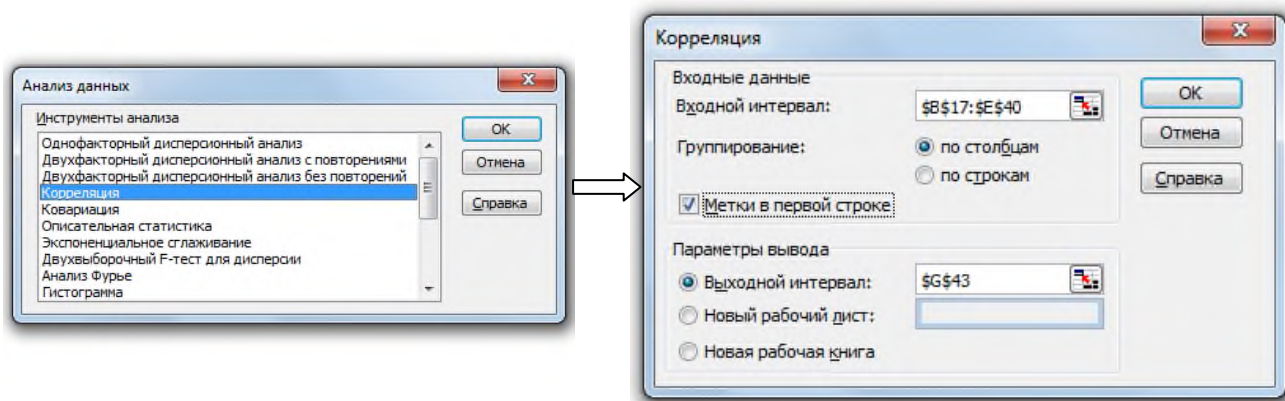


Рисунок 6 – Вызов и вид диалогового окна функции «Корреляция»

Если коэффициент корреляции приближается к нулю, то степень линейной связи между двумя рассматриваемыми переменными очень мала, то есть линейной зависимости нет. Однако в этом случае между переменными  $X$  и  $Y$  может существовать нелинейная зависимость. В таком случае предполагают иной вид зависимости и подбирают соответствующее уравнение. В этом случае, для оценки тесноты связи между ними нельзя применять линейный коэффициент корреляции, а следует определять корреляционное отношение, которое определяет долю колебаний  $Y$ , обусловленную изменением  $X$ .

Для вычисления корреляционного отношения  $\rho_{Y/X}$  диапазон изменения  $X$  разбивают на  $\varphi$  интервалов, от 1 до  $\varphi$ . При выборе количества интервалов можно ориентироваться на следующее соотношение:  $\varphi \approx 3,322259 \cdot \lg N + 1$ . Предварительно проводят сортировку параметров  $X$  и  $Y$  за весь опытный период (по параметру  $X$  с целью достижения минимальных колебаний  $X$  в пределах каждого интервала). В каждом интервале (от 1 до  $\varphi$ ) подсчитывают число  $Z_\varphi$  попавших в него точек и интервальное среднее  $\bar{Y}_\varphi$ . После чего рассчитывают корреляционное отношение по формуле:

$$\rho_{y/x} = \frac{\sqrt{\sum_{\varphi=1}^{\varphi} Z_\varphi \cdot (\bar{Y}_\varphi - \bar{Y})^2}}{\sqrt{\sum_{i=1}^N (Y_i - \bar{Y})^2}}.$$

Корреляционное отношение может изменяться в пределах от 0 до 1. Если величина корреляционного отношения превышает абсолютное значения коэффициента корреляции, то можно сделать вывод о существовании нелинейной зависимости между параметрами  $X$  и  $Y$ .

По величине коэффициента корреляции и корреляционного отношения можно качественно охарактеризовать тесноту связи между исследуемыми

параметрами: если значение коэффициента корреляции или корреляционного отношения менее 0,2 – связи нет, если 0,2-0,5 – связь слабая, если 0,5-0,75 – связь средняя, если более 0,75 – связь тесная.

Окончательный вывод о том, существует ли статистически значимая зависимость между X и Y делают на основании сравнения расчетного коэффициента корреляции (корреляционного отношения) с критическим ( $r_a$ ), который определяют по таблицам [9] при числе наблюдений N и вероятности p. Если  $r_{xy} > r_a$ , то для данного числа наблюдений N коэффициент корреляции  $r_{xy}$  является значимым, то есть с вероятностью p можно утверждать, что существует зависимость Y от X. Аналогичное сравнение проводят и для корреляционного отношения.

Значимость коэффициента парной корреляции (корреляционного отношения) можно оценивать также по t-критерию Стьюдента, величина которого определяется по уравнению

$$t = \frac{|r_{xy}| \cdot \sqrt{N-2}}{\sqrt{1-r_{xy}^2}}.$$

Значение t-критерия Стьюдента, характеризующего значимость коэффициента парной корреляции, автоматически определяется дополнительной функции MS Excel «Регрессия», входящей в подменю «Анализ данных».

Статистически значимая линейная зависимость существует при условии, что t-критерий Стьюдента больше табличного значения, найденного при заданном уровне значимости  $\alpha$  и числе степеней свободы  $f = N-2$ . Табличное значение t-критерия Стьюдента ( $t_a$ ) также можно определить с помощью встроенной функции MS Excel - СТЬЮДРАСПОБР(вероятность  $\alpha$ ; степени свободы f). Значимость корреляционного отношения  $\rho_{y/x}$  проверяется аналогичным образом.

Если случайная величина Y зависит от нескольких случайных величин  $X_1, X_2, X_3$  и т.д., то в этом случае наличие зависимости Y от X проверяется с помощью коэффициента множественной корреляции R. Если количество независимых переменных равно двум, R может быть рассчитано по формуле:

$$R_{YX_1X_2} = \sqrt{\frac{r_{YX_1}^2 + r_{YX_2}^2 - 2 \cdot r_{YX_1} \cdot r_{YX_2} \cdot r_{X_1X_2}}{1 - r_{X_1X_2}^2}},$$

где  $r_{YX_1}, r_{YX_2}, r_{X_1X_2}$  – величины парных коэффициентов корреляции для всех возможных комбинаций зависимости переменных Y,  $X_1$  и  $X_2$ .

При количестве влияющих факторов более двух расчеты существенно

осложняются, поэтому их удобно проводить с помощью дополнительной функции MS Excel «Регрессия», входящей в подменю «Анализ данных».

Коэффициент множественной корреляции принимает только положительные значения, поэтому с его помощью нельзя установить знак зависимости. Теснота связи определяется так же, как и по коэффициенту парной корреляции: чем ближе значение  $R$  к 1, тем меньше рассеяние экспериментальных точек от линейной зависимости и тем теснее связь между исследуемыми параметрами.

Значимость коэффициента множественной корреляции оценивается по критерию Фишера, определяемому из выражения

$$F = \frac{N - n - 1}{n} \cdot \frac{R_{YX_1 X_2 \dots X_n}^2}{1 - R_{YX_1 X_2 \dots X_n}^2},$$

где  $N$  – объем выборки;

$n$  – число независимых переменных (факторов).

Величина критерия Фишера автоматически определяется при проведении регрессионного анализа посредством дополнительной функции MS Excel «Регрессия», входящей в подменю «Анализ данных».

Если расчетное значение критерия Фишера превышает его критическую величину, определяемую при количестве независимых переменных (факторов)  $k_1 = n$ , числе степеней свободы  $k_2 = N - n$  и выбранном уровне значимости  $\alpha$ , то делают вывод о значимости коэффициента множественной корреляции и существовании статистически значимой зависимости между функцией  $Y$  и переменными  $X$ .

Теоретическое (табличное) значение критерия Фишера можно определить с помощью встроенной функции MS Excel - FРАСПОБР(вероятность  $\alpha$ ;  $k_1$ ;  $k_2$ ).

### 5.2.3 Регрессионный анализ

Регрессионный анализ применяют при необходимости количественной оценки зависимости между двумя или более параметрами, статистическая связь между которыми установлена корреляционным анализом. Регрессионный анализ заключается в поиске коэффициентов в уравнениях регрессии и в статистической оценке результатов расчета по полученным уравнениям.

В начале регрессионного анализа необходимо задаться видом зависимости, которые подразделяют на линейные и нелинейные (степенные, логарифмические и др.). При выборе зависимости выбирают ту, которая с наибольшей

точностью описывает исследуемую зависимость, а также такую, которая не противоречит теоретическим представлениям об исследуемой зависимости. Оценку вида парных зависимостей обычно проводят графическим методом, строя, так называемое, корреляционное поле, на которое наносятся все экспериментальные значения исследуемой пары факторов X и Y. Для построения корреляционного поля удобно использовать встроенную функцию MS Excel «Мастер диаграмм» (из меню «Вставка»), позволяющую строить «точечную диаграмму» взаимосвязи анализируемой пары значений.

В регрессионном анализе используют линейные парные, линейные множественные и нелинейные уравнения регрессии.

При проведении парного линейного регрессионного анализа определяют уравнение вида  $Y = a + b \cdot X$ , коэффициенты которого рассчитываются по уравнениям:

$$b = r_{XY} \cdot \frac{\sigma_Y}{\sigma_X};$$

$$a = \bar{Y} - b \cdot \bar{X}.$$

Множественный линейный регрессионный анализ заключается в отыскании коэффициентов a и  $b_j$  в уравнении вида  $Y = a + b_1 X_1 + b_2 X_2 + \dots + b_p X_p$ , что выполняется, как правило, методом наименьших квадратов.

Самым простым вариантом множественного регрессионного анализа является случай, когда функция зависит от двух переменных. В этом случае коэффициенты в уравнении регрессии  $Y - \bar{Y} = b_1 \cdot (X_1 - \bar{X}_1) + b_2 \cdot (X_2 - \bar{X}_2)$ , определяются по формулам:

$$b_1 = \frac{\sigma_Y}{\sigma_{X_1}} \cdot \frac{(r_{YX_1} - r_{YX_2} \cdot r_{X_1X_2})}{(1 - r_{X_1X_2}^2)};$$

$$b_2 = \frac{\sigma_Y}{\sigma_{X_2}} \cdot \frac{(r_{YX_2} - r_{YX_1} \cdot r_{X_1X_2})}{(1 - r_{X_1X_2}^2)}.$$

При количестве факторов больше двух определение уравнение регрессии значительно усложняется, поэтому его проводят с помощью специализированных программ на ЭВМ. Это же относится и к случаю проведения множественного нелинейного регрессионного анализа. Парный нелинейный регрессионный анализ можно выполнить «вручную» по методике парного линейного анализа, предварительно проведя линеаризующие преобразования с исследуемыми параметрами (с целью приведения нелинейной зависимости к линейному виду).

После получения уравнения регрессии проводят проверку его адекватно-



сти по критерию Фишера. Проверка основана на сопоставлении общей дисперсии  $Y$  ( $\sigma_Y^2$ ) с остаточной дисперсией ( $\sigma_{Y_{оос}}^2$ ). Остаточная дисперсия представляет собой показатель ошибки предсказания результатов расчета с помощью уравнения регрессии:

$$\sigma_{Y_{оос}}^2 = \frac{\sum_{i=1}^N (Y_{расч,i} - Y_i)^2}{N - 2}.$$

где  $Y_{расч,i}$  – значение переменной  $Y$ , полученное по уравнению регрессии при  $i$ -ом значении  $X$ .

Гипотеза о статистической значимости уравнения регрессии подтверждается при выполнении условия:

$$F = \frac{\sigma_Y^2}{\sigma_{Y_{оос}}^2} > F_a.$$

Критическое значение критерия Фишера ( $F_a$ ) определяется по таблице F-распределения Фишера для уровня значимости  $\alpha$  и числа степеней свободы  $k_1 = N - 1$  и  $k_2 = N - n$  ( $n$  – число коэффициентов (факторов) уравнения регрессии). Его можно определить также с помощью встроенной функции MS Excel - ФРАСПОБР(вероятность  $\alpha$ ;  $k_1$ ;  $k_2$ ).

#### **5.2.4 Использование дополнительных функций MS Excel**

Использование возможностей функции MS Excel «Регрессия» для упрощения статистической обработки результатов эксперимента, особенно при анализе многофакторных процессов, было подробно изложено в п. 4.4.4.

## **6 Варианты заданий и требования к оформлению работы**

Индивидуальное задание на выполнение домашнего задания / контрольной работы обучающиеся получают от преподавателя на первом практическом занятии в виде массива данных пассивного эксперимента (в электронном виде) и темы исследования.

Примерные темы исследования:

1 Изучение влияния технологии нагрева металла под прокатку на энергосиловые параметры процесса прокатки.

2 Исследование влияния длительного подстуживания раската на промежуточном рольганге на механические свойства проката.

3 Исследования влияния параметров деформации на механические свойства стали.

4 Изучение влияния скорости деформации металла на параметры прокатки толстых листов.

5 Исследование влияния параметров ускоренного охлаждения с прокатного нагрева на механические свойства штрипсов.

Выполненная работа оформляется в виде пояснительной записки в соответствии с правилами оформления ВКР, действующими в филиале. Пояснительная записка должна иметь следующую структуру: титульный лист, содержание, введение, основная часть в соответствии с индивидуальной темой исследования, заключение.

В разделе «Введение» формулируются основные задачи, решаемые при выполнении исследования. В основной части проводится статистическая обработка данных в соответствии с рекомендациями настоящего пособия (структура основной части выбирается обучающимися). В разделе «Заключение» приводятся основные результаты проведенного анализа экспериментальных данных.

Подготовленная пояснительная записка сдается на проверку на кафедру металлургических технологий и оборудования. Домашнее задание / контрольная работа считаются выполненными, если они зачтены преподавателем, ведущим занятия.

## 7 Пример выполнения работы

Рассмотрим пример выполнения работы на тему «Исследование влияния параметров нагрева и прокатки на величину отсортировки листового проката».

С целью определения параметров, влияющих на отсортировку листового проката из стали 10Г2ФБЮ, был выполнен анализ производственных данных. Усредненные производственные данные представлены в таблице 8.

Таблица 8 - Параметры нагрева и прокатки для стали 10Г2ФБЮ

Параметр	Фактические значения		
	мин	макс	среднее
Температура посада, °С	0	500	436,4
Общее время нагрева сляба, мин	216	406	264,5
Температура выгрузки, °С	1022,8	1279,7	1227,0
Температура после последнего прохода ДУО, °С	953	1058	1020,0
Толщина подката из ДУО, мм	62	67	63,2
Время подстуживания после ДУО, мин	111,4	331,6	256,1
Температура начала прокатки на КВАРТО, °С	732,15	1053,4	892,8
Температура конца прокатки КВАРТО, °С	790	823	807,0
Толщина листа, мм	14	16	15,41
Обжатие в ДУО за проход, %	8,14	14,96	10,61
Обжатие в КВАРТО за проход, %	12,98	15,33	14,01
Температура начала УКО, °С	719,1	797,1	779,2
Температура конца УКО, °С	549,3	644,8	617,7
Скорость охлаждения, К/с	17,2	25,6	20,6

Из таблицы 8 видно, что из всех контролируемых параметров нагрева и прокатки статистически значимо отличаются температура посада сляба; общее время нахождения заготовки в печи; время подстуживания после ДУО; температура начала прокатки на КВАРТО; температура конца УКО.

Выполнен парный корреляционный анализ выборки, усредненной по плавкам (84 плавки), в ходе которого стандартными средствами MS Excel определяли коэффициенты детерминации ( $R^2$ ) для парных зависимостей. При проведении корреляционного анализа учитывались только те параметры, значимость которых была установлена в ходе проведения дисперсионного анализа. В результате, в число параметров, которые статистически значимо влияют на отсортировку проката из стали 10Г2ФБЮ вошли:

- температура посада заготовки;
- общая продолжительность нагрева заготовки;
- величина обжатия за проход в ДУО;
- суммарное обжатие в ДУО;

- температуры начала и конца прокатки в клети КВАРТО.

Для иллюстрации влияния параметров производства на его результаты на рисунках 6-10 приведены графики парных зависимостей для наиболее значимых параметров.

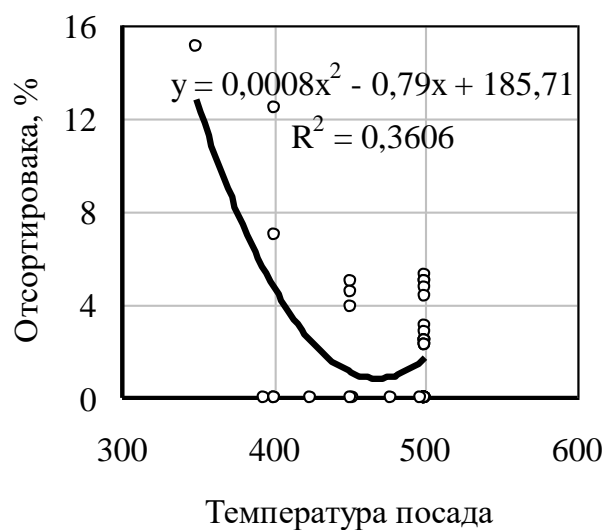


Рисунок 6 - Влияние температуры проката слябовой заготовки на отсортировку листового проката из стали 10Г2ФБЮ

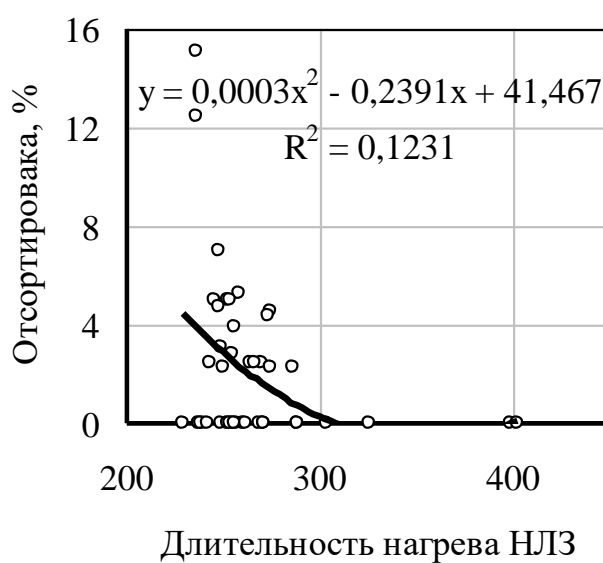


Рисунок 7 - Влияние продолжительности нагрева слябовой заготовки на отсортировку листового проката из стали 10Г2ФБЮ

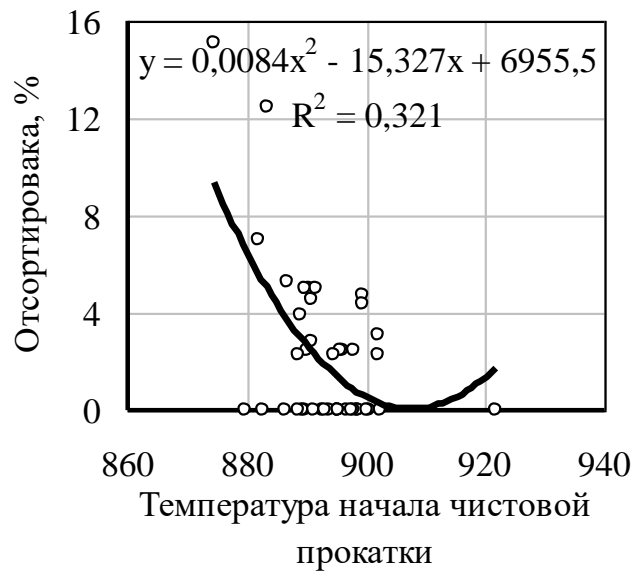


Рисунок 8 - Влияние температуры начала чистовой прокатки на отсортировку листового проката из стали 10Г2ФБЮ

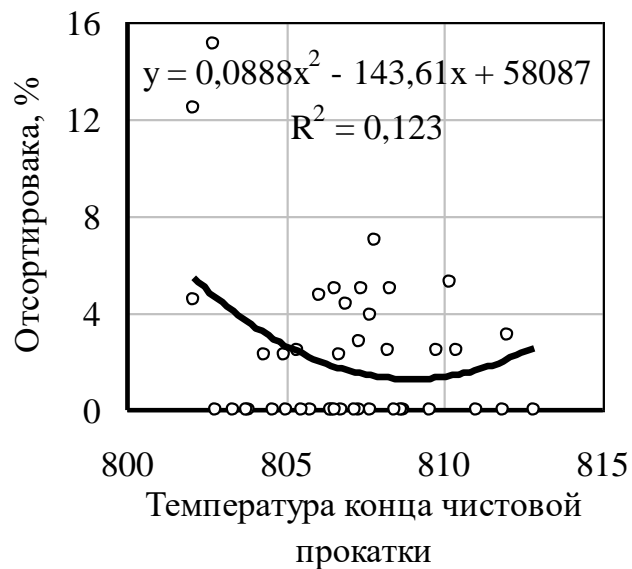


Рисунок 9 - Влияние температуры конца чистовой прокатки на отсортировку листового проката из стали 10Г2ФБЮ

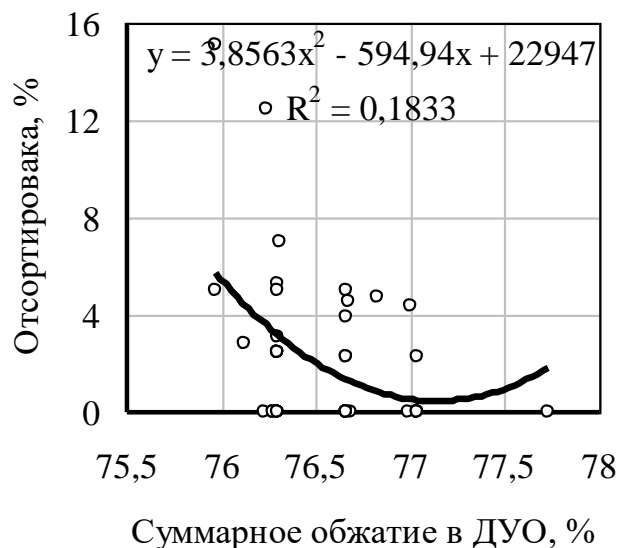


Рисунок 10 - Влияние суммарного обжатия в клети ДУО на отсортировку листового проката из стали 10Г2ФБЮ

Таким образом, снижение величины отсортировки проката из стали 10Г2ФБЮ толщиной 14-16 мм обеспечивается при следующих условиях:

- максимальная температура посадки и время нагрева заготовки, что обеспечивает более равномерный прогрев и позволяет минимизировать внутренние напряжения при прокатке;
- увеличение суммарного обжатия в черновых проходах (уменьшение толщины подката) способствует устранению несплошностей в слябовой заготовке и снижает отсортировку по УЗК;
- увеличение температуры начала чистой прокатки способствует снижению отсортировки проката.

Отсортировка листового проката из стали 10Г2ФБЮ в равной степени определяется условиями ее нагрева, степенью обжатия в клети ДУО и температурой начала прокатки в КВАРТО. Толщина проката не показала существенного влияния в связи с тем, что регрессионный анализ проводился для условий производства листа толщиной 14-16 мм. При расширении диапазона исследования по толщине листа, этот параметр следует повторно проверить на значимость влияния.

Можно установить, что при текущем качестве непрерывнолитой заготовки для получения листового проката толщиной 14-16 мм без отсортировки по неметаллическим включениям и УЗК необходимо производить горячий посад слябов с температурой 450-500 °С и нагревать их не менее 270 мин, повышать суммарную величину обжатия в клети ДУО (до 78 %) и температуру начала прокатки в клети КВАРТО до 920-930 °С.

## Библиографический список

1. Гецов Г. Как писать обзоры // Наука и жизнь, 1984. №1. С.106-107.
2. Штремель М.А. Инженер в лаборатории (Организация труда). – М.: Металлургия, 1983. – 128 с.
3. ГОСТ 7.32-2017 «Система стандартов по информации, библиотечному и издательскому делу. Отчет о научно-исследовательской работе. Структура и правила оформления»
4. ГОСТ Р7.0.5-2008 Библиографическая ссылка. Общие требования и правила составления. Введен 2009.01.01. – М.: Стандартинформ, 2008. – 19 с.
5. Ковшов В.Н. Постановка инженерного эксперимента. – Киев: Вища школа, 1982. – 120 с.
6. Адлер Ю.П., Маркова Е.В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. – М.: Наука, 1976. – 280 с.
7. Таблицы планов эксперимента для факторных и полиномиальных моделей (Справочное издание) / Бродский В.З., Бродский Л.И., Голикова Т.И. и др. – М.: Металлургия, 1982. – 752 с.
8. Львовский Е.Н. Статистические методы построения эмпирических формул: учеб. пособ. для вузов. – М.: Высшая школа, 1988. – 239 с.
9. Кнотек М., Войта Р., Шефц И. Анализ металлургических процессов методами математической статистики. – М.: Металлургия, 1968. – 212 с.
10. Измайлова А.С., Торшина А.В. Методические указания по оформлению выпускных квалификационных работ. – Новотроицк: НФ НИТУ «МИСИС», 2022. – 86 с.

ШАПОВАЛОВ АЛЕКСЕЙ НИКОЛАЕВИЧ  
БИРЮКОВА ОЛЕСЯ ДМИТРИЕВНА  
КУНИЦИНА НАТАЛЬЯ ГЕННАДЬЕВНА

## ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА

Методические указания  
по выполнению домашнего задания / контрольной работы  
для студентов направления подготовки 22.03.02 Metallurgy  
Профиль – Обработка металлов давлением,  
всех форм обучения

Подписано в печать 26.12.2023 г.		
Формат 60x90 $\frac{1}{16}$ Рег. № 265	Печать цифровая Тираж 10 экз.	Уч.-изд.л. 3,0

Национальный исследовательский технологический университет «МИСИС»  
Новотроицкий филиал  
462359, Оренбургская обл., г. Новотроицк, ул. Фрунзе, 8.  
E-mail: [nf@misis.ru](mailto:nf@misis.ru)  
Контактный тел. 8 (3537) 679729.