**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ОРЕНБУРГСКИЙ ГОСУДАРСТВЕННЫЙ АГРАРНЫЙ УНИВЕРСИТЕТ»**

Кафедра «Математика и теоретическая механика»

МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

МАТЕМАТИЧЕСКАЯ ТЕОРИЯ ПЛАНИРОВАНИЯ ЭКСПЕРИМЕНТОВ

Направление подготовки 27.03.04 Управление в технических системах

Профиль образовательной программы Системы и средства автоматизации технологических процессов

Форма обучения очная

СОДЕРЖАНИЕ

1. КОНСПЕКТ ЛЕКЦИЙ

1.1 Лекция № 1 (2 часа).

Тема: «Основные понятия планирования эксперимента»

1.1.1 Вопросы лекции:

1. Эксперимент, его виды.
2. План эксперимента, этапы планирования.
3. Функция отклика.

1.1.2 Краткое содержание вопросов:

1. Эксперимент, его виды

Эксперимент – совокупность операций совершаемых над объектом исследования с целью получения информации о его свойствах.

Эксперимент, в котором исследователь по своему усмотрению может изменять условия его проведения, называется активным экспериментом. Если исследователь не может самостоятельно изменять условия его проведения, а лишь регистрирует их, то это пассивный эксперимент.

План эксперимента – совокупность данных определяющих число, условия и порядок проведения опытов.

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

− стремление к минимизации общего числа опытов;

− одновременное варьирование всеми переменными, определяющими процесс, по специальным правилам – алгоритмам;

− использование математического аппарата, формализующего многие действия экспериментатора;

− выбор четкой стратегии, позволяющей принимать обоснованные решения после каждой серии экспериментов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям, совокупность действий направленных на разработку стратегии экспериментирования (от получения априорной информации до получения работоспособной математической модели или определения оптимальных условий).

В зависимости от решаемых экспериментальным путем задач используются различные приемы организации эксперимента и методы обработки данных. В связи с этим методы планирования эксперимента можно условно выделить несколько видов.

Отсеивающий эксперимент используется на стадии предварительных исследований для выявления существенных факторов, т.е. для определения, какие из множества входных факторов оказывают влияние на выходной параметр, а какие нет.

Дисперсионный анализ – это математический метод, предназначенный для изучения объектов с факторами качественного типа. Он позволяет разложить общую дисперсию выходной величины на отдельные компоненты, обусловленные влиянием отдельных факторов. При дисперсионном анализе решаются вопросы о влиянии определенных факторов на интересующий нас показатель и какова сила влияния этого факторов.

Факторный анализ (определяет главные компоненты) относится к методам пассивного эксперимента. Суть метода заключается в переходе от пространства коррелированных входных факторов к пространству независимых переменных, называемых главными компонентами.

Изучение механизма явлений. Используется анализ для определения зависимости выходных параметров объекта исследования от входных факторов. Выделяют следующие группы планов регерессионного анализа: простейшая модель в виде полинома первой степени – линейной модель, планы второго порядка; планы третьего порядка и прочие планы не полиномиальных моделей, например, в виде ряда Фурье, а также ортогональные полиномы.

Корреляционный анализ применяют, если один из факторов качественный, а другие количественные.

Дискриминантный анализ применяют для разделения (классификации) выходного параметра на классы; используют в теории распознавания образов. Решение задачи дискриминантного анализа связано с ситуациями, когда выходной параметр имеет качественный характер.

Поиск оптимальных условий имеет основной целью найти условие, т.е. значения входных факторов, при которых выходной параметр объекта принимает экстремальные значения. Экстремальные планы делятся на одномерные и многомерные, градиентные и не градиентные.

Адаптивная оптимизация (эволюционное планирование) используется в тех случаях, когда под влиянием процессов старения, изменения свойств сырья, дрейфа и других неконтролируемых причин оптимум в области управляющих факторов может смещаться. Поиск оптимума нередко усложняется недопустимостью варьирования факторов в широких пределах из-за опасности раскачивания процесса и риска получения бракованной продукции.

Цель планирования эксперимента – нахождение таких условий и правил проведения опытов, при которых удается получить надежную и достоверную информацию об объекте с наименьшей затратой труда, а также представить эту информацию в компактной и удобной форме с количественной оценкой точности.

2. План эксперимента, этапы планирования

Экспериментальные исследования включают в себя различные формы эмпирического познания: *наблюдение, сравнение, контроль, измерение*.

Правильно выбранный план проведения эксперимента *позволяет*:

· отобрать наиболее существенные факторы, влияющие на объект исследования и, тем самым ускорить процесс исследований, снизить его стоимость

· разработать научно обоснованную программу исследований, дающую достоверную оценку результатов эксперимента при снижении затрат на его проведение.

Экспериментом называется строгая последовательность заранее определенных действий, которая ведет к получению одной или множества величин, представляющих результат эксперимента.

Эксперимент должен *обладать* свойством *воспроизводимости*, повторяемости. Это означает, что при повторении экспериментов его результаты должны быть близкими предыдущим.

Соотношение полезного эффекта, полученного в результате эксперимента, и затрат на его проведение определяет *эффективность* эксперимента.

Структура планирования эксперимента должна охватывать следующие основные *задачи*:

· Выбор цели эксперимента,
· Выбор исходных данных,
· Выбор оборудования,
· Проведение эксперимента,
· Обработка результатов,
· Анализ результатов.

Задача выбора цели эксперимента важна и определяет все аспекты его планирования.

Для проведения эксперимента *нужно*:

· Иметь представление о диапазоне, характеристике измеряемой величины.

· Обосновано выбрать метод измерения, наиболее подходящий для решения поставленной задачи.

· Выбрать средство измерения для данного метода.

· Продумать и составить рациональную методику измерения для данного метода и средства.

· Оценить, с какой точностью нужно производить измерения.

· Знать какие погрешности и вследствие чего могут возникнуть, искать способы исключения уменьшения погрешности.

Целесообразно строить исследования сложного технологического объекта по следующим основным этапам.

· Ознакомление с технологией и аппаратурным оформлением производственного процесса, изучение текущей технической документации и результатов проведенных ранее экспериментов. На основе полученных данных оцениваются числовые характеристики случайных переменных, автокорреляционные и взаимные корреляционные функции, строятся двухмерные поля корреляции и т.д.; на основе этого оцениваются сведения организации основного эксперимента.

· Планирование и проведение основного эксперимента, сбор и обработка основного массива экспериментально-статистических данных с целью получения математического описания. На этом этапе уточняются точка съема данных, включаются в рассмотрение новые переменные, выбирается центр основного эксперимента, определяется интервал времени для обеспечения независимости соседних измерений, оценивается необходимое время эксперимента и объем статистических данных.

· Анализ и интерполяция полученной математической модели. Здесь выявляются существенные переменные, отсеиваются несущественные переменные, оценивается работоспособность математической модели.

· Совершенствование математического описания с целью повышения его работоспособности.

· Использование усовершенствованной и работоспособной математической модели для непосредственного управления процессом.

Условия правильной организации пассивного эксперимента во времени следующие:

· одновременные замеры всех намеченных переменных выполняются во всех намеченных точках объекта;

· последовательные измерения должны осуществляться строго через равные промежутки времени.

3. Функция отклика

Пусть интересующее нас свойство (Y) объекта зависит от нескольких (n) независимых переменных ($X_1, X_2, \ldots, X_n$) и мы хотим выяснить характер этой зависимости – $Y=F(X_1, X_2, \ldots, X_n)$, о которой мы имеем лишь общее представление. Величина Y – называется «отклик», а сама зависимость $Y=F(X_1, X_2, \ldots, X_n)$ – «функция отклика».

Отклик должен быть определен количественно. Однако могут встречаться и качественные признаки Y. В этом случае возможно применение рангового подхода. Пример рангового подхода – оценка на экзамене, когда одним числом оценивается сложный комплекс полученных сведений о знаниях студента.

Независимые переменные $X_1, X_2, \ldots, X_n$ – иначе факторы, также должны иметь количественную оценку. Если используются качественные факторы, то каждому их уровню должно быть присвоено какое-либо число. Важно выбирать в качестве факторов лишь независимые переменные, т.е. только те которые можно изменять, не затрагивая другие факторы. Факторы должны быть однозначными. Для построения эффективной математической модели целесообразно провести предварительный анализ значимости факторов (степени влияния на функцию), их ранжирование и исключить малозначащие факторы.

Диапазоны изменения факторов задают область определения *Y*. Если принять, что каждому фактору соответствует координатная ось, то полученное пространство называется факторным пространством. При n=2 область определения Y представляется собой прямоугольник, при n=3 – куб, при n >3 – гиперкуб.

При выборе диапазонов изменения факторов нужно учитывать их совместимость, т.е. контролировать, чтобы в этих диапазонах любые сочетания факторов были бы реализуемы в опытах и не приводили бы к абсурду. Для каждого из факторов указывают граничные значения

$$X_{i\min} \le X_i \le X_{i\max}, \ i = 1, ..., n.$$

Регрессионный анализ функции отклика предназначен для получения ее математической модели в виде уравнения регрессии

$$Y = F(X_1, X_2, ..., X_n; B_1, B_2, ..., B_m) + e,$$

где B_1, ..., B_m – некоторые коэффициенты; e – погрешность.

Среди основных методов планирования, применяемых на разных этапах исследования, используют:

- планирование отсеивающего эксперимента, основное значение которого выделение из всей совокупности факторов группы существенных факторов, подлежащих дальнейшему детальному изучению;
- планирование эксперимента для дисперсионного анализа, т.е. составление планов для объектов с качественными факторами;
- планирование регрессионного эксперимента, позволяющего получать регрессионные модели (полиномиальные и иные);
- планирование экстремального эксперимента, в котором главная задача – экспериментальная оптимизация объекта исследования;
- планирование при изучении динамических процессов и т.д.

Инициатором применения планирования эксперимента является Рональд А. Фишер, другой автор известных первых работ – Френк Йетс. Далее идеи планирования эксперимента формировались в трудах Дж. Бокса, Дж. Кифера. В нашей стране – в трудах Г.К. Круга, Е.В. Маркова и др.

В настоящее время методы планирования эксперимента заложены в специализированных пакетах, широко представленных на рынке программных продуктов, например: StatGrapfics, Statistica, SPSS, SYSTAT и др.

1.2 Лекция № 2 (2 часа).

Тема: «Обработка результатов измерений»

1.2.1 Вопросы лекции:

1. Прямые и косвенные измерения
2. Критерии оценки грубых погрешностей
3. Определение числа повторностей опыта

1.2.2 Краткое содержание вопросов:

1. Прямые и косвенные измерения

Задача обработки результатов измерений заключается в нахождении приближенного значения или оценки или оценки измеряемой величины X и указания его среднего квадратического отклонения. Если измерения проводились по одной и той же методике средствами измерения одинаковой точности при постоянных внешних условиях, то такие измерения называют прямые равноточные измерения.

Для таких измерений находят: 1) среднее арифметическое значений; 2) среднее квадратичное отклонение; 3) ошибку средней; 4) доверительный интервал, используя критерий Стьюдента.

В некоторых случаях одну и ту же величину необходимо измерить, используя различные методы и средства измерений – прямые неравноточные измерения. В этом случае находят: 1) среднее взвешенное; 2) среднее квадратическое отклонение.

При косвенных измерениях уравнения имеют вид $A = f(X;Y)$, т.е. искомую величину не измеряют, а вычисляют по результатам измерений других величин, при этом погрешность зависит не только от погрешностей величин Х, У, но и от вида функциональной зависимости. Тогда приближенное значение измеряемой величины находят по формуле $\overline{a} = f(\overline{x};\overline{y})$, где $\overline{x}$, $\overline{y}$ - средние арифметические значения, полученные по результатам прямых измерений.

2. Критерии оценки грубых погрешностей

В процессе измерений, последующей обработки данных, а также формализации результатов в виде математической модели, возникают погрешности и теряется часть информации, содержащейся в исходных данных. Применение методов планирования эксперимента позволяет определить погрешность математической модели и судить о ее адекватности. Если точность модели оказывается недостаточной, то применение методов планирования эксперимента позволяет модернизировать математическую модель с проведением дополнительных опытов без потери предыдущей информации и с минимальными затратами.

Источником грубых погрешностей являются ошибки, допущенные во время измерений:

- неправильный отсчет по шкале измерительного прибора;
- неправильные записи результатов наблюдений.

Вопрос о грубых ошибках решается общими методами проверки статистических гипотез.

3. Определение числа повторностей опыта

Большинство экспериментальных исследований с применением математических методов планирования экспериментов имеет своей целью определение оптимальных условий протекания исследуемого процесса, обеспечение максимальной эффективности исследуемого объекта. Конечный результат выдается в виде доверительного интервала $y = \overline{y} + \varepsilon(\overline{y})$, в котором $\varepsilon(\overline{y})$ может задаваться заранее величиной относительной доверительной ошибки среднеарифметической оценки результата опыта: $\alpha = \dfrac{\varepsilon(\overline{y})}{\overline{y}} \cdot 100\%$.

Если сравнить полученные оценки

а) $y = \overline{y} + \varepsilon(\overline{y}) = 98{,}2 \pm 8{,}9$

б) $y = \overline{y} + \varepsilon(\overline{y}) = 99{,}15 \pm 2{,}38$,

то более точным будет прогнозирование по оценке б), чем по а). Для получения результата б) приходиться проводить исследования при большем числе повторностей опытов, при использовании приборов более высокого класса точности, т.е. следует ожидать увеличения затрат на само исследование. Величина $\varepsilon(\overline{y})$ задается с учетов всех этих условий. Обычно бывает достаточно принять $\varepsilon(\overline{y}) = (0{,}05 \div 0{,}1)\overline{y}$ и доверительную вероятность $P = 0{,}95$. Неизвестная $\overline{y}$ оценивается приблизительно.

Определение числа повторностей при заданной величине $\varepsilon(\overline{y})$ может осуществляться различными методами:

1. При известном среднем квадратическом отклонении σ число повторностей опыта вычисляется по формуле: $m = \frac{t^2(P)\sigma^2}{\varepsilon^2(\bar{y})}$, где $t(P)$ – параметр, определяемый из таблиц в зависимости от принятой доверительной вероятности P.

2. Если в распоряжении исследователя имеется лишь экспериментальная оценка дисперсии $s^2(y_k)$ с числом степеней свободы f, то используя границы доверительного интервала оценки дисперсии $\sigma^2{}_{\max}$ и $\sigma^2{}_{\min}$, с заданной вероятностью P найти доверительный интервал для искомого числа повторностей, т.е. $m_{\min} \le m \le m_{\max}$:

$$m_{\max} = \frac{t^2(P)\sigma_{\max}{}^2}{\varepsilon^2(\bar{y})},\ m_{\min} = \frac{t^2(P)\sigma^2{}_{\min}}{\varepsilon^2(\bar{y})}.$$

$\sigma^2{}_{\min} = s^2(y_k) \cdot Z_1^2$, $\sigma^2{}_{\max} = s^2(y_k) \cdot Z_2^2$.

Параметры Z_1^2, Z_2^2 берутся из таблицы, определяющей величину доверительной оценки дисперсий в зависимости от принятой доверительной вероятности P и числа степеней свободы f оценки дисперсии $s^2(y_k)$. К реализации приходится принимать $m = m_{\max}$, что гарантирует с вероятностью P получение заданной доверительной ошибки $\varepsilon(\bar{y})$.

3. Если экспериментатор лишь приступает к исследованию, то определение необходимого числа повторностей для получения заданной величины $\varepsilon(\bar{y})$ следует осуществлять одновременно с последовательной постановкой опытов по определению $s^2(y_k)$. Если осуществить $m_1 = 3 \div 5$, то рассчитав $s^2(\bar{y}) = \frac{1}{m} s^2(y_k)$ и $t(P_1; f_1) = \frac{\varepsilon(\bar{y})}{s(\bar{y})}$ по таблица Стьюдента при $f_1 = m_1 - 1$ можно определить доверительную вероятность P_1. Если она будет меньше той, которая требуется, то надо увеличивать число повторностей и периодически производить соответствующие расчеты. Увеличение числа повторностей прекращают, как только достигается заданная величина P.

1. 3 Лекция № 3 (2 часа).

Тема: «МНК для однофакторного эксперимента»

1.3.1 Вопросы лекции:

1. Формализация экспериментальных данных
2. МНК и его применение для однофакторного эксперимента

1.3.2 Краткое содержание вопросов:

1. Формализация экспериментальных данных

При использовании методов планирования эксперимента необходимо найти ответы на 4 вопроса:

- Какие сочетания факторов и сколько таких сочетаний необходимо взять для определения функции отклика?
- Как найти коэффициенты B_0, B_1, …, B_m?
- Как оценить точность представления функции отклика?
- Как использовать полученное представление для поиска оптимальных значений Y?

Геометрическое представление функции отклика в факторном пространстве X_1, X_2, …, X_n называется поверхностью отклика (рис. 1).

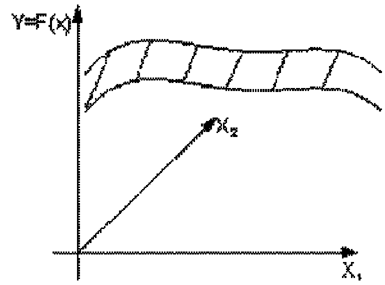


Рис. 1. Поверхность отклика

Если исследуется влияние на Y лишь одного фактора X_1, то нахождение функции отклика – достаточно простая задача. Задавшись несколькими значениями этого фактора, в результате опытов получаем соответствующие значения Y и график $Y = F(X)$ (рис. 2).

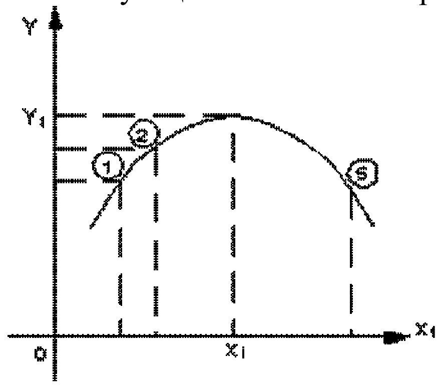


Рис. 2. Построение функции отклика одной переменной по опытным данным

По его виду можно подобрать математическое выражение функции отклика. Если мы не уверены, что опыты хорошо воспроизводятся, то обычно опыты повторяют несколько раз и получают зависимость с учетом разброса опытных данных.

Если факторов два, то необходимо провести опыты при разных соотношениях этих факторов. Полученную функцию отклика в 3^X-мерном пространстве (рис. 1) можно анализировать, проводя ряд сечений с фиксированными значениями одного из факторов (рис. 3). Вычлененные графики сечений можно аппроксимировать совокупностью математических выражений.

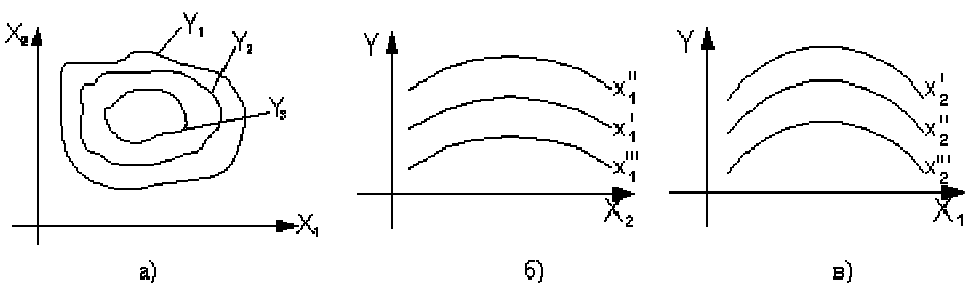


Рис. 3. Сечения поверхности отклика при фиксированных откликах (а) и переменных (б, в)

При трех и более факторах задача становится практически неразрешимой. Если и будут найдены решения, то использовать совокупность выражений достаточно трудно, а часто и не реально.

2. МНК и его применение для однофакторного эксперимента

Пусть в результате опыта зависимость между переменными x и y выражена в виде таблицы:

x	x_1	x_2	. . .	x_i	. . .	x_n
y	y_1	y_2	. . .	y_i	. . .	y_n

В системе координат xOy построим точки $M_1(x_1;y_1), M_2(x_2;y_2)$, ... , $M_i(x_i;y_i)$, ..., $M_n(x_n;y_n)$ (рис.).

Предположим, что указанные точки расположены в достаточной близости от некоторой прямой (l), уравнение которой $y = ax + b$.

Как найти значения параметров «a» и «b» прямой, пользуясь приведенной выше таблицей?

Рассмотрим точку $M_i(x_i;y_i)$, которая в общем случае не лежит на прямой, и точку $N_i(x_i;ax_i + b)$, которая лежит на прямой (l). Разность ординат $(ax_i + b) - y_i$ характеризует отклонение точки $M_i(x_i;y_i)$ от прямой (l). Эта разность будет положительной, если рассматриваемая точка расположена ниже прямой и отрицательной, если точка расположена выше прямой. Выберем значения параметров a и b таким образом, чтобы сумма квадратов этих разностей была наименьшей. В этом состоит способ наименьших квадратов. Итак, определим наименьшее значение функции двух переменных:

$$S = f(a;b) = (ax_1 + b - y_1)^2 + (ax_2 + b - y_2)^2 + ... + (ax_i + b - y_i)^2 + ... + (ax_n + b - y_n)^2 \quad (1)$$

Пользуясь необходимыми условиями существования экстремума, будем иметь: $\frac{\partial S}{\partial a} = 0$ и $\frac{\partial S}{\partial b} = 0$. Или:

$$\frac{\partial S}{\partial a} = 2(ax_1 + b - y_1)\cdot x_1 + 2(ax_2 + b - y_2)\cdot x_2 + ... + 2(ax_i + b - y_i)\cdot x_i + ...$$
$$... + 2(ax_n + b - y_n)\cdot x_n = 0$$
$$\frac{\partial S}{\partial b} = 2(ax_1 + b - y_1) + 2(ax_2 + b - y_2) + ... + 2(ax_i + b - y_i) + ... + 2(ax_n + b - y_n) = 0$$

После простейших преобразований получим следующую систему двух линейных уравнений относительно неизвестных a и b:

$$\begin{cases} a(x_1^2 + x_2^2 + ... + x_n^2) + b(x_1 + x_2 + ... + x_n) - (x_1y_1 + x_2y_2 + ... + x_ny_n) = 0 \\ a(x_1 + x_2 + ... + x_n) + bn - (y_1 + y_2 + ... + y_n) = 0 \end{cases}$$

Или, введя символы суммы, получим:

$$\begin{cases} a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i + bn = \sum_{i=1}^{n} y_i \end{cases} \quad (2)$$

Систему (2) можно представить иначе, если ввести средние значения величин x, y, x^2, xy. Пусть

$$\overline{x} = \frac{\sum_{i=1}^{n} x_i}{n}; \quad \overline{y} = \frac{\sum_{i=1}^{n} y_i}{n}; \quad \overline{xy} = \frac{\sum_{i=1}^{n} x_i y_i}{n}; \quad \overline{x^2} = \frac{\sum_{i=1}^{n} x_i^2}{n}.$$

Деля почленно уравнения (2) на n и подставляя средние значения $\overline{x}, \overline{y}, \overline{x^2}, \overline{xy}$, получим систему:

$$\begin{cases} a\overline{x^2} + b\overline{x} = \overline{xy} \\ a\overline{x} + b = \overline{y} \end{cases} \quad (3)$$

Решая систему (2) или (3), получим такие значения параметров a и b, при которых уравнение $y = ax + b$ наилучшим образом выражает зависимость между переменными x и y по их значениям, данным в таблице.

Способом наименьших квадратов можно пользоваться и тогда, когда точки $M_1(x_1; y_1), M_2(x_2; y_2)$, ... , $M_i(x_i; y_i)$, …, $M_n(x_n; y_n)$ расположены в достаточной близости от некоторой параболы $y = ax^2 + bx + c$. В этом случае полагают, что переменные x и y связаны между собой квадратической функцией $y = ax^2 + bx + c$. Нужно найти значения a, b и c так, чтобы сумма квадратов отклонений была наименьшей.

Аналогично случаю линейной зависимости между переменными x и y, можно получить систему трех уравнений с тремя неизвестными a, b и c:

$$\begin{cases} a\sum_{i=1}^{n} x_i^4 + b\sum_{i=1}^{n} x_i^3 + c\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2 y_i \\ a\sum_{i=1}^{n} x_i^3 + b\sum_{i=1}^{n} x_i^2 + c\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i y_i \\ a\sum_{i=1}^{n} x_i^2 + b\sum_{i=1}^{n} x_i + c\cdot n = \sum_{i=1}^{n} y_i \end{cases} \qquad (4)$$

Если ввести средние значения величин $x, y, x^2, x^3, x^4, xy, x^2 y$, то система (4) примет вид:

$$\begin{cases} a\overline{x^4} + b\overline{x^3} + c\overline{x^2} = \overline{x^2 y} \\ a\overline{x^3} + b\overline{x^2} + c\overline{x} = \overline{xy} \\ a\overline{x^2} + b\overline{x} + c = \overline{y} \end{cases} \qquad (5)$$

1. 4 Лекция № 4 (2 часа).

Тема: «Обобщение МНК»

1.4.1 Вопросы лекции:

1. Симметричный и равномерный план для однофакторного эксперимента
2. Проверка адекватности полученного уравнения.
3. Обобщение метода наименьших квадратов на многофакторный линейный случай

1.4.2 Краткое содержание вопросов:

1. Симметричный и равномерный план для однофакторного эксперимента

Задача в значительной степени упростится, если при планировании эксперимента, можно будет обеспечить условие: $\sum C_u = 0$.

При натуральной размерности факторов выполнить условие $\sum C_u = 0$ невозможно, т. к. в этом случае величина фактора должна иметь как положительные значения, так и отрицательные. Если же точку отсчета величины фактора перенести в середину диапазона изменения фактора (центр эксперимента) $C_0 = \dfrac{C_{\max} + C_{\min}}{2}$, то появляется возможность удовлетворить условию $\sum C_u = 0$ в виде $\sum_{u=1}^{N} C'_u = 0$, где $C'_u = C_u - C_0$.

Для равномерного плана $C_u - C_{(u-1)} = \lambda = const$,
где λ – интервал варьирования фактора.

Условие $\sum C_u = 0$ может быть выполнено, если для обозначения величины фактора использовать безразмерные выражения: $x_u = \frac{C_u}{\lambda} = \frac{C_u - C_0}{\lambda}$, отсюда легко увидеть, что условие $\sum_{u=1}^{N} x_u = 0$ эквивалентно условию $\sum C_u = 0$ и такие планы называют симметричными.

При составлении плана диапазон фактора ориентировочно ограничивают величинами C_{min} и C_{max}, назначенными после изучения литературы по теме исследования. От опыта к опыту предусматривают такое изменение величины фактора, которое позволило бы достоверно уловить имеющимися в распоряжении исследователя приборами изменение выхода процесса y'_u.

С учетом величины λ и диапазона ($C_{max} - C_{min}$) определяют число опытов, округляя его до нечетного N:

$$N = \frac{C_{max} - C_{min}}{\lambda} + 1$$

Затем определяют величины факторов в каждом из N опытов и уточняют исследуемый диапазон фактора $C_N - C_1$:

$$C_u = C_0 + \lambda\left(u - \frac{N+1}{2}\right) = C_0 + \lambda x_u$$

где x_u – безразмерное выражение фактора, аналогичное полученному по соотношению $x_u = u - \frac{N+1}{2} = \frac{C_u - C_0}{\lambda}$.

Число опытов эксперимента может быть четным или нечетным, и, как правило, должно быть больше числа коэффициентов N' уравнения.

Чем больше разность (N – N'), тем с большей точностью можно получить оценки коэффициентов данного уравнения и тем в большей степени эти оценки будут освобождены от влияния случайных неуточненных факторов.

2. Проверка адекватности полученного уравнения

Для проверки гипотезы адекватности модели необходимо сопоставить достигнутую точность модели с величиной, характеризующей точность наблюдений. Если ошибки, характеризующие точность модели, превосходят ошибки наблюдений, то гипотеза адекватности модели отклоняется. Для проверке гипотезы об адекватности необходимо, чтобы в заданных точках были проведены повторные наблюдения. Только в этом случае возможно провести оценку дисперсии ошибок наблюдений, связанную с ошибками проведения экспериментов.

Допустим в каждой точке наблюдений проведено m_i экспериментов. Тогда y_{ij} - это $j-$е наблюдение в $i-$й точке. Вычислим для каждой точке среднее значение: $\overline{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$. Если $\hat{y}-$оценка среднего значения наблюдений, то $Q_1 = \sum_{i=1}^{N}\left(\overline{y_i} - \hat{y}_i\right)^2$ обусловлена неадекватностью модели, она характеризует, насколько средние по наблюдениям в точках значения выхода отличаются от значений, предсказываемых моделью. Сумма квадратов $Q_2 = \sum_{i=1}^{N}\sum_{j=1}^{m_i}\left(y_{ij} - \overline{y_i}\right)^2$ характеризует ошибку проведенных экспериментов. Обозначим: $\varphi_1 = N - k - 1$; $\varphi_2 = N(m-1)$. Если гипотеза адекватности истина, то $s^2{}_r = \frac{Q_1}{\varphi_1}$ и $s^2{}_e = \frac{Q_2}{\varphi_2}$ будут несмещенными оценками дисперсии наблюдений и величина $F = \frac{s^2{}_r}{s^2{}_e}$ - является случайной величиной, имеющей распределение Фишера. Далее проверяется гипотеза по критерию Фишера. При отклонении гипотезы адекватности считается, что модель неадекватна и необходимо выбирать другую модель. Может возникнуть необходимость в проведении дополнительных наблюдений для проверки новой гипотезы.

3. Обобщение метода наименьших квадратов на многофакторный линейный случай

Пусть имеется k факторов и известно, что отклик и факторы связаны линейно $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k$. Выпишем для этого случая матрицы X, Y, B:

$$X_{N\times(k+1)} = \begin{pmatrix} x_{01} & x_{11} & \cdots & x_{k1} \\ x_{02} & x_{12} & \cdots & x_{k2} \\ \cdots & \cdots & \cdots & \cdots \\ x_{0N} & x_{1N} & \cdots & x_{kN} \end{pmatrix},\ Y_{N\times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \cdots \\ y_N \end{pmatrix},\ B_{(k+1)\times 1} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_k \end{pmatrix}.$$

Исходная система линейных уравнений имеет вид $Y = X \cdot B$, откуда неизвестная матрица $B = \left(X^T X\right)^{-1} X^T Y$.

Матрицу системы нормальных уравнений можно записать в виде:

$$X^T X = \begin{pmatrix} \sum x_0^2 & \sum x_0 x_1 & \sum x_0 x_2 & \cdots & \sum x_0 x_k \\ \sum x_0 x_1 & \sum x_1{}^2 & \sum x_1 x_2 & \cdots & \sum x_1 x_k \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sum x_0 x_k & \sum x_1 x_k & \sum x_2 x_k & \cdots & \sum x_k{}^2 \end{pmatrix},\ X^T Y = \begin{pmatrix} \sum y x_0 \\ \sum y x_1 \\ \cdots \\ \sum y x_k \end{pmatrix}.$$

Чтобы получить ответ, т.е. вектор B, остается обратить матрицу $X^T X$ и умножить обратную матрицу на $X^T Y$.

1. 5 Лекция № 5 (2 часа).

Тема: «Полный факторный эксперимент»

1.5.1 Вопросы лекции:

1. Двухуровневые планы многофакторного эксперимента.
2. План ПФЭ 2^2 и его геометрическое изображение.
3. План ПФЭ 2^3 и его геометрическое изображение.
4. Многомерные ПФЭ типа 2^n.

1.5.2 Краткое содержание вопросов:

1. Двухуровневые планы многофакторного эксперимента

Изучение процесса функционирования объекта позволяет выявить факторы, оказывающие существенное влияние на функцию отклика. Выбор существенных переменных потенциально определяет степень достижения адекватности получаемой модели: отсутствие в исходном перечне существенных параметров, да еще и произвольно меняющихся в ходе эксперимента, не позволяет правильно решить задачу оптимизации; включение несущественных параметров усложняет модель, вызывает значительное увеличение объема экспериментов, хотя по результатам исследования несущественность соответствующих параметров будет выявлена.

Для каждой переменной следует определить диапазон и характер изменения (непрерывность или дискретность). Ограничения на диапазон изменений могут носить принципиальный или технический характер. Принципиальные ограничения факторов не могут быть нарушены при любых обстоятельствах. Эти ограничения задаются исходя из физических представлений (например, емкость устройств памяти всегда имеет положительное значение). Второй тип ограничений связан с технико-экономическими соображениями, например, с наличием соответствующего аппаратно-программного комплекса, принятой технологией обработки информации.

Выделение области изменения факторов является не формальной задачей, а основывается на опыте исследователя. В рамках области допустимых значений факторов необходимо выделить начальную область планирования эксперимента. Этот выбор включает определение основного (нулевого) уровня как исходной точки построения плана и интервалов варьирования. Интервал варьирования задает относительно основного уровня значения фактора, при которых будут производиться эксперименты. Обычно интервалы являются симметричными относительно центрального значения. Интервал варьирования должен отвечать двум ограничениям: его применение не должно приводить к выходу фактора за пределы области допустимых значений; он должен быть больше погрешности задания значений фактора (в противном случае уровни фактора станут не различимыми). В пределах этих ограничений выбор конкретного значения является неформальной процедурой, учитывающей ориентировочную информацию о кривизне поверхности функции отклика.

Фактор должен быть управляемым, т.е. экспериментатор может поддерживать его постоянное значение в течение всего опыта. Для фактора необходимо указать его конкретные значения и средства контроля. Сам фактор должен быть первичным, ибо сложно управлять фактором, который в свою очередь является функцией других факторов. Для каждого фактора следует указать точность его задания и поддержания в ходе эксперимента.

Одновременное изменение факторов предполагает их совместимость, что означает осуществимость и безопасность всех их сочетаний. Необходимо также обеспечить независимость изменения каждого фактора, что означает возможность установления любого значения фактора вне связи со значениями других факторов.

Цель исследования, требуемая точность получаемых результатов, имеющиеся ресурсы ограничивают множество допустимых моделей функции отклика (с усложнением модели и повышением точности оценки показателей резко возрастает объем необходимых опытов) и соответственно предопределяют план проведения экспериментов.

При экспериментальном исследовании объекта качество полученных результатов и произведенные затраты определяются организацией эксперимента. В связи с этим, помимо исследования эффективных методов обработки экспериментальных данных, необходимо оптимизировать сам эксперимент по следующим параметрам: числу учитываемых входных факторов, интервалам варьирования и числу уровней варьирования факторов, сочетанием их и последовательностью проведения опытов.

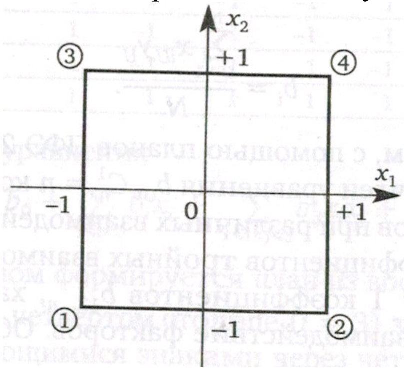
При проведении эксперимента необходимо включать в рассмотрение все факторы, которые могут влиять на выходной параметр. В ходе эксперимента последовательно можно устанавливать несколько значений уровней факторов. Такой эксперимент называется *полным факторным экспериментом* (ПФЭ). Эксперимент, в котором встречаются не все возможные сочетания уровней факторов, называется *дробным факторным экспериментом (ДФЭ)*.

Если факторы варьируются на (р) уровнях, то число опытов для ПФЭ будет $N = p^k$, где к – число факторов. Поэтому выбор сочетаний уровней факторов, их числа и порядка постановки опытов является целью планирования эксперимента.

Ортогональным называется план с попарно ортогональными вектор-столбцами матрицы планирования. Использование критерия ортогональности имеет целью упрощение вычислений и получение независимых оценок коэффициентов. При этом замена любого коэффициента в модели не изменяет оценок остальных коэффициентов.

2. План ПФЭ 2^2 и его геометрическое изображение

В ряде случаев реальные объекты имеют нелинейные зависимости. Один из часто встречающихся видов нелинейности связан с эффектом взаимодействия, когда значение одного фактора зависит от значения другого фактора. Для того, чтобы найти оценку коэффициента при парном взаимодействии, необходимо, пользуясь правилом перемножения столбцов, получить столбец произведения двух факторов.



На рисунке представлена матрица планирования эксперимента с варьированием двух факторов по двум уровням («+1» - верхний уровень, «-1» – нижний уровень сигнала).

№ опыта	x_0	x_1	x_2	x_1x_2	y
1	+1	+1	+1	+1	y_1
2	+1	-1	+1	-1	y_2
3	+1	+1	-1	-1	y_3
4	+1	-1	-1	+1	y_4

Матрица планирования должна обладать следующими *свойствами*:

- симметрией, т.е. сумма элементов столбца каждого фактора должна быть равна нулю относительно центра эксперимента (центра между уровнями);

- нормальностью, т.е. сумма квадратов элементов каждого столбца должна быть равна числу опытов;

- ротабельностью, т.е. равенством точности предсказания значений отклика на равных расстояниях от центра эксперимента.

3. План ПФЭ 2^3 и его геометрическое изображение

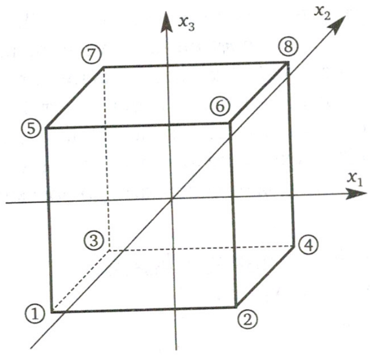
Пусть функция отклика $y = f(x_1; x_2; x_3)$. Все различные комбинации уровней переменных $x_1; x_2; x_3$ в ПФЭ 2^3 представлены в таблице:

Матрица независимых переменных								наблюдение
x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	
1	-1	-1	-1	1	1	1	-1	y_1
1	1	-1	-1	-1	-1	1	1	y_2
1	-1	1	-1	-1	1	-1	1	y_3
1	1	1	-1	1	-1	-1	-1	y_4
1	-1	-1	1	1	-1	-1	1	y_5
1	1	-1	1	-1	1	-1	-1	y_6
1	-1	1	1	-1	-1	1	-1	y_7
1	1	1	1	1	1	1	1	y_8

Для плана ПФЭ 2^3 число факторов $n=3$ и выполняется $N=2^3=8$ опытов. Уравнение может содержать до восьми членов:

$$y=b_0+b_1x_1+b_2x_2+b_3x_3+b_{12}x_1x_2+b_{13}x_1x_3+b_{23}x_2x_3+b_{123}x_1x_2x_3 .$$

Произведение $x_ix_j(i\neq j)$ назовем парным взаимодействием или взаимодействием первого порядка факторов x_i и x_j. Произведение $x_1x_2x_3$ назовем тройным взаимодействием или взаимодействием второго порядка факторов $x_1;x_2;x_3$. Столбцы $x_1;x_2;x_3$ образуют план эксперимента, а другие столбцы носят вспомогательный характер. Геометрическое изображение плана ПФЭ 2^3 с указанием номеров точек плана в факторном пространстве представлено на рисунке (точки плана располагаются в вершинах куба):



4. Многомерные ПФЭ типа 2^n

Эксперимент, в котором уровни каждого фактора комбинируются со всеми уровнями других факторов, есть ПФЭ.

План эксперимента задает множество точек проведения экспериментов

$$x^i=\left(x_1^i,\,x_2^i,...,\,x_k^i\right),\ i=1,2,...,N,$$

а совокупность значений факторов во всех *N* точках плана эксперимента образует матрицу плана $\overline{D}=\left(x_j^i\right)$.

Строки матрицы соответствуют опытам, столбцы – факторам, элемент матрицы x_j^i задает значение *j*-го фактора в i-м опыте. На практике широкое распространение получили представления функции отклика в виде линейной по параметрам модели регрессионного анализа. Вследствие влияния на результаты экспериментов случайных воздействий истинные значения коэффициентов модели можно определить только приближенно. Оценку вектора неизвестных параметров β находят с помощью метода наименьших квадратов по результатам экспериментов.

Формируя специальным образом матрицу плана, можно получить независимые оценки компонент вектора неизвестных параметров b, которые будут характеризовать вклад каждого фактора в значение функции отклика.

Итак, задача заключается в определении общей формы записи функции отклика y, в большинстве случаев вид которой, получаемый из теоретических соображений, является сложным для практического применения, а при неполном знании объекта вообще неизвестен. По данным причинам функцию целесообразно представить в универсальном, удобном для практического применения виде, чему соответствует полиномиальный. Тогда системой базисных функций будет совокупность степенных функций с целыми неотрицательными показателями степени. Полиномиальная форма представления функции отклика примет вид

$$y=b_0+b_1x_1+...+b_kx_k+b_{12}x_1x_2+b_{13}x_1x_3+...+b_{k-1k}x_{k-1}x_k+b_{11}x_k^2+...+\varepsilon ,$$

где ε – случайная величина, характеризующая ошибку опыта.

Такая функция отклика линейна относительно неизвестных коэффициентов и будет полностью определена, если заданы степень полинома и коэффициенты. Степень полинома обычно устанавливается исследователем априорно и уточняется в ходе исследования. На практике наибольшее распространение получили полиномы первого и второго порядка, т.е. линейные и квадратичные модели. Коэффициенты полинома принято называть эффектами факторов.

Применение ТПЭ основано на ряде допущений.

1. Функция отклика содержит в своем составе неслучайную и случайную составляющие. Оценки коэффициентов формируются путем усреднения результатов наблюдений. Поэтому при достаточно большом количестве наблюдений можно считать, что случайная составляющая ε распределена по нормальному закону с нулевым математическим ожиданием и дисперсией σ^2, не зависящей от значений факторов, что позволяет получить несмещенную оценку математического ожидания функции отклика в конкретной точке плана. Иначе говоря, результаты, полученные путем усреднения повторных опытов в каждой точке плана, представляют собой независимые, нормально распределенные случайные величины.

2. Факторы $x_1, x_2, \dots, x_k$ отличаются пренебрежимо малыми ошибками по сравнению с ошибкой в определении величины y (учет помех в задании факторов приводит к трудноразрешимым проблемам в оценке коэффициентов функции отклика). Возникновение ошибки в определении значения функции отклика объясняется не столько погрешностью измерений, сколько влиянием на результат работы системы неучтенных или случайных факторов.

3. Дисперсии среднего значения функции отклика в различных точках равны друг другу (выборочные оценки дисперсии однородны), поэтому при многократных повторных наблюдениях над величиной y при некотором наборе значений $x_{1u}, x_{2u}, \dots, x_{ku}$, получаемая оценка дисперсии среднего значения не будет отличаться от оценки дисперсии, полученной при многократных наблюдениях для любого другого набора значений независимых переменных $v_{1s}, v_{2s}, \dots, v_{ks}$.

Указанные допущения позволяют использовать для расчетов коэффициентов полинома МНК, применение которого, вообще говоря, не требует, чтобы результаты наблюдения были нормально распределены. Этот метод в любом случае дает решение, минимизирующее сумму квадратов отклонений результатов наблюдения от значений функции отклика. Допущение о нормальном распределении используется, например, при проверке адекватности функции отклика.

1. 6 Лекция № 6 (2 часа).

Тема «Дробный факторный эксперимент»

1.6.1 Вопросы лекции:

1. Выбор дробности плана.
2. Реплики от ПФЭ.

1.6.2 Краткое содержание вопросов:

1. Выбор дробности плана.

При увеличении числа входных факторов общее число опытов в ПФЭ резко увеличивается. При построении линейных моделей имеется возможность сокращения числа опытов за счет потери информации, несущественной для данного вида моделей.

Пользуясь таким планированием, можно вычислить четыре коэффициента для рассматриваемого примера. Если имеется основание считать, что эффект парного взаимодействия отсутствует, то последний столбец можно представить в виде варьирования нового фактора x_3.

Так как x_3 в данном плане как взаимодействие x_1x_2, то оценка четвертого коэффициента является смешанной. Поставив только четыре опыта для оценки влияния трех факторов, была использована только половина всех возможных для данного примера вариантов. На основании данного примера можно сделать вывод о том, что с целью уменьшения числа опытов часть факторов целесообразно варьировать как взаимодействие.

Разрешающей способностью такой *полуреплики* называют число независимых переменных, входящих в определенный контраст. Дробные реплики широко применяются при получении линейных моделей. Целесообразность их возрастает по мере роста числа факторов. Если план обладает избыточностью опытов для нахождения оценок коэффициентов регрессионного уравнения, то он называется насыщенным.

При увеличении числа рассматриваемых факторов для уменьшения числа экспериментов можно воспользоваться и другими специальными методами, например, планом в виде *латинского квадрата*, *греко-латинского квадрата* и др. Планы второго порядка используют для нахождения коэффициентов нелинейных моделей второго порядка.

2. Реплики от ПФЭ

С ростом количества факторов k число точек плана в ПФЭ растет по показательной функции 2^k. Планы ПФЭ позволяют получить несмещенные оценки градиента функции отклика в центральной точке, но в случае применения линейного полинома оказываются недостаточно эффективными по количеству опытов при большом числе независимых переменных, так как остается слишком много степеней свободы на проверку адекватности модели. Например, при $k = 5$ на проверку адекватности линейной модели остается 26 степеней. Хотя большое количество опытов и приводит к существенному снижению погрешности в оценке коэффициентов, все же такое число степеней свободы для проверки адекватности является чрезмерным.

Таким образом, в случаях, когда используются только линейные приближения функции отклика, количество опытов следует сократить, используя для планирования так называемые *регулярные дробные реплики* от ПФЭ, содержащие подходящее число опытов и сохраняющие основные свойства матрицы планирования. Реплика, включающая только половину экспериментов ПФЭ, называется полурепликой, включающая четвертую часть опытов – четвертьрепликой и т. д. Краткое обозначение указанных дробных реплик 2^{k-1}, 2^{k-2} соответственно.

Построение регулярной дробной реплики или проведение *дробного факторного эксперимента* (ДФЭ) типа 2^{k-p} предусматривает отбор из множества k факторов $k-p$ основных, для которых строится план ПФЭ. Этот план дополняется p столбцами, которые соответствуют остальным факторам. Каждый из этих столбцов формируется по специальному правилу, а именно, получается как результат поэлементного умножения не менее двух и не более $k-p$ определенных столбцов, соответствующих основным факторам. Иначе говоря, в дробных репликах p линейных эффектов приравнены к эффектам взаимодействия. Но именно такое построение матрицы планирования и позволяет обеспечить ее симметричность, ортогональность и нормированность.

Таблица 1

Матрица планирования				Вектор результатов
x_0	x_1	x_2	x_3	y
+1	–1	–1	+1	y_1
+1	–1	+1	–1	y_2
+1	+1	–1	–1	y_3
+1	+1	+1	+1	y_4

Правило образования каждого из p столбцов ДФП называют *генератором плана*. Каждому дополнительному столбцу соответствует свой генератор (для плана типа 2^{k-p} должно быть задано p различных генераторов). Генератор задается как произведение основных факторов, определяющее значение элементов соответствующего дополнительного столбца матрицы планирования. Примером записи генератора для плана 2^{3-1} служит выражение $x_3 = x_1x_2$, табл. 1. Матрица планирования ДФП типа 2^{k-p} содержит $k + 1$ столбец и $N = 2^{k-p}$ строк.

Запись плана в виде 2^{k-p} не дает полной характеристики регулярной дробной реплики, так как основные эффекты можно приравнять к различным эффектам взаимодействия. Правило смешивания, определяющее коррелированные основные эффекты и эффекты взаимодействия, удобно описывать с помощью *определяющего контраста реплики*. Определяющий контраст полуреплики получается путем умножения генерирующего соотношения на его же левую часть, а так как для любой кодированной переменной x_i^2=1, то левая часть формулы определяющего контраста всегда равна единице и обозначается I. В частности, для ДФП типа 2^{3-1} и генераторе $x_3 = x_1x_2$ имеет место определяющий контраст $I = x_1x_2x_3$ (генератор умножается на переменную x_3, следовательно, $x_3\,x_3 = I = x_1\,x_2\,x_3$).

Чтобы определить, с какими параметрами смешана оценка коэффициента данного фактора, следует умножить обе части определяющего контраста на этот фактор. Учитывая равенство x_i^2=1, получим порядок смешивания оценок коэффициентов при использовании конкретного плана. В рассматриваемом примере для плана 2^{3-1} и определяющего контраста $I = x_1x_2x_3$ порядок смешивания факторов следующий:

$$x_1 = x_1^2\,x_2\,x_3 = x_2\,x_3;\;\; x_2 = x_1\,x_2^2\,x_3 = x_1\,x_3;\;\; x_3 = x_1\,x_2\,x_3^2 = x_1\,x_2\,.$$

Планы дробных реплик строят различным образом, но так, чтобы соблюдались основные свойства матрицы планирования. Например, ДФП 2^{3-1} можно представить одной из двух полуреплик, генераторами которых являются $x_3 = x_1x_2$ и $x_3 = -\,x_1x_2$ соответственно. Определяющие контрасты этих полуреплик: $x_3^2 = I = x_1x_2x_3$ и $x_3^2 = I = -\,x_1x_2x_3$.

В этих полурепликах смешивание факторов задается соотношениями:

$$\text{а) } x_1 = x_2x_3\,,\;\; x_2 = x_1x_3\,,\;\; x_3 = x_1x_2\,;\;\; \text{б) } x_1 = -\,x_2x_3\,,\;\; x_2 = -\,x_1x_3\,,\;\; x_3 = -\,x_1x_2\,.$$

Реализовав обе полуреплики путем совместной обработки результатов экспериментов можно получить раздельные оценки для линейных эффектов и эффектов взаимодействия (такой вариант плана соответствует ПФЭ).

Разрешающая способность полуреплик (возможность раздельного определения коэффициентов уравнения) зависит от генерирующих соотношений. Так, если для плана 2^{4-1} выбрать генерирующее соотношение $x_4 = x_1x_2$, то получим реплику с контрастом $I = x_1x_2x_4$ и разрешающей способностью $x_1 = x_2x_4$ и т.д. Здесь линейные эффекты определяются совместно с парными взаимодействиями. Очевидно, что в первую очередь следует пренебречь взаимодействием более высоких порядков из-за их более низкой вероятности существования по сравнению с парными. У полуреплики с контрастом $I = x_1x_2x_3x_4$ или равноценным $I = -\,x_1x_2x_3x_4$ линейные эффекты будут определяться совместно уже только с тройными взаимодействиями, что повышает точность оценок параметров модели (потенциально величина смещения в оценке коэффициента уменьшается). С ростом количества независимых переменных растет разрешающая способность полуреплик, позволяя оценивать раздельно сначала линейные эффекты, затем парные, тройные взаимодействия и т. д.

Реплики можно строить высокой степени дробности, сокращая тем самым количество экспериментов. Пусть необходимо изучить влияние пяти переменных и известно, что все эффекты взаимодействия пренебрежимо малы. Для линейного приближения следует определить шесть коэффициентов, что потребует применения плана с количеством точек не менее шести. Ближайшее большее число, соответствующее целой степени 2, равно восьми, это дает возможность получить дробную реплику, эквивалентную ПФЭ 2^3, т. е. реплику 2^{5-2} или четвертьреплику. Для построения четвертьреплики необходимы два генерирующих соотношения. В целях построения такой реплики целесообразно пожертвовать тройным и одним из двойных взаимодействий. Пусть этим двойным взаимодействием будет x_1x_2. Тогда можно построить четыре различные четвертьреплики, каждая из которых задается двумя генерирующими соотношениями:

а) $x_4 = x_1x_2$, $x_5 = x_1x_2x_3$; б) $x_4 = x_1x_2$, $x_5 = -x_1x_2x_3$; в) $x_4 = -x_1x_2$, $x_5 = x_1x_2x_3$; г) $x_4 = -x_1x_2$, $x_5 = -x_1x_2x_3$.

Определяющие контрасты каждой четвертьреплики задаются двумя соотношениями:

а) $I = x_1x_2x_4$, $I = x_1x_2x_3x_5$; б) $I = x_1x_2x_4$, $I = -x_1x_2x_3x_5$;
в) $I = -x_1x_2x_4$, $I = x_1x_2x_3x_5$; г) $I = -x_1x_2x_4$, $I = -x_1x_2x_3x_5$.

Из этой совокупности четвертьреплик следует выбрать только одну, например, выберем реплику, задаваемую первой парой генерирующих соотношений. Матрица планирования ДФП получается из матрицы ПФЭ 2^{k-p} для k–p основных факторов добавлением p столбцов, элементы которых вычисляются по соответствующим генерирующим соотношениям, табл. 2.

Таблица 2

Матрица планирования						Вектор результатов
x_0	x_1	x_2	x_3	x_4	x_5	
+	–	–	–	+	–	y_1
+	+	–	–	–	+	y_2
+	–	+	–	–	+	y_3
+	+	+	–	+	–	y_4
+	–	–	+	+	+	y_5
+	+	–	+	–	–	y_6
+	–	+	+	–	–	y_7
+	+	+	+	+	+	y_8

Для полной характеристики разрешающей способности четвертьреплик вводят обобщающие определяющие контрасты, третий компонент которых получается путем перемножения попарно первых двух контрастов. Для выбранной четвертьреплики обобщающий определяющий контраст $I = x_1x_2x_4 = x_1x_2x_3x_5 = x_3x_4x_5$.

Все совместные оценки находятся путем умножения обобщающего определяющего контраста последовательно на x_1, x_2 и т.д. В рассматриваемом случае совместные оценки задаются соотношениями: $x_1 = x_2x_4 = x_2x_3x_5 = x_1x_3x_4x_5$, $x_2 = x_1x_4 = x_1x_3x_5 = x_2x_3x_4x_5$, …, $x_5 = x_1x_2x_4x_5 = x_1x_2x_3 = x_3x_4$.

Разрешающая способность выбранной четвертьреплики невысокая – все линейные эффекты определяются совместно с парными взаимодействиями. Этой репликой можно пользоваться для оценки линейных эффектов при условии равенства нулю соответствующих парных взаимодействий. Если такой уверенности нет, то следует применить полуреплику (что требует в два раза большего количества точек плана эксперимента по сравнению с четвертьрепликой) с генерирующим соотношением $x_5 = x_1x_2x_3x_4$, пользуясь которым, можно разделить все линейные эффекты и парные взаимодействия.

Построение обобщающего определяющего контраста для реплик более высокой степени дробности производится аналогично четвертьреплике: исходные контрасты сначала перемножаются попарно, получаются контрасты первого уровня; затем контрасты первого уровня снова перемножаются попарно, получаются контрасты второго уровня и так далее, пока не будет исчерпана возможность перемножения. Если получается два и более одинаковых контрастов, то из них оставляется только один. Обобщающий определяющий контраст составляется путем перечисления выражений для всех сформированных контрастов.

1. 7 Лекция № 7 (2 часа).

Тема: «Критерии оптимальности планов»

1.7.1 Вопросы лекции:

1. Типы планов эксперимента
2. Критерии оптимальности и их выбор
3. D-оптимальные планы

1.7.2 Краткое содержание вопросов:

1. Типы планов эксперимента

Среди всех классов планов основное внимание в практической работе уделяется ортогональным и ротатабельным планам.

Ортогональным называется план, для которого выполняется условие парной ортогональности столбцов матрицы планирования, в частности, для независимых переменных, где N – количество точек плана эксперимента, k – количество независимых факторов. При ортогональном планировании коэффициенты полинома определяются независимо друг от друга – вычеркивание или добавление слагаемых в функции отклика не изменяет значения остальных коэффициентов полинома. Для ортогональных планов эллипсоид рассеяния ориентирован в пространстве так, что направления его осей совпадают с направлениями координат пространства параметров.

Использование *ротатабельных* планов обеспечивает для любого направления от центра эксперимента равнозначность точности оценки функции отклика (постоянство дисперсии предсказания) на равных расстояниях от центра эксперимента. Это особенно важно при решении задач поиска оптимальных значений параметров на основе градиентного метода, так как исследователь до начала экспериментов не знает направление градиента и поэтому стремится принять план, точность которого одинакова во всех направлениях. В ряде случаев при исследовании поверхности отклика требуется униморфность модели, а именно, соблюдение постоянства значений дисперсии ошибки в некоторой области вокруг центра эксперимента. Выполнение такого требования целесообразно в тех случаях, когда исследователь не знает точно расположение области поверхности отклика с оптимальными значениями параметров.

По соотношению между количеством оцениваемых неизвестных параметров модели и количеством точек плана эксперимента все планы подразделяются на три класса: *ненасыщенные* – количество параметров меньше числа точек плана; *насыщенные* – обе величины одинаковы; *сверхнасыщенные* – количество параметров больше числа точек плана. Метод наименьших квадратов применяют только при ненасыщенном и насыщенном планировании, и он не применим для сверхнасыщенного планирования.

Для некоторых планов важную роль играет свойство *композиционности*. Так, композиционные планы для построения полиномов второго порядка получают добавлением некоторых точек к планам формирования линейных функций. Это дает возможность в задачах исследования сначала попытаться построить линейную модель, а затем при необходимости, добавив наблюдения, перейти к моделям второго порядка, использую ранее полученные результаты и сохраняя при этом некоторое заданное свойство плана, например его ортогональность. Кроме рассмотренных критериев в планировании экспериментов вполне естественно применяется критерий минимума числа экспериментов, т.е. среди всех планов желательно выбирать такой, который требует минимального числа опытов при соблюдении требований к качеству оценки функции или ее параметров.

2. Критерии оптимальности и их выбор

В настоящее время используется свыше 20 различных критериев оптимальности планов, которые подразделяются на две основные группы. К первой группе относят критерии, связанные с ошибками оценок коэффициентов, а ко второй – с ошибкой оценки поверхности отклика. Далее будут охарактеризованы только те критерии, которые наиболее часто применяются при решении задач оптимизации, описания поверхности отклика и оценки влияния факторов.

Критерии первой группы представляют интерес для задач оптимизации, выделения доминирующих (наиболее значимых) параметров на начальных этапах решения оптимизационных задач или для выявления несущественных параметров в задачах восстановления закономерности функционирования объекта. Геометрическое истолкование свойств ошибок коэффициентов связано со свойствами эллипсоида их рассеяния, определяемого математическим ожиданием и дисперсией значений ошибок. Пространственное расположение, форма, и размер эллипсоида полностью зависят от плана эксперимента.

Критерии второй группы используются при решении задач описания поверхности отклика, определения ограничений на значения параметров. Основным здесь является критерий *G*-оптимальности, который позволяет построить план с минимальным значением наибольшей ошибки в описании функции отклика. Применение *G*-оптимального плана дает уверенность в том, что в области планирования нет точек с чрезмерно большой ошибкой описания функции.

3. D-оптимальные планы

Критерию *D*-оптимальности соответствует минимальный объем эллипсоида рассеяния ошибок (минимум произведения всех дисперсий коэффициентов полинома). В соответствующем плане эффекты факторов максимально независимы друг от друга. Этот план минимизируют ожидаемую ошибку предсказания функции отклика. Критерию *A*-оптимальности соответствует план с минимальной суммарной дисперсией всех коэффициентов. Критерию *E*-оптимальности – план, в котором максимальная дисперсия коэффициентов будет минимальна.

Выбор критерия зависит от задачи исследования, так при изучении влияния отдельных факторов на поведение объекта применяют критерий *E* -оптимальности, а при поиске оптимума функции отклика – *D*-оптимальности. Если построение *D*-оптимального плана вызывает затруднения, то можно перейти к *A*-оптимальному плану, построение которого осуществляется проще.

Между критериями оптимальности и методами построения оптимальных планов экспериментов существует жесткая связь. Построение планов производится или с использованием каталогов планов или с использованием непосредственно методов планирования экспериментов, что является непростой задачей и требует достаточно высокой квалификации исследователя в области ТПЭ.

Поиск оптимальных значений параметров является одной из важных задач, решаемых при создании новых технических систем, управлении производством или технологическими процессами. В соответствии с теорией эффективности необходимо:

- сформировать критерий эффективности (функцию отклика в терминах ТПЭ). В большинстве случаев эффективность определяется совокупностью показателей, характеризующих частные свойства исследуемой системы и выполняемой ею операции. Критерий эффективности строится на множестве значений частных показателей с использованием теории полезности или методов векторной оптимизации. В некоторых случаях критерий эффективности удается построить на множестве значений одного показателя, переведя все остальные показатели в разряд ограничений;
- выделить управляемые и неуправляемые параметры (факторы) системы и среды, оказывающие существенное влияние на критерий эффективности;
- определить ограничения на значения параметров.

Задача оптимизации заключается в нахождение экстремума функции отклика в области допустимых значений параметров.

Реализация задачи оптимизации, основанная на применении ТПЭ, как и любой задачи экспериментального исследования, начинается с определения объекта анализа, цели исследования, изучении сущности исследуемого процесса, анализе имеющихся ресурсов, возможности проведения экспериментов с изучаемым объектом в необходимом диапазоне изменения факторов.

1. 8 Лекция № 8 (2 часа).

Тема: «Элементы регрессионного анализа и оптимальное планирование»

1.8.1 Вопросы лекции:

1. Линейная регрессия.
2. Проверка гипотез при использовании линейной регрессии
3. Интервальные оценки при линейной регрессии.
4. Многофакторная линейная регрессия.

1.8.2 Краткое содержание вопросов:

1. Линейная регрессия.

При пассивном эксперименте исследователь не имеет возможности воздействовать на изучаемый объект. Вследствие того, что на значение выходного параметра помимо входного фактора влияет фактор случайности. Эта зависимость не является однозначно определенной и в простейшем виде может быть представлена аддитивной математической моделью: $y = F(x) + E$, где Е – случайные отклонения.

Если экспериментальные данные нанести на график, то значимость (у) от (х) будет диффузной (расплывчатой). Такая не вполне определенная зависимость называется *регрессионной*, а зависимость приведенного вида представляет собой *регрессию*. Линия, наилучшим образом выравнивающая зависимость средних значений называется *уравнением регрессии*. Регрессия – это закон изменения условного математического ожидания выходной величины в зависимости от изменения входной величины. Для линейного случая уравнение регрессии имеет вид: $F_{x,y} = \alpha x + b$, где α – коэффициенты регрессии.

Задача изучения механизма явления считается решенной, если найдена регрессионная модель и оценены статистические характеристики случайных отклонений. Регрессионная модель является приближенной оценкой истинной функциональной зависимости.

Регрессионная модель – это аналитическое описание линии регрессии. Если экспериментальные данные нанести на график, то линии регрессии можно аппроксимировать прямой, участком синусоиды, параболы, эллипса и т.п. Большую трудность при подборе вида аппроксимирующих функций оказывает разброс экспериментальных данных относительно предполагаемой кривой. Этот разброс определяется фактором случайности на выходной параметр.

Когда зависимости плохо интерпретируемы, используют быстрые и простые графические методы, например, метод контура, метод медианных центров.

Метод наименьших квадратов (МНК) – более строгий метод подбора эмпирических зависимостей – является общей частью корреляционного и регрессионного анализов, задача которых получение коэффициентов уравнения регрессии.

Как корреляционный, так и регрессионный анализы состоят из двух частей: расчет коэффициентов уравнения регрессии методом МНК и статистическая оценка результатов. Если МНК можно применять при любых статистических данных, распределенных по любому закону плотности вероятности, то дать статистическую оценку полученным коэффициентам и уравнению регрессии можно лишь на основе определенных теоретических предпосылок. Корреляционный и регрессионный анализы различаются теоретическими предпосылками, т.е. способами статистической оценки.

Задачи регрессионного анализа - исследование тесноты статистической связи между входными и выходными случайными переменными с помощью регрессионных уравнений и коэффициентов. В простейшем случае получают двумерное регрессионное уравнение, в котором участвует лишь одна переменная.

Регрессионный анализ позволяет решать более широкий класс задач, чем корреляционный анализ, и получать оценки коэффициентов нелинейности уравнений регрессии.

Теоретические предпосылки для регрессионного анализа являются менее жесткими:

· случайные помехи должны иметь нормальный закон распределения;

· дисперсии выхода во всех точках факторного пространства должны быть однородными;

· все соседние измерения по каждой входной переменной должны быть независимыми;

· случайные помехи на выходе объекта в каждом опыте должны быть независимы друг от друга, а также от значений входных переменных и коэффициентов уравнений регрессии.

Таким образом, при регрессионном анализе не требуется, чтобы все входные переменные были распределены нормально.

Термин «линейный регрессионный анализ» используют, когда рассматриваемая функция линейно зависит от оцениваемых параметров (от независимых переменных зависимость может быть произвольной). Теория оценивания неизвестных параметров хорошо развита именно в случае линейного регрессионного анализа. Если же линейности нет и нельзя перейти к линейной задаче, то, как правило, хороших свойств от оценок ожидать не приходится.

2. Проверка гипотез при использовании линейной регрессии

Критерий Стьюдента (t-критерий). Критерий позволяет найти вероятность того, что оба средних значения в выборке относятся к одной и той же совокупности. Данный критерий наиболее часто используется для проверки гипотезы: «Средние двух выборок относятся к одной и той же совокупности».

При использовании критерия можно выделить два случая. В первом случае его применяют для проверки гипотезы о равенстве генеральных средних двух независимых, несвязанных выборок (так называемый двухвыборочный t-критерий). В этом случае есть контрольная группа и экспериментальная (опытная) группа, количество испытуемых в группах может быть различно.

Во втором случае, когда одна и та же группа объектов порождает числовой материал для проверки гипотез о средних, используется так называемый парный t-критерий. Выборки при этом называют зависимыми, связанными.

а) случай независимых выборок

Статистика критерия для случая несвязанных, независимых выборок равна: $t_{эмп} = \dfrac{\overline{x} - \overline{y}}{\sigma_{x-y}}$, где $\overline{x}$, $\overline{y}$ - средние арифметические в экспериментальной и контрольной группах, σ_{x-y} - стандартная ошибка разности средних арифметических.

Подсчет числа степеней свободы осуществляется по формуле: $k = n_1 + n_2 - 2$. При численном равенстве выборок $k = 2n - 2$. Далее необходимо сравнить полученное значение $t_{эмп}$ с теоретическим значением t - распределения Стьюдента. Если $t_{эмп}<t_{крит}$, то гипотеза H_0 принимается, в противном случае нулевая гипотеза отвергается и принимается альтернативная гипотеза.

Рассмотрим пример использования t-критерия Стьюдента для несвязных и неравных по численности выборок. Здесь могут возникнуть такие вопросы:

1. Что если полученное в опыте значение t окажется меньше табличного? Тогда надо принять нулевую гипотезу.

2. Доказано ли преимущество экспериментального метода? Не столько доказано, сколько показано, потому что с самого начала допускается риск ошибиться в пяти случаях из ста (p=0,05). Наш эксперимент мог быть одним из этих пяти случаев. Но 95% возможных случаев говорит в пользу альтернативной гипотезы, а это достаточно убедительный аргумент в статистическом доказательстве.

б) случай связанных (парных) выборок

В случае связанных выборок с равным числом измерений в каждой можно использовать более простую формулу t-критерия Стьюдента.

Число степеней свободы k определяется по формуле $k = n - 1$. Рассмотрим пример использования t-критерия Стьюдента для связных и, очевидно, равных по численности выборок. Если $t_{эмп}<t_{крит}$, то нулевая гипотеза принимается, в противном случае принимается альтернативная.

Критерий Фишера позволяет сравнивать величины выборочных дисперсий двух независимых выборок. Для вычисления $F_{эмп}$ нужно найти отношение дисперсий двух выборок, причем так, чтобы большая по величине дисперсия находилась бы в числителе, а меньшая – в знаменателе. Формула вычисления критерия Фишера такова: $F_{эмп} = \frac{\sigma^2{}_x}{\sigma^2{}_y}$, где $\sigma^2{}_x$, $\sigma^2{}_y$ - дисперсии первой и второй выборки соответственно.

Так как, согласно условию критерия, величина числителя должна быть больше или равна величине знаменателя, то значение $F_{эмп}$ всегда будет больше или равно единице.

Число степеней свободы определяется также просто:

$k_1 = n_2 - 1$ для первой выборки (т.е. для той выборки, величина дисперсии которой больше) и $k_2 = n_2 - 1$ для второй выборки.

В Приложении 1 критические значения критерия Фишера находятся по величинам k_1 (верхняя строчка таблицы) и k_2 (левый столбец таблицы). Если $t_{эмп}>t_{крит}$, то нулевая гипотеза принимается, в противном случае принимается альтернативная.

3. Интервальные оценки при линейной регрессии.

Кроме точечных оценок углового коэффициента и отрезка на оси ординат оказывается возможным построить и интервальные оценки этих параметров.

Пример: построить доверительный интервал для линейной регрессии по следующим данным:

X	20	22	24	26	28	30	32	34	36	38	40	42
y	8,91	9,83	10,75	11,67	12,59	13,51	14,43	15,35	16,27	17,19	18,11	19,03

Еще одним полезным понятием в простой линейной регрессии является интервал предсказания, т.е. интервальная оценка среднего последующих наблюдений при некотором значении $x = x_0$.

4. Многофакторная линейная регрессия.

Во многих задачах регрессии встречается более одной независимой переменой. Например, выход продукта металлургической продукции может зависеть от температуры, гидростатического давления и концентрации углерода, в этом случае необходимы, по крайней мере, три независимые переменные. Общая задача подбора модели

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$$

называется задачей множественной линейной регрессии. Неизвестные параметры обычно находятся по методу наименьших квадратов.

Для нахождения оптимальных планов для оценивания параметров регрессии возможно использование последовательного, минимаксного или байесовского подходов.

2. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ПРОВЕДЕНИЮ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

2.1 Практическое занятие № 1 (2 часа).

Тема: «Определение точечных оценок»

2.1.1 Задание для работы:

1. Выборка и ее характеристики.
2. Определение несмещенных оценок с минимальной дисперсией
3. Методы нахождения оценок

2.1.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. Выход сахара (в кг) на каждый центнер сахарной свеклы в 20 хозяйствах составил: 15,3; 16,4; 17,1; 16,8; 14,5; 16,5; 14,7; 16,4; 17,2; 15,8; 16,3; 14,9; 16,1; 15,9; 14,6; 17,0; 16,3; 16,7; 17,6; 16,0. Построить гистограмму частот. Определить выборочную среднюю и выборочное среднее квадратическое отклонение.

2. Урожайность двух сортов A и B пшеницы, возделываемых на трех участках с одинаковыми условиями роста и развития, характеризуется следующими таблицами:

а) сорт A:

Урожайность	18	19	20
Площадь участка	15	25	15

б) сорт B:

Урожайность	17	19	22
Площадь участка	20	20	13,3

Какой из двух сортов является лучшим (сравнить дисперсии)?

3. Найти выборочную среднюю и выборочное среднее квадратическое отклонение роста обследованных студентов по следующим данным (в качестве вариант принять середины интервалов):

Рост студента (x_i)	154 – 158	158 – 162	162 – 166	166 – 170	170 – 174	174 – 178	178 – 182
Число студентов (n_i)	10	14	26	28	12	8	2

4. Для случайной величины X – времени безотказной работы элемента дано эмпирическое распределение среднего времени работы в часах:

x_i	5	15	25	35	45	55	65
n_i	365	245	150	100	70	45	25

Найти методом моментов точечную оценку неизвестного параметра показательного распределения.

5. Случайная величина X (отклонение контролируемого размера изделия от номинала) подчинена нормальному закону распределения с неизвестными параметрами a и σ. Ниже приведено эмпирическое распределения отклонения от номинала $n = 200$ изделий:

x_i	0,3	0,5	0,7	0,9	1,1	1,3	1,5	1,7	1,9	2,1	2,3
n_i	6	9	26	25	30	26	21	24	20	8	5

Найти методом моментов точечные оценки параметров a и σ нормального распределения.

6. Случайная величина X (число появлений события A в m независимых испытаниях) подчинена биномиальному закону распределения с неизвестным параметром p. Ниже приведено эмпирическое распределение числа появлений события A в 1000 независимых испытаниях (в первой строке указано число x_i появлений события в одном опыте из $m = 10$ испытаний; во второй строке приведена частота n_i – число опытов, в которых наблюдалось x_i появлений события A):

x_i	0	1	2	3	4	5	6	7
n_i	2	3	10	22	26	20	12	5

Найти методом наибольшего правдоподобия точечную оценку неизвестного параметра p биномиального распределения.

2.1.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.

2.2 Практическое занятие № 2 (2 часа).

Тема: «Обработка результатов измерений»

2.2.1 Задание для работы:

1. Определение истинного значения измеряемой величины.
2. Определение грубых ошибок.

2.2.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. Произвести обработку результатов измерений сопротивления одноомной катушки сопротивления при $p = 0{,}95$. Значения сопротивления следующие: 1,000309; 1,000391; 1,000395; 1,000392; 1,000389; 1,000396; 1,000388; 1,000389; 1,000393; 1,000394.

2. Выполнено четыре серии измерений одной и той же величины в различных условиях и получены следующие значения: $x_1 = 10{,}24, \sigma_1 = 0{,}054$; $x_2 = 9{,}98, \sigma_2 = 0{,}125$; $x_3 = 10{,}07, \sigma_3 = 0{,}059$; $x_4 = 10{,}33, \sigma_4 = 0{,}057$. Найти средневзвешенное значение и произвести оценку точности.

3. Определить среднее квадратичное отклонение вычисленной длины волны модулированного светового потока λ, если известны скорость света $c = 2999792{,}5\ км/с$ со средним квадратическим отклонением $\sigma_c = 0{,}4\ км/с$ и частота $f = 10000\ кГц$ со средним квадратическим отклонением $\sigma_f = 0{,}15\ кГц$.

4. Определить, содержится ли грубая погрешность в результатах шестикратного взвешивания изделия: 72,361; 72,352; 72,357; 72,346; 72,344; 72,340 (г) при доверительной вероятности $p = 0{,}975$.

2.2.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.

2.3 Практическое занятие № 3 (2 часа).

Тема: «МНК»

2.3.1 Задание для работы:

1. Определение параметров МНК

2.3.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. Получены данные между длиной колоса (X) и числом зерен (Y) в нем:

X	7	8	11	12	9	7	9	10	7	8	10	13	13	14	12	7	9	8	9	10
Y	15	20	28	28	23	17	23	25	16	21	23	28	31	32	30	16	25	20	23	25

По данным составить корреляционную таблицу; построить эмпирическую линию регрессии и записать уравнение теоретической линии регрессии, используя метод наименьших квадратов.

2. Методом наименьших квадратов найти уравнение прямой регрессии Y на X по данным, приведенным в корреляционной таблице:

X \ Y	16	26	36	46	56	n_x
20	4					4
25	6	8				14
30		10	32	4		46
35			3	12	1	16
40			9	6	5	20
n_y	10	18	44	22	6	100

2.3.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.

2.4 Практическое занятие № 4 (2 часа).

Тема: «МНК для многофакторного эксперимента»

2.4.1 Задание для работы:

1. МНК при обработке результатов многофакторного эксперимента

2.4.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. Произведено по четыре испытания на каждом из трех уровней фактора F . При уровне значимости 0,05 проверить нулевую гипотезу о равенстве групповых средних. Предполагается, что выборки извлечены из нормальных совокупностей с одинаковыми дисперсиями. Результаты испытаний приведены в таблице.

Номер испытания	Уровни фактора		
	F_1	F_2	F_3
1	38	20	21
2	36	24	22
3	35	26	31
4	31	30	34

2. Дана матрица планирования в кодированном виде:

Номер опыта	x_0	x_1	x_2	x_3	x_4	x_5	x_6	x_7	y
1	1	1	1	1	1	1	1	1	12,8
2	1	1	-1	-1	-1	-1	1	1	14,2
3	1	-1	-1	1	1	-1	-1	1	16,4
4	1	-1	1	-1	-1	1	-1	1	10,1
5	1	-1	-1	-1	1	1	1	-1	18,3
6	1	-1	1	1	-1	-1	1	-1	14
7	1	1	1	-1	1	-1	-1	-1	17,6
8	1	1	-1	1	-1	1	-1	-1	19,8

Провести обработку данных матричным способом. Записать уравнение результата многофакторного эксперимента

2.4.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.

2.5 Практическое занятие № 5 (2 часа).

Тема: «Полный факторный эксперимент»

2.5.1 Задание для работы:

1. Построение плана ПФЭ 2^2.
2. Построение плана ПФЭ 2^3.

2.5.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. Пусть в результате проведения экспериментов по плану ПФЭ 2^2, т.е. при изменении двух факторов, получены опытные значения: $y_1 = 6, y_2 = 3, y_3 = 4, y_4 = 7$. Построить план ПФЭ 2^2.

2. Пусть в результате проведения экспериментов по плану ПФЭ 2^3, т.е. при изменении трех факторов, получены опытные значения: $x_1 = 7, x_2 = 4, x_3 = 5, x_4 = 8, x_5 = 3$, $x_6 = 9, x_7 = 2, x_8 = 6$. Построить план ПФЭ 2^3.

3. Исследуя выход слитка металла из кристаллизатора в процессе непрерывного литья, было выбрано три фактора (температура, расход воды и процентная концентрация добавки) на верхнем и нижнем уровнях каждый. Построить модель в соответствии с комбинациями обработок плана 2^3 по восьми наблюдениям:

У (выход)	32	36	57	46	65	57	48	68
x_1 (температура)	-1	-1	-1	1	1	-1	1	1
x_2 (давление)	-1	-1	1	-1	1	1	-1	1
x_3 (концентрация)	-1	1	-1	-1	-1	1	1	1

2.5.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.

2.6 Практическое занятие № 6 (2 часа).

Тема: «Дробный факторный эксперимент»

2.6.1 Задание для работы:

1. Построение реплик от ПФЭ.

2.6.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. Построить план ДФЭ 2^{4-1} и определить полином

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{14}x_1x_4 + b_{23}x_2x_3 + b_{24}x_2x_4 + b_{34}x_3x_4$$

Значения функции отклика: 10, 4, 8, 7, 9, 8, 5, 6,5.

2. Планирование эксперимента при отработке центробежных топливных форсунок.

При проектировании центробежных топливных форсунок камер сгорания газотурбин используются зависимости, связывающие основные гидравлические параметры – коэффициент расхода и корневой угол распыла топлива – с геометрической характеристикой форсунки

$$A = \frac{\pi r_c R}{n f_{вх}} \sin\beta$$

где r_c – радиус сопла, R – радиус закручивания, β – угол между закручивающими каналами и осью форсунки, $f_{вх}$ – площадь закручивающих пазов, n – количество пазов.

Однако в процессе экспериментальной доводки опытного газотурбинного двигателя и в дальнейшем, при его серийном производстве, возникают задачи, требующие для своего решения учета таких конструктивных факторов, которые непосредственно не входят в указанную геометрическую характеристику форсунки.

Предполагается, что на гидравлические параметры форсунок влияют пять конструктивных факторов, не входящих в геометрическую характеристику:

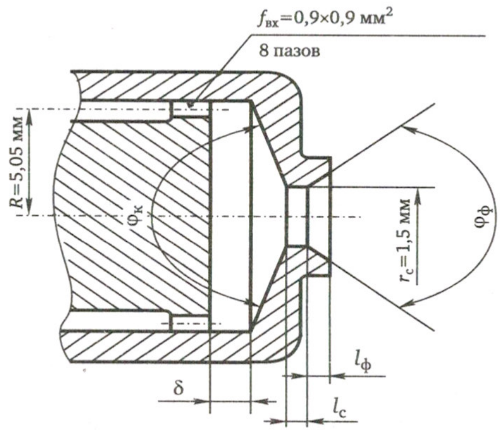
$X_1 = \frac{l_c}{r_c}$ – отношение длины к радиусу соплового отверстия;

$X_2 = \frac{l_ф}{r_ф}$ – отношение длины выходной фаски к радиусу соплового отверстия;

$X_3 = \varphi_ф$ – угол выходной фаски соплового отверстия, рад;

$X_4 = \varphi_к$ – угол корпуса камеры закручивания, рад;

$X_5 = \frac{\delta}{R}$ – отношение длины камеры закручивания к радиусу закручивания.



Составить матрицу планирования дробного факторного эксперимента 2^{5-2} с заменой $x_4 = x_1 x_2$ и $x_5 = x_1 x_2 x_3$ с функциями отклика G – расход через форсунку (244, 292, 303, 255, 247, 252, 262, 254) и α – угол распыла топлива (117, 92, 97, 95, 109, 101, 115, 99).

2.6.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.

2.7 Практическое занятие № 7 (2 часа).

Тема: «Критерии оптимальности планов»

2.7.1 Задание для работы:

1. Применение метода оптимизации

2.7.2 Краткое описание проводимого занятия:

Выполнить следующие задания:

1. При исследовании процесса переэтерификации метилметакрилата диметиламиноэтанолом необходимо подобрать условия, обеспечивающие максимальный выход целевого продукта и высокую производительность процесса. Для определения оптимального режима варьировались значения трех факторов: X_1 – количество катализатора, X_2 – молярное соотношение реагентов, X_3 – время реакции. В качестве выходных показателей фиксировали: y_1 – выход продукта от теоретического на взятый аминоспирт, y_2 – конверсия аминоспирта, y_3 – производительность процесса.

Наилучшие исходные условия процесса характеризовались значениями факторов $X_1 = 4{,}0$, $X_2 = 1{,}4$, $X_3 = 3{,}0$, при которых $y_1 = 74{,}4$, $y_2 = 78{,}7$, $y_3 = 27{,}6$.

В окрестностях этой точки был спланирован эксперимент типа 2^{3-1} с $1 = -x_1 x_2 x_3$ для определения линейных коэффициентов при трех факторах. Условия кодирования факторов и матрица планирования с полученными результатами приведены в таблицах:

Уровни факторов	X_1	X_2	X_3	y_1	y_2	y_3
0	4	1,4	3	74,4	78,7	27,6
-1	3	1,3	2,5			
+1	5	1,5	3,5			
Шаг варьирования	1	0,1	0,5			

Номер опыта	X_1	X_2	X_3	y_1	y_2	y_3
1	-	-	-	45	48,3	23,5
2	+	-	+	84,6	87	31,8
3	-	+	+	59,1	70	20,4
4	+	+	-	82,4	82,6	35,7

2.7.3 Результаты и выводы:

В результате проведения практического занятия студент освоит основные понятия и методы математической теории планирования экспериментов; научится планировать эксперимент по изучению сложных объектов при решении профессиональных задач; овладеет навыками планирования эксперимента и основными приемами сбора, представления и анализа экспериментальных данных.