МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Кафедра анализа данных и искусственного интеллекта**

**КУРСОВАЯ РАБОТА**

**ДРОБНЫЕ 2n–p ФАКТОРНЫЕ ПЛАНЫ**

Работу выполнила \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_Н.В.Федорова

(подпись)

Направление подготовки   09.03.03 Прикладная информатика

Направленность (профиль) Прикладная информатика в экономике

Научный руководитель

Д-р. техн. наук, проф.  \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.А.Халафян

(подпись)

Нормоконтролер

канд. физ.-мат. наук, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Г.В.Калайдина

(подпись)

Краснодар

2023

Разрыв страницы

РЕФЕРАТ

Курсовая работа 29 страниц, 1 рисунок, 8 таблиц, 5 источников.

ДРОБНЫЙ ФАКТОРНЫЙ ПЛАН, ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ, ГЕНЕРАТОР ПЛАНА, ОПРЕДЕЛЯЮЩИЙ КОНТРАСТ, СМЕШАННЫЙ ЭФФЕКТ, ОПТИМАЛЬНОСТЬ ПЛАНА, ПОЛУРЕПЛИКИ

Данная работа посвящена изучению планов дробных факторных экспериментов. Выявлению их достоинств и недостатков в сравнении с полными факторными экспериментами. Будут рассмотрены все этапы проведения дробного факторного эксперимента, что позволит лучше понять все нюансы построения плана. При реализации плана это позволит добиться наилучших результатов в конкретных обстоятельствах.

Материалы работы будут полезны как аналитикам строящим модель по полученным данным, так лицам ответственным за составление плана эксперимента и его проведение.

Итак целью работы является изучение и практическая реализация дробного факторного эксперимента.

Для этого выполним следующие задачи.

* Изучим критерии оптимальности планов эксперимента.
* Изучим метод построения ДФЭ.
* Рассмотрим методы статистического анализа планов ДФЭ.

СОДЕРЖАНИЕ

[Введение 4](#_Toc136174347)

[1 Полный факторный эксперимент как основа ДФЭ 5](#_Toc136174348)

[1.1 Общее представление о ПФЭ 5](#_Toc136174349)

[1.2 Оптимальность плана 5](#_Toc136174350)

[1.3 ДФЭ общее представление 8](#_Toc136174351)

[1.4 Построение ДФЭ 12](#_Toc136174352)

[2 Статистический анализ результатов эксперимента 17](#_Toc136174353)

[2.1 Проверка гипотезы о виде закона распределения 17](#_Toc136174354)

[2.2 Исключение резко выделяющихся значений 19](#_Toc136174355)

[2.3 Дисперсия параметра оптимизации 20](#_Toc136174356)

[2.4 Проверка однородности дисперсий 21](#_Toc136174357)

[2.5 Вычисление оценок коэффициентов уравнения регрессии 23](#_Toc136174358)

[2.6 Проверка значимости коэффициентов регрессии 24](#_Toc136174359)

[2.7 Проверка значимости математической модели 25](#_Toc136174360)

[Заключение 28](#_Toc136174361)

[Список использованных источников 29](#_Toc136174362)

ВВЕДЕНИЕ

Зачастую при возникновении необходимости исследования влияния некоторых факторов на какой–либо процесс не уделяют особого внимания составлению плана действий. Выполняется собор всех возможных данных и проводится анализ. Но сам по себе правильно проведенный статистический анализ недостаточен для достижения научной достоверности, поскольку качество любой информации, получаемой в результате анализа данных, зависит от качества самих данных, по которым проводится опыт.

Одна из простейших разновидностей планов экспериментов это дробные двухуровневые факторные планы. Которые мы и рассмотрим.

В матрицу неграмотно составленного плана могут входить наблюдения, отражающие влияние только одного фактора, в то время как изучение комбинаций нескольких факторов было бы гораздо эффективнее. Матрица дробного факторного эксперимента заполняется как раз такими наблюдениями. Кроме того, большое количество данных не всегда даёт лучший результат, для того, кто проводит эксперимент это не только довольно затратно, но и в некоторых случаях это может привести к потере точности результатов анализа. Дробные факторные эксперименты позволяют сохранять точность, при гораздо меньшем количестве опытов.

Мы рассмотрим как процесс построения дробных факторных экспериментов, так и их анализ.

# 1 Полный факторный эксперимент как основа ДФЭ

## 1.1 Общее представление о ПФЭ

Методы планирования экспериментов используются для выявления влияния некоторых факторов на исследуемый процесс и дальнейший поиск оптимальных уровней факторов для достижения требуемого результата.

Рассматриваются планы, факторы которых принимают только два значения, иначе говоря, каждый фактор имеет два уровня. Значения уровней можно обозначать по–разному 0 и 1, да и нет, наиболее часто используемый вариант это +1 и –1.

Допустим исследуется влияние n факторов на процесс, каждый фактор принимает два значения, тогда для изучения влияния всех факторов и их взаимодействий на процесс нам необходимо перебрать все возможные комбинации значений уровней факторов, совокупность которых составит полный факторный план – ПФП. Из комбинаторики нам известно, что количество – N таких комбинаций вычисляется по формуле размещения с повторениями .

Таким образом полный факторный двухуровневый план – это множество всех точек в n–мерном пространстве, координаты которых являются +1 или –1.

Для того чтобы лучше понять суть дробных факторных экспериментов для начала разберём более подробно полные факторные планы.

## 1.2 Оптимальность плана

Итак, математическая модель объекта – это функция, связывающее параметр оптимизации y с факторами y = f(x1, x2, ..., xk). Множество точек x (1) , ..., x (N) факторного пространства, в которых проводится эксперимент, представляется с помощью плана эксперимента:

 (1)

где k – число факторов, N – число опытов (точек факторного пространства). Точка называется центром плана.

  (2)

Если центр плана совпадает с началом координат, то план называется центральным. Матрица M называется информационной матрицей плана.

 (3)

где – оценка дисперсии параметра y. Обратная к M матрица M–1 – дисперсионная матрица плана.

Выбор соответствующего плана эксперимента позволяет обеспечить математической модели разные свойства. Наиболее распространенными являются следующие критерии оптимальности планов эксперимента.

* Ортогональность. План, у которого дисперсионная матрица M–1 диагональна, называется ортогональным (иными словами, удовлетворяет критерию ортогональности). В этом случае полученные оценки коэффициентов регрессии независимы (не смешаны) и имеют одинаковые дисперсии.

 (4)

* Рототабельность. План, при котором дисперсия выходного параметра Y зависит только от расстояния от центра плана, называется ротатабельным.[5] В этом случае точность предсказания значений Y одинакова на равных расстояниях от центра эксперимента и не зависит от направления.

 (5)

Где R – радиус окружности (сферы, гиперсферы), на которой располагаются точки плана.

* Нормировка.

 (6)

* А–оптимальнось. План, при котором дисперсионная матрица имеет минимальный след (минимальную сумму диагональных элементов), называется А–оптимальным. В этом случае достигается минимальная средняя дисперсия оценок коэффициентов.
* D–оптимальность. План, при котором достигается минимальное значение определителя дисперсионной матрицы, называется D–оптимальным. В этом случае достигается минимум обобщённой дисперсии коэффициентов.
* E–оптимальность. План, при котором достигается минимум максимального собственного значения дисперсионной матрицы, называется E–оптимальным.
* G–оптимальность. План, при котором достигается минимальная величина максимальной дисперсии выходного параметра y, называется G–оптимальным.

Свойство ортогональности для ПФЭ позволяет упростить вычисления и получить независимые (несмешаные) оценки коэффициентов модели.

 (7)

Это означает, что замена нулем любого коэффициента в уравнении модели не изменит оценок остальных коэффициентов. Это важно, когда точный вид модели не известен и требуется по экспериментальным данным отобрать факторы, существенно влияющие на зависимую переменную. Если условие ортогональности не выполняется, после исключения незначимых коэффициентов необходимо пересчитывать оценки оставшихся коэффициентов и их дисперсии. При этом могут измениться выводы относительно значимости коэффициентов.

Матрица, удовлетворяющая условиям симметричности, нормировки и ортогональности, называется оптимальной. Матрица планирования ПФЭ является оптимальной для линейных математических моделей. Если математическая модель содержит взаимодействия факторов, то свойство ротатабельности не выполняется, т. е. точность предсказания значений Y будет завесить от расстояния от центра плана.

Теперь мы имеем достаточное представление от полного факторного эксперимента, значит можем перейти к пояснению того, что собой представляет и зачем нужен дробный факторный эксперимент – ДФЭ.

## 1.3 ДФЭ общее представление

Число опытов, проводимых по плану полного факторного эксперимента, очень быстро возрастает при увеличения числа факторов, количество опытов при исследовании влияния 3 фактов на отклик количество опытов равно 8, при 4 – 16, а при 10 – 1024. Следовательно, даже при небольших n применение полных факторных планов является весьма проблематичным. К тому же количество опытов зачастую прямо пропорционально стоимости эксперимента.

Кроме того, если заранее известно, что в модели не должно быть взаимодействия факторов или мы предполагаем, что ими можно пренебречь, то в математической модели необходимо определить лишь коэффициенты при самих факторах.

 (8)

При этом известно, что для определения n коэффициентов уравнения регрессии необходимо чтобы количество проводимых опытов N было на 1 больше. Т. е. необходимо соблюдать условие насыщенности плана [2]:

 (9)

В случае же, когда N – n>;1 план не является насыщенным, он может быть близок к насыщенному по мере приближения разни­цы N – n к единице. При большой разнице, когда N – n> 3 план называется сверх насыщенным. Оценить однозначно параметры модели можно только в первых двух случаях, в сверхнасыщенном эксперименте оценить однозначно все параметры модели невозможно.

С помощью полного факторного эксперимента можно определить коэффициенты неполного квадратичного полинома, в качестве примера используем примера используем модель с тремя факторами:

 (10)

Значит, в предположении, что в модели не учитывается взаимодействие факторов полный факторный эксперимент даёт нам избыточную информацию и не позволяет однозначно оценить параметры модели.

Для сокращения количества опытов применяют дробные факторные эксперименты – ДФЭ, в нашем случает типа 2n–k.

 ДФЭ – это часть полного факторного эксперимента, в котором такое же количество факторов но меньшее число опытов.

При к =1 число опытов в плане ДФЭ в два раза меньше, чем в плане ПФЭ, поэтому такие планы называют полурепликой плана ПФЭ. При к =2 для плана ДФЭ N=2n–2  называют четвертьрепликой.

Графически возможно представить только планы в которых не более 3 факторов.

Изображение планов ПФЭ 23 и ДФЭ 23–1 в факторном пространстве (трехмерном) представлено на рисуснке 1. План ПФЭ 23 представлен кубом с восемью узлами (точками плана), а возможные планы ДФЭ 23–1 – проекциями этого куба на три плоскости, т.е. из восьми узлов выбираются четыре. Из куба можно также выбрать четыре точки из восьми, не лежащие в одной плоскости, и сформировать план ДФЭ 23–1.



Рисунок 1– Графическое представление ПФЭ и ДФЭ с 3 факторами

Важно понимать, что нельзя составить дробный факторный эксперимент просто оставив в плане половину опытов полного факторного эксперимента, ведь в таком случае может перестать быть оптимальным потеряв одно или несколько из важных свойств. Поясним на примере факторного эксперимента с тремя независимыми переменным, представленного в таблице 1.

Таблица 1 – ПФЭ с 3 факторами

|  |  |  |  |
| --- | --- | --- | --- |
| № | x1 | x2 | x3 |
| 1 | 1 | –1 | 1 |
| 2 | 1 | –1 | –1 |
| 3 | 1 | 1 | 1 |
| 4 | 1 | 1 | –1 |
| 5 | –1 | –1 | 1 |
| 6 | –1 | –1 | –1 |
| 7 | –1 | 1 | 1 |
| 8 | –1 | 1 | –1 |

В таблице представлен полный факторный эксперимент 23, составим 2 варианта дробного факторного эксперимента 23–1 используя в качестве наблюдений строки 1–4 и 3–6, планы представлены в таблицах 1 и 2 ниже. Для обоих планов проверим сохраняется ли свойство ортогональности.

Таблица 2 – ДФЭ с 3 факторами составленный из 1–4 строк ПФЭ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № | x1 | x2 | x3 | x1x2 | x1x3 | x2x3 |
| 1 | 1 | 1 | –1 | 1 | –1 | –1 |
| 2 | 1 | –1 | –1 | –1 | –1 | 1 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | –1 | 1 | –1 | 1 | –1 |
|  |  |  | сумма xixj | 0 | 0 | 0 |

Таблица 3 – ДФЭ с 3 факторами составленный из 3–6 строк ПФЭ

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| № | x1 | x2 | x3 | x1x2 | x1x3 | x2x3 |
| 3 | 1 | 1 | 1 | 1 | 1 | 1 |
| 4 | 1 | –1 | 1 | –1 | 1 | –1 |
| 5 | –1 | 1 | –1 | –1 | 1 | –1 |
| 6 | –1 | –1 | –1 | 1 | 1 | 1 |
|  |  |  | сумма xixj | 0 | 4 | 0 |

Как видно из таблицы для первого варианта матрицы эксперимента свойство ортогональности сохраняется, а вот во втором варианте в таблице 3 полуреплики матрица плана не ортогональна.

## 1.4 Построение ДФЭ

Для построения дробного факторного плана типа 2n–p из множества n факторов отбирают n – p основных факторов, для которых строится полный факторный план с таблицей Xn–p.

Этот план затем заполняют р столбцами, которые соответствуют оставшимся факторам. Каждый из этих р столбцов получается, как результат поэлементного перемножения не менее двух и не более n – p определенных столбцов, соответствующих основным факторам. Такой способ заполнения значений переменных позволяет плану сохранять свойство ортогональности.

 Способ образования каждого из р столбцов дробного факторного плана называется генератором плана.

В таком случае получается, что каждый из p неосновных факторов представляет собой также взаимодействие некоторых основных переменных, значит найденные коэффициенты уравнения регрессии будут являться совместными оценками линейных эффектов и эффектов взаимодействия. Оценки, которые не позволяют их разделить, называют смешанными. [1]

В случае для плана 24–1 в качестве генератора x4, допустим было выбрано взаимодействие эффектов x1x2, т.е. x4=x1x2. Значит оценка эффекта x4 будет смешана с оценкой эффекта взаимодействия x1x2, обозначается это так:

 (12)

Где – оценка смешанного эффекта x4, – истинная оценка эффекта и – оценка эффекта взаимодействия.

Легко заметить, что для произведения x1x2x4 выполняются соотношения +1 = x1x2x4 или –1 = x1x2x4. Символическое обозначение произведения столбцов, равного +1 или –1, называется определяющим контрастом. В приведённом примере 1= x1x2x4 – определяющий контраст факторов x1, x2, x4.

Контраст помогает определять смешанные эффекты. Чтобы определить, какой эффект смешан с данным, надо умножить обе части определяющего контраста на столбец, соответствующий данному эффекту. Т.е. x1=x2x4 и x2=x1x4.

Значит смешанной окажется оценка не только эффекта x4, но и эффектов x1 и x2.

 (13)

Если вводится не один, а несколько дополнительных факторов, то каждому дополнительному фактору будет соответствовать свой генератор плана. В этом случае для определения смешанности оценок используют обобщающий контраст, который строится из отдельных контрастов, а также их произведений во всевозможных сочетаниях.

Допустим мы составляем ДФЭ 25–2, в котором 3 основных фактора. В качестве генераторов плана возьмём соотношения x4=x1x2 и x5=x2x3, их определяющие контрасты 1=x1x2x4 и 1=x2x3x5. Вычислим обобщающий контраст, перемножив определяющие. 1=x1x2x4x2x3x5. Очевидно произведение x2\*x2=1, получаем обобщающий контраст 1=x1x3x4x5

Число максимальное число факторов s, входящих в определяющий контраст, называется разрешающей способностью ДФЭ. Генераторы ДФЭ с максимальной разрешающей способностью называются главными и при построении ДФЭ они являются предпочтительными. В качестве генераторов обычно выбирают незначимые взаимодействия. Для нашего примера в обоих генераторах использовалось только 2 фактора, значит разрешающая способность плана 3. Однако в качестве генератора x4 или x5 мы могли выбрать соотношение x1x2x3, тогда разрешающая способность плана была бы равна 4.

Итак, проводя ДФЭ вследствие сокращения числа опытов мы получаем смешанные оценки эффектов уравнения регрессии и теряем их точность. Причём чем больше число p из формулы 2n–p, тем более смешанные мы получаем оценки. Для того чтобы получить более точные оценки коэффициентов уравнения регрессии проводят повторный эксперимент, при котором задаются противоположные генерирующие соотношения, если в первом плане генераторами служили соотношения x1x2 и x1x3, то во втором эксперименте генераторами послужат –x1x2 и –x1x3. Составив 2 уравнения регрессии получают 2 варианта прогнозных значений отклика , тогда , соответственно для четверть реплик можно получить ,. Если же требуется уточнить сами коэффициенты уравнения, то поступают следующим образом.

Допустим в плане 23–1 генератором x3 служит произведение x1x2, тогда коэффициенты уравнения регрессии будут совместными оценками линейных эффектов и двойных взаимодействий.

 (14)

Если мы положим x3=–x1x2, то получим вторую половину матрицы 23. В этом случае совместные оценки имеют вид:

 (15)

При реализации обеих полуреп­лик можно получить раздельные оценки для линейных эффектов и эффектов взаимодействия, как и в полном факторном эксперименте 23. Объединение этих двух полуреплик и есть полный факторный эксперимент 23.

Теперь мы можем получить более точные и раздельные оценки эффектов уравнения.

 (16)

Аналогично для плана 25–2, где x4=x1x2, x5=x1x2x3. Для определения смешанности будем использовать контрасты 1=x1x2x4=x1x2x3x5 и обобщающий контраст 1=x3x4x5. Для решения  получаем смешанность оценок наподобие первого коэффициента:

 (17)

Избавиться от части оценок коэффициентов взаимодействия факторов можно, задав в качестве генераторов x4=–x1x2, x5=–x1x2x3, тогда получим оценки решения .

 (18)

Аналогично примеру для 23–1 получаем такие оценки.

 (19)

Что очевидно мало эффективно, поэтому при создании ДФЭ так важно задать генераторы, позволяющие устремить оценки эффектов взаимодействия эффектов, смешанные с основными эффектами, к нулю.

#

# 2 Статистический анализ результатов эксперимента

Для того чтобы облегчить понимание методов и формул, которые будут приведены ниже, будем применять их к ДФЭ 25–2 гираторами служат x4=x1x2, x5=x1x3. И допустим мы уже провели серию каких–либо экспериментов, повторив каждый опыт 2 раза. Полученные результаты представлены в таблице 4 ниже.

Таблица 4 – Эксперимент ДФЭ с 5 факторами

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| № | x1 | x2 | x3 | x4=x1x2 | x5=x1x3 | y1 | y2 |
| 1 | –1 | –1 | –1 | 1 | 1 | 5,707 | 2,830 |
| 2 | 1 | –1 | –1 | –1 | –1 | 11,098 | 11,494 |
| 3 | –1 | 1 | –1 | –1 | 1 | 7,874 | 9,210 |
| 4 | 1 | 1 | –1 | 1 | –1 | 16,832 | 17,646 |
| 5 | –1 | –1 | 1 | 1 | –1 | 11,722 | 11,587 |
| 6 | 1 | –1 | 1 | –1 | 1 | 5,150 | 4,818 |
| 7 | –1 | 1 | 1 | –1 | –1 | 11,910 | 13,611 |
| 8 | 1 | 1 | 1 | 1 | 1 | 7,378 | 9,878 |

Итак, мы составили план ДФЭ и провели опыты, теперь необходимо провести анализ полученных результатов.

Большинство критериев, которые используются при статистическом анализе результатов, основаны на предположении, что соответствующие выборки имеют нормальное распределение.

## 2.1 Проверка гипотезы о виде закона распределения

Проверяется гипотеза H0 о том, что выборка значений yij, j = 1, ...n взята из генеральной совокупности, распределённой по нормальному закону с параметрами . Статистика критерия (расчетное значение критерия) вычисляется по формуле:

 (20)

Если dрасч<dкр, где dкр=kα,n–2 – двухсторонний квантиль распределения Колмогорова Kn–2, то гипотеза H0 принимается на уровне значимости α, иначе гипотеза H0 отвергается. Таблица двухсторонних квантилей распределения Колмогорова можно найти в (О.Н. Любимова, 2017) в прил. 5.

Для вычисления теоретических частот нормального закона распределения используются формулы:

 (21)

Где fi – эмпирическая частота наблюдения.

Посчитаем статистику критерия для нашего примера. С помощью приведённых формул вычисляем теоретические частоты, соответствующие нормальному закону распределения, для выборок y1 и y2, результаты представлены в таблице 5.

Таблица 5 – Расчёт значений статистики d

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| y1 |  |  |  | y2 |  |  |  |
| 5,1498 | 1 | 1,1066 | 0,1066216 | 2,8304 | 1 | 0,619 | 0,381 |
| 5,707 | 1 | 1,1214 | 0,1214278 | 4,818 | 1 | 0,6401 | 0,3599 |
| 7,378 | 1 | 1,155 | 0,1549864 | 9,2098 | 1 | 0,6639 | 0,3361 |
| 7,8739 | 1 | 1,1617 | 0,1616539 | 9,8781 | 1 | 0,6646 | 0,3354 |
| 11,098 | 1 | 1,1663 | 0,1663156 | 11,494 | 1 | 0,663 | 0,337 |
| 11,722 | 1 | 1,1594 | 0,159432 | 11,587 | 1 | 0,6628 | 0,3372 |

Продолжение таблицы 5

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 11,91 | 1 | 1,1569 | 0,1568795 | 13,611 | 1 | 0,654 | 0,346 |
| 16,832 | 1 | 1,018 | 0,018018 | 17,646 | 1 | 0,6165 | 0,3835 |
| d1=0,166315578003055 | d2=0,383485194046529 |

Вычислив статистики критерия, сравним их с критическими значениями.
 d1 и d2 < kα=0.1,6, значитмы можем принять гипотезу о том, что обе выборки имеют нормальное распределение. Это утверждение позволит нам провести дальнейший анализ, не теряя точности значений.

## 2.2 Исключение резко выделяющихся значений

Гипотеза о том, что все значения выборки yi, i = 1, ...n принадлежат одной и той же генеральной совокупности, распределенной по нормальному закону с (неизвестными) параметрами α и σ2, проверяется с помощью критерия W. Статистика критерия имеет вид:

 (21)

В данном случае tкр =Wα,n – двухсторонний квантиль распределения Wn, который можно найти в (О.Н. Любимова, 2017) в прил. 2. Если tрасч < tкр, то гипотеза H0 принимается, иначе отвергается.

Если критерий отвергает H0, проверяемое значение yi называют резко выделяющимся значением или грубой ошибкой и исключают из выборки. Для исправленной выборки проверка проводится снова до тех пор, пока гипотеза H0 не будет принята.

Используя данные приведённого примера, согласно формулам, считаем дисперсии и находим максимальный модуль разности I наблюдения переменной и её среднего. переменных y1 и y2, промежуточные вычисления значения t представлены в таблице 6.

Таблица 6 – Расчёт значений статистики t

|  |  |  |
| --- | --- | --- |
|   | y1 | y2 |
| Сред | 9,70887 | 10,13432 |
|  | 7,123 | 7,512 |
| Sy | 15,30925 | 22,14343 |
| ti | 0,49739 | 0,362652 |

Вычислив расчётное значение t статистики, сравним её с критическим значением по таблице, W0.05, 8 =2,273. Оба расчётных значения оказались меньше критического, значит в выборках нет грубых ошибок.

## 2.3 Дисперсия параметра оптимизации

По терминологии теории планирования эксперимента дисперсия параметра оптимизации – это также дисперсия воспроизводимости эксперимента, а также дисперсия погрешности одного наблюдения. Речь идет об оценке дисперсии.[4]

Если каждый из N опытов в эксперименте проводился n раз, то дисперсия воспроизводимости эксперимента вычисляется по формуле (22).

 (22)

где yij – j–е значение функции отклика в точке № i плана эксперимента. Если опыты не дублировались, т.е. n = 1, то вычисления производятся по формуле (23).

 (23)

Если число повторных наблюдений ni для разных опытов различно, то для вычисления используется формула (24).

 (24)

Где – дисперсия i–го опыта, fi – число степеней свободы в i–м опыте, fi = n**i** – 1, i = 1, ..., N.

## 2.4 Проверка однородности дисперсий

Опыт считается воспроизводимым, если дисперсии значений yi не отличаются значимо друг от друга. Оценку дисперсии для каждой точки факторного пространства (для каждого из k опытов) вычисляют по формуле (25).

 (25)

Гипотезу H0:  однородности (равенства) дисперсий проверяют или с помощью M–критерия Бартлетта (когда не все ni равны между собой) или с помощью критерия Кокрена (когда все ni одинаковы, n1 =... = nk = n).[4]

Статистика критерия Бартлетта вычисляется по формулам (26).

 (26)

Где – объединенная дисперсия, а  – Дисперсия группы j.

Эта тестовая статистика соответствует распределению хи–квадрат с k–1 степенями свободы, т.е. В ~ χ2 (k–1), критическое значени можно найти в (Е.А. Трофимова, 2018) в приложении – таблица 2.

Если  то мы принимаем нулевую гипотезу, иначе можем сделать вывод, что не все группы имеют одинаковую дисперсию.

Критерий Бартлетта чувствителен к распределениям выборок, по которым была вычислена дисперсия. В частности, если все оценки построены по выборкам из совокупностей, распределения которых отличны от нормальных, то критерий может с большой вероятностью отвергнуть гипотезу Н0, когда она верна.

Статистика критерия Кокрена вычисляется по формуле (27).

 (27)

Где Gкр=gα,n–1,k – верхний квантиль распределения Кокрена Gn–1,K, соответствующее значениеможно найти в (О.Н. Любимова, 2017) в прил. 3..Если Gрасч < Gкр, гипотеза об однородности дисперсий принимается, в противном случае – отвергается. Если гипотеза Н0 отвергается, тогда эксперимент необходимо повторить, изменив условия его проведения.

 В нашем примере в выборках y1 и y2 одинаковое количество наблюдений (так было изначально и к тому же нам не пришлось исключать грубые ошибки) поэтому для проверки однородности дисперсий мы можем использовать метод Кокрена. В предыдущем подразделе мы расчитали , тогда . Далее α = 0.05, n–1 = 7, k=2, значит значение соответствующее Gкр=g0.05,7,2=0,8159. G<Gкр – гипотеза об однородности дисперсий принимается.

## 2.5 Вычисление оценок коэффициентов уравнения регрессии

Пусть в общем случае верна математическая модель вида (28).

 (28)

Необходимо с помощью ДФЭ 2s==N, (s=n–p), вычислить оценки коэффициентов, т.е. найти зависимость (29).

 (29)

Где  – статистические оценки соответствующих коэффициентов математической модели.

Благодаря предварительной стандартизации масштаба факторов и ортогональности матрицы плана, расчет оценок коэффициентов регрессии можно выполнить по простым формулам (30) и (31).

 (30)

 (31)

При этом x0=1, и в приведённых формулах подразумевается, что каждый из N опытов был проведён m раз, поэтому можно рассчитать по формуле (31).

 (31)

Вычислим коэффициенты уравнения регрессии без взаимодействия факторов, т.е. оценки эффектов b0–b5. Так как каждый из опытов мы проводили дважды, то нам необходимо сначала рассчитать , результат в таблице 7.

Таблица 7 – Значения

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| i | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|  | 4,269 | 11,296 | 8,542 | 17,239 | 11,655 | 4,984 | 12,761 | 8,628 |

И теперь используя эти значения мы вычислим коэффициенты уравнения регрессии. Уравнение имеет вид (32).

 (32)

## 2.6 Проверка значимости коэффициентов регрессии

Гипотезу о статистической значимости (отличии от нуля) коэффициентов регрессии проверяют по критерию Стьюдента. Статистику критерия вычисляют по формуле (33).

 (33)

При сохранении свойства рототабельности плана, все экспериментальные точки факторного пространства одинаково удалены от центра эксперимента и оценки всех коэффициентов уравнения регрессии независимо от их величины имеют одинаковую погрешность вычисляемую по формуле (34).

 (34)

Где – оценка дисперсии выходного параметра y.

Критическое значение – двухсторонний квантиль распределения Стьюдента TN1, критическое значени можно найти в (Е.А. Трофимова, 2018) в приложении – таблица 1. Если tрасч > tкр, гипотеза о значимости коэффициента bj принимается, в противном случае коэффициент считается незначимым и его можно убрать из модели.

Вычислим все необходимые значения для проверки значимости коэффициентов: , , tкр=t0.95, 16 =2,12. Наблюдаемые значения t приведены в таблице 8.

Таблица 8 – Наблюдаемые значения t значимости коэффициентов регрессии

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| t0 | t1 | t2 | t3 | t4 | t5 |
| 4,8441 | 0,3 | 0,913 | 0,203 | 0,257 | 1,619 |

Статистически значимым можно считать только оценку эффекта b0, остальные коэффициенты можно приравнять к 0.

## 2.7 Проверка значимости математической модели

Рассчитаем какую долю изменчивости исходного отклика описывает составленная модель, т.е. рассчитаем коэффициент детерминации R2.

 (35)

Тогда детерминация вычисляется по формуле (36).

 (36)

Однако коэффициент детерминации множественной регрессии обладает одним недостатком – он неизбежно увеличивается при добавлении в регрессию новой переменной, даже если влияние последней на у несущественно. С целью исправления обозначенного недостатка вычисляют скорректированный (adjusted) коэффициент детерминации.

 (37)

где p – количество факторов включённых в модель.

Для проверки гипотезы Н0 о статистической незначимости уравнения регрессии в целом используется F– критерий Фишера. Статистика критерия рассчитывается следующим образом:

(38)

Если окажется, что , то гипотеза Н0 отвергается с вероятностью ошибки α, иначе говорят о том что уравнение регрессии статистически значимо. Критическое значение можно найти в (Е.А. Трофимова, 2018) в приложении – таблица 4.

Выясним какую долю изменчивости исходного отклика описывает составленное уравнение регрессии. Для расчёта коэффициента детерминации были рассчитаны ESS=3.829, TSS=126.408, RSS=122.579.

Тогда R2= 0,969 и – составленная модель описывает около 80% изменчивости исходного отклика. Проверим статистическую значимость. Каждая переменная содержит 8 наблюдений (n=8), в уравнении регрессии 5 факторов (p=5), значит верно уравнение (39).

(39)

 (40)

Итак – полученное уравнение регрессии статистически значимо.

Как правило при подборе подходящей модели, вначале проверяют адекватность линейного (по факторам) уравнения регрессии (41).

 (41)

Если гипотеза об адекватности подтверждается, то в качестве математической модели выбирают линейную; если отклоняется – добавляют эффект взаимодействия с наибольшим значащим коэффициентом и вновь проверяют гипотезу, и так до тех пор, пока остаются степени свободы для вычисления статистики критерия. Если в результате модель все же оказалась неадекватной, это говорит о том, что тип математической модели выбран неудачно и модель должна быть уточнена.

ЗАКЛЮЧЕНИЕ

Икак в ходе работы были выполнены поставленные задачи, а именно:

* Были изучены критерии оптимальности планов эксперимента.
* Рассмотрены преимущества и недостатки построения ДФЭ.
* Составлен пример ДФЭ, проведён его статистический анализ.

Дробный факторный эксперимент позволяет в несколько раз сократить количество приводимых опытов, что может существенно снизить затраты на проведение эксперимента. Однако при включении в модель взаимодействия факторов необходимо следить за тем, чтобы план оставался насыщенным.

Вследствие сокращения числа опытов мы можем получить только смешанные оценки эффектов уравнения регрессии, причём чем «меньше» становится реплика, т.е. по мере увеличения числа генераторов плана оценки будут только сильнее «смешиваться» друг с другом. А для того, чтобы разделить оценки смешанных эффектов или получить более точные прогнозные значения оклика, необходимо провести ещё одну серию опытов с отрицательными генераторами плана, что увеличивает количество опытов в плане как минимум в 2 раза. Поэтому при составлении ДФЭ для получения желаемой точности прогноза необходимо выбирать: либо исходя из априорной информации задать генерирующие соотношения так, чтобы оценки эффектов взаимодействия факторов, смешанные с оценками главных факторов, были близки к нулю, либо перейти к плану с большим количеством наблюдений.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Лекции по МОД. Дробные факторные эксперименты: сайт / studfile.net –URL: https://studfile.net/preview/4001711/page:9/ (дата обращения: 13.5.2023)
2. Дробный факторный эксперимент: сайт / studme.org –URL: https://studme.org/232711/matematika\_himiya\_fizik/drobnyy\_faktornyy\_eksperiment (дата обращения: 13.5.2023)
3. Теория вероятности и математическая статистика: учебное пособие / Е.А. Трофимова, Н.В. Кисляк, Д.В. Гилёв – Екатеринбург: Издательство Уральского университета, 2018 – ISBN 978-5-7996-2317-3 (дата обращения: 13.5.2023).
4. Построение и проверка регрессионных моделей при обработке результатов факторных экспериментов: практикум / О.Н. Любимова, В.В. Сиськов – Владивосток: ФГАОУ ВО «ДВФУ», 2017 – ISBN 978-5-7444-4100-5
5. Основные понятия рототабельности ЦКП: сайт / studme.org –URL: https://studme.org/232733/matematika\_himiya\_fizik/osnovnye\_ponyatiya\_rotatabelnosti (дата обращения: 13.5.2023).