

Лекция 1.

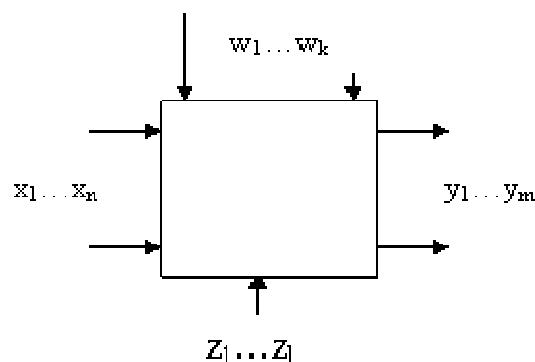
Содержание лекции: Введение.

основные понятия теории планирования экспериментов.

лекции: изучить основные термины теории планирования экспериментов.

Цель

Особенность систем управления - их сложность, которая проявляется в значительном числе и многообразии параметров, определяющих течение процессов, их взаимное влияние и связи. Поэтому изучение процессов проводится с помощью модели – упрощенной системы, отражающей отдельные, ограниченные в нужном направлении характеристики рассматриваемого процесса (математическое и физическое моделирование). Объект на языке математики описывается в виде некоторой системы уравнений и функциональных соотношений между отдельными параметрами модели. При отсутствии или ограниченном объеме теоретических сведений о моделируемом объекте, когда неизвестен даже ориентировочный вид соотношений, описывающих его свойства, уравнения математического описания могут представлять собой систему эмпирических зависимостей, полученных в результате обследования действующего объекта. Математическая теория, лежащая в основе построения различных статистических моделей – теория вероятностей, изучающая закономерности массовых случайных явлений. С каждым из подобных явлений может быть связано определенное событие – осуществляющееся или нет в результате опытов. При решении задач статистической обработки экспериментальных данных качественное описание случайных явлений в терминах событий, когда отмечается лишь факт его наличия или отсутствия, недостаточно. Поэтому результаты опытов представляются количественно в виде некоторой случайной величины. Статистическое описание результатов наблюдений составляет содержание математической статистики. Для получения статистических моделей предназначены методы планирования экспериментов. Основная особенность любой экспериментальной модели – то, что подобная модель не может точно описать поведение объекта в любом конкретном опыте. Статистическая модель описывает поведение объекта в среднем, характеризуя неслучайные свойства объекта, которые в полной мере могут проявиться лишь при многократном повторении опытов в неизменных условиях. Из различных статистических моделей наибольший практический интерес вызывают *регрессионные* модели. Регрессионный анализ представляет собой математический аппарат идентификации сложных объектов исследования, метод обработки результатов наблюдения при активном и пассивном эксперименте. Идея метода заключается в минимизации функционала несоответствия выходов модели и объекта. Решение этой задачи приводит к *методу наименьших квадратов*. Теория планирования эксперимента формулирует приемы и способы оптимальной организации экспериментирования при исследовании объектов самой различной физической природы. Теория исходит из абстрактной схемы сложной системы, называемой «черным ящиком» (Рисунок 1.1). Считается, что исследователь может наблюдать входы и выходы «черного ящика» и по результатам наблюдений определять зависимость между входами и выходами. Входы могут меняться детерминировано (контролируемые, управляемые) или случайно (контролируемые, но не управляемые).



сложной системы

Рисунок 1.1 – Схема

На этой схеме

$X = (x_1 \dots x_n)$ – контролируемые и управляемые параметры объекта;

$Z = (z_1 \dots z_l)$ – входные контролируемые, но не управляемые параметры;

$W = (w_1 \dots w_k)$ – неконтролируемые, случайным образом изменяющиеся параметры, «шум» объекта;

$Y = (y_1 \dots y_m)$ – наблюдаемые выходы объекта.

Комплекс параметров $X = (x_1 \dots x_n)$ называется основным, он определяет условия эксперимента. В качестве случайных рассматриваются обычно параметры, которые по тем или иным причинам невозможно учесть (или очень трудно). Такое подразделение на основные и случайные параметры условно. Случайным будет любой параметр, не вошедший в основной комплекс, даже если он хорошо изучен. Входные переменные $x_1 \dots x_n$ называются *факторами*. Выходная переменная – зависимая y называется *откликом* (реакцией). Фиксированное значение фактора называется уровнем. *Уровни* – те значения, которые устанавливаются для каждого фактора при определении условий проведения экспериментов. Факторы могут отличаться по числу уровней. *Факторное пространство* – это множество факторов, значения которых исследователь может контролировать в ходе подготовки и проведения эксперимента. **Целью эксперимента** является нахождение функции y , при этом предполагается, что значение отклика складывается из двух составляющих:

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

где $f(x_1 \dots x_n)$ – функция отклика (неслучайная функция факторов); $e(x_1 \dots x_n)$ – ошибка эксперимента (случайная величина); $x_1 \dots x_n$ – определенное сочетание уровней факторов из факторного пространства.

Очевидно, что y является случайной переменной, так как зависит от случайной величины $e(x_1 \dots x_n)$. Дисперсия $D[y]$, которая характеризует точность измерения, равна дисперсии ошибки опыта $D[y] = D[e]$.

В условиях эксперимента факторы могут варьировать, благодаря чему исследовать влияние фактора на наблюдаемую переменную. Если влияние некоторого фактора на наблюдаемую переменную изменяется при изменении уровня некоторого другого фактора, говорят, что между факторами существует *взаимодействие*. Суть анализа заключается в разложении общей вариации случайной величины на независимые слагаемые – *эффекты*, каждый из которых характеризует влияние того или иного фактора (*главный эффект*) или их взаимодействия (*эффект взаимодействия*). *Эксперимент* – система операций, воздействий и наблюдений, направленных на получение информации об объекте при исследовательских испытаниях.

Опыт – отдельная элементарная часть эксперимента (воспроизведение исследуемого явления в определенных условиях проведения эксперимента).

Пассивный эксперимент – когда экспериментатор находится в положении пассивного наблюдателя; здесь факторы контролируемые, но не управляемые.

Активный эксперимент – факторы в эксперименте контролируемые и управляемые, и экспериментатор может в соответствии с планом эксперимента их целенаправленно изменять.

В дальнейшем рассматриваем вопросы, связанные с активным экспериментом. *План эксперимента* – совокупность данных, определяющих число, условия и порядок проведения опытов.

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям. То есть это целенаправленное управление, которое реализуется в условиях неполного знания механизма изучаемого явления.

Область действия – область возможных значений факторов x при эксперименте.

Область планирования – область значений x , в которых находятся точки, отвечающие условиям проведения опытов используемого плана эксперимента (это часть области действия или совпадает с ней). Область планирования может задаваться *интервалами* возможного изменения факторов $x_{i\min} \leq x_i \leq x_{i\max}$, $i=1, \dots, n$. Характерные точки области планирования – $x_{i\min}$, $x_{i\max}$, основной уровень – x_{i0} . Обычно вектор x_0 задает центр области планирования (центр эксперимента). В ее окрестности и располагаются все точки плана.

Основной уровень – который представляет наибольший интерес в данный момент. Обычно

$$x_{i0} = \frac{x_{i\min} + x_{i\max}}{2}.$$

Интервал (шаг) варьирования фактора x_i :

$$\Delta x_i = \frac{x_{i\max} - x_{i\min}}{2}.$$

Тогда $x_{i\max} = x_{i0} + \Delta x_i$, $x_{i\min} = x_{i0} - \Delta x_i$.

Обычно вводится нормализация (стандартизованный масштаб):

$$x_i = \frac{x_i - x_{i0}}{\Delta x_i}, i = 1, 2, \dots, n.$$

Здесь начало координат – центр эксперимента, единица измерения – шаг варьирования факторов. Это облегчает математические выкладки.

Матрица плана – стандартная форма записи условий проведения эксперимента в виде матрицы, (i, j) –ый элемент которой – уровень j -го фактора в i -том опыте. Может иметь совпадающие строки.

Матрица спектра плана – матрица, состоящая из всех строк матрицы плана, отличающихся уровнем хотя бы одного фактора. Все строки этой матрицы различны.

Матрица дублирования – квадратная матрица, диагональные элементы которой равны числу параллельных опытов в соответствующих точках спектра плана.

Таким образом, *план эксперимента* – совокупность точек $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_N$, где \bar{x} – вектор входа, N –

$$\sum_{g=1}^N m_g = N$$

число наблюдений, с числом m_1, m_2, \dots, m_N ($\bar{g} =$) наблюдений в каждой точке, то есть матрица спектра

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \dots & \dots & \dots & \dots \\ x_{N1} & x_{N2} & \dots & x_{Nn} \end{pmatrix}$$

совместно с матрицей дублирования экспериментов

$$\begin{pmatrix} m_1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & m_N \end{pmatrix}$$

Лекция 2. Статистическое оценивание

Содержание лекции: свойства статистических оценок характеристик случайных величин; метод максимального правдоподобия, проверка статистических гипотез.

Цель лекции: изучить основные определения и методику применения статистических критериев для проверки гипотез относительно характеристик случайных величин.

2.1 Статистические оценки и их свойства

Назначение статистических методов заключается в том, чтобы по выборке ограниченного объема N , то есть по некоторой части генеральной совокупности, высказать обоснованное суждение о ее свойствах в целом. Подобное суждение может быть сформировано путем *оценивания* выборочных параметров (характеристик) генеральной совокупности с помощью каких-либо подходящих *функций от наблюдений*.

Функция $G = g(x_1, x_2, \dots, x_n)$ от элементов выборки из рассматриваемой генеральной совокупности называется *статистикой*. Фиксированная статистика, подходящая для оценивания того или иного параметра генеральной совокупности, называется *статистической оценкой* (выборочным значением) данного параметра. Произвольная статистика сама является случайной величиной и как любая случайная величина может быть описана с помощью плотности вероятности или числовых характеристик. В общем случае она зависит от параметров генеральной совокупности и объема выборки N . Любая оценка параметра генеральной совокупности является случайной величиной. Свойства (вероятностные) оценки параметра могут быть определены также с помощью выборочной функции распределения и ее характеристик. Для оценивания одного и того же параметра можно использовать различные оценки (статистики). Чтобы выбрать наилучшую из них, необходимо сформулировать некоторые требования к свойствам оценок, желательные с точки зрения практики.

Оценка $\hat{\theta}$ параметра θ называется *состоятельной*, если при неограниченном увеличении объема N выборки значение оценки (с вероятностью 1) стремится к своему теоретическому значению, то есть $\hat{\theta} \rightarrow \theta$, когда $N \rightarrow \infty$. Это означает, что с ростом N выборочное распределение все в большей степени концентрируется вокруг θ и точность оценки безгранично возрастает. В частности, для состоятельной оценки справедливо

$$\lim D\{\hat{\theta}(X_1, X_2, \dots, X_N)\} \text{ при } N \rightarrow \infty.$$

Оценка $\hat{\theta}$ называется *несмещенной*, если для любого объема N выборки математическое ожидание оценки $\hat{\theta}$ равно оцениваемому параметру θ : $m_{\hat{\theta}} = M\{\hat{\theta}\} = \theta$. Для несмещенной оценки характерно отсутствие систематической погрешности. Несмещенная оценка $\hat{\theta}_{\text{эф}}$ называется *эффективной*, если среди

всех оценок параметра она обладает наименьшей дисперсией:

$$D\{\hat{\theta}_{\text{ФП}}\} = \sigma_{\hat{\theta}_{\text{ФП}}}^2 = M\{(\hat{\theta} - \theta)^2\} = \min.$$

2. 2 Метод максимального правдоподобия

Свойства оценок различных параметров θ во многом определяются видом закона распределения исследуемой генеральной совокупности, а также способом оценивания. Известны общие методы, позволяющие во многих случаях решить задачу нахождения по экспериментальным данным наилучших оценок неизвестных параметров. *Метод максимального правдоподобия* является наиболее важным общим методом оценивания неизвестных параметров по экспериментальным данным.

Пусть x_1, x_2, \dots, x_N – выборка из генеральной совокупности с плотностью вероятности $f_x(x; \theta)$, зависящей от постоянного неизвестного параметра θ . Элементы выборки статистически независимы, поэтому можем записать выборочную плотность вероятности для случайной выборки N

$$f_{x_1, x_2, \dots, x_N}(x_1, x_2, \dots, x_N; \theta) = f_{x_1}(x_1; \theta) f_{x_2}(x_2; \theta) \cdot \dots \cdot f_{x_N}(x_N; \theta). \quad (2.1)$$

До эксперимента функция правдоподобия выражает предпочтительность различных значений θ при конкретной выборке. После того, как произведен эксперимент, значения выборки известны – это некоторые числа, а параметр θ неизвестен. Тогда плотность вероятности, полученная при подстановке выборочных значений x_1, x_2, \dots, x_N в плотность вероятности (2.1), называется *функцией правдоподобия* $L(\theta)$ для параметра θ :

$$L(\theta) = f_{x_1}(x_1; \theta) f_{x_2}(x_2; \theta) \cdot \dots \cdot f_{x_N}(x_N; \theta).$$

Эта функция выражает предпочтительность различных значений θ при той конкретной выборке, которая фактически имеет место.

Метод максимального правдоподобия состоит в том, что в качестве оценки $\hat{\theta}$ неизвестного параметра θ , выбирается его значение, максимизирующее функцию правдоподобия. Во многих случаях

максимум этой функции находится из уравнения $\frac{dL(\theta)}{d\theta} = 0$.

Иногда удобнее находить максимум не самой функции $L(\theta)$, а ее логарифма $l(\theta) = \ln L(\theta)$. Тогда получаем уравнение

$$\frac{dl(\theta)}{d\theta} = \frac{d}{d\theta} \left[\sum_{i=1}^N \ln f_{x_i}(x_i; \theta) \right]_{=0}.$$

Для определения оценки используется процедура максимизации функции правдоподобия, так как эта функция выражает предпочтительность различных значений θ при конкретной выборке. Такая оценка дает предпочтительные значений, так как при этом вероятность получения выборки максимальна.

2.3 Оценки параметров

Оценка неизвестного параметра распределения, определяемая одним числом, называется *точечной*.

Они представляют собой числа, полученные путем подстановки выборочных значений x_1, x_2, \dots, x_n в формулу для оценивания искомого параметра. Именно для таких оценок используются понятия состоятельности, несмещенности, эффективности. Точечные оценки не дают информации о степени близости $\hat{\theta}$ к соответствующему теоретическому параметру θ . Поэтому более информативный способ оценивания неизвестных параметров заключается не в определении единичного точечного значения, а в построении интервала, где с заданной степенью достоверности будет находиться оцениваемый параметр, то есть в построении *интервальной оценки* параметра θ .

2.4 Статистические гипотезы

Статистическая гипотеза – определенное предположение относительно свойств генеральной совокупности, из которой извлекается выборка. *Критерий статистической гипотезы* есть правило, позволяющее отвергнуть или принять данную гипотезу на основании выборки. При построении такого правила используются некоторые функции результатов наблюдения, они называются *статистиками для проверки гипотез*. Все возможные значения подобных статистик для проверки той или иной гипотезы делятся на две группы: *область принятия гипотезы* и *критическая область*. Критическая область состоит из всех значений статистики, при которых принимается решение отвергнуть принимаемую гипотезу как ложную.

Проверка гипотезы сводится к выяснению, попадает или нет значение используемой статистики в критическую область: если не попадает – гипотеза принимается как не противоречащая результатам наблюдения, если попадает – гипотеза отвергается. Так как эти решения базируются на статистиках, найденных по выборкам ограниченного объема всегда возможны ошибки.

Ошибка первого рода – вероятность отвергнуть правильную гипотезу, она называется *уровнем значимости* α .

Ошибка второго рода – вероятность принять неверную гипотезу β , она связана с понятием мощности. *Мощность критерия* есть вероятность отбросить неверную гипотезу и равна $1 - \beta$. Ясно, что при построении статистических критериев желательно, чтобы вероятность принять неверное решение была бы достаточно мала.

2.5 Критерии значимости

Критерии значимости – это критерии, с помощью которых проверяются гипотезы об абсолютных значениях параметров или о соотношениях между ними для генеральных совокупностей с известной (с точностью до параметров) функцией распределения вероятностей. При проверке подобных гипотез, если распределение генеральной совокупности определяется целой группой параметров, а проверяемая гипотеза касается лишь их части, остальные параметры полагаются известными или же вычисляются по данным выборки. С помощью критериев значимости выясняется:

а) не противоречат ли результаты наблюдений предположению о том, что некоторые из параметров генеральной совокупности равны определенным числам;

б) если имеются две или большее количество выборок, существуют ли объективные основания считать, что они являются представителями одной и той же генеральной совокупности, или же экспериментальные данные говорят о наличии неслучайных различий между выборками, свидетельствующих о том, что они извлечены из разных генеральных совокупностей.

Пусть некоторая оценка $\hat{\theta}$ вычислена по выборке объема N . Пусть есть причины считать, что истинное значение оцениваемого параметра θ , то есть его значение в генеральной совокупности равно θ_0 . Это проверяемое предположение часто называют *нулевой гипотезой* и обозначают $H_0: \theta = \theta_0$. Гипотеза, которая противопоставляется нулевой и в пользу которой склоняется исследователь, отвергая проверяемую нулевую гипотезу, называется *альтернативной гипотезой* H_1 ,

Для проверки гипотез надо знать функцию плотности распределения вероятности $f_{\theta}(\hat{\theta})$, построенную теоретически в предположении, что нулевая гипотеза верна. Если она получена, то с ее помощью легко найти такую зону, случайное попадание в которую выборочного значения $\hat{\theta}$ при справедливой нулевой гипотезе маловероятно. Подобную зону малой вероятности можно использовать в качестве критической области, отвергая нулевую гипотезу, если $\hat{\theta}$ попадает в нее. Конечно, при этом сохраняется возможность совершить ошибку первого рода, хотя вероятность данной ошибки, равная q , по условию мала и определяется выбором уровня значимости.

2.6 Критерии согласия

Критерием согласия называется критерий гипотезы о том, что генеральная совокупность имеет распределение предполагаемого типа. Среди различных критериев согласия наиболее употребителен универсальный критерий согласия Пирсона χ^2 .

Лекция 3. Построение регрессионных моделей объектов

Содержание лекции:

основные элементы регрессионного анализа.

Цель лекции:

изучить понятия регрессионного анализа и методы оценивания параметров регрессионной модели.

3.1 Элементы регрессионного анализа

Статистические методы планирования активного эксперимента являются одним из способов математического описания статистики сложных объектов исследования. Задача отыскания оценок коэффициентов уравнения по результатам опытов в N точках факторного пространства является типичной задачей множественного регрессионного анализа (при выполнении некоторых предпосылок). Регрессионный анализ включает методы получения регрессионной зависимости (зависимости среднего значения какой-либо величины от некоторой другой величины), связывающей значения отклика объекта факторами при наличии аддитивной помехи случайного характера.

Будем рассматривать такие объекты, для которых случайную составляющую можно отделить от регулярной, то есть

$$y = \varphi(\bar{x}) + e.$$

Наиболее распространены модели, линейные по параметрам, то есть

$$y = \varphi(\bar{x}) + e = \sum_{j=0}^d \beta_j f_j(\bar{x}) dx = \bar{f}^T(\bar{x}) \bar{\beta} + e,$$

где $\bar{x} = \{x_1, \dots, x_n\}$ - факторы, $\bar{\beta} = \{\beta_1, \dots, \beta_n\}$ - вектор неизвестных коэффициентов регрессии, $\bar{f}(\bar{x}) = \{f_0(\bar{x}), f_1(\bar{x}), \dots, f_d(\bar{x})\}$ - вектор базисных функций.

Система базисных функций должна быть выбрана до проведения эксперимента и последующих расчетов. Теоретически считается, что вид модели известен и требуется по экспериментальным данным с помощью N опытов найти неизвестные $\bar{\beta}$. Реально же адекватную (соответствующую объектам исследования) модель приходится подбирать методом проб и ошибок. Практически чаще применяются полиномиальные регрессионные модели. Обычно, чем больше число факторов, тем меньше максимальный порядок подбираемой модели.

Для построения таких моделей вводятся некоторые предпосылки относительно статистических свойств случайной составляющей e :

1. Результаты наблюдений y_1, \dots, y_n отклика в N точках факторного пространства представляют собой нормально распределенные случайные величины, то есть на них воздействуют нормально распределенные случайные помехи e с нулевым математическим ожиданием $M\{e\} = 0$ и постоянной дисперсией $\sigma_e^2 = const$. Постоянство дисперсии означает, что интенсивность шума не меняется в различных наблюдениях. Данное требование необходимо для корректного выполнения всех статистических процедур в регрессионном анализе.

2. Значения помехи в различных наблюдениях некоррелированы. Для обеспечения этого требования используется рандомизация опытов.

3. Дисперсии $\sigma^2\{y_g\}$ $g=1, \dots, N$ равны. Иными словами получаемые в результате многократных повторных наблюдений над величиной y_g в точке x_g выборочные оценки $S_g^2\{y\}$ однородны, а дисперсия $\sigma^2\{y_g\}$ не зависит от математического ожидания $M\{y_g\}$, то есть не отличается от дисперсии $\sigma^2\{y_g\}$ полученной при повторных наблюдениях в любых других точках факторного пространства (воспроизводимость с равной точностью). Это означает постоянство дисперсии помехи равной дисперсии воспроизводимости $\sigma_e^2 = const$.

4. Независимые управляемые факторы x_1, \dots, x_n измеряются с пренебрежимо малыми ошибками по сравнению с ошибками в определении y . То есть вклад, вносимый случайными ошибками измерений факторов \bar{x} в дисперсию воспроизводимости, должен быть пренебрежимо мал по сравнению с действием других неконтролируемых факторов, образующих E .

Это предположение можно проверить лишь после получения регрессионной модели, так как степень влияния ошибки измерения некоторого фактора зависит не только от величины ошибки, но и от того, насколько сильно сам фактор влияет на y . Чем сильнее его влияние, тем в большей степени скажется и ошибка в его измерении. Поэтому, эта предпосылка просто постулируется, и лишь затем, если найденная регрессионная модель окажется неработоспособной, начинают анализировать, не связано ли это с наличием существенных ошибок в измерении x .

5. Вектор-столбцы базисных функций должны быть линейно-независимыми, то есть

$$\sum_{g=1}^N [af_j(\bar{x}_g) + bf_i(\bar{x}_g)] = 0 \text{ при } \forall a \text{ и } b \neq 0.$$

Выполнение этого требования необходимо для получения отдельных оценок всех коэффициентов регрессии β , входящих в регрессионную модель. В активном эксперименте это обеспечивается соответствующим выбором плана эксперимента. В пассивном эксперименте его выполнение должно быть

проверено в ходе расчетов. Рассматриваемое условие ограничивает и общее количество коэффициентов, входящих в регрессионную модель: оно не должно быть больше числа опытов, то есть $d+1 \leq N$.

Регрессионный анализ включает в себя:

а) оценивание неизвестных коэффициентов β регрессионной модели, а также дисперсии воспроизводимости σ_e^2 .

б) статистический анализ результатов (выяснение точности оценивания отдельных коэффициентов регрессии, регрессионной модели в целом, проверки ее адекватности и работоспособности).

3.2 Точечные оценки параметров регрессионной модели

Имеем N наблюдений над значениями факторов x_1, \dots, x_n и отклика y .

$$\bar{x} = \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \dots \\ \bar{x}_N \end{pmatrix} = \begin{pmatrix} x_{11} \dots x_{1n} \\ x_{21} \dots x_{2n} \\ \dots \\ x_{N1} \dots x_{Nn} \end{pmatrix}; \quad y = \begin{pmatrix} y_1 \\ y_2 \\ \dots \\ y_N \end{pmatrix}.$$

Значения базисной функции в каждом из наблюдений

$$\bar{F} = \begin{pmatrix} \bar{f}_1(\bar{x}) \\ \bar{f}_2(\bar{x}) \\ \dots \\ \bar{f}_N(\bar{x}) \end{pmatrix} = \begin{pmatrix} f_{11}(\bar{x}) \dots f_{1d}(\bar{x}) \\ f_{21}(\bar{x}) \dots f_{2d}(\bar{x}) \\ \dots \\ f_{N1}(\bar{x}) \dots f_{Nd}(\bar{x}) \end{pmatrix}.$$

Используя эту информацию, необходимо оценить неизвестные коэффициенты β_0, \dots, β_d , то есть найти соответствующие значения оценок этих коэффициентов $\bar{b} = \{b_0, \dots, b_d\}$. Эти оценки – случайные величины, так как результаты измерений сопровождаются случайными ошибками.

Значения \bar{b} находятся из условия минимума суммы квадратов отклонений измеренных значений отклика от полученных с помощью моделей, то есть путем минимизации следующей суммы (невязки):

$$S = \sum_{g=1}^N [y_g - \varphi(x_g, \bar{\beta})]^2 \rightarrow \min$$

Выбор критерия зависит от характера помех, то есть от их статистических свойств. При нормальном распределении помех наибольшую точность дает среднеквадратичный критерий. А задача решается просто ввиду гладкости S (метод называется *методом наименьших квадратов*). Условия минимума гладкой функции: равенство нулю производных по всем неизвестным параметрам. Из этого условия получим систему линейных алгебраических уравнений, решая которую можно получить искомые оценки b_0, \dots, b_d . Эта система называется *системой нормальных уравнений* и в матричной форме имеет вид:

$$(\bar{F}^T \cdot \bar{F}) \cdot \bar{b} = (\bar{F}^T \cdot y).$$

Матрица системы $\bar{F}^T \cdot \bar{F} = \bar{\phi}$ называется *информационной матрицей Фишера*. Она симметрична относительно главной диагонали. Для существования единственного решения нормальной системы

необходимо и достаточно, чтобы $\det \|\phi\| \neq 0$. Это условие сводится к условию линейной независимости вектор-столбцов матрицы \bar{F} (одна из предпосылок регрессионного анализа). Решить систему нормальных уравнений можно любым известным способом. В общем виде решение можно записать так:

$$\bar{b} = \phi^{-1} \cdot (\bar{F}^T \cdot y)$$

Пусть C_{jk} – элемент матрицы ϕ^{-1} : $C_{jk} = \phi_{kj} / |\phi|$, где ϕ_{kj} – алгебраическое дополнение элемента f_{kj} в ϕ . Тогда

$$b_j = \sum_{k=0}^d C_{jk} \left(\sum_{r=1}^N y_r \cdot f_{rk} \right), \quad j = \overline{0, d}$$

Из этого выражения видно, что значения коэффициентов зависят от количества членов уравнений регрессии. То есть увеличение или уменьшение числа членов уравнения влияет на значения коэффициентов всех включенных в уравнение членов. Такая неопределенность в оценивании коэффициентов регрессии крайне затрудняет их физическую интерпретацию.

Если же матрица Фишера диагональна (столбцы матрицы \bar{F} попарно ортогональны), то система нормальных уравнений распадается на независимые уравнения, а коэффициенты получаются независимо.

Таким образом, для получения независимых оценок коэффициентов \bar{b} столбцы матрицы \bar{F} должны быть линейно независимы и ортогональны. Линейная независимость ограничивает общее количество коэффициентов, входящих в регрессионную модель: $d+1 \leq N$ (например, три коэффициента, два фактора – минимум четыре опыта); ортогональность позволяет получить независимые оценки коэффициентов (независимо от числа членов уравнения регрессии).

3.3 Свойства оценок параметров регрессионной модели

Приведем свойства оценок параметров регрессионной модели:

1. Математическое ожидание оценок $M\{b_j\} = \beta_j, j = \overline{0, d}$, то есть оценки b_j являются несмещенными.

2. Дисперсии оценок коэффициентов равны

$$\sigma_{b_j}^2 = \sigma_e^2 \cdot C_{jj}$$

и минимальны в классе линейных несмещенных оценок.

3. Оценки b_0, \dots, b_d подчиняются совместному $(d+1)$ -мерному распределению с вектором математических ожиданий $M\{\bar{b}\} = \bar{\beta}$.

Итак, кроме оценок коэффициентов регрессии b_j в общем случае необходимы оценки дисперсии воспроизводимости σ_e^2 .

Лекция 4. Статистический анализ результатов

Содержание лекции: статистический анализ результатов построения регрессионной модели.

Цель лекции: изучить методику оценивания качества полученной модели.

Статистический анализ позволяет выяснить качество полученной регрессионной модели: точность, адекватность и работоспособность.

4.1 Оценка дисперсии воспроизводимости

Если заранее известно, что регрессионная модель адекватна объекту исследования, то в качестве оценки дисперсии σ_{ε}^2 можно использовать так называемую остаточную дисперсию

$$S_{ост}^2 = \frac{1}{N - (d + 1)} \sum_{g=1}^N (y_g - \hat{y}_g)^2,$$

где \hat{y}_g - предсказанное по уравнению регрессии значение отклика в опыте с номером g . Действительно, для адекватной модели единственной причиной различий между фактическим значением y_g в g -ом опыте и предсказанным по уравнению \hat{y}_g может быть влияние шума. Поэтому $S_{ост}^2$ - оценка σ_{ε}^2 :

$$S_{\varepsilon}^2 = S_{ост}^2; \nu = N - (d + 1), \nu - \text{число степеней свободы}$$

(Величина зависящая от элементов выборки и входящая в формулу дисперсии называется связью. Разность между объемом выборки N и числом связей называется числом степеней свободы).

Если же вид адекватной модели неизвестен, то проводим проверку воспроизводимости эксперимента, то есть проверку предпосылки регрессионного анализа об однородности выборочных дисперсий S_g^2 . Иными словами, хотим проверить гипотезу о равенстве генеральных дисперсий

$$\sigma^2\{y_1\} = \sigma^2\{y_2\} = \dots = \sigma^2\{y_N\} \text{ при опытах в точках } \bar{x}_1, \bar{x}_2, \dots, \bar{x}_N.$$

Оценки дисперсий находятся по известной формуле

$$S_g^2 = \frac{1}{m-1} \sum_{k=1}^m (y_{gk} - \bar{y}_g)^2, \nu_{loc} = m - 1.$$

Для проверки гипотезы об однородности дисперсий S_g^2 используется критерий Кохрена.

Полученные в результате решения задачи оценки коэффициентов \bar{b} (и предсказательные значения отклика \hat{y}) тем предпочтительнее, чем меньше интегральные характеристики точности, например

дисперсия оценок коэффициентов регрессии $\sigma_{b_j}^2$. Так как метод обработки задан (метод наименьших квадратов), система базисных функций выбрана, то эти характеристики зависят лишь от наблюдаемых во всех опытах значениях входов. При пассивном эксперименте извлечь пользу из этой зависимости нельзя.

Другое дело, когда эксперимент активный, можно предусмотреть при какой комбинации значений \bar{x} следует проводить каждый отдельный опыт, то есть можно заранее составить *план* эксперимента. Для рассматриваемых регрессионных моделей, линейных по параметрам, планы не зависят от результатов измерений отклика, поэтому могут быть найдены заранее.

4.2 Проверка значимости коэффициентов регрессии

После определения оценок \bar{b} коэффициентов регрессии нужно проверить гипотезы об их значимости. В результате проверки устанавливается, статистически значимы или незначимы отличия от нуля оценок коэффициентов регрессии. То есть выясняется, обусловлено ли отличие b_i от 0 чисто случайными обстоятельствами, влиянием помехи или же это отличие неслучайно и вызвано тем, что в теоретической модели присутствует соответствующий коэффициент регрессии $b_j \neq 0$. Проверка значимости коэффициентов регрессии чаще всего производится раздельно, поочередно для каждой оценки b_i . Итак, для каждого коэффициента проверяется нуль-гипотеза $\beta = 0$. Проверку значимости проводят с помощью критерия *Стьюдента*. Если в результате проверки оценку b_j считают статистически незначимой, то надо принять решение: оставить или отбросить незначимый коэффициент. Формально – необходимо отбросить, фактически – решение принимается после анализа, особенно если t_j близко к критическому. Уверенность специалистов, что данный фактор необходим, может быть доводом для того, чтобы оставить. И если первоначальная система базисных функций выбрана правильно, то отбрасывание тех, чьи коэффициенты незначительно отличаются от нуля по t -критерию, ведет к смещению оценок остальных коэффициентов регрессии.

Статистическая незначимость оценки b_j коэффициента регрессии может быть обусловлена следующими причинами:

1) данный j -й фактор не имеет функциональной связи с откликом y , то есть $\beta_j = 0$;

2) уровень x_{j0} базового режима \bar{x}_0 находится в точке частного экстремума функции отклика по

$$\beta_j = \frac{\partial y}{\partial x_j} = 0;$$

фактору x_j и тогда

3) интервал варьирования Δx_j выбран малым;

4) вследствие влияния неуправляемых и неконтролируемых факторов велика ошибка воспроизводимости эксперимента.

4.3. Проверка адекватности регрессионной модели.

Чтобы проверить гипотезу об адекватности математического описания опытным данным, достаточно оценить отклонение предсказанной по полученному уравнению регрессии величины отклика

\hat{y}_g от результатов наблюдений y_g в одних и тех же g -х точках факторного пространства. Проверяется модель со значимыми коэффициентами. Эта проверка возможна, если получена оценка дисперсии шума S_e^2 и выполнено условие $N > d^*$ (d^* - число отобранных коэффициентов). Действительно, если $N = d^* + 1$, то проверять адекватность нет смысла, так как в этом случае $y_g = \bar{y}_g, g = \overline{1, N}$, то есть поверхность, соответствующая модели, проходит точно через все экспериментальные точки. В этом случае не остается степеней свободы для проверки гипотезы об адекватности математического описания.

Рассеяние результатов наблюдений вблизи уравнения регрессии, оценивающего истинную функцию отклика, можно охарактеризовать с помощью дисперсии адекватности

$$S_{ad}^2 = \frac{m}{N-d} \sum_{g=1}^N (y_g - \hat{y}_g)^2,$$

где d - число членов аппроксимирующего полинома. Дисперсия адекватности определяется числом степеней свободы $\nu_{ad} = N - d$.

Идея проверки адекватности состоит в следующем. Для адекватной модели остаточная дисперсия (дисперсия адекватности) является оценкой дисперсии шума σ_e^2 ($S_{ocm}^2 = S_{ad}^2 = S_e^2$). Если же модель неадекватна, то $S_{ocm}^2 = S_{ad}^2$ будет оценкой σ_e^2 плюс некоторая компонента, обусловленная неадекватностью.

То есть надо S_{ocm}^2 сопоставить с какой-либо оценкой S_e^2 дисперсии шума σ_e^2 . При адекватной модели отличие между ними должно быть числом случайным; при неадекватной S_{ocm}^2 должно быть значимо больше S_e^2 . Это сопоставление проверяется с помощью *F-критерия Фишера*

Если гипотеза об адекватности отвергается, необходимо переходить к более сложной форме математического описания либо, если это возможно, проводить эксперимент с меньшим интервалом варьирования Δx_j . Максимальная величина интервала варьирования определяется условием адекватного описания объекта в области варьирования. Если при больших интервалах варьирования математическая модель неадекватна, то возникают систематические ошибки в определении коэффициентов, для уменьшения которых требуется сузить область варьирования. Однако с уменьшением интервала варьирования появляется целый ряд новых трудностей: растет отношение помехи к полезному сигналу, что приводит к необходимости увеличивать число параллельных опытов для выделения полезного сигнала на фоне шума, то есть уменьшаются абсолютные значения оценок b_j коэффициентов, величины которых непосредственно зависят от Δx_j , и оценки коэффициентов могут стать статистически незначимыми. Для выбора интервала варьирования проводят предварительные эксперименты. Интервал варьирования можно выбирать равным 0.05-0.3 от допустимого диапазона варьирования факторов, то есть область варьирования составляет примерно 10-60% от всего диапазона. Начальную точку варьирования (базовую точку) выбирают возможно ближе к центру области факторного пространства, в которой ищется математическое описание объекта.

4.4 Анализ работоспособности модели.

С помощью этого выясняется практическая возможность использования регрессионной модели для решения определенной задачи. (предсказание значения отклика, управление объектом и т.д.) Этот анализ необходим, так как иногда даже адекватные модели оказываются бесполезными из-за низкой точности. Анализ включает в себя несколько процедур:

а) исследование остатков – разности между наблюдаемыми значениями отклика и предсказанными по уравнению $l_g = y_g - \hat{y}_g$;

Если регрессионная модель подобрана правильно, то есть, адекватна функции отклика, то остатки являются проявлением действия шума. Так как о свойствах шума были сделаны определенные предположения, то теперь можно выяснить - не противоречат ли свойства остатков этим предположениям. Подобный анализ проводится графически (зависимость l_g от времени, y_g , x_g и т.д.). Просматривая остатки, можно понять какова погрешность предсказания в точке, где проводились измерения, устраивает ли величина погрешности, при каких комбинациях факторов погрешность больше и т.д.

Исследование остатков плодотворно тогда, когда число наблюдений больше количества коэффициентов. Если это не так, то для проверки качества модели желательно провести дополнительные эксперименты;

б) вычисление *коэффициента детерминации* R^2 ; этот коэффициент показывает, какая доля из общего рассеяния экспериментальных значений отклика относительно своего среднего обусловлена регрессионной зависимостью.

При $R^2 = 0$ поверхность отклика y определяется полностью случайными возмущениями, влияние факторов не обнаружено; при $R^2 = 1$ регрессионная кривая проходит через все экспериментальные точки. Очевидно, чем ближе R к единице, тем лучше уравнение регрессии описывает изменение y (это при $N > d + 1$), а малое R всегда говорит о низкой точности уравнения.

Лекция 5. Критерии оптимальности планов

Содержание лекции: основные принципы планирования; критерии оптимальности планов регрессионного анализа.

Цель лекции: изучить принципы планирования экспериментов и критерии оптимальности планов.

5.1 Основные принципы планирования экспериментов

Принципы планирования экспериментов направлены на повышение эффективности экспериментирования, то есть на получение максимума информации при минимуме опытов и состоят в следующем:

1. *Отказ от полного перебора возможных состояний.*

Для получения исчерпывающей информации о свойствах функции отклика необходимо бесконечное число опытов во всех точках области планирования. Это практически нереально. Вообще-то необходимо увеличение числа уровней и факторов. Но на практике обычно задаются выбором вида функции отклика и выбор числа уровней варьирования по каждому фактору связан с этой функцией. Эту функцию выбирают, используя следующий принцип.

2. *Принцип постепенного усложнения математической модели (принцип последовательного планирования).*

В отсутствие информации нет смысла сразу строить сложную модель, надо начать с простейшей. Если же модель непригодна, необходим следующий этап: постановка новых дополнительных опытов, позволяющих получить более сложную модель и т.д. (то есть, если обратиться к широко используемым полиномиальным моделям, то надо начинать с линейной модели и т. д.

3. Принцип сопоставления с шумом.

Если случайная помеха значительна, то разумно ли тратить силы на получение сложной модели? Например, расточительно подключать приборы высокого класса точности для измерения переменной, отягощенной большой случайной ошибкой. То есть точность получаемой модели должна быть сопоставима с интенсивностью случайной помехи. Чем меньше уровень помехи, тем более точной (и, следовательно, сложной) должна быть модель; и чем выше уровень помехи, тем в большей степени можно ожидать, что более простая и менее точная модель окажется работоспособной.

Так как многие реальные объекты характеризуются высоким уровнем помех, получили большое распространение полиномиальные регрессионные модели (степени 1 и 2). С теоретической точки зрения эти модели менее содержательны, чем теоретические модели типа дифференциальных уравнений. Но с практической точки зрения являются весьма эффективным, а иногда и единственным средством изучения сложных объектов.

4. Принцип рандомизации (принцип приведения к случайности).

Этот принцип состоит в такой организации эксперимента, которая позволяет сделать случайными (рандомизировать) систематически действующие помехи неподдающиеся учету. То есть, не в силах учесть действие неслучайных переменных, искусственно создаем в эксперименте случайную ситуацию. Например, систематические ошибки возможны из-за каких-то погрешностей при настройке экспериментальной установки. И если в серии из восьми опытов проводить первые четыре опыта на нижнем уровне первого фактора, то погрешность будет одинаковой для всех четырех опытов (если опыты проводить подряд).

5. Принцип оптимальности планирования эксперимента.

Это центральный принцип в теории планирования экспериментов. План должен обладать некоторыми оптимальными свойствами с точки зрения определенного заранее выбранного критерия оптимальности. Критерии могут формулироваться по-разному, зависеть от типа решаемой задачи, но принцип «меньше опытов – больше информации» сохраняется.

Таким образом, необходим такой план, который соответствовал бы экстремальным значениям выбранного показателя эффективности плана. Эти показатели называются *критериями оптимальности* регрессионных планов.

5.2 Критерии оптимальности планов

Существует две группы критериев оптимальности: к первой группе относятся критерии, связанные с точностью оценивания коэффициентов регрессии; вторую группу составляют критерии, связанные с предсказательными свойствами эмпирического уравнения регрессии. Рассмотрим несколько критериев из обеих групп:

1 группа) Наиболее часто используемый критерий - *критерий D – оптимальности*: план ε^* называется *D – оптимальным*, если

$$\det \phi(\varepsilon^*) = \max_{\varepsilon \in \{\varepsilon\}} \det \phi(\varepsilon),$$

где $\{\varepsilon\}$ - множество планов заданной области планирования.

D – оптимальный план максимизирует величину определителя матрицы Φ или, что тоже самое – минимизирует обобщенную дисперсию оценок коэффициентов регрессии. С геометрической точки зрения это означает минимизацию объема эллипсоида рассеяния данных оценок.

Критерий ортогональности. План ε^* называется *ортогональным* если матрица Φ (а, следовательно, и Φ^{-1}) диагональная, то есть столбцы матрицы \bar{F} являются попарно ортогональными векторами.

Доказано, что при заданных значениях диагональных элементов матрицы Φ дисперсии оценок коэффициентов регрессии минимальны для ортогонального плана ε^* . Оценки коэффициентов регрессии получаются независимо, что облегчает их анализ, резко сокращает все вычисления.

2 группа) План называется *G - оптимальным*, если

$$\max_{x \in H} \sigma^2 \{y(x, \varepsilon^*)\} = \min_{\varepsilon \in (\varepsilon)} \max_{x \in H} \sigma^2 \{y(x, \varepsilon)\}.$$

То есть G – оптимальный план минимизирует максимальную дисперсию предсказания выхода модели по уравнению регрессии в области H .

Критерий ротатабельности. План ε^* называется *ротатабельным*, если дисперсия предсказания отклика по уравнению регрессии, полученному с помощью этого плана, постоянна на фиксированном расстоянии от центра эксперимента, то есть

$$\sigma^2 \{y(\bar{x}, \varepsilon^*)\} = h(\rho),$$

где $h(\rho)$ – некоторая функция расстояния ρ от центра планирования: $\rho = \sqrt{x_i^2}$.

Для ротатабельного планирования точность предсказания одна и та же на равном расстоянии от центра эксперимента, независимо от того, на каком расстоянии находятся точки \bar{x} , где вычисляются $y(\bar{x})$.

Кроме этих критериев желательны и некоторые другие свойства планов (насыщенность и композиционность).

План называется *насыщенным*, если число наблюдений равно числу неизвестных параметров регрессионной модели. Желательно, чтобы любой план был близок к насыщенному, особенно на начальном этапе, когда требуется получить хотя бы приближенное представление об объекте, но зато с минимальными затратами.

План называется *композиционным*, если в его спектр в качестве составной части входят точки спектра плана, который был реализован при построении более простой модели. То есть, если модель полученная на предшествующем этапе не устраивает и требуется ее усовершенствование, например от полинома k -й степени переходим к полиному $(k+1)$ -й степени, то точки спектра первого этапа используются и в дальнейшем.

Композиционность плана – важное условие эффективности практической реализации принципа постепенного усложнения модели.

План называется *центральным*, если его центр расположен в начале координат.

Желательны еще простота вычислений коэффициентов моделей, симметрия в расположении точек плана и минимальное число уровней варьирования по каждому из переменных.

Лекция 6. Планы первого порядка

Содержание лекции: планирование первого порядка - полный факторный и дробный факторный эксперимент.

Цель лекции: изучить принципы планирования экспериментов для построения линейных и неполных степенных моделей.

С точки зрения структуры модели объектов могут быть линейно-параметризованными и нелинейно-параметризованными, то есть линейными и нелинейными по параметрам. Мы рассматриваем линейно-параметризованные модели. Такие модели могут быть линейными по факторам, неполно-квадратичными по факторам и квадратичными по факторам. Для рассматриваемых регрессионных моделей, линейных по параметрам, планы не зависят от результатов измерения отклика, поэтому могут быть найдены заранее.

На практике чаще применяются полиномиальные регрессионные модели. Для построения таких моделей предназначены планы первого порядка. Для определения оценок неизвестных коэффициентов регрессии подобных моделей каждый из факторов должен варьироваться, по крайней мере, на двух уровнях.

Пример X_i^0 - базовый (основной) уровень, ΔX_i - шаг варьирования фактора X_i . Экспериментальные точки должны располагаться внутри или на границе гиперпараллелепипеда

$X_i^0 - \Delta X_i \leq X_i \leq X_i^0 + \Delta X_i, i = \overline{1, n}$. Для упрощения расчетов перейдем от обычного масштаба к

$$x_i = \frac{X_i - X_i^0}{\Delta X_i}, i = \overline{1, n}.$$

нормализованному (стандартизованному):
имеем

Для нормализованных переменных

$$\varphi(\vec{x}) = \beta_0 + \beta_1 x_1 + \dots + \beta_n x_n + \beta_{12} x_1 x_2 + \dots + \beta_{n-1n} x_{n-1} x_n + \beta_{11} x_1^2 + \dots + \beta_{nn} x_n^2 + \dots$$

$$\beta_i = B_i \Delta X_i, i = \overline{1, n}; \beta_0 = B_0 + \sum_{i=1}^n B_i X_i^0; -1 \leq x_i \leq +1, i = \overline{1, n}.$$

Известно несколько разновидностей планов первого порядка - однофакторный эксперимент, полный факторный эксперимент, дробный факторный эксперимент.

а) *Однофакторный эксперимент* предусматривает поочередное варьирование каждого из факторов, в то время как все остальные факторы стабилизируются на некотором уровне. Для получения линейной модели достаточно двух уровней для каждого фактора. Будем считать, что каждый фактор принимает значения $x_{i\pm} = x_i^0 + \Delta x_i; x_{i\mp} = x_i^0 - \Delta x_i$, а все другие факторы в этот момент стабилизированы на своем

базовом уровне $x_j^0, j \neq i$. Для нормализованных переменных $x_{i\pm} = +1, x_{i\mp} = -1, x_j = 0$. Геометрически точки плана располагаются в середине сторон ограничивающего квадрата. Рассматриваемый план является ротатабельным. С точки зрения других критериев оптимальности, однофакторный эксперимент неудовлетворителен. Поэтому не находит широкого применения.

б) *Полным факторным экспериментом* (ПФЭ) называется эксперимент, реализующий все возможные неповторяющиеся комбинации n независимых управляемых факторов на всех уровнях их изменения. Если каждый фактор изменяется на двух уровнях, то общее число таких комбинаций $N = 2^n$. Такие планы обозначаются ПФЭ 2^n . 2^n - число элементов спектра. Геометрически точки плана ПФЭ 2^n расположены в вершинах n -мерного куба.

ПФЭ 2^n позволяет в принципе оценить 2^n коэффициентов регрессионной модели при 2^n базисных функциях. Первые $(n+1)$ базисных функций очевидны: $f_0(\bar{x}) = 1, f_1(\bar{x}) = x_1, \dots, f_n(\bar{x}) = x_n$. Так как в произвольный полином могут быть включены степени x_i или различные комбинации их произведений, а сами эти факторы в ПФЭ 2^n принимают значения, равные +1 или -1, то и столбец, отвечающий любой базисной функции, состоит из +1 и -1. Напомним, что в число предпосылок регрессионного анализа входит условие линейной независимости столбцов численных значений этих базисных функций (столбцов матрицы \bar{F}). Тогда в матрице \bar{F} не должно быть полностью совпадающих или полностью противоположных (по знаку) столбцов. Отсюда следует, что в числе базисных функций не может быть функции вида x_i^k , где $k > 1$ целое, так как при четном k столбец состоит из +1, а при нечетном $x_i^k = x_i$. В то же время любые комбинации произведений факторов могут быть в числе базисных функций. Таким образом, с помощью ПФЭ 2^n возможно получение регрессионных моделей в виде неполных степенных полиномов.

Общее количество коэффициентов $1 + n + C_n^2 + C_n^3 + \dots + C_n^n = 2^n$, то есть план ПФЭ является насыщенным. ПФЭ 2^n является также ортогональным планом.

Дисперсии оценок коэффициентов вычисляются по формулам $\sigma_{b_j}^2 = \frac{\sigma_e^2}{N}$, то есть все коэффициенты рассчитываются с равной точностью.

Дисперсия предсказания отклика по линейной регрессионной модели, полученной с помощью ПФЭ 2^n

$$\sigma^2\{\hat{y}(\bar{x}, \bar{b})\} = D\{b_0 + \sum_{i=1}^n b_i x_i\} = \sigma_{b_0}^2 + \sigma_{b_i}^2 \sum_{i=1}^n x_i^2 = \frac{\sigma_e^2}{N} (1 + \sum_{i=1}^n x_i^2) = \frac{\sigma_e^2}{N} (1 + \rho^2)$$

Для фиксированного ρ дисперсия является постоянной, то есть план ПФЭ 2^n является ротатабельным. Он также является A-, E-, D-, G-оптимальным. Следовательно, при достаточной простоте построения ПФЭ 2^n обладает хорошими свойствами.

С практической точки зрения ПФЭ присущ существенный недостаток, связанный с быстрым ростом количества опытов при увеличении числа факторов n (например, при $n=10$, количество опытов $N=1024$).

Вместе с тем, трудно представить функцию отклика, для которой квадратичные члены x_j^2 не существенны, но в то же время играют роль взаимодействия высокого порядка (вида $x_1 x_2 \dots x_n$). То есть эти модели для $n > 5$ вряд ли находят практическое применение. Как правило, исследователь на первых порах

ограничивается линейной моделью с дополнительными парными (в крайнем случае, тройными) взаимодействиями. В подобной ситуации ПФЭ становится избыточным. Возникает необходимость в построении более экономной с точки зрения числа опытов разновидности планов. Поэтому были разработаны дробные факторные эксперименты (ДФЭ).

в) *Дробный факторный эксперимент.* Во многих практических задачах идентификации влияние взаимодействий (произведений факторов) второго и высших порядков отсутствует или пренебрежимо мало. Кроме того, на первых порах часто нужно получить в первом приближении лишь линейную аппроксимацию изучаемого уравнения связи при минимальном количестве опытов. В других задачах, кроме основных факторов, необходимо учитывать лишь некоторые взаимодействия из множества возможных взаимодействий, входящих в планы ПФЭ. Поэтому неэффективно использовать ПФЭ для оценивания коэффициентов лишь при линейных членах и некоторых парных произведениях из-за реализации большого числа вариантов варьирования (2^n), в особенности при большом числе факторов n .

Эксперимент, реализующий часть (дробную реплику) полного факторного эксперимента, называется *дробным факторным экспериментом* (ДФЭ). В эксперименте по плану ДФЭ пропущены некоторые сочетания уровней факторов. Сокращение перебора уровней приводит к потере информации. Поэтому при дробном факторном эксперименте важно так спланировать эксперимент, чтобы терялась наименее существенная, при данной постановке задачи, информация. Особенно широко используется дробный факторный эксперимент, в котором теряется информация о взаимодействиях изучаемых факторов. Это правомерно в тех случаях, когда эффекты взаимодействия заведомо отсутствуют или настолько малы, что их можно не учитывать. То есть мы должны иметь дополнительную априорную информацию о свойствах изучаемого объекта. Эта информация должна быть сформирована в виде списка *существенных переменных* – факторов и их взаимодействий, отобранных для возможного включения в модель. Конечно, в ходе дальнейшего исследования может оказаться, что часть этих переменных не влияет на отклик. В списке существенных переменных всегда присутствует фиктивная переменная x_0 , которой соответствует коэффициент β_0 . Качество результатов, полученных с помощью ДФЭ, зависит от правильности списка «существенных переменных». Практическая процедура заключается в следующем:

1. Из числа x_1, x_2, \dots, x_n факторов выбираются k ведущих, $k=n-p$. Пусть, например, первые x_1, x_2, \dots, x_k . Для них записывается план ПФЭ 2^k , то есть задается программа изменения каждого из ведущих факторов.

2. Для оставшихся $p=n-k$ факторов в качестве программы изменения в ходе эксперимента выбираются столбцы, соответствующие тем или иным произведениям ведущих факторов.

Соотношение, приводящее один из столбцов $x_{k+1}, x_{k+2}, \dots, x_n$ к столбцу определенного произведения факторов x_1, x_2, \dots, x_k называется *генерирующим соотношением*. Всего должно быть p генерирующих соотношений. Выбор их произволен, но нельзя использовать существенные переменные (иначе в матрице \bar{F} будут совпадающие столбцы).

3. Проверить пригодность найденного спектра плана. Применение ДФЭ всегда связано со *смешиванием*, то есть с совместным оцениванием нескольких теоретических коэффициентов математической модели. Если в матрице \bar{F} нет полностью совпадающих или полностью противоположных (в смысле знака) столбцов, то спектр плана пригоден. Если такие столбцы есть, то надо выбрать другое генерирующее соотношение или изменить набор ведущих факторов.

Для этого используется понятие *определяющего соотношения*. Выражение определяющего соотношения (ОС) получается перемножением, которое получается из генерирующего соотношения и имеет вид: слева 1, справа – произведение левой и правой частей генерирующего соотношения. Число определяющих соотношений равно числу генерирующих соотношений. Если определяющих соотношений несколько, то записывается *обобщенное определяющее соотношение* (ООС), которое равно почленному произведению всех определяющих соотношений.

Знание определяющего соотношения позволяет найти всю систему совместных оценок без изучения матрицы планирования ДФЭ. Соотношения, задающие эти оценки, можно найти, последовательно перемножая каждую переменную из списка существенных переменных на определяющее соотношение (с

учетом того, что $x_f^{2k} = 1, x_f^{2k+1} = x_f$). Получаем $d+1$ равенство, каждое из которых содержит 2^p переменных (факторов и их взаимодействий). Действия всех переменных, входящих в определенное равенство, смешаны между собой. Им соответствуют одинаковые столбцы матрицы \bar{F} . Если какие-либо из указанных равенств содержат хотя бы двух переменных из списка существенных переменных, то план не пригоден и должен быть изменен.

7 Лекция. Задача оптимизации в экспериментальных исследованиях

Содержание лекции: изучение основных принципов поисковой оптимизации при наличии аддитивных случайных помех.

Цель лекции: изучить постановку задачи поисковой оптимизации и принципы решения их.

7.1 Постановка задача оптимизации в экспериментальных исследованиях

Задача оптимизации – одна из наиболее важных и распространенных задач, встречающихся в практике научных и инженерных исследований как теоретического, так и прикладного характера. Во многих случаях основной целью является определение наилучших в некотором смысле условий, решений, значений параметров, уровней факторов. Такие оптимизационные задачи возникают, например:

- 1) при управлении различными технологическими процессами, агрегатами, где необходимо достижение максимума производства при наилучшем качестве и минимуме затрат;
- 2) при проектировании различных инженерных устройств, приборов, когда необходимо подобрать параметры таким образом, чтобы получить наилучшие эксплуатационные характеристики;
- 3) при создании новых образцов продукции с наилучшими свойствами (например, сплав с максимальной прочностью);
- 4) задачи чисто вычислительного характера, определение регрессионной модели методом наименьших квадратов или численное построение плана экспериментов в соответствии с выбранным критерием.

Задача оптимизации формулируется следующим образом: найти значения управляемых факторов $x_1 = x_1^*, x_2 = x_2^*, \dots, x_n = x_n^*$ объекта исследования, при которых его отклик y (целевая функция, критерий оптимизации) достигает своего экстремального значения (минимума или максимума)

$$y^* = \text{ext}_x y(x).$$

Экстремальные точки во многих случаях находятся с учетом определенных ограничений на \bar{x} .

Все методы оптимизации можно разбить на два класса:

1) теоретические методы, когда задача оптимизации определена с математической точки зрения и допускает применение аналитических методов оптимизации – дифференциальное или вариационное исчисление, линейное, динамическое, целочисленное программирование;

2) экспериментальные методы, используемые в условиях, когда функция отклика $y(\bar{x})$ неизвестна и имеется возможность измерить значения y при различных x_1, x_2, \dots, x_n . Такая ситуация может быть не только при исследовании физических объектов, но и в задачах теоретического плана, когда аналитические

методы непригодны и используются численные методы (эксперименты проводятся на вычислительной машине).

Отличием методов оптимизации в задачах исследования реальных объектов от процедур чисто вычислительного плана является наличие неконтролируемых факторов – шума случайного характера. Рассматриваемые методы оптимизации делятся на две группы: поисковые и методы, основанные на предварительном получении эмпирической модели объекта, описывающей его поведение в области оптимума.

В поисковых методах осуществляется последовательное локальное изучение поверхности отклика. Экстремальное значение достигается с помощью последовательных процедур, включающих в себя:

а) определение по результатам специально спланированных экспериментов направления движения из некоторой начальной точки, в окрестности которой проводится эксперимент. Это направление зависит от локальных свойств поверхности отклика вблизи данной точки и определяется таким образом, чтобы движение в найденном направлении приводило к значению отклика, более близкому к оптимальному, чем в предыдущей точке;

б) организацию движения в найденном направлении;

в) многократное повторение указанных этапов до достижения точки оптимума.

Вторая группа методов предусматривает изучение не локальных, а общих свойств поверхности отклика в определенной области факторного пространства, где предположительно находится оптимум. Для методов оптимизации данной группы наибольший интерес представляет проблема планирования эксперимента, обеспечивающего наивысшую точность в определении координат оптимума.

Большинство объектов имеют ограничения, и в этих условиях применение методов поиска имеет свои особенности. Имеется два типа ограничений: факторные и функциональные.

Факторными называются ограничения, накладываемые на допустимые значения отдельных факторов:

$$x_{i\min} \leq x_i \leq x_{i\max}, i = 1, \dots, n \quad (1)$$

Ограничения этого типа фактически всегда присутствуют при решении любой оптимизационной задачи.

Ограничения на значения совокупности нескольких факторов

$$A_{j\min} \leq \varphi_j(x_1, \dots, x_n) \leq A_{j\max}, j = 1, 2, \dots, k. \quad (2)$$

Такие ограничения имеют место, если какие-либо комбинации значений факторов x_i оказываются недопустимыми. Например,

$$x_1 + x_2 + x_3 \leq A_{\max}, x_1 x_2 \leq A_{\min}, \sqrt{\sum x_i^2} \leq A \quad \text{и т.д.}$$

Приведенные типы ограничений полностью известны, заданы в аналитической форме.

Функциональные ограничения – это ограничения, накладываемые на целевые функции (при многооткликовом объекте исследования), характеризующие качественные или количественные стороны объекта:

$$y_{j\min} \leq y_j \leq y_{j\max}, \quad (3)$$

p – номер отклика.

То есть требуется получить экстремальное значение одной функции отклика при ограничениях, наложенных на другие функции отклика (например, объем выпуска продукции должен быть не ниже плана, при этом процент примесей в готовой продукции не должен быть выше допустимого). В общем случае неравенства (1-3) определяют в факторном пространстве область допустимых значений Ω . Если точка экстремума находится вне области Ω , то задача оптимизации формулируется как поиска условного экстремума, когда требуется найти максимум y на подмножестве $x \in \Omega$, то есть наилучшее значение на границе области ограничений.

7.2 Решение задачи оптимизации при наличии помех

Для решения задачи оптимизации можно применить два принципиально различных подхода:

1) если известна или есть возможность найти n -факторную математическую модель для той части факторного пространства, где расположен экстремум функции отклика, то задачу оптимизации решают аналитическим или численным методом;

2) если математическое описание не получено по каким-либо причинам, то осуществляют экспериментальный поиск точки экстремума.

В первом случае используют известное из математического анализа свойство функций, имеющих экстремум: в точке экстремума первая производная этой функции обращается в нуль. Приравняв нулю все частные производные функции отклика по каждому из n факторов, получают систему уравнений, которую можно решить аналитически или численно.

Во многих практических случаях аналитическая зависимость функции отклика неизвестна или ее нахождение представляет собой сложную задачу. Тогда задачу оптимизации проще решать с помощью второго подхода, то есть с помощью экспериментального анализа. Для этого сначала с помощью специально спланированных пробных опытов изучают характер поверхности отклика в районе первоначально выбранной (*базовой, основной*) точки факторного пространства. Затем совершают «рабочее» движение в сторону экстремума, причем направление движения определяют по результатам пробных опытов. Такое движение может осуществляться путем ряда этапов, которые могут объединяться в «циклы» (последовательная процедура).

Методы поисковой оптимизации различаются способами постановки пробных опытов в окрестности начальной (базовой) точки и определения направления движения к экстремуму, а также способами организации самого рабочего движения к экстремуму. Выбор базовой точки является отдельной задачей. Обычно базовая точка соответствует номинальному режиму ведения технологического процесса $\bar{x}_0 = (x_{10}, x_{20}, \dots, x_{n0})$. Иногда эту точку выбирают в центре области, которую желательно исследовать, либо в центре области ограничений, если они имеются (при таком выборе базовой точки все направления являются равноправными, а это важно в случае, когда заведомо ничего неизвестно о том, где, хотя бы примерно, расположен экстремум).

Для повышения надежности результатов в каждой запланированной точке факторного пространства выполняют по несколько параллельных опытов. Разные поисковые методы в разных условиях обладают различной помехоустойчивостью. Под *помехоустойчивостью* метода понимается его способность правильно оценивать направление рабочего движения, а также способность быстро и точно приводить рабочую точку в область экстремума, несмотря на наличие помех.

Среди методов поисковой оптимизации можно отметить следующие:

- *метод Гаусса-Зейделя*, который предусматривает поочередное нахождение частных экстремумов целевой функции по каждому фактору $x_1, x_2, \dots, x_n, i = 1, 2, \dots, n$. при этом на каждом i -м этапе стабилизируют $n-1$ факторов и варьируют только один. i -й фактор. Задачу поиска экстремума методом Гаусса-Зейделя решают в несколько этапов, объединенных в циклы;

- *градиентные методы* имеют несколько разновидностей, различающихся правилами выбора ступеней варьирования и рабочих шагов на каждом этапе движения к экстремуму. Сущность стратегии всех этих разновидностей состоит в том, что на каждом этапе вокруг очередной базовой точки организуют пробные эксперименты, по результатам которых оценивают новое направление градиента, после чего в этом направлении совершают один рабочий шаг. Среди градиентных методов можно отметить метод простого градиента, метод крутого восхождения (метод Бокса-Уилсона);

- *симплексный метод* позволяет совмещать пробные опыты для определения направления движения с рабочим движением по поверхности отклика в области оптимума. Напомним, что симплексом называется выпуклая фигура, образованная $n+1$ вершинами в пространстве n факторов, причем эти $n+1$ вершин не принадлежат одновременно ни одному из подпространств из $n-1$ факторов. Основная идея симплексного метода состоит в следующем. Если во всех $n+1$ вершинах симплекса поставить опыты и измерить отклик, то (при не слишком большом уровне шумов) по величине отклика в вершинах можно судить, в каком направлении следует двигаться, чтобы приблизиться к экстремуму;

- метод случайного поиска заключается в том, что точку каждого пробного опыта для изучения поверхности отклика в районе базовой (начальной) точки выбирают случайным образом. Опыты производят в исходной (начальной) точке и в случайно выбранной пробной точке, измерения отклика в них сравнивают и, если ищется максимум, совершают рабочий шаг в направлении возрастания целевой функции.

Лекция 8. Планы второго порядка

Содержание лекции: планирование второго порядка для получения математического описания в виде полинома второй степени

Цель лекции: изучить методы планирования экспериментов для построения регрессионных моделей в виде квадратичных полиномов

8.1 Структура планов второго порядка

Планы второго порядка предназначены для получения регрессионной модели в виде полного квадратичного полинома (полинома второй степени):

$$y(x) = a_0 + \sum_{i=1}^n a_i x_i + \sum_{i=1}^n a_{ii} x_i^2 + \sum_{\substack{i,j=1 \\ i < j}}^n a_{ij} x_i x_j \quad (1)$$

Такие планы используются, когда заранее известно, что объект обладает существенно нелинейными свойствами, либо когда использование планирования первого порядка не позволило получить адекватную регрессионную модель и появилась необходимость ее усложнить.

Планы второго порядка более сложные по структуре, требуют при своей реализации увеличенного количества опытов. Модель (1) содержит $k = 1 + 2n + C_n^2 = \frac{(n+1)(n+2)}{2}$ членов, что в $\frac{n+2}{2}$ раза больше, чем в линейной модели. Следовательно, возрастает минимально необходимое количество точек в спектре плана. Для получения квадратичной зависимости каждый фактор должен изменяться, по крайней мере, на трех уровнях.

Все планы второго порядка делятся на две большие группы: симметричные и несимметричные.

Симметричным называется такой план, для которого

$$\sum_{i=1}^N x_{ig} = \Lambda_2; \sum_{i=1}^N x_{ig}^2 x_{jg} = \Lambda_{22}, i \neq j; \sum_{i=1}^N x_{ig}^4 = \Lambda_4;$$

$$\sum_{i=1}^N x_{ig}^3 = \sum_{i=1}^N x_{ig} x_{jg} + \sum_{i=1}^N x_{ig} x_{jg} x_{kg} = 0, i \neq j$$

для любых $i, j, k=1, 2, \dots, n; i \neq j, N$ -число точек плана.

То есть для всех факторов x_i или указанных комбинаций нескольких факторов записанные суммы равны определенным постоянным, причем константы одни и те же независимо от номера фактора. Несимметричные планы этим свойством не обладают.

Симметричные планы обладают большей упорядоченностью в расположении точек. Для них характерно подобие в программе изменения каждого из факторов, а в некоторых случаях и полная симметрия. Для симметричных планов справедливы достаточно простые соотношения для оценок коэффициентов регрессии, их выборочных дисперсий и ковариаций. Это связано с тем, что информационная матрица Фишера имеет своеобразную блочную структуру, причем большинство ее элементов равно нулю, а остальные равны Λ_2, Λ_{22} . Несимметричные планы во многом менее удобны, чем симметричные. Но зато более экономны по необходимому количеству опытов.

При использовании планов второго порядка весьма важным оказывается вопрос об области планирования эксперимента, то есть той части факторного пространства, где должны располагаться точки плана. Можно указать три варианта ее задания:

1) область планирования естественная, полученная путем достройки планов первого порядка. При этом новые, дополнительные точки, координаты которых определяются с учетом тех или иных критериев оптимальности, могут выходить за пределы области планирования на первом этапе, то есть за пределы гиперкуба $-1 < x_i < +1, i=1, 2, \dots, n$. Размеры области планирования, получающиеся путем дополнения, зависят от вида критерия оптимальности и количество факторов n ;

2) область планирования - гиперкуб $|x_{ig}| \leq 1, i=1, \dots, n, g=1, \dots, N;$

3) область планирования – единичный гипершар $x_1^2 + x_2^2 + \dots + x_n^2 \leq 1$.

Два последних варианта в отличие от первого предусматривают, что область планирования выбирается заранее, исходя от особенностей решаемой задачи. Область в виде гиперкуба более естественна с практической точки зрения.

При необходимости план всегда можно перенастроить для иной области планирования. Например, чтобы получить план, вписанный в гиперкуб, если имеется план, заданный в естественной области, для

каждого фактора x_i находят максимальное по модулю значение и заменяют x_i на $\frac{x_i}{\max x_{ij}}$.

8.2 Симметричные планы второго порядка

Центральные композиционные (то есть строящиеся последовательно) планы (ЦКП) состоят из трех блоков:

1) основу или ядро плана составляют точки ПФЭ 2^n или ДФЭ 2^{n-p} (количество точек $N_\Phi = 2^n$);

2) так называемые «звездные» точки, расположенные на координатных осях на расстоянии $\pm \alpha$ от центра эксперимента; общее число таких точек $N_\alpha = 2n$;

3) опыты в центре плана, то есть нулевые (центральные) точки; число таких точек $N_0 \geq 1$.

Общее число точек ЦКП $N = N_{\phi} + N_{\alpha} + N_0$.

Произвольный симметричный план ЦКП:

Составляющие части ЦКП	g	x ₁	x ₂	...	x _n	Число точек
Ядро плана	1	-1	-1	..	-1	2 ^{n-p}
	2	+1	-1	...	-1	
	...	-1	+1	...	-1	
	2 ^{n-p}	
«Звездные» точки	2 ^{n-p} + 1	-α				2n
	2 ^{n-p} + 2	+ α				
	...		- α			
	...		+ α			
	...				- α	
	2 ^{n-p} + 2n				+ α	
Центральные точки	2 ^{n-p} + 2n + 1	0			0	N ₀
				
	2 ^{n-p} + 2n + N ₀	0			0	

Конкретные значения α и N₀ выбираются, исходя из тех или иных критериев оптимальности регрессионных экспериментов.

8.2.1 Ортогональные центральные композиционные планы (ОЦКП)

В случае ОЦКП критерием оптимальности плана является ортогональность столбцов матрицы планирования. В силу ортогональности все оценки коэффициентов определяются независимо друг от друга. Как правило, в ОЦКП N₀=1. По критерию ортогональности матрица Фишера должна быть диагональной, для чего необходимо принять меры для обеспечения попарной ортогональности столбцов, отвечающих свободному члену β₀ и квадратичным коэффициентам β_{ii, i=1, ..., n}, а также столбцов, отвечающих квадратичным членам между собой.

Ортогонализацию столбцов x₀ и x_i² производят с помощью преобразования

$$\bar{x}_i^2 = x_i^2 - \frac{1}{N} \sum_{i=1}^N x_{iF}^2 = x_i^2 - (\bar{x}_i^2)$$

где N – общее число строк матрицы планирования.

То есть регрессионная модель ищется в виде

$$\varphi(x) = \beta'_0 + \sum_{i=1}^n \beta_i x_i - \sum_{i=1}^{n-1} \sum_{j=1}^n \beta_{ij} x_i x_j + \sum_{i=1}^n \beta_{ii} (x_i^2 - \bar{x}_i^2); \quad \beta'_0 = \beta_0 + \sum_{i=1}^n \beta_{ii} \bar{x}_i^2$$

Условие ортогональности справедливо при любом α. Но при произвольном α остаются неортогональными столбцы матрицы F, отвечающие различным квадратичным переменным. Поэтому в ОЦКП значение α выбирается как раз из условия ортогональности именно этих столбцов. Отсюда выводится уравнение для расчета величины звездного плеча

$$4\alpha^4 + 4N_\phi \alpha^2 - N_\phi(N_\alpha + N_0) = 0,$$

где α – величина звездного плеча, N_ϕ – число точек ПФЭ (ДФЭ), N_0 – число центральных точек, N_α – число «звездных» точек

Параметры ОЦКП для $n=2,3,4$:

n	α	N_α	N_0	N_ϕ	N
2	1,0	4	1	4	9
3	1,215	6	1	8	15
4	1,414	8	1	16	25

Точность оценивания различных групп коэффициентов регрессии неоднородна (не как в ПФЭ): наиболее точно оценивается свободный член A_0 , несколько менее точно – линейные коэффициенты, еще хуже – коэффициенты A_{ij} при парных взаимодействиях, наименее точно – квадратичные коэффициенты A_{ii} . Вычисление оценок коэффициентов регрессии, как обычно, производится методом наименьших квадратов.

8.2.2 Ротатабельные центральные композиционные планы (РЦКП)

ОЦКП не удовлетворяет критерию ротатабельности планирования, причем при фиксированном расстоянии от центра планирования дисперсии предсказания по разным направлениям могут различаться весьма значительно, что с позиции экспериментатора крайне нежелательно.

РЦКП строится исходя из требования критерия ротатабельности. Критерий оптимальности в этом случае

$$\sigma^2 \{ \hat{y} \} = \text{const} \quad \text{при } R = \text{const}.$$

Величину звездного плеча α вычисляют по формуле $\alpha = 2^{\frac{n-1}{4}}$.

Ротатабельные планы оптимальны и в том смысле, что они позволяют минимизировать систематические ошибки, связанные с неадекватностью представления результатов исследования полиномами второго порядка.

Число центральных точек N_0 не влияет на свойство ротатабельности и выбирается таким образом, чтобы обеспечивать равную точность предсказания выходной величины \hat{y} внутри области планирования.

Необходимые значения α_0 и N_0 приведены в таблице

n	α	N_α	N_0	N_ϕ	N
2	1,414	4	5	4	13
3	1,682	6	6	8	20
4	2,000	8	7	16	31

Ротатабельный план не требует ортогонализации, поэтому никаких преобразований переменных при составлении матрицы планирования не производят.

Список литературы

1. Красовский Г.И., Филаретов Г.Ф. Планирование эксперимента. – Минск: изд. БГУ, 1982.
2. Хартман К. и др. Планирование эксперимента в исследовании технологических процессов. – М.: Мир, 1977.
3. Химмельблау Д. Анализ процессов статистическими методами. – М.: Мир, 1972.
4. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. – М.: Мир, 1987.
5. Круг Г.К. и др. Планирование эксперимента в задачах нелинейного оценивания и распознавания образов. – М.: Наука, 1981.