

Министерство науки и высшего образования РФ

Томский государственный университет
систем управления и радиоэлектроники

Е.Ю. Костюченко, А.Ю. Якимук

НАУЧНО-ТЕХНИЧЕСКИЙ СЕМИНАР

Учебно-методическое пособие
для студентов направлений подготовки
10.00.00 Информационная безопасность

Томск
2022

УДК 004.056
ББК 32.973.26-018.2
К 64

Костюченко, Евгений Юрьевич

К 64 Научно-технический семинар: учебно-методическое пособие / Е.Ю. Костюченко, А.Ю. Якимук. – Томск: Томск. гос. ун-т систем упр. и радиоэлектроники, 2022. – 44 с.

Настоящее учебно-методическое пособие содержит описания лабораторных и самостоятельных работ по дисциплине «Научно-технический семинар» для направлений подготовки, входящих в укрупненную группу специальностей и направлений 10.00.00 Информационная безопасность.

УДК 004.056
ББК 32.973.26-018.2

© Е.Ю. Костюченко, А.Ю. Якимук, 2022
© Томск. гос. ун-т систем упр. и радиоэлектроники, 2022

СОДЕРЖАНИЕ

Введение.....	4
ЛАБОРАТОРНАЯ РАБОТА №1	
Определение критериев подобия и параметров модели	5
ЛАБОРАТОРНАЯ РАБОТА №2	
Дисперсионный анализ.....	12
ЛАБОРАТОРНАЯ РАБОТА №3	
Корреляционный анализ.....	21
ЛАБОРАТОРНАЯ РАБОТА №4	
Регрессионный анализ	26
ЛАБОРАТОРНАЯ РАБОТА №5	
Полный факторный эксперимент	35
Литература	44

Введение

Целью преподавания дисциплины является формирование системного представления о методах научных исследований и обучение студентов принципам использования научных методов проведения активного и пассивного экспериментов.

Задачи изучения дисциплины:

- Дать общее представление о процессе научного исследования;
- Дать общее представление о методологии и организации научного исследования;
- Научить решать задачи разработки конструкций и их технологических процессов производства строгим математическим путём.

ЛАБОРАТОРНАЯ РАБОТА №1

Определение критериев подобия и параметров модели

1. Цель работы

Цель работы состоит в практическом освоении одного из способов определения критериев подобия и параметров модели, используемых при моделировании конструкций в процессе их разработки и производства.

2. Общие положения

Сложность конструкций электронных вычислительных средств (ЭВС) и протекающих в них физических процессов, а также сложность процессов, необходимых для производства ЭВС, часто не позволяют исследовать их в реальных условиях. Выходом является исследование на различного рода моделях, некоторым образом подобных моделируемому объекту или процессу. В основе такой методики изучения лежит теория подобия и моделирования, являющаяся, по сути дела, методологией эксперимента, без которого, к сожалению, пока не удаётся обойтись при проектировании и производстве ЭВС.

Теория подобия и моделирования широко используется для моделирования электрических, тепловых, механических и других процессов, при исследовании надёжности сложных систем и управлении их качеством на этапах проектирования и экспериментальной отработки. На базе этой теории исследуются свойства конструкций при изменении их размеров в процессе микроминиатюризации при условии сохранения основных электрических параметров и надёжности. При проведении испытаний ЭВС в несколько различных условиях или режимах теория подобия используется для того, чтобы установить, являются ли математически подобными процессы, протекающие в этих условиях или режимах. Можно привести ещё массу примеров.

В основе теории подобия лежат три теоремы. Кроме того, имеют место дополнительные положения:

- 1) о подобии сложных систем;
- 2) о подобии нелинейных систем;
- 3) о подобии анизотропных или неоднородных систем;
- 4) о подобии физических явлений при отсутствии геометрического подобия;
- 5) об условиях подобия при вероятностном характере изучаемых явлений.

Из этого перечня также следует, что теория подобия применима к широкому классу явлений и систем.

Если говорить о видах подобия, то наряду с основными (полным, неполным и приближённым) существуют и особые виды: квазиподобие, эквивалентное, функциональное, кибернетическое и интегральное подобие. По отношению к этим видам методы подобия начинают уже применяться не только как средство, направляющее постановку эксперимента и облегчающее обработку его результатов, но и как средство, облегчающее решение уравнений.

3. Теоретические сведения

3.1. Способы определения критериев подобия

Одной из главных задач теории подобия и моделирования является определение критериев подобия. Эти критерии находятся как для классического подобия, так и для особых видов подобия.

Используются, главным образом, три способа:

- 1) на основе анализа уравнений (способ интегральных аналогов);

- 2) на основе анализа размерностей;
- 3) на основе констант подобия.

Эти способы можно применять, когда имеется математическое описание процесса. Достоинством второго способа по отношению к двум остальным является то, что он может применяться и в тех случаях, когда известны только параметры исследуемого процесса, а его математическое описание отсутствует.

Во втором способе имеется три метода: метод П-теоремы (прежде он проходил под названием способ П-теоремы), метод нулевых степеней и метод исключения размерностей

Рассмотрим более подробно метод П-теоремы.

Известно, что при описании системы (или процесса) m -параметрами, из которых k имеют независимые между собой размерности, количество критериев подобия в такой системе будет $m - k$.

Каждый критерий в рассматриваемом способе определяется по формуле

$$P_i = \frac{Z_i}{Z_1^{x_1} \cdot Z_2^{x_2} \cdot \dots \cdot Z_k^{x_k}},$$

где Z_1, \dots, Z_k - независимые по размерностям переменные; Z_i - зависимая переменная; $i = \overline{k+1, m}$; x_1, \dots, x_k - неизвестные показатели степени, которые следует определить для каждого критерия.

Приведённая методика требует вначале выбора k независимых переменных. Их количество не может быть больше числа основных единиц размерностей используемой системы измерений - q , т.е. $k \leq q$. Выбор числа независимых переменных проводится следующим образом.

Размерности всех m переменных записываются через основные единицы размерностей в соответствующих степенях. Из показателей степеней основных единиц размерностей составляется матрица $\|A\|$, которая содержит

m строк и q столбцов. Далее следует найти ранг этой матрицы, который определяет количество независимых переменных. Рангом матрицы называют наибольший порядок, который могут иметь её миноры (определители D), не обращающиеся в нуль.

Вначале исследуются на равенство нулю определители q -го порядка, составленные из матрицы $\|A\|$. Таких определителей будет

$$C_D = \frac{m!}{q!(m-q)!}.$$

Если хотя бы один из них окажется не равным нулю, то количество независимых переменных и ранг матрицы равны q . Если все определители оказываются равными нулю, то, следовательно, ранг матрицы $\|A\|$ меньше q и число независимых переменных – $k < q$. Для дальнейших исследований принимают $k = q - 1$.

Затем для определения ранга матрицы $\|A\|$ исследуют частичные матрицы $\|B\|$ размером $k \times q$, составленные из матрицы $\|A\|$. Таких частичных матриц будет

$$C_B = \frac{m!}{k!(m-k)!}.$$

В этих частичных матрицах анализируются определители k -го порядка. Если окажется, что существует хотя бы один из них, не равный нулю, то следовательно ранг матрицы $\|A\|$ равен k . В k независимых переменных входят те из m переменных, которым соответствуют строки определителя, не равного нулю.

Если в частичных матрицах $\|B\|$ окажется несколько определителей k -го порядка, не равных нулю, то имеем ряд наборов из k независимых переменных. Если все определители k -го порядка оказываются равными нулю, то

принимают $k = q - 2$ и исследование продолжается подобным описанному способом.

Для определения ранга матрицы удобно поступать следующим образом. Вначале в матрице находится минор наименьшего порядка (например, второго), не равный нулю. Затем находятся миноры большего порядка по правилу окаймления.

Определив ряд наборов, состоящих из k -независимых переменных, выбирают один из наборов. Для определения наборов независимых переменных используйте программу, установленную в ЭВМ. Из всех наборов следует выбрать один и проверить его вручную на независимость (определитель не равен нулю).

Теперь можно записать $(m - k)$ критериев подобия и перейти к определению степеней независимых переменных, стоящих в знаменателе того или иного критерия. Для этого переменные в числителе и в знаменателе критерия подобия записываются через их размерности. Затем составляется система уравнений, каждое из которых соответствует одной основной единице размерностей и представляет собой алгебраическую сумму показателей степеней для этой основной единицы в числителе и в знаменателе критерия. Эта сумма приравнивается нулю. Например,

$$P_i = \frac{l}{m^{x_1} \cdot w^{x_2} \cdot F^{x_3}}.$$

Записывая через размерности, получаем

$$P_i = \frac{[M^0 L^1 T^0]}{[M^1 L^0 T^0]^{x_1} \cdot [M^0 L^0 T^{-1}]^{x_2} \cdot [M^1 L^1 T^{-2}]^{x_3}}.$$

Система уравнений имеет вид:

$$\text{для } M: 0 = X_1 + X_3$$

$$\text{для } L: 1 = X_3$$

для T : $0 = -X_2 - 2X_3$.

Отсюда находим: $X_1 = -1$, $X_2 = -2$, $X_3 = 1$. Следовательно, критерий подобия будет

$$\Pi_i = \frac{l}{m^{-1} \cdot w^{-2} \cdot F} = \frac{mlw^2}{F}.$$

Для получения удобных видов критериев подобия найденные критерии можно перемножать и делить, возводить в степень, а также умножать их на постоянный коэффициент.

3.2. Определение параметров модели

Это определение проводится на основе первой теоремы теории подобия. Вначале выбираются базисные величины параметров оригинала. Базисная величина – это один из параметров данного наименования, который выбирается в качестве единицы измерения для остальных величин той же размерности. Например, в оригинале имеется несколько индуктивностей. Величина одной из них выбирается базисной. С остальными индуктивностями базисная величина связывается масштабными соотношениями. Таким образом, для оригинала известны все базисные величины и масштабные соотношения.

Выбираются произвольно, в разумных пределах, базисные величины k независимых параметров модели. Подставляя в критерии подобия известные базисные величины оригинала и модели и приравнивая одинаковые критерии, получаем систему уравнений, из которой определяем остальные базисные величины модели. Пользуясь масштабными соотношениями для оригинала типа m_{Z_i} и базисными величинами параметров модели, например, $Z_{iом}$, находим остальные параметры модели по соотношениям

$$Z_{iм} = m_{Z_i} \cdot Z_{iом}.$$

4. Порядок выполнения работы

1. Получить у преподавателя вариант задания.
2. Найти: количество определителей в матрице $\|A\|$, критерии подобия методом П-теоремы.
3. Сформулировать выводы.
4. Сдать отчёт преподавателю.

5. Вопросы для самопроверки

1. Какие вопросы позволяет решать теория подобия и моделирования.
2. Содержание трёх теорем подобия.
3. Какие существуют виды подобия.
4. Какие имеются способы и методы определения критериев подобия и в чём они состоят.
5. Каковы достоинства способа на основе анализа размерностей.
6. Что такое ранг матрицы.
7. Что такое независимые переменные.
8. Как определяются независимые переменные.
9. Как определяются показатели степеней независимых переменных в критериях подобия.
10. Что такое базисные величины параметров оригинала и модели.
11. Как определяются неизвестные базисные величины модели.
12. Как определяются остальные параметры модели.
13. Чем определяется количество видов записи критериев подобия в разных способах.
14. Как можно получить удобную форму критерия подобия.

ЛАБОРАТОРНАЯ РАБОТА №2

Дисперсионный анализ

1. Цель работы

Цель работы состоит в практическом освоении дисперсионного анализа, как метода, используемых для обработки опытных данных, полученных при проведении пассивного эксперимента.

2. Общие положения

Дисперсионный анализ является методом пассивного эксперимента. Это эксперимент, который может выполняться как на действующем оборудовании в производственных условиях, так и в лабораторных условиях.

В результате проведения дисперсионного анализа получают оценку значимости влияния фактора на функционирование объекта или характер процесса

3. Теоретические сведения

Сущность дисперсионного анализа заключается в разложении суммарной дисперсии, наблюдаемой в опыте на две: факторную, связанную с действием изучаемого фактора, и остаточную, обусловленную техникой эксперимента, и их сравнением.

Модель однофакторного дисперсионного анализа имеет следующий вид:

$$y_{ji} = \bar{Y} + T_j + \varepsilon_{ji}, \quad i = \overline{1, n_j}, \quad j = \overline{1, k},$$

где

y_{ji} – i -е наблюдение над параметром y , когда фактор А находится на j -том уровне,

\bar{Y} – общее среднее значение параметра y ,

T_j – эффект j -го уровня фактора А, т.е. вклад фактора А в величину функции y , когда фактор находится на j -ом уровне,

ε_{ji} – случайная ошибка в i -ом наблюдении на j -ом уровне,

k – число уровней фактора A ,

n_i – число наблюдений функции y на j -ом уровне фактора A .

Основные предположения дисперсионного анализа:

- 1) случайные величины ε_{ji} независимы и нормально распределены с параметрами $M(\varepsilon_{ji}) = 0$, $D(\varepsilon_{ji}) = \sigma^2 = const$ (дисперсии однородны и не зависят от уровня фактора);
- 2) для модели с фиксированными уровнями (т.е. фактор A принимает только данные наблюдаемые значения) $\sum_{i=1}^k T_j = 0$;
- 3) для модели со случайными уровнями величины T_j – независимые нормально распределенные случайные величины с параметрами $M(\varepsilon_{ji}) = 0$, $D(\varepsilon_{ji}) = \sigma^2 = const$.

Из модели получим основное уравнение дисперсионного анализа. Для этого переносим \bar{Y} в левую часть модели и добавляем $\pm \bar{y}_j$ — среднее арифметическое по j -у уровню:

$$y_{ji} - \bar{Y} = T_j + \varepsilon_{ji} \equiv \bar{y}_j - \bar{Y} + y_{ji} - \bar{y}_j.$$

Отклонение наблюдения параметра y от общего среднего определяется отклонением среднего по уровню от общего среднего и отклонением наблюдения от среднего по уровню.

Разность $\bar{y}_j - \bar{Y}$ — обусловлена действием фактора A , разность $y_{ji} - \bar{y}_j$ — обусловлена техникой эксперимента.

Возводим обе части тождества в квадрат и суммируем по всем уровням и наблюдениям.

$$\begin{aligned} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ji} - \bar{Y})^2 &= \sum_{j=1}^k \sum_{i=1}^{n_j} [(y_j - \bar{Y}) + (y_{ji} - \bar{y}_j)]^2 = \\ &= \sum_{j=1}^k \sum_{i=1}^{n_j} (y_j - \bar{Y})^2 + \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2 \end{aligned}$$

$$\text{т. к. } 2 \sum_{j=1}^k \sum_{i=1}^{n_j} (y_j - \bar{Y})(y_{ji} - \bar{y}_j) = 2 \sum_{j=1}^k (y_j - \bar{Y}) \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j) = 0$$

Полученное основное уравнение дисперсионного анализа представляет собой разложение суммы квадратов отклонений от общего среднего на сумму квадратов отклонений средних по уровню от общего среднего и сумму квадратов отклонений внутри испытаний (правило разложения вариации).

Это уравнение можно записать в следующем виде:

$$SS_{\text{общ}} = SS_A + SS_{\text{ош}}.$$

Суммы квадратов отклонений, деленные на соответствующее число степеней свободы, являются несмещенными оценками дисперсий:

$$\begin{aligned} S_{\text{общ}}^2 &= \frac{SS_{\text{общ}}}{f_{\text{общ}}} = \frac{1}{N-1} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ji} - \bar{Y})^2, \\ S_A^2 &= \frac{SS_A}{f_A} = \frac{1}{k-1} \sum_{j=1}^k \sum_{i=1}^{n_j} (\bar{y}_j - \bar{Y})^2 = \frac{1}{k-1} \sum_{j=1}^k n_j (\bar{y}_j - \bar{Y})^2, \\ S_{\text{ош}}^2 &= \frac{SS_{\text{ош}}}{f_{\text{ош}}} = \frac{1}{N-k} \sum_{j=1}^k \sum_{i=1}^{n_j} (y_{ji} - \bar{y}_j)^2, \end{aligned}$$

где $N = \sum_{j=1}^k n_j$ – общее число наблюдений. Если по всем уровням одинаковое число наблюдений n , то $N = k \cdot n$.

Таким образом в дисперсионном анализе, кроме общей дисперсии, вычисляют еще другие оценки рассеяния: дисперсию между выборками S_A^2 , которая основана на колебаниях частных средних (средних по выборке) вокруг общей средней, и дисперсию внутри выборки $S_{\text{ош}}^2$, которая основана на колебании значений параметра вокруг частной средней внутри отдельных выборок.

Схему дисперсионного анализа можно представить в виде следующей таблицы:

Источник изменчивости (дисперсии)	Число степеней свободы	Сумма квадратов	Дисперсии (средний квадрат)	Критерий
Различие между уровнями (выборками)	$f_A = k - 1$	SS_A	$S_A^2 = \frac{SS_A}{f_A}$	$F_{расч} = \frac{S_A^2}{S_{ош}^2}$
Различие внутри уровней (выборок)	$f_{ош} = N - k$	$SS_{ош}$	$S_{ош}^2 = \frac{SS_{ош}}{f_{ош}}$	$\alpha =$ $F_{табл} =$ $F_{расч} \geq F_{табл}$
Сумма	$f_{общ} = N - 1$	$SS_{общ}$	$S_{общ}^2 = \frac{SS_{общ}}{f_{общ}}$	влияет Фактор не влияет

Заключительным этапом дисперсионного анализа является сравнение двух дисперсий S_A^2 и $S_{ош}^2$ с помощью F -критерия Фишера

$$F_{расч} = \frac{S_A^2}{S_{ош}^2}.$$

Расчетное значение F -критерия сравнивается с его табличным значением, найденным по заданному уровню значимости α (или $q\%$, $\alpha = q/100$) или доверительной вероятности (надежности вывода) $P = 1 - \alpha$ и соответствующем числе степеней свободы f_A и $f_{ош}$:

$$F_{табл}(\alpha, f_A, f_{ош}).$$

Если $F_{расч} \geq F_{табл}$, то гипотеза H_0 , заключающаяся в том, что различие между сравниваемыми величинами S_A^2 и $S_{ош}^2$ отсутствует, отвергается. Это значит, что различие между дисперсиями S_A^2 и $S_{ош}^2$ существенно и изучаемый фактор A влияет на признак качества (исследуемый параметр y), т.е. влияние фактора A значимо.

Если $F_{расч} < F_{табл}$, то в этом случае принимается нулевая гипотеза H_0 : различие между сравниваемыми дисперсиями случайное и фактор не влияет на выходной параметр.

Для уверенности в правильности применения F -критерия необходимо, чтобы остаточная дисперсия $S_{ош}^2$ была установлена достаточно основательно, на числе степеней свободы не менее 10, т.е. должно быть выполнено условие $N - k = k(n - 1) \geq 10$.

Табличное значение F -критерия больше единицы, поэтому числитель $F_{расч}$ должен быть больше знаменателя, т.е. если $S_A^2 < S_{ош}^2$, то они меняются местами.

На практике обычно используют преобразованные формулы для нахождения сумм квадратов:

$$SS_A = \frac{1}{n} \sum_{j=1}^k (\sum_{i=1}^{n_j} y_{ji})^2 - \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}^2;$$

$$SS_{ош} = \sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}^2 - \frac{1}{n} \sum_{j=1}^k (\sum_{i=1}^{n_j} y_{ji})^2;$$

$$SS_{общ} = SS_A + SS_{ош} = \sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}^2 - \frac{1}{N} \sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}^2.$$

При разном числе наблюдений на уровнях n_j :

$$SS_A = \sum_{j=1}^k \frac{1}{n_j} (\sum_{i=1}^{n_j} y_{ji})^2 - \frac{1}{N} (\sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji})^2;$$

$$SS_{\text{ош}} = \sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}^2 - \sum_{j=1}^k \frac{1}{n_j} (\sum_{i=1}^{n_j} y_{ji})^2.$$

Для выполнения дисперсионного анализа результаты наблюдений удобнее сводить в таблице следующего вида:

Уровень фактора	Результаты наблюдений (значения признака качества)						Сумма по уровню (выборке)	Квадрат суммы по уровню	Сумма квадратов наблюдений по уровню	Среднее значение (в выборке)
	1	2	...	i	...	n _j				
1	y ₁₁	y ₁₂	...	y _{1i}	...	y _{1n_j}	$\sum_{i=1}^{n_j} y_{1i}$	$\left(\sum_{i=1}^{n_j} y_{1i}\right)^2$	$\sum_{i=1}^{n_j} y_{1i}^2$	
2	y ₂₁	y ₂₂	...	y _{2i}	...	y _{2n_j}	$\sum_{i=1}^{n_j} y_{2i}$	$\left(\sum_{i=1}^{n_j} y_{2i}\right)^2$	$\sum_{i=1}^{n_j} y_{2i}^2$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
j	y _{j1}	y _{j2}	...	y _{ji}	...	y _{jn_j}	$\sum_{i=1}^{n_j} y_{ji}$	$\left(\sum_{i=1}^{n_j} y_{ji}\right)^2$	$\sum_{i=1}^{n_j} y_{ji}^2$	$\bar{y}_j =$
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	
k	y _{k1}	y _{k2}	...	y _{ki}	...	y _{kn_j}	$\sum_{i=1}^{n_j} y_{ki}$	$\left(\sum_{i=1}^{n_j} y_{ki}\right)^2$	$\sum_{i=1}^{n_j} y_{ki}^2$	
							$\sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}$	$\sum_{j=1}^k \left(\sum_{i=1}^{n_j} y_{ji}\right)^2$	$\sum_{j=1}^k \sum_{i=1}^{n_j} y_{ji}^2$	$\bar{Y} =$

(поменять ориентацию страницы)

4. Порядок выполнения работы

1. Получить у преподавателя вариант задания.
2. Заполнить таблицу опытных и расчётных данных, проведя необходимые расчёты.
3. Заполнить таблицу «Схема дисперсионного анализа».
4. Определить $F_{\text{расч}}$.
5. Задавшись уровнем значимости α (или доверительной вероятностью P), найти табличное значение критерия $F_{\text{табл}}(\alpha, f_A, f_{\text{ост}})$.
6. Сделать выводы о принятии нулевой или альтернативной гипотезы, т.е. не влияет или влияет исследуемый фактор (значим или не значим).
7. Сдать отчёт преподавателю.

5. Вопросы для самопроверки

1. Цель дисперсионного анализа.
2. Какие суммы квадратов отклонений используются в дисперсионном анализе?
3. В чём заключается нулевая гипотеза H_0 ?
4. В чём заключается альтернативная гипотеза H_1 ?
5. Какой критерий используется в дисперсионном анализе?
6. По каким данным должны находиться значения критерия в таблицах?
7. Что такое схема дисперсионного анализа?
8. В чём сущность дисперсионного анализа?
9. Что представляет собой основное уравнение дисперсионного анализа?
10. В чём проявляется влияние фактора на выходной параметр?
11. В чём проявляется влияние техники эксперимента?
12. Почему расчётное значение F -критерия должно быть больше единицы и как это обеспечить?

13. Какие дисперсии определяются в дисперсионном анализе, и чем они обусловлены?

14. Какие степени свободы выбираются для определения соответствующих дисперсий?

15. Каково минимальное число наблюдений в дисперсионном анализе?

ЛАБОРАТОРНАЯ РАБОТА №3

Корреляционный анализ

1. Цель работы

Цель работы состоит в практическом освоении корреляционного анализа, как метода, используемых для обработки опытных данных, полученных при проведении пассивного эксперимента.

2. Общие сведения

Корреляционный анализ является методом пассивного эксперимента. Это эксперимент, который может выполняться как на действующем оборудовании в производственных условиях, так и в лабораторных условиях.

В результате проведения корреляционного анализа получают сведения о форме и тесноте связи между переменными.

3 Теоретические сведения

3.1 Построение корреляционной таблицы

Корреляционная таблица строится на основе массива опытных данных. Диапазоны значений y и x делятся на l и k интервалов соответственно. В каждом интервале записывается среднее значение y_i или x_i для данного интервала. По массиву опытных данных определяются частоты появления пар значений y и x в соответствующих интервалах и заносятся в таблицу.

Для упрощения процедуры вычислений используется метод «условного нуля», при котором новые (кодированные) средние значения y_j^{\cdot} и x_i^{\cdot} в интервалах определяются по формулам:

$$y_j^{\cdot} = \frac{y_j - y_0}{d_y},$$

$$x_i^{\cdot} = \frac{x_i - x_0}{d_x}.$$

где y_0 и x_0 – значения середин интервалов, которым соответствует наибольшие частоты m_y и m_x ;

d_y и d_x – ширина интервалов для x и y .

Значения y_0 и x_0 могут быть выбраны также в средних интервалах диапазонов изменения y и x .

Если в таблице введены кодированные значения y_j^{\cdot} и x_i^{\cdot} , то действительные значения соответствующих величин определяются по формулам:

$$\bar{x} = x_0 + d_x * \bar{x}^{\cdot};$$

$$\bar{y} = y_0 + d_y * \bar{y}^{\cdot};$$

средние квадратические отклонения:

$$S_x = d_x * S_x^{\cdot};$$

$$S_y = d_y * S_y^{\cdot};$$

$$S_{\bar{y}_x} = d_y^{\cdot} * S_{\bar{y}_x}^{\cdot};$$

ковариация:

$$C_{xy} = d_x * d_y * C_{xy}^{\cdot}.$$

3.2. Определение коэффициента корреляции, корреляционного отношения и их оценок

Для вычисления коэффициента корреляции r_{xy} и корреляционного отношения $\eta_{y/x}$ необходимо составить расчётную таблицу, в которую записываются все вспомогательные вычисления (см. Приложение).

После заполнения таблицы определяются условные:

средние значения

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k m_x * x_i,$$

$$\bar{y} = \frac{1}{n} \sum_{j=1}^e m_y * y_j,$$

ковариация

$$C_{xy} = \frac{1}{n} \sum_{i=1}^k (x_i \sum_{j=1}^e m_{xy} * y_j) - \bar{x} * \bar{y},$$

средние квадратические отклонения

$$S_x = \sqrt{\frac{1}{n} \sum_{i=1}^k m_x * (x_i)^2 - (\bar{x})^2},$$

$$S_y = \sqrt{\frac{1}{n} \sum_{j=1}^e m_y * (y_j)^2 - (\bar{y})^2}.$$

среднее квадратическое отклонение величин условных средних арифметических значений \bar{y}_x от общей средней \bar{y}

$$S_{\bar{y}_x} = \sqrt{\frac{1}{n} \sum_{i=1}^k m_x * (\bar{y}_x)^2 - (\bar{y})^2},$$

где $\bar{y}_x = \frac{1}{m_x} \sum_{j=1}^e m_{xy} * y_j$

m_{xy} – частоты в клетках.

Затем определяются действительные значения найденных величин по формулам, приведённым в разделе 3.2.1.

Для нахождения коэффициента корреляции и корреляционного отношения используются следующие выражения (при $n \geq 25$):

$$r_{xy} = \frac{C_{xy}}{S_x * S_y}.$$

$$\eta_{y/x} = \frac{S_{\bar{y}_x}}{S_y}.$$

Оценка достоверности коэффициента корреляции и корреляционного отношения производится следующим образом. Определяются их средние квадратические ошибки:

для коэффициента корреляции

$$\sigma_r = \frac{1-r_{xy}^2}{\sqrt{n}},$$

для корреляционного отношения

$$\sigma_\eta = \frac{1-\eta_{y/x}^2}{\sqrt{n}}.$$

Затем определяются отношения $\frac{r_{xy}}{\sigma_r}$ и $\frac{\eta_{y/x}}{\sigma_\eta}$.

Если эти отношения оказываются больше трёх (у ряда авторов > 4), то можно считать, что достоверности r_{xy} и $\eta_{y/x}$ доказаны. В противном случае достоверности r_{xy} и $\eta_{y/x}$ должны быть поставлены под сомнение, пока более глубокие исследования не опровергнут это мнение.

При достаточно большом n (практически при $n \geq 50$), действительные (генеральные) значения лежат в пределах:

для коэффициента корреляции

$$r_{xy} - t_\alpha * \sigma_r \leq r_{\text{ген}} \leq r_{xy} + t_\alpha * \sigma_r,$$

для корреляционного отношения

$$\eta_{y/x} - t_\alpha * \sigma_\eta \leq \eta_{y/x \text{ ген}} \leq \eta_{y/x} + t_\alpha * \sigma_\eta,$$

где t_α – определяется по таблице интеграла вероятностей (функции Лапласа):

$$\Phi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt.$$

Для этого задаются уровнем значимости q (%) или $\alpha = q$ (%) / 100, определяют доверительную вероятность (надёжность вывода) $\rho = 1 - \alpha = \Phi(t)$, по которой в таблице для $\Phi(t)$ находят t_α .

4. Порядок выполнения работы

1. Получить у преподавателя вариант задания.
2. Заполнить расчётную таблицу.
3. Определить коэффициент корреляции и корреляционное отношение.
4. Сделать оценки коэффициента корреляции и корреляционного отношения
5. Сделать выводы о линейности модели.
6. Сдать отчёт преподавателю.

5. Вопросы для самопроверки

1. Цель корреляционного анализа?
2. Для чего используется метод «условного нуля» при обработке экспериментальных данных?
3. Что такое коэффициент корреляции?
4. Что такое корреляционное отношение?
5. Что такое ковариация?
6. Как оценивается достоверность коэффициента корреляции?
7. Укажите свойства коэффициента корреляции?
8. Укажите свойства корреляционного отношения?
9. Что такое корреляционное поле?

ЛАБОРАТОРНАЯ РАБОТА №4

Регрессионный анализ

1. Цель работы

Цель работы состоит в практическом освоении регрессионного анализа, как метода, используемого для обработки опытных данных, полученных при проведении пассивного эксперимента.

2. Общие положения

Регрессионный анализ является методом пассивного эксперимента. Это эксперимент, который может выполняться как на действующем оборудовании в производственных условиях, так и в лабораторных условиях.

В результате проведения регрессионного анализа получают модель процесса в виде уравнения регрессии (линейного, квазилинейного или степенного).

3 Теоретические сведения

Прежде, чем выбирать регрессионную модель, необходимо путём анализа коэффициента корреляции и корреляционного отношения установить форму связи между переменными: линейная или нелинейная. Для этого приведём некоторые свойства коэффициента корреляции:

- 1) Если $r_{xy} = \pm 1$, то x и y связаны линейной функциональной связью вида $y = b_0 + b_1 * x$;
- 2) Если $r_{xy} = 0$, то между x и y нет линейной корреляционной связи, но криволинейная возможна;
- 3) Чем ближе коэффициент корреляции к ± 1 , тем точнее и сильнее линейная корреляционная связь;
- 4) Чем ближе коэффициент корреляции к нулю, тем слабее линейная корреляционная связь.

Некоторые свойства корреляционного отношения:

- 1) Если $\eta_{y/x} = 0$, то корреляционная связь между x и y отсутствует.
- 2) Если $\eta_{y/x} = 1$, то связь между x и y криволинейная функциональная (при этом должно быть $r_{xy} = 0$).
- 3) Всегда $\eta_{y/x} \geq |r_{xy}|$.
- 4) При $\eta_{y/x} \approx |r_{xy}|$ связь между x и y является точной линейной корреляционной.

Таким образом, если после определения r_{xy} и $\eta_{y/x}$, они оказываются приблизительно равными, а их левые доверительные границы достаточно велики, то можно считать, что имеется линейная корреляционная связь между переменными x и y .

Однако после такой визуальной оценки необходимо проверить гипотезу равенства между собой абсолютной величины коэффициента корреляции и корреляционного отношения. Эта оценка проводится с помощью критерия

$$T_r = \frac{|\theta_r - 1|}{\sigma_{\theta r}},$$

$$\text{где } \theta_r = \frac{n - f_y - 2}{f_y - 2} * \frac{\eta_{y/x}^2 - r_{xy}^2}{1 - \eta_{y/x}^2}, \sigma_{\theta r} = \sqrt{\frac{2(n-4)}{(f_y-2)(n-f_y-4)}}$$

где f_y – число интервалов признака y в корреляционной таблице.

Если $T_r > 3$, то отклонение $\eta_{y/x}$ от r_{xy} существенно. Если $T_r < 3$, то расхождение между $\eta_{y/x}$ и r_{xy} случайно и корреляционную связь можно признать точно линейной.

Затем необходимо построить корреляционное поле, на которое наносятся эмпирические и расчётные линии регрессии.

Для построения эмпирических линий регрессий y на x и x на y необходимо вычислить соответственно средние значения \bar{y}_x для средних значений x в интервалах по формуле

$$\bar{y}_{x \text{ эмп}} = \frac{1}{m_x} \sum_{j=1}^e m_{xy} * y_j,$$

и средние значения \bar{x}_y для средних значений x в интервалах по формуле

$$\bar{x}_{y \text{ эмп}} = \frac{1}{m_y} \sum_{i=1}^k m_{xy} * x_i.$$

Вычисленные значения $\bar{x}_{y \text{ эмп}}$ и $\bar{y}_{x \text{ эмп}}$ заносим в таблицы (см. ниже) и на корреляционное поле, на котором полученные точки соединим отрезками прямых.

Для вычислений эмпирических линий регрессий можно воспользоваться соответственно строкой 6 и столбцом 5 расчетной таблицы, а действительные значений \bar{y}_x и \bar{x}_y найти по формулам

$$\bar{y}_{x \text{ эмп}} = y_0 + d_y * \bar{y}_{xi} \text{ и } \bar{x}_{y \text{ эмп}} = x_0 + d_x * \bar{x}_{yj}.$$

После построения эмпирических кривых регрессий на корреляционном поле, ищем расчётные кривые регрессий в виде

$$y_x = b_0 + b_1 * x \text{ и } x_y = a_0 + a_1 * y.$$

Коэффициенты a и b определяются по методу наименьших квадратов из системы нормальных уравнений (например, для регрессии \bar{y}_x):

$$\begin{cases} nb_0 + b_1 \sum x_i = \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i * y_i \end{cases}$$

Однако в данном случае, при наличии ранее найденных значений, соответствующих величин, удобнее определить коэффициенты регрессии по следующим формулам:

$$b_1 = r_{xy} \frac{S_y}{S_x},$$

$$b_0 = \bar{y} - b_1 \bar{x},$$

$$a_1 = r_{xy} \frac{S_x}{S_y},$$

$$a_0 = \bar{x} - a_1 \bar{y},$$

которые получаются из системы уравнений путём несложных преобразований. Для этого оба уравнения системы делятся на n :

$$\begin{cases} b_0 + b_1 \bar{x} = \bar{y} \\ b_0 \bar{x} + b_1 \bar{x}^2 = \bar{x} * \bar{y} \end{cases}$$

Решая эту эквивалентную систему, находим

$$b_1 = \frac{\bar{x} * \bar{y} - \bar{x} * \bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{C_{xy}}{S_x^2} = \frac{C_{xy}}{S_x * S_y} * \frac{S_y}{S_x} = r_{xy} \frac{S_y}{S_x},$$

где $C_{xy} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y})$ – эмпирический корреляционный момент (ковариация).

Для оценки точности нахождения точечных оценок (проверки гипотезы достоверности) коэффициентов регрессии вычисляют критерии

$$t_{b_1} = \frac{b_1}{S_{b_1}} \text{ и } t_{b_0} = \frac{b_0}{S_{b_0}},$$

$$\text{где } S_{b_0} = S_{\text{ост}} \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{(n-1)S_x^2}}, S_{b_1} = \frac{S_{\text{ост}}}{\sqrt{(n-1)S_x^2}},$$

$$\text{где } S_{\text{ост}} = \sqrt{\frac{\sum_i^n (y_i - \bar{y}_x)^2}{n-2}} = \sqrt{\frac{n-1}{n-2} S_y^2 (1 - r_{xy}^2)}.$$

Полученные критерии сравниваются с табличными, которые находятся по таблицам t -распределения Стьюдента при выбранной доверительной вероятности ρ и числе степеней свободы $k = n - 2$. Если $|t_b|_{\text{расч}} \geq t_b \text{ табл}$, то можно считать коэффициент регрессии достоверным (значимым). Если для коэффициента регрессии оказывается $|t_b|_{\text{расч}} < t_b \text{ табл}$, то его следует считать незначимым, равным нулю.

После определения и проверки коэффициентов регрессии записываем уравнения регрессии через их найденные значения и вычисляем линии регрессии $\bar{x}_{y \text{ теор}}$ и $\bar{y}_{x \text{ теор}}$ для средних значений x_i и y_j в интервалах:

$$\bar{y}_{x \text{ теор}} = b_0 + b_1 * x_i$$

$$\bar{x}_{y \text{ теор}} = a_0 + a_1 * y_j$$

Полученные значения вносим в таблицы следующего вида:

x_i				
$\bar{y}_{x \text{ эксп}}$				
$\bar{y}_{x \text{ теор}}$				

y_j				
$\bar{x}_{y \text{ эксп}}$				
$\bar{x}_{y \text{ теор}}$				

По таблицам на корреляционном поле строим эмпирические и теоретические линии регрессии.

Доверительные интервалы для генеральных коэффициентов регрессии β находятся по формулам:

$$b_0 - t_{\alpha/2, n-2} * S_{b_0} < \beta_0 < b_0 + t_{\alpha/2, n-2} * S_{b_0},$$

$$b_1 - t_{\alpha/2, n-2} * S_{b_1} < \beta_1 < b_1 + t_{\alpha/2, n-2} * S_{b_1},$$

где $t_{\alpha/2, n-2}$ – определяется по таблицам t -распределения Стьюдента по числу степеней свободы $n - 2$ и доверительной вероятности

$$\rho(|t| < t_{\alpha/2, n-2}) = 1 - \alpha.$$

Доверительные интервалы для условных математических ожиданий $M(Y/X = x^*)$ определяются для оценки уклонения усреднённых значений зависимой переменной $\bar{y}_x = b_0 + b_1 * x^*$, рассчитанных по эмпирическому уравнению регрессии, от соответствующих условных математических ожиданий $M(Y/X = x^*)$, задаваемых генеральным уравнением регрессии $M(Y/X = x^*) = \beta_0 + \beta_1 * x^*$ при любом фиксированном значении независимой переменной x^* . Доверительный интервал, накрывающий с доверительной вероятностью $\rho = 1 - a$ условное математическое ожидание, находится по формуле

$$b_0 + b_1 * x^* - t_{\frac{a}{2}, n-2} * S_{\text{ост}} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)S_x^2}}$$

$$< M(Y/X = x^*) <$$

$$b_0 + b_1 * x^* + t_{\frac{a}{2}, n-2} * S_{\text{ост}} \sqrt{\frac{1}{n} + \frac{(x^* - \bar{x})^2}{(n-1)S_x^2}}$$

Наконец, последняя проверка при регрессионном анализе состоит в анализе соответствия математической модели (в нашем случае линейного уравнения регрессии) экспериментальным данным, если до проведения регрессионного анализа не был выполнен корреляционный анализ.

Проверка соответствия заключается в проверке нулевой гипотезы $H_0: \beta_1 = 0$ против альтернативной гипотезы $H_1: \beta_1 \neq 0$. Если принимается нулевая гипотеза, то считается, что уравнение регрессии Y на X либо имеет нелинейный вид, либо эти переменные являются независимыми случайными величинами.

Одним из методов проверки нулевой гипотезы является дисперсионный анализ. Для удобства все вычисления располагаются в следующей таблице:

Источник изменчивости	Суммы квадратов	Число степеней свободы	Средние квадраты	Отношения средних квадратов
Линейная регрессия	$B = b_1^2 \sum (x_i - \bar{x})^2 = b_1^2 S_x^2 (n - 1)$			
Остаток	$C = A - B =$			
Полная сумма	$A = \sum (y_i - \bar{y})^2 = (n - 1) S_y^2$			

Табличное значение критерия $F_{\text{табл}}$ определяется при выбранном уровне значимости α и степенях свободы 1, $(n - 2)$. Если $F_{\text{набл}} \geq F_{\text{табл}}$, то гипотеза $H_0: \beta_1 = 0$ отвергается, т.е. существует линейная связь между Y и X .

4. Порядок выполнения работы

1. Получить вариант задания у преподавателя.
2. Определить коэффициента регрессии моделей \bar{y}_x и \bar{x}_y .
3. Провести оценки точности их определения (для b_0 и b_1).
4. Построить эмпирические и расчётные линии регрессии.
5. Найти доверительные интервалы для условного математического ожидания при выбранном фиксированном значении x^* .
6. Проверить соответствие модели экспериментальным данным с помощью дисперсионного анализа.
8. Сформулировать выводы.
9. Сдать отчёт преподавателю.

5. Вопросы для самопроверки

1. Как определяются эмпирические линии регрессии?
2. Как определяются расчётные линии регрессии?
3. Какие регрессии имеют место при двух взаимосвязанных переменных?

4. Как оценивается точность определения коэффициентов регрессии?
5. Как определяется соответствие модели экспериментальным данным?
6. Какие суммы квадратов используются в регрессионном анализе?

ЛАБОРАТОРНАЯ РАБОТА №5

Полный факторный эксперимент

1. Цель работы

Цель работы состоит в практическом освоении методики планирования и обработки результатов активного эксперимента на примере полного факторного эксперимента.

2. Общие сведения

Активный эксперимент выполняется с целью получения математического описания (модели) исследуемого процесса или объекта. Он проводится, как правило, в лабораторных условиях, на этапе проектирования объекта или технологического процесса, поскольку его применение в производственных условиях может привести к браку выпускаемой продукции.

Широта области применения активного эксперимента частично видна из вариантов заданий для настоящей лабораторной работы, с которыми следует предварительно ознакомиться: это отработка схем, в том числе по надёжности, и технологических процессов; описание поверхности отклика в исследуемой точке и определение составляющих градиента при оптимизации и т.п.

Методы активного эксперимента рассматриваются в науке «Планирование эксперимента»: полный факторный эксперимент (ПФЭ), дробный факторный эксперимент (ДФЭ) и др.

Модель объекта находится в форме линейного, неполного квадратичного или уравнения регрессии (полинома) второй и более высоких степеней. Далее, соответственно, приводятся формы линейного и неполного квадратичного уравнений регрессии:

$$\hat{y} = \hat{b}_0 + \sum_{i=1}^k \hat{b}_i * x_i,$$

$$\hat{y} = \hat{b}_0 + \sum_{i=1}^k \hat{b}_i * x_i + \sum_{i < e} \hat{b}_{ie} * x_i * x_e.$$

Выполнение активного эксперимента разбивается на три этапа:

- 1) составление плана эксперимента;
- 2) выполнение эксперимента;
- 3) обработка результатов эксперимента.

В основе планирования эксперимента лежат концепции рандомизации и оптимального использования факторного пространства.

3. Теоретические сведения

3.1. Планирование и выполнение эксперимента

В зависимости от характера исследуемого процесса и степени описывающего его полинома различают планы первого, второго и более высоких порядков.

В основе ПФЭ лежит перебор в последовательно проводимых рандомизированных экспериментах всех возможных сочетаний значений (уровней) факторов (параметров).

В случае, если процесс или объект описывается линейной или неполной квадратичной моделью, достаточно в эксперименте менять факторы в двух уровнях: верхнем и нижнем, которые симметричны относительно базового (основного, нулевого) уровня фактора. Число опытов в ПФЭ составляет $N = 2^k$, где k – число факторов.

Прежде, чем составлять план, необходимо выбрать основной уровень и интервалы изменения факторов. План эксперимента составляется в виде таблицы, называемой матрицей планирования. Каждая строка матрицы соответствует различным опытам, а каждый столбец – значениям фактора в том или ином опыте. Эти значения представляются в кодированном виде: +1 соответствует верхнему уровню фактора, -1 – нижнему уровню фактора. В матрицу вводится фиктивная переменная x_0 , которая служит для определения свободного члена уравнения регрессии, и во всех опытах она равна +1.

Фиктивная переменная позволяет записывать и использовать, например, линейное уравнение регрессии, в удобной компактной форме:

$$\hat{y} = \hat{b}_0 * x_0 + \sum_{i=1}^k \hat{b}_i * x_i = \sum_{i=0}^k \hat{b}_i * x_i.$$

Для того, чтобы исключить влияние неконтролируемых факторов на результаты эксперимента, проводятся серии параллельных опытов, которые рандомизируются во времени, т.е. устанавливается случайная последовательность проведения опытов в каждой серии с помощью таблицы случайных чисел. Эти серии опытов заносятся в матрицу планирования. После составления плана (матрицы) проводится эксперимент, результаты которого (значения выходного параметра) заносятся в матрицу. Естественно, что в эксперименте значения независимых факторов (параметров) устанавливаются в действительных величинах (натуральном масштабе) в соответствии с их кодированными значениями в матрице.

После проведения всех опытов приступают к обработке результатов с помощью таблицы.

3.2. Обработка результатов эксперимента

Обработка результатов эксперимента содержит:

- 1) проверку воспроизводимости опытов;
- 2) расчёт коэффициентов регрессии;
- 3) проверку значимости коэффициентов регрессии;
- 4) построение математической модели;
- 5) проверку гипотезы об адекватности математической модели результатам эксперимента;
- 6) построение математической модели для действительных значений (натурального масштаба) факторов (параметров).

3.2.1 Проверка воспроизводимости опытов

По измеренным значениям выходного параметра y_j в серии параллельных опытов для каждой j -й строки матрицы планирования (каждого сочетания факторов) вычисляются среднее значение выходного параметра в j -м опыте и выборочная дисперсия выходного параметра в j -м опыте:

$$\bar{y}_j = \frac{1}{m} \sum_{g=1}^m y_{jg},$$

$$S_j^2 = \frac{1}{m-1} \sum_{g=1}^m (y_{jg} - \bar{y}_j)^2,$$

где g – номер серии параллельных опытов, m – количество серий параллельных опытов.

Из полученных выборочных дисперсий находится максимальная выборочная дисперсия по всем опытам $S_{j \max}^2$.

Для проверки воспроизводимости опытов используется критерий Кохрена, который служит для оценки случайности отличия дисперсий, если их более двух: отношение максимальной выборочной дисперсии к сумме всех выборочных дисперсий:

$$G_{\text{расч}} = \frac{S_{j \max}^2}{\sum_{j=1}^N S_j^2},$$

где j – номер опыта (строка таблицы), N – количество опытов (число строк).

Расчётное значение критерия Кохрена сравнивается с его табличным значением $G_{\text{крит}}$, которое находится при числе степеней свободы $f_1 = m - 1$ и $f_2 = N$ и заданном уровне значимости α , % .

Если $G_{\text{расч}} < G_{\text{крит}}$, то гипотеза об однородности дисперсий принимается. Это значит, что условия проведения опытов одинаковы и отклонения выходного параметра случайны.

В случае $G_{\text{расч}} > G_{\text{крит}}$ отклонение максимальной дисперсии от остальных не случайно и, следовательно, опыты не воспроизводимы, т.е. имеется доминирующий фактор.

3.2.2 Расчёт коэффициентов регрессии

В предположении, что план эксперимента ортогональный, коэффициенты регрессии определяются по формуле

$$\hat{b}_i = \frac{1}{N} \sum_{i=1}^N x_{ij} * \bar{y}_j,$$

где i – номер фактора, x_{ij} – значение i -го фактора в j -ом опыте.

Поскольку факторы в опытах имеют кодированные значения +1 или -1, то для вычисления коэффициента регрессии b_i нужно при суммировании просто значению \bar{y}_j в j -м опыте приписывать знак i -го фактора в этом опыте. Аналогично определяются коэффициенты b_{ie} для эффектов взаимодействия.

3.2.3 Проверка значимости коэффициентов регрессии

Вначале находятся оценка дисперсии воспроизводимости опытов по выборочным дисперсиям выходного параметра и оценка дисперсии ошибки определения коэффициентов регрессии по следующим формулам:

$$S_y^2 = \frac{1}{N} \sum_{j=1}^N S_j^2,$$

$$S^2\{b\} = \frac{1}{N^2 * m} \sum_{j=1}^N S_j^2 = \frac{S_y^2}{N * m}.$$

Эта оценка (или среднеквадратическое отклонение $S\{b\}$) одинакова для всех коэффициентов регрессии.

Для проверки значимости коэффициентов регрессии используется t -критерий Стьюдента:

$$t_{i \text{ расч}} = \frac{|\hat{b}_i|}{\sqrt{S^2\{b\}}} = \frac{|\hat{b}_i|}{S\{b\}}.$$

Расчётное значение t -критерия сравнивается с табличным $t_{\text{крит}}$, определяемым при числе степеней свободы $f = N(m - 1)$ и уровне значимости α , %. $t_{\text{крит}}$ одинаково для всех коэффициентов в данном эксперименте.

Если найденный для коэффициента регрессии критерий $t_{i \text{ расч}} > t_{\text{крит}}$, то данный коэффициент регрессии является статистически значимым. Незначимы будут коэффициенты, для которых $t_{i \text{ расч}} < t_{\text{крит}}$.

3.2.4 Построение математической модели

В соответствии с результатами проверки значимости коэффициентов регрессии строится линейное или неполное квадратичное (квазилинейное) уравнение регрессии, из которого исключаются члены, содержащие незначимые коэффициенты, если таковые имеются:

$$\hat{y} = \hat{b}_0 + \sum_{i=1}^k \hat{b}_i * x_i + \sum_{i < e} \hat{b}_{ie} * x_i * x_e.$$

Если все коэффициенты регрессии оказались значимыми, то проверке на адекватность подлежат линейная и неполная квадратичная модели.

3.2.5 Проверка гипотезы об адекватности математической модели результатам эксперимента

Вначале находится оценка по принятой модели выходного параметра y_j в каждом опыте (строке). Для этого в уравнение регрессии подставляются кодированные значения (+1, -1) факторов, которые они имеют в соответствующем опыте. Если проверяются две модели (линейная и квазилинейная), то результаты вычислений во всех строках таблиц при проверке адекватности записываются в виде дроби: в числителе для первой модели, в знаменателе для второй модели.

Вычисляется оценка дисперсии неадекватности последующей формуле:

$$S_{\text{ад}}^2 = \frac{1}{N-d} \sum_{j=1}^N (\bar{y}_j - \hat{y}_j)^2,$$

где d – число линейных членов аппроксимирующего полинома (уравнения регрессии).

Проверка адекватности проводится по F -критерию Фишера, который служит для оценки случайности отклонения дисперсий, если их не более двух:

$$F_{\text{расч}} = \frac{S_{\text{ад}}^2}{S_y^2}.$$

При вычислении F -критерия числитель всегда должен быть больше знаменателя, так как в таблицах F -критерия его значения больше единицы. Если оказалось, что $S_{\text{ад}}^2 < S_y^2$, то они меняются местами.

Расчетное значение F -критерия сравнивается с табличным $F_{\text{крит}}$, которое находится при заданных уровнях значимости α , % и степенях свободы $f_1 = N - d$ и $f_2 = N(m - 1)$.

Если $F_{\text{расч}} < F_{\text{крит}}$, то принимается гипотеза об адекватности построенной математической модели результатам эксперимента. Если $F_{\text{расч}} > F_{\text{крит}}$, то гипотеза отвергается: линейная (или квазилинейная) модель не адекватна результатам эксперимента. В этом случае переходят к построению плана для квадратичной модели и т.д.

3.2.6 Построение математической модели для действительных значений (натурального масштаба) факторов (параметров)

Для такого построения в адекватную математическую модель подставляются кодированные значения факторов в следующем виде:

$$x_i = \frac{\tilde{x}_i - \tilde{x}_{i0}}{\Delta\tilde{x}_i},$$

где \tilde{x}_i – натуральное (действительное) значение фактора, \tilde{x}_{i0} – базовый (основной) уровень фактора, $\Delta\tilde{x}_i$ – шаг варьирования фактора.

Значения \tilde{x}_{i0} и $\Delta\tilde{x}_i$ известны, так как они выбираются перед составлением плана эксперимента.

После преобразований уравнение регрессии приводится к следующему, например, виду:

$$y = \tilde{b}_0 + \tilde{b}_1 * \tilde{x}_1 + \tilde{b}_2 * \tilde{x}_2 + \tilde{b}_{12} * \tilde{x}_1 * \tilde{x}_2,$$

где $\tilde{b}_i, \tilde{b}_{ie}$ – новые значения коэффициентов регрессии.

4. Порядок выполнения работы

1. Ознакомиться со всеми вариантами заданий.
2. Получить у преподавателя вариант задания.
3. Составить план эксперимента в соответствии с заданием.
4. Провести эксперимент.
5. Обработать результаты эксперимента. При обработке результатов эксперимента уровни значимости $\alpha, \%$ выбирать в зависимости от имеющихся таблиц для G -, t - и F -критериев.
6. Сформулировать выводы.
7. Сдать отчёт преподавателю.

5. Вопросы для самопроверки

1. С какой целью проводится активный эксперимент.
2. В каких условиях проводится активный эксперимент.
3. Какие существуют методы планирования эксперимента.
4. Что такое план эксперимента.
5. Когда используются планы первого порядка.
6. Какие этапы имеют место в активном эксперименте.
7. Какие концепции лежат в основе планирования эксперимента.
8. Чему соответствуют строки и столбцы матрицы эксперимента.

9. В каком виде представляются значения факторов в матрице планирования.

10. В скольких уровнях изменяются факторы для плана первого порядка.

11. Что такое кодированное значение фактора.

12. Для чего проводятся серии параллельных опытов.

13. Какие проверки осуществляются при обработке результатов эксперимента.

14. Какой критерий используется для проверки воспроизводимости опытов и что он собой представляет.

15. Как рассчитываются коэффициенты регрессии при выполнении условия ортогональности плана.

16. Что такое дисперсия воспроизводимости опытов.

17. Что такое дисперсия ошибки определения коэффициентов регрессии.

18. Какой критерий используется для проверки значимости коэффициентов регрессии и что он собой представляет.

19. Что такое дисперсия неадекватности и как она определяется.

20. По какому критерию проводится проверка адекватности и что он собой представляет.

21. По каким исходным данным находятся табличные значения критериев, используемых для проверок.

22. Как перейти в уравнении регрессии от кодированных значений факторов к натуральным.

Литература

1. Решетников М.Т. Планирование эксперимента и статистическая обработка данных. Учебное пособие. – Томск: Томский гос.ун-т систем управления и радиоэлектроники, 2000 г. 232 с.
2. Семенов, С. А. Планирование и обработка результатов эксперимента : учебное пособие / С. А. Семенов. — 2-е изд., пер. и доп. — Москва : РТУ МИРЭА, 2021. — 48 с.
3. Серафинович Л.П. Планирование эксперимента: Учебное пособие. – Томск: Изд-во ВСпектр, 2006. – 128 с.
4. Степанишин, В. В. Научное исследование. Подготовка научно-исследовательской работы : учебно-методическое пособие / В. В. Степанишин, Г. В. Кондратов, А. М. Жариков. — Москва : МГАВМиБ им. К.И. Скрябина, 2021. — 47 с.