

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ.....	2
1. ПРАКТИЧЕСКАЯ РАБОТА № 1. СТАТИСТИЧЕСКИЕ КРИТЕРИИ И ИХ ПРИМЕНЕНИЕ.....	2
1.1. Краткие теоретические сведения.....	2
1.1.1. Распределение Стьюдента (t-критерий).....	3
1.1.2. Распределение Пирсона.....	5
1.1.3. Распределение Фишера (F-критерий)	6
1.1.4. Распределение Кохрена (G-критерий)	7
1.1.5. τ -распределение (τ -критерий).....	8
1.2. Задание на практическую работу	9
1.3. Контрольные вопросы	10
2. ПРАКТИЧЕСКАЯ РАБОТА № 2. ДИСПЕРСИОННЫЙ АНАЛИЗ	11
2.1. Краткие теоретические сведения.....	11
2.2. Задание на практическую работу	14
2.3. Контрольные вопросы	14
3. ПРАКТИЧЕСКАЯ РАБОТА № 3. ОДНОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ	15
3.1. Краткие теоретические сведения.....	15
3.2. Задание на практическую работу	19
3.2. Контрольные вопросы	20
4. ПРАКТИЧЕСКАЯ РАБОТА № 4. МНОГОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ	23
4.1. Краткие теоретические сведения.....	23
4.2. Задание на практическую работу	26
4.3. Контрольные вопросы	27
5. ЛИТЕРАТУРА	27

Введение

В практике специалиста — технолога машиностроительного производства, эксперимент играет исключительно важную роль.

Разработчик новой технологии обращается к эксперименту, чтобы проверить правильность положенных в основу решений.

Исследователь, изучающий свойства какого-либо технологического объекта, прибегает к эксперименту, чтобы обосновать допущения и доказать правомерность принятой модели.

Квалифицированный исследователь при работе с моделью, реализованной на ЭВМ, обязательно применит современные приемы, разработанные в теории эксперимента, чтобы сэкономить время исследования и получить результаты в наиболее полном и удобном для использования виде.

Заводской инженер, ответственный за качество продукции, организует и проводит сложный эксперимент — контрольные типовые, ресурсные и другие испытания.

Развитие теории эксперимента вместе со стремительно прогрессирующими техническими средствами его проведения открывает широкие возможности для повышения эффективности всех видов экспериментальных исследований.

Цель работы — обучение практическим навыкам обработки экспериментальных данных.

1. ПРАКТИЧЕСКАЯ РАБОТА № 1. СТАТИСТИЧЕСКИЕ КРИТЕРИИ И ИХ ПРИМЕНЕНИЕ

1.1. Краткие теоретические сведения

Решение статистических задач проводится на основе ряда специальных распределений с известными (табулированными) квантилями. Общий подход состоит в том, что вводится случайная величина, связанная с изу-

чаемой и удобная для решения той или иной конкретной задачи.

1.1.1. Распределение Стьюдента (t-критерий).

Распределение было получено Госсетом (псевдоним Стьюдента) в 1908 г. Зависит от объема выборки N или числа степеней свободы $f=N-1$, с которым определена выборочная дисперсия S^2 (среднее квадратическое отклонение (CKO) выборки S), и от заданной вероятности ответа, определяемой параметром уровня значимости q . Формула критерия:

$$t = \frac{\bar{x} - a}{s} \sqrt{N}, \quad (1)$$

где a — генеральное среднее исследуемой случайной величины X ;

\bar{x} — выборочное математическое ожидание X .

Распределение симметрично относительно начала координат, т. е.

$$t_{q/2} = -t_{1-q/2}.$$

Более полные таблицы квантилей распределения Стьюдента для уровня значимости $q \neq 0,05$ содержатся в [1].

Таблица 1

Квантили распределения Стьюдента при $q=0,05$

$f=N-1$	1	2	3	4	5	6	7	8	9	10	20	30	∞
$t_{1-q/2}$	12,71	4,30	3,18	2,78	2,57	2,45	2,37	2,31	2,26	2,23	2,09	2,04	1,96

Распределение Стьюдента дает возможность находить генеральное среднее или проверять статистические гипотезы при очень малых выборках.

Пример 1. Известны три значения нормально распределенной случайной величины X : $x_1 = 10$; $x_2 = 9,5$; $x_3 = 10,2$. Требуется оценить генеральное среднее с вероятностью $p = 0,95$ (задача первого типа).

Определим выборочное ($N = 3$) математическое ожидание \bar{x}

$$\bar{x} = \frac{\sum_{i=1}^N x_i}{N} = \frac{10 + 9,5 + 10,2}{3} = 9,9.$$

Определим выборочное СКО при $f = N-1=2$:

$$s = \sqrt{\frac{\sum_{i=1}^N (x_i - \bar{x})^2}{N-1}} = \sqrt{\frac{(10-9,9)^2 + (9,5-9,9)^2 + (10,2-9,9)^2}{2}} = 0,36.$$

Запишем выражение для t в соответствии с (1):

$$t = \frac{\bar{x} - a}{s} \sqrt{N} = \frac{9,9 - a}{0,36} \sqrt{3}.$$

По табл. 1 определим квантильные границы $t_{1-q/2} = 4,30$, $t_{q/2} = -4,30$ и запишем неравенство

$$-4,30 < \frac{9,9 - a}{0,36} \sqrt{3} < 4,30.$$

Решив неравенство относительно a , получим с доверительной вероятностью $p = 0,95$

$$9,0 < a < 10,8.$$

Пример 2. Проверить гипотезу, состоящую в том, что нормально распределенная случайная величина X имеет генеральное математическое ожидание $a = 10$ на основании результатов двух испытаний: $x_1=8,1$, $x_2=8,9$ (задача второго типа).

По результатам испытаний определяем $\bar{x} = 8,5$ и $s = 0,57$.

Вычисляем по (1) значение критерия Стьюдента t :

$$t = \frac{8,5 - 10}{0,57} \sqrt{2} = -3,71.$$

Выбираем уровень значимости $q = 0,05$ и по табл. 1 находим для $f=N-1=1$ границы критических областей гипотезы $t_{1-q/2} = 12,71$, $t_{q/2} = -12,71$.

Гипотеза не отвергается на уровне значимости $q=0,05$, поскольку $t = -3,71$ лежит вне критических областей. Отметим, что очень малая информация ($N = 2$) и низкий уровень значимости не дают оснований отвергнуть плохую на взгляд гипотезу. Если бы, например, те же результаты

($\bar{x} = 8,5$ и $s = 0,57$) были получены при $N = 4$ ($f = 3$), то, как легко видеть, гипотезу следовало отвергнуть на том же уровне значимости. При выборе более жесткого уровня значимости, $q = 0,2$ и $T_{1-q/2} = 3,08$, т. е. гипотеза также должна быть отвергнута.

1.1.2. Распределение Пирсона

Распределение Пирсона (χ^2 -критерий) удобно для оценки генеральной дисперсии σ^2 по выборочной s^2 . В 1900 г. Пирсон ввел случайную величину

$$\lambda^2 = \frac{(N-1) \cdot S^2}{\sigma^2} \quad (2)$$

и нашел ее распределение, зависящее лишь от $f = N - 1$. Оно несимметрично, следовательно, $\lambda_{q/2}^2 = \lambda_{1-q/2}^2$. Некоторые квантили приведены в табл.2. Более полные таблицы содержатся в справочниках [1].

Таблица 2

Квантили распределения Пирсона λ_{1-q}^2

f	1	2	3	4	5	6	7	8	9	10	15	20
$\lambda_{1-0,95}^2$	0,0039	0,103	0,352	0,71	1,14	1,63	2,17	2,73	3,32	3,94	7,3	10,9
$\lambda_{1-0,05}^2$	3,8	6,0	7,8	9,5	11,1	12,6	14,1	15,5	16,9	18,3	25,0	31,4

Критерий Пирсона позволяет решать статистические задачи о дисперсиях.

Пример 3. Найти с вероятностью $p = 0,9$ доверительный интервал для σ^2 нормально распределенной случайной величины, если при $f = 5$, $S^2 = 1$. Определив по табл.2 квантильные границы $\lambda_{0,05}^2 = 1,14$ и $\lambda_{0,95}^2 = 11,1$, имеем $1,14 < fs^2 / \sigma^2 < 11,1$ или после преобразований $0,45 < \sigma^2 < 4,39$

Пример 4. Проверить гипотезу, состоящую в том, что генеральная дисперсия нормально распределенной случайной величины $\sigma^2 = 2$, если обработка данных десяти опытов ($f = 9$) дала $S^2 = 4$. В связи с тем, что $s^2 \geq \sigma^2$,

имеет смысл использовать одностороннюю оценку сверху. Выбрав $q = 0,05$, определим по табл. 2 $\lambda_{0,95}^2 = 16,9$. Обнаруживаем, что $\lambda^2 = \frac{9 \cdot 4}{2} = 18 > 16,9$, т. е. гипотеза должна быть отвергнута на уровне значимости 0,05, поскольку рассчитанная величина λ^2 оказалась в критической области.

1.1.3. Распределение Фишера (F-критерий)

Используется для проверки однородности двух выборочных дисперсий s_1^2 и s_2^2 (обычно принимают $s_1^2 > s_2^2$ и используют односторонние оценки), найденных соответственно с f_1 и f_2 , с целью установить или отвергнуть их принадлежность одной генеральной совокупности. Фишером введена случайная величина

$$F = \frac{s_1^2}{s_2^2} \quad (3)$$

и построено ее распределение, зависящее от f_1 и f_2 . Некоторые квантили распределения Фишера для $q=0,05$ приведены в табл. 3.

Таблица 3

Квантили распределения Фишера $F_{1-0,05}$

f_2	f_1								
	1	2	3	4	5	6	12	24	∞
1	164,4	199,5	215,7	224,6	230,2	234,0	244,9	249,0	254,3
2	18,5	19,2	19,2	19,3	19,3	19,3	19,4	19,5	19,5
3	10,1	9,6	9,3	9,1	9,0	8,9	8,7	8,6	8,5
4	7,7	6,9	6,6	6,4	6,3	6,2	5,9	5,8	5,6
5	6,6	5,8	5,4	5,2	5,1	5,0	4,7	4,5	4,4
6	6,0	5,1	4,8	4,5	4,4	4,3	4,0	3,8	3,7
7	5,6	4,7	4,4	4,1	4,0	3,9	3,6	3,4	3,2
8	5,3	4,5	4,1	3,8	3,7	3,6	3,3	3,1	2,9
9	5,1	4,3	3,9	3,6	3,5	3,4	3,1	2,9	2,7
10	5,0	4,1	3,7	3,5	3,3	3,2	2,9	2,7	2,5
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

Пример 5. Проверить гипотезу об однородности двух выборочных дисперсий нормально распределенной случайной величины. Результаты опытов: $s_1^2=0,9$ при $f_3=3$ и $s_2^2=0,2$ при $f_2=5$.

Выбрав $q=0,05$, находим по табл. 3 $F_{1-q}=5,4$. Из данных опытов имеем $F=0,9/0,2=4,5$, следовательно, $F < F_{1-q}$ и гипотеза об однородности s_1^2 и s_2^2 , т. е. о принадлежности их одной генеральной совокупности, характеризуемой σ^2 , не отвергается на уровне значимости 0,05

1.1.4. Распределение Кохрена (G-критерий)

Используется для проверки однородности k выборочных дисперсий, найденных с одинаковыми числами степеней свободы $f_i=N-1$. Кохреном введена случайная величина

$$G = \frac{s_{j,\max}^2}{\sum_{j=1}^k s_j^2}, \quad (4)$$

где $s_{j,\max}^2$ — наибольшая из сравниваемых дисперсий; и построено ее распределение, зависящее от f_j и k . В табл. 4 приведены некоторые квантили G_{1-q} для $q=0,05$.

Таблица 4

Квантили распределения Кохрена $C_{1-0,05} \times 10^2$

k	f_j										
	1	2	3	4	5	6	7	8	9	10	∞
2	100	98	94	91	88	85	83	82	80	79	50
3	97	87	80	75	71	68	65	63	62	60	33
4	91	77	68	63	59	56	54	52	50	49	25
5	84	68	60	54	51	48	46	44	42	41	20
6	78	62	53	48	44	42	40	38	37	36	17
7	73	56	48	43	40	37	35	34	33	32	14
8	68	52	44	39	36	34	32	30	29	28	13
9	64	48	40	36	33	31	29	28	27	26	11
10	60	45	37	33	30	28	27	25	24	24	10
20	39	24	22	19	17	16	15	14	14	13	5
∞	0	0	0	0	0	0	0	0	0	0	0

Пример 6 Проверить гипотезу об однородности выборочных дисперсий $s_1^2=3$, $s_2^2=5$, $s_3^2=15$, $s_4^2=2$ и $s_5^2=4$, каждая из которых определена с $f_j=4$ на уровне значимости $q=0,05$.

По табл. 4 находим для односторонней оценки при $f_j=4$ и $k=5$ $G_{0,95} = 0,54$. Из опытных данных в соответствии с (4) имеем

$$G = \frac{15}{3+5+15+2+4} = 0,52 < 0,54$$

т. е. на уровне значимости 0,05 гипотеза об однородности дисперсий не отвергается.

Отметим здесь, что наилучшей оценкой k -однородных дисперсий служит дисперсия, определяемая как

$$s^2 = \left(\sum_{j=1}^k s_j^2 \right) / k \quad (5)$$

с числом степеней свободы $f = kf_j = k(N-1)$. Эта оценка может использоваться для определения доверительного интервала для генеральной дисперсии σ^2 .

1.1.5. τ -распределение (τ -критерий)

Используется для проверки однородности наблюдений, исключения грубых ошибок или выбросов. Квантили распределения случайной величины

$$\tau = \frac{|x_{ip} - \bar{x}|}{s}, \quad (6)$$

зависящего лишь от объема выборки N , по которой определяются \bar{x} и s , приведены для вероятности $p=0,95$ в табл. 5.

Таблица 5

Квантили τ -распределения при $q=0,05$

N	3	4	5	6	7	8	9	10	15	20
τ_{1-q}	1,41	1,69	1,87	2,00	2,17	2,24	2,29	2,29	2,49	2,62

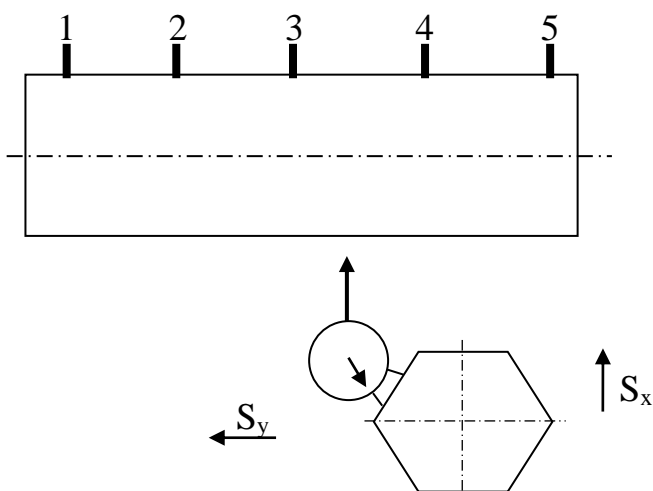
Пример 7. На одном из пяти одинаковых агрегатов (третьем), выполняющих однотипные технологические операции, были внедрены мероприятия по экономии электроэнергии. Оценить их эффективность, если зарегистрированное месячное потребление энергии каждым агрегатом составляет $x_1 = 10$, $x_2 = 12$, $x_3 = 8$, $x_4 = 9$, $x_5 = 11$.

Как и предполагалось, на третьем агрегате расход энергии минимален, т. е. $x_3 = 8$. Выбран $q=0,05$, найдем по табл. 5 что, $\tau_{0,95} = 1,87$. По опыт-

ным данным определим $\bar{x} = 10$, $s = 1,58$ и вычислим $\tau = \frac{|8-10|}{1,58} = 1,27$.

Поскольку $\tau = 1,27 < 1,87$, т. е. не попадает в критическую область гипотезы, нет оснований считать, что достигнутый результат неслучаен, т.е. обусловлен проведенными мероприятиями. Он вполне может быть объяснен естественным разбросом в потреблении электроэнергии отдельными агрегатами, и оснований для выплаты премии в этом случае статистика не дает.

1.2. Задание на практическую работу



В протоколе испытаний точности позиционирования привода подачи токарного станка с ЧПУ зафиксированы данные измерений фактической погрешности положения револьверной головки (см. схему).

Рис. 1. Схема измерений

Данные измерений

№ сечения	Отклонения, мкм					
	1	2	3	4	5	6
1	5	8	-3	2	-6	16
2	10	14	16	8	7	2
3	-4	-8	9	8	6	-7
4	11	11	8	14	-3	6
5	-14	-3	11	12	8	-4

Составить отчет об испытаниях, ответив с вероятностью $p=0,95$ на следующие вопросы:

1. Нет ли грубых ошибок в результатах измерений?
2. Зависит ли точность привода от номера сечения?
3. Оцените общую точность позиционирования привода.
4. Можно ли считать незначимым отличие погрешности от нуля?
5. Оцените наиболее вероятное отклонение позиционирования.

1.3. Контрольные вопросы

1. Что означает понятие «однородность дисперсий»?
2. В чем различие генеральных и выборочных параметров случайной величины?
3. Определите понятие «квантиль».
4. Назначение критерия Стьюдента.
5. Назначение критерия Пирсона.
6. Назначение критерия Фишера.
7. Назначение критерия Пирсона.
8. Каким образом применяется t -критерий для выявления грубых результатов.

2. ПРАКТИЧЕСКАЯ РАБОТА №2. ДИСПЕРСИОННЫЙ АНАЛИЗ

2.1. Краткие теоретические сведения

Поставим задачу определить факт влияния факторов X на функцию цели Y в условиях действия случайных факторов W .

Задача решается с помощью дисперсионного анализа (ДА), который позволяет выявить, приведет ли изменение X к значимому изменению Y , фактически является ли X варьируемым фактором.

Результат опыта Y можно представить как сумму 2-х слагаемых:

$$Y = Y' + Y'',$$

где Y' — часть результата, обусловленная влиянием X ;

Y'' — часть результата, обусловленная влиянием W .

По свойствам дисперсии:

$$DY = DY' + DY'',$$

где DY — общая дисперсия результатов наблюдений;

DY' — дисперсия Y , обусловленная влиянием X ;

DY'' — дисперсия Y , обусловленная влиянием W .

Если оценить однородность общей дисперсии результатов DY и дисперсии от влияния случайных факторов DY'' , то в случае принятия гипотезы компонента DY' (дисперсия от влияния X) оказывается незначительной, или X на Y не влияет.

Таблица 6

План проведения дисперсионного анализа

№ опыта	Уровни X_j				
	X_1	X_2	X_3	...	X_M
1	Y_{11}	Y_{12}	Y_{13}	...	X_{1M}
2	Y_{21}	Y_{22}	Y_{23}	...	X_{2M}
3
$i = 1 \dots K$	$j = 1 \dots M$...
K	Y_{K1}	Y_{K2}	Y_{K3}	...	Y_{KM}

Итак, дисперсионный анализ сводится к определению дисперсии ре-

зультатов опытов DY'' , обусловленных влиянием случайных факторов W . Данная задача может быть решена только путем запланированного повторения K опытов на M уровнях. В табл. 6 приведен стандартный план построения исследований.

Для исключения систематических погрешностей данных измерений последовательность проведения $N = K \cdot M$ опытов определяется с помощью процедуры **рандомизации** (табл. 7).

Таблица 7

Рандомизация последовательности проведения опытов

№ п.п.	Генератор случайных чисел от 1 до $M \cdot K$							
Уровень	1	1	1	1	2	2	...	K
Опыт №	1	2	3	M	2	3	...	M
Порядок проведения	5	2	1	3	8	4	...	24

Обработка данных строится по следующему алгоритму.

1. Определяем выборочное математическое ожидание по каждому уровню фактора X

$$\bar{Y}_j = \frac{\sum_{i=1}^K Y_{i,j}}{K}.$$

2. Определяем выборочные дисперсии для каждого уровня фактора X

$$Sy_j^2 = \frac{\sum_i (Y_{i,j} - \bar{Y}_j)^2}{K - 1}.$$

3. Проверим однородность полученных выборочных дисперсий по критерию Кохрена

$$G = \frac{\max(Sy_j^2)}{\sum_{j=1}^M Sy_j^2}.$$

Если $G < G_{\text{табл}}$ (табл. 4), гипотеза однородности принимается.

В случае, когда гипотеза однородности отвергается, необходимо прекратить расчет и рекомендовать:

- 1) Увеличить количество опытов;
- 2) Проверить надежность измерительных средств и принять меры к повышению стабильности поведения объекта.
4. Определим дисперсию от влияния случайных факторов

$$DY'' = \frac{\sum_{j=1}^M S_{y_j}^2}{M}.$$

5. Определим математическое ожидание всех наблюдений

$$\bar{Y} = \frac{\sum_i^K \sum_j^M Y_{i,j}}{M \cdot K}.$$

6. Определим общую дисперсию всех наблюдений

$$DY = \frac{\sum_i^K \sum_j^M (Y_{i,j} - \bar{Y})^2}{M \cdot K - 1}.$$

7. Проверим дисперсии DY'' и DY на однородность по критерию Фишера.

$$F_p = \frac{DY}{DY''}.$$

Табличное значение критерия Фишера определяем по табл. 3 для

$$f_1 = M \cdot (K-1), \quad f_2 = M \cdot K - 1$$

Если $F_p > F_{\text{табл}}$, то гипотеза однородности дисперсий DY и DY'' отвергается и, следовательно, с вероятностью p (см. табл. Фишера) управляемый фактор X оказывает значимое влияние на функцию цели Y . В противном случае необходимо сделать противоположный вывод.

2.2. Задание на практическую работу

Определить факт влияния скорости подачи S (мм/мин) на точность позиционирования привода Δ по имеющимся экспериментальным данным.

S , мм/мин	Δ , мкм					
	Номера опытов					
	1	2	3	4	5	6
1	5	8	−3	2	−6	16
10	10	14	16	8	7	2
100	−4	−8	9	8	6	−7
500	11	11	8	14	−3	6
1000	−14	−3	11	12	8	−4

2.3. Контрольные вопросы

1. Что такое рандомизация и зачем она проводится?
2. Назначение и сущность дисперсионного анализа.
3. Почему необходима однородность дисперсий по каждому уровню фактора?
4. Можно ли применить G-критерий для вывода о влиянии X на Y .
5. Почему необходимы повторения опытов на каждом уровне X ?

3. ПРАКТИЧЕСКАЯ РАБОТА № 3. ОДНОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

3.1. Краткие теоретические сведения

Познакомившись с основными целями и приемами дисперсионного анализа, поставим следующую задачу: научиться определять, как влияет фактор X (пока один) на функцию цели Y в условиях действия случайных факторов W .

Заметим, что из двух величин X и Y , между которыми устанавливается связь, лишь одна Y случайная. Если бы обе величины были неслучайными, то мы бы имели дело с обычной функциональной зависимостью, изучаемой в математическом анализе. Если, наоборот, рассматриваются две случайные величины X и Y , то для оценки связи между ними можно воспользоваться коэффициентом корреляции [7].

В нашем случае (X — неслучайная величина, Y —случайная) зависимость $Y = F(X)$ называют *регрессией* или *уравнением регрессии*. Эта зависимость в отличие от функциональной отражает связь между *наиболее вероятным значением* случайной величины Y и величиной X .

Основой для нахождения регрессии служит экспериментальный материал — N точек в плоскости Y, X . Их расположение пока что несущественно: они могут быть получены, как в рассмотренном примере, или могут появиться в результате установки X на некоторых фиксированных уровнях и проведения на каждом уровне одного или нескольких измерений Y .

Получение регрессии включает три этапа.

Этап I: постулируется вид уравнения регрессии, включающего, кроме X и Y , несколько неизвестных параметров (коэффициентов), которые находятся по опытным данным.

Этап II: определяются неизвестные коэффициенты, входящие в постулированное уравнение регрессии.

Этап III: проводится статистический анализ регрессии с целью доказательства ее адекватности экспериментальным данным, которые она представляет.

Этап I выполняется обычно на основе априорных соображений (например, нужна линейная модель), предшествующего опыта, интуиции и других неформализуемых приемов.

Обычно (но не обязательно) уравнение регрессии постулируется в виде полинома. Наиболее распространенными и удобными регрессионными моделями можно считать следующие: модель 1 порядка $Y=B_0+B_1 X$; модель II порядка: $Y=B_0+B_1 X+B_{11} X^2$.

В практике могут использоваться любые другие модели, в том числе и нелинейные относительно коэффициентов (см. табл. 8).

Таблица 8

Виды уравнений регрессии и их линеаризующие преобразования

№ п/п	Функции	Линеаризующие преобразования			
		Преобразования переменных		Выражения для величин B_0 и B_1 .	
		Y'	X'	B_0'	B_1'
1.	$Y = B_0 + \frac{B_1}{X}$	Y	$\frac{1}{X}$	B_0	B_1
2.	$Y = \frac{1}{B_0 + B_1 \cdot X}$	$\frac{1}{Y}$	X	B_0	B_1
3.	$Y = \frac{X}{B_0 + B_1 \cdot X}$	$\frac{X}{Y}$	X	B_0	B_1
4.	$Y = B_0 \cdot B_1^X$	$\ln(Y)$	X	$\ln(B_0)$	$\ln(B_1)$
5.	$Y = B_0 \cdot e^{B_1 \cdot X}$	$\ln(Y)$	X	$\ln(B_0)$	B_1
6.	$Y = \frac{1}{B_0 + B_1 \cdot e^{-X}}$	$\frac{1}{Y}$	e^{-X}	B_0	B_1
7.	$Y = B_0 \cdot X^{B_1}$	$\ln(Y)$	$\ln(X)$	$\ln(B_0)$	B_1
8.	$Y = B_0 + B_1 \cdot \ln(X)$	Y	$\ln(X)$	B_0	B_1
9.	$Y = \frac{B_0}{B_1 + X}$	$\frac{1}{Y}$	X	$\frac{B_1}{B_0}$	$\frac{1}{B_0}$
10.	$Y = \frac{B_0 \cdot X}{B_1 + X}$	$\frac{1}{Y}$	$\frac{1}{X}$	$\frac{B_1}{B_0}$	$\frac{1}{B_0}$
11.	$Y = B_0 \cdot e^{\frac{B_1}{X}}$	$\ln(Y)$	$\frac{1}{X}$	$\ln(B_0)$	B_1

12.	$Y = B_0 + B_1 \cdot X^n$	Y	X^n	B_0	B_1
-----	---------------------------	-----	-------	-------	-------

В результате замены переменных, все указанные в табл. 8 функции могут быть преобразованы в линейную модель первого порядка.

Отметим, что неудачно выбранная (постулированная) модель может вызвать непроизводительную трату времени, хотя и не приводит к принципиальным ошибкам в результате, поскольку ее адекватность проверяется на последнем этапе и плохая модель не используется. Поэтому на этапе выбора вида регрессии надо использовать всю имеющуюся информацию. Например, если известно, что Y и X обратно пропорциональны, можно выбрать в качестве фактора $X' = 1/X$. Если априорной или какой-либо другой информации нет, следует начинать с простейшей модели (I порядка) — даже если она и окажется неадекватной, потребует для получения меньших усилий.

Этап II — нахождение неизвестных коэффициентов по опытным данным — обычно выполняют на основе метода наименьших квадратов (МНК). Рассмотрим этот широко используемый в науке и технике метод и научимся его применять.

Пусть постулирована линейная регрессия и требуется найти B_0 и B_1 по N опытным точкам. В любом u -м опыте имеем

$$Y_u = \beta_0 + \beta_1 X_u + \varepsilon_u,$$

где Y_u , X_u — значения случайной величины Y и неслучайного аргумента X в u -м опыте; β_0 , β_1 — «истинные» значения коэффициентов регрессии, соответствующие модели $Y_{\text{ист}} = \beta_0 + \beta_1 X$;

ε_u — ошибка, объясняющая отличие наблюдаемого значения от «истинного» и обусловленная воздействием на Y случайных факторов.

Очевидно, что нам никогда не удастся найти β_0 и β_1 , а коэффициенты B_0 и B_1 , которые мы должны определить, есть некоторые *наилучшие оценки* β_0 и β_1 . Метод наименьших квадратов определяет следующее содержание термина «наилучшие»:

$$\sum_{i=1}^N (Y_i - Y_{i \text{ уcm}})^2 \Rightarrow \min ,$$

т. е. такие, что сумма квадратов отклонений наблюдаемых Y от «истинных» $Y_{\text{ист}}$ минимальна.

Для линейной регрессии следует другая запись условия МНК

$$S = \sum_{i=1}^N (Y_i - \beta_0 - \beta_1 X_i)^2 = \sum_{i=1}^N \varepsilon_i^2 \Rightarrow \min ,$$

непосредственно позволяющая найти оценки B_0 и B_1 . Продифференцировав S по β_0 и β_1 и приравняв полученные частные производные нулю, имеем следующую систему линейных алгебраических уравнений:

$$\begin{cases} B_0 N + B_1 \sum_{i=1}^N X_i = \sum_{i=1}^N Y_i \\ B_0 \sum_{i=1}^N X_i + B_1 \sum_{i=1}^N X_i^2 = \sum_{i=1}^N (Y_i \cdot X_i), \end{cases}$$

решая которую, любым способом, определяем неизвестные коэффициенты уравнения регрессии B_0 и B_1 .

Отметим, что изложенная процедура может быть осуществлена в любых случаях — МНК всегда позволит провести линию, соответствующую постулированному выражению и наилучшим образом (в указанном выше смысле) отвечающую экспериментальным точкам.

Этап III — анализ адекватности регрессии, в результате которого мы должны выяснить, удачно ли выбрана исходная модель, т. е. адекватно ли отражает полученная регрессия экспериментальный материал, положенный в ее основу.

Проверку адекватности постулированной модели произведем путем сравнения двух дисперсий, одна из которых S_Y^2 характеризует влияние на Y случайных факторов, а вторая, дисперсия адекватности S_{ad}^2 , оценивает разброс опытных значений Y относительно линии регрессии и вычисляется как

$$S_{ad}^2 = \frac{\sum_{i=1}^N (Y_i - y(X_i))^2}{N - K}$$

с числом степеней свободы $f_{ad} = N - K$ (в модели по N опытам определены K коэффициентов).

Составим отношение

$$F_p = \frac{S_Y^2}{S_{ad}^2}$$

и сравним его с квантильной границей критерия Фишера F_T при $f_1 = N-1$ и $f_2 = f_{ad}$. Если $F < F_{\text{табл.}}$, то гипотеза об однородности сравниваемых дисперсий не отвергается на уровне значимости q , следовательно, наблюдаемый разброс Y и $y(X_i)$ может быть объяснен естественным влиянием случайных факторов и модель неадекватно описывает процесс. Если же $F > F_{\text{табл.}}$ то S_{ad}^2 значимо меньше S_y^2 , т. е. постулированная модель статистически значимо описывает процесс и может быть признана адекватной. Заметим, что в случае, когда дисперсия адекватности стремится к нулю, расчетный критерий Фишера стремится к бесконечности. Таким образом, дисперсия адекватности может быть признана основным параметром, оценивающим качество построенной модели.

3.2. Задание на практическую работу

Исследуется влияние скорости подачи на шероховатость поверхности при обработке стальной заготовки на токарном станке. В качестве постоянных параметров эксперимента используются: глубина резания $t = 0,5$ мм, диаметр обрабатываемой поверхности $d = 50$ мм, частота вращения шпинделя $n = 1000$ мин⁻¹, Материал инструмента — Т15К6, материал детали — Сталь 20.

Выполнить однофакторный регрессионный анализ по полученным

экспериментальным данным. Форма регрессионной зависимости определяется по табл. 8. Номер варианта необходимо получить у преподавателя.

3.2. Контрольные вопросы

1. Метод наименьших квадратов в регрессионном анализе.
2. Особенности выбора вида регрессионной зависимости на первом этапе.
3. Определить систему уравнений для вычисления неизвестных коэффициентов регрессионной зависимости вида $Y = B_0 \cdot B_1^x$.
4. Определите понятие дисперсии адекватности модели.
5. Какая из нескольких полученных регрессионных моделей должна быть признана наилучшей?

№ п.п.	Ra , мкм	S , мм/мин
1	9,54	340
2	0,74	65
3	6,13	245
4	0,51	20
5	3,09	185
6	7,79	278
7	2	135
8	0,72	40
9	5,3	235
10	1,29	60
11	5,86	235
12	1,59	80
13	1,37	90
14	1,94	100
15	1,59	100
16	2,21	120
17	3	140
18	2,23	160
19	3,87	180
20	3,97	200
21	4,93	220
22	5,6	265
23	6,76	280
24	9,41	300
25	8,76	315
26	11,2	350

4. ПРАКТИЧЕСКАЯ РАБОТА № 4. МНОГОФАКТОРНЫЙ РЕГРЕССИОННЫЙ АНАЛИЗ

4.1. Краткие теоретические сведения

Наиболее интересные и практически важные задачи обычно предполагают получение зависимости Y от *нескольких воздействующих факторов* $\{X \rightarrow X_1, X_2, \dots, X_i, \dots, X_n\}$. Такие задачи приводят к многофакторному регрессионному анализу.

Очевидно, что иметь дело с несколькими разнородными независимыми переменными, которые выражаются совершенно различными именнованными числами и варьируются в различных пределах и т. п., по крайней мере, неудобно. Для того чтобы избавиться от этого неудобства, перейдем к кодированным факторам, которые будем обозначать строчными буквами x с соответствующими индексами и которые будут связаны с исходными истинными факторами X соотношениями

$$x = \frac{X - X_{cp}}{X_{cp} - X_{min}} = \frac{X - X_{cp}}{X_{max} - X_{cp}},$$

где X_{min} , X_{max} — минимальное и максимальное значения фактора, выбранные внутри диапазона его изменения (варьирования); X_{cp} — среднее значение фактора, определяемое как

$$X_{cp} = \frac{X_{min} + X_{max}}{2}.$$

После выполнения операции кодирования мы имеем дело с безразмерными факторами, варьирующимися около нуля в небольших пределах (часто в активном эксперименте выбираются пределы $-1 \dots +1$).

Постулируем модель следующего вида (коэффициенты при кодированных факторах будем обозначать строчными буквами b):

$$\hat{Y} = b_0 + b_1 x_1 + \dots + b_n x_n + b_{12} x_1 x_2 + b_{13} x_1 x_3 + \dots \\ \dots + b_{n-1, n} x_{n-1} x_n + b_{123} x_1 x_2 x_3 + \dots + b_{12 \dots n} x_1 x_2 \dots x_n + b_{11} x_1^2 + \dots + b_{nn} x_n^2.$$

В модель входят нулевой член, члены вида $b_i x_i$ (линейная регрессия),

члены вида $b_{ijk}x_i x_j x_k$, отражающие взаимодействия факторов (неполная квадратичная регрессия), и, наконец, члены вида $b_{ii}x_i^2$ (квадратичная регрессия). Очевидно, что при значительных n запись вида уравнения становится весьма громоздкой и неудобной. Не изменяя вида модели, запишем ее в более компактной форме

$$\hat{Y} = \sum_{i=0}^m b_i x_i,$$

где $x_0 \equiv 1$; x_i при $i > n$ обозначает произведения и квадраты факторов; b_i при $i > n$ — коэффициенты при произведениях и квадратах факторов; m — полное число членов полинома, не считая нулевого (всего $m+1$ член).

Переходим теперь к определению коэффициентов регрессии b_i по данным опытов. Как и в случае одномерной регрессии (см. п. 3.1), в любом u -м опыте имеем

$$Y_u = \beta_0 x_{0u} + \beta_1 x_{1u} + \dots + \beta_m x_{mu} + \varepsilon_u,$$

где Y_u — значение Y в u -м опыте $x_{0u}, x_{1u}, \dots, x_{mu}$ — значения независимых переменных x_0, \dots, x_m , в u -м опыте; β_0, \dots, β_m — «истинные» значения коэффициентов регрессии, оценки которых b_0, \dots, b_m входят в модель; ε_u — ошибка в u -м опыте.

Результаты всех N опытов запишем в виде матричного уравнения

$$Y = X\beta + \varepsilon,$$

где

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_u \\ \vdots \\ Y_N \end{bmatrix}; \quad X = \begin{bmatrix} x_0 & x_{11} & \dots & x_{m1} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{0u} & x_{1u} & \dots & x_{mu} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ x_{0N} & x_{1N} & \dots & x_{mN} \end{bmatrix}; \quad \beta = \begin{bmatrix} \beta_0 \\ \vdots \\ \beta_i \\ \vdots \\ \beta_m \end{bmatrix}; \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_u \\ \vdots \\ \varepsilon_N \end{bmatrix}.$$

Размер вектора наблюдений, или матрицы-столбца Y и вектора ошибок, или матрицы-столбца ε — $(N \times 1)$; размер вектора коэффициентов $(m+1 \times 1)$; размер матрицы X $(N \times m+1)$.

Условие МНК запишем в виде

$$S = (Y - X\beta)^2 \rightarrow \min.$$

Нормальные уравнения получим, продифференцировав по β :

$$\frac{dS}{d\beta} = -2x_i(Y - x\beta)$$

(дифференцирование выполняется по обычным правилам, но перед скобкой записывается x_i , чтобы умножение имело смысл) и приравняв производную нулю:

$$X^T X B = X^T Y$$

Тогда

$$B = (X^T \cdot X)^{-1} X^T Y,$$

что и является матрицей МНК—оценок коэффициентов регрессии.

Обозначив $X^T X = C$, получим $B = C^{-1} X^T Y$.

Отметим, что в выражение для B входит обратная матрица C^{-1} . Исходная матрица C определена на основе матрицы X , отражающей значения независимых переменных в опытах. Если мы сами задаем независимые переменные в каждом опыте, т.е. проводим активный эксперимент, мы имеем возможность позаботиться о том, чтобы сконструировать C , а, следовательно, и C^{-1} некоторым наилучшим образом.

Перейдем к третьему этапу регрессионного анализа и получим выражения для дисперсии оценок коэффициентов B .

Введем в рассмотрение матрицу дисперсии коэффициентов:

$$DB = C^{-1} \cdot DY,$$

где DY — дисперсия наблюдений

В общем случае DB — квадратная матрица размером $(1+m) \times (1+m)$; диагональные элементы — дисперсии соответствующих коэффициентов, недиагональные ковариации, представляющие собою меру неопределенно-

сти, возникающей в силу зависимости друг от друга коэффициентов регрессии. Матрица **DB** может быть использована для оценки значимости коэффициентов регрессии.

$$t_i = \frac{B_i}{DB_{i,i}}, \quad i = 0, 1, \dots, m$$

где t_i — расчетный критерий Стьюдента в задаче о гипотезе равенства нулю i — го коэффициента модели.

Проверка адекватности регрессии выполняется способом, описанным в п. 3.1.

4.2. Задание на практическую работу

Исследуется влияние диаметра обработки d , скорости резания V , по-

Ra, мкм	S, мм/мин	d , мм	V , м/мин
10,7	43	80	0,1
9,6	28	80	0,1
8,9	74	60	0,05
8,0	48	60	0,05
11,4	32	60	0,15
10,6	24	60	0,15
13,9	28	40	0,3
14,8	36	40	0,3
14,8	36	40	0,3
18,3	36	40	0,5
18,0	48	20	0,5
10,0	48	34	0,1
8,7	69	54	0,05
14,1	66	30	0,2
14,8	56	30	0,25
16,5	63	30	0,3

дачи S на шероховатость поверхности при обработке стальной заготовки на сверлильном станке. В качестве постоянных параметров эксперимента используются: материал детали — сталь 40, материал инструмента Р6М5.

Построить модель процесса с помощью многофакторного регрессионного анализа. Конкретный вид модели задается преподавателем. Экспериментальные данные представлены в таблице.

4.3. Контрольные вопросы

1. Какую информацию о процессе можно получить при анализе значимости коэффициентов регрессионной модели?
2. Составьте возможный вид уравнения регрессии, состоящего из 4 управляемых факторов и 20 коэффициентов.
3. Объясните необходимость перехода к относительным факторам.
4. Поясните смысл и необходимость проведения активного эксперимента при моделировании исследуемого процесса.
5. Каким образом осуществляется обратный переход к декодированным факторам в регрессионной многофакторной модели?

5. Литература

1. Автоматизация экспериментальных исследований (организация эксперимента). Чиченев Н.А. Под ред. академика АН КазССР П.И. Полухина: Учебное пособие для вузов.- М.: Металлургия, 1983, — 256 с.
2. Адлер Ю. П., Маркова. Е. В., Грановский Ю.В. Планирование эксперимента при поиске оптимальных условий. Изд.2-е, М.: Наука, 1976, — 279 с.
3. Денисов В.К. Математическое обеспечение системы ЭВМ-экспериментатор (регрессионный и дисперсионный анализы). М.: Наука, 1977, — 251 с.
4. Шенк Х. Теория инженерного эксперимента: Пер, с англ./ Ред. Бусленко Н.П., М.: Мир, 1972, — 381 с.
5. Ильинский Я.Ф. Элементы теории эксперимента, М.: Моск. энерг. ин-т, 1983, — 92 с.
6. Круг Г.К., Сосулин Ю.А. Планирование эксперимента в задачах идентификации и экстраполяции. М.: Наука, 1977, — 208 с.

Учебно-методическое издание

Составитель

Алексей Романович ЛЕБЕДЕВ

АВТОМАТИЗАЦИЯ ОБРАБОТКИ
ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

Методические указания к практическим работам №1 — №4
по дисциплине «Автоматизация обработки экспериментальных данных»
для студентов 5-го курса
специальности 151001 всех форм обучения

Ответственный за выпуск

Зав. кафедрой «Технология автоматизированного машиностроения»
д-т. техн. наук, профессор М.П. Шишкарёв

Подписано в печать _____

Формат 60х84/16

Бумага офсетная. Объем _____ усл. п.л., _____

уч.-изд. л.

Заказ № _____ Тираж _____ экз.

Редакционно-издательский отдел РГАСХМ ГОУ

344023, г. Ростов-на-Дону, ул. Страны Советов, 1

Отпечатано в копировально-множительном бюро РГАСХМ ГОУ