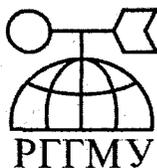


Министерство образования Российской Федерации
Российский Государственный Гидрометеорологический
Университет

Бекряев В. И.

ПРАКТИКУМ
ПО ОСНОВАМ ТЕОРИИ ЭКСПЕРИМЕНТА



РГГМУ

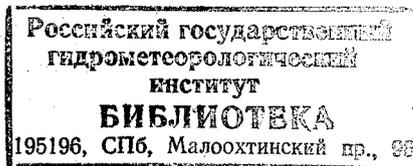
Санкт-Петербург
2003

УДК 551.509.3; 519.24; 389.001

Бекряев В. И. Практикум по основам теории эксперимента. - СПб.; РГГМУ, 2003 – 72 с.

Практикум содержит тематические задания по различным разделам учебной дисциплины «Основы теории эксперимента». Предназначен для более глубокого изучения теоретических основ и приобретения навыков обработки и анализа результатов измерений, математического планирования экспериментов.

Рекомендуется для студентов, магистров и аспирантов гидрометеорологических университетов.



© Российский государственный гидрометеорологический университет (РГГМУ), 2003
В. И. Бекряев

ОГЛАВЛЕНИЕ

	Стр.	
	Предисловие	4
Задание 1.	Общие сведения из теории измерений	5
Задание 2.	Распределение случайных погрешностей	7
Задание 3.	Исследование зависимости выборочных среднего и среднего квадратического отклонения от объема выборки	9
Задание 4.	Доверительные границы погрешности результата измерений и среднего квадратического отклонения	11
Задание 5.	Анализ эмпирических распределений.	15
Задание 6.	Часть I. Графические методы анализа	18
	Анализ эмпирических распределений.	
Задание 7.	Часть II. Проверка гипотезы о нормальности распределения результатов измерений	22
	Погрешности косвенных измерений	
Задание 8.	Суммирование систематических и случайных погрешностей	25
Задание 9.	Оценка однородности и равнозначности рядов измерений.	28
	Часть I. Параметрические методы	
Задание 10.	Оценка однородности и равнозначности рядов измерений.	33
	Часть II. Непараметрические методы	
Задание 11.	Дисперсионный анализ	38
Задание 12.	Корреляция между случайными величинами	41
Задание 13.	Автокорреляционная и спектральная функции	45
Задание 14.	Интерполяция и сглаживание	48
Задание 15.	Планирование активного эксперимента при изучении поверхности отклика.	53
	Часть I. Планирование эксперимента при построении функции отклика	
Задание 16.	Планирование активного эксперимента при изучении поверхности отклика.	57
	Часть II. Планирование эксперимента при поиске экстремума	
Задание 17.	Построение дискриминантной функции	62
	Приложения	
	Литература	71

ПРЕДИСЛОВИЕ

Выпускник технического университета, профессионально связанный с постановкой эксперимента, выполнением измерений, использованием измерительной информации, неизбежно сталкивается с обработкой и анализом результатов измерений, оценками их точности, принятием решений в условиях неопределенности. Эти задачи решаются методами прикладной математической статистики. Являющаяся ветвью математической статистики теория планирования эксперимента дает последовательно четкую стратегию подготовки и проведения опытов, обеспечивающую заданную точность при минимальном числе испытаний. Теория эксперимента как учебная дисциплина предназначена для более глубокого изучения и практического использования методов прикладной статистики.

Настоящий практикум является дополнением и продолжением учебного пособия «Основы теории эксперимента», 2002 г. (автор Бекряев В. И.). Он построен по принципу тематических заданий. Такая форма позволяет учащемуся получить более полное представление о различных разделах курса. Большинство заданий содержит элементы численного статистического моделирования. Выполнение их ориентировано на использовании ПК. В заданиях, как правило, не используются конкретные ряды измерений. «Результаты экспериментов» имитируются с помощью простых моделей. Предполагается, что исполнителю как экспериментатору неизвестна зависимость исследуемой величины от контролируемых факторов. В то же время как имитатор он «знает» конкретную модель объекта. При этом статистические характеристики генеральной совокупности являются заранее известными. Это обстоятельство позволяет выявить достоинства и недостатки методов анализа выборок ограниченного объема.

ОБЩИЕ СВЕДЕНИЯ ИЗ ТЕОРИИ ИЗМЕРЕНИЙ

Проблема обеспечения единства измерений в масштабах страны и мирового сообщества требует однозначности и определенности терминологии, унификации единиц и аттестации средств измерений.

Целью работы является изучение терминологической базы теории измерений, структуры, содержания и правил использования системы единиц СИ, способов выражения классов точности средств измерений.

При выполнении задания учащемуся предлагается проработать содержание государственных стандартов [7, 8, 9] и познакомиться с пояснениями и комментариями, представленными в изданиях [2, 10, 13, 14]:

Содержание работы

1. Изучение терминов и определений в метрологии. Изучите ГОСТ 16263-70 [7]. Некоторые из определений приводятся в книгах [2, 13, 14]. Сравните определения, данные в стандарте [7] с тем, как они уточняются и развиваются в словаре-справочнике [10]. Выпишите или запомните (необязательно наизусть, но сохраняя основные положения) определения следующих ключевых терминов:

- физическая величина (ФВ),
- размер ФВ, значение ФВ, истинное значение ФВ, действительное значение ФВ;
- единица физической величины,
- система единиц физических величин: основные, дополнительные и производные единицы, когерентная система единиц;
- измерения физической величины,
- однократные и многократные, прямые, косвенные, совместные и совокупные, ряд измерений;
- средства измерительной техники (технические средства измерений),
- рабочие и метрологические средства измерений, меры, измерительные приборы и преобразователи, измерительные установки и системы;
- принцип, метод и методика измерений,
- виды методов измерений;
- погрешности результатов измерений,
- абсолютные и относительные погрешности, статические и динамические, методические, инструментальные и субъективные, систематические, случайные и грубые погрешности;
- погрешности средств измерения,
- абсолютные, относительные, приведенные, основные и дополнительные;
- эталоны единиц физических величин,
- первичный и государственный эталоны, рабочий эталон, образцовые и рабочие средства измерений, передача размера единиц;
- метрологическая служба.

2. Изучение системы единиц СИ ГОСТ 8.417-81 [8]. Используйте литературные источники [2, 13, 14].

2.1. Приготовьте ответы на следующие контрольные вопросы.

1. Перечислите основные достоинства СИ.
2. Дайте определение основных и дополнительных единиц СИ.
3. Как образуются производные единицы СИ.
4. Перечислите виды внесистемных единиц, приведите примеры единиц каждого вида.
5. Назовите регламентированные стандартом дольные и кратные приставки к единицам СИ.
6. Перечислите основные правила образования, использования и написания наименований и обозначений единиц СИ.

2.2. Выполните упражнения.

2.2.1. Запишите с использованием кратных и дольных единиц СИ следующие выражения:

- Среднее расстояние от Земли до Солнца $R = 150000000$ км =
- Масса атмосферы Земли $M = 5,3 \cdot 10^{18}$ кг =
- Заряд электрона $e = 1,602 \cdot 10^{-19}$ Кл =
- Электрическая постоянная $\epsilon_0 = 8,85 \cdot 10^{-12}$ Ф/м =

– Магнитная постоянная $\mu_0 = 12,56 \cdot 10^{-7}$ Гн/м =

– Гравитационная постоянная $\gamma = 6,67 \cdot 10^{-11}$ м³/(кг·с²).

2.2.2. Представьте следующие выражения в форме $n \cdot 10^x$.

- | | |
|-------------|---------------------------|
| 1. 12,0 ГОм | 4. 0,12 мкКл ² |
| 2. 120 пФ | 5. 500 фСм/м |
| 3. 1,20 мкг | 6. 0,5 Эбк |

2.2.3. Найдите и исправьте ошибки — отступления от ГОСТа (если они есть) в следующих выражениях:

- температура воздуха 15° С,
- давление воздуха 1010 мб,
- влажность воздуха 85%,
- скорость ветра 11 $\frac{\text{м}}{\text{с}}$ с порывами до 17 метров/с,
- направление ветра 150 град.,
- дальность видимости 300 м \pm 20 %,
- высота нижней границы облаков 1,2 км.,
- интенсивность осадков 14 мм в час,
- напряженность электрического поля 0,03МВ/м.

2.2.4. Найдите и исправьте отступления от ГОСТа (если они есть).

- | | |
|--|--|
| 1. $c_p = 1005$ Дж/кг·К | 6. 5 кМВт=5 МкВт |
| 2. $L_0 = 2500$ кДж/кг | 7. 4,7 мегом = 4,7 МОм |
| 3. 1См = 1 А ² ·кг ⁻¹ ·с ⁻³ /м ² | 8. $W = 10^{-3} UIt$ кВ·час, если U в В, I в А, а t в ч. |
| 4. 27 ± 2 мV | 9. 1 Н·км = 1 ньютон на километр. |
| 5. 1200 пF | |

2.2.5. Просклоняйте по падежам наименование единиц с численными значениями

- 996 гектопаскалей,
- 48 кандел,
- 20 вольт,
- 14 метров в секунду,
- 18 промилле.

3. Классы точности средств измерений (ср. изм.). Изучите ГОСТ 8.401-80 [9]. Используйте литературу [2, 13, 14].

3.1. Приведите ответы на следующие вопросы.

1. Что такое класс точности ср. изм.?
2. Как выражаются абсолютные погрешности ср. изм.?
3. Что такое приведенная погрешность ср. изм.?
4. Как выражаются приведенная и относительная погрешности ср. изм.?
5. Как обозначаются классы точности на шкалах и щитках средств измерений?

3.2. Решите задачи.

1. В распоряжении экспериментатора имеются два вольтметра. Класс точности первого $\textcircled{2,0}$, шкала 0...100 В, класс точности второго 1.0, шкала -50...100 В. Какой из этих приборов следует использовать, чтобы получить минимальную абсолютную погрешность, если предполагаемое значение напряжения около 60 В?

2. Имеются два прибора, предназначенные для измерения величины x . Шкалы приборов одинаковы: 0... x_k . Класс точности одного из них $\textcircled{1,0}$, другого 0,5/0,25. При каком значении x абсолютные погрешности обоих приборов окажутся равными?

Литература: [2], раздел 1, [7], [8], [9], [10], [13], [14].

РАСПРЕДЕЛЕНИЕ СЛУЧАЙНЫХ ПОГРЕШНОСТЕЙ

Целью работы является закрепление теоретических сведений о функции распределения, ее свойствах, способах выражения, видах распределений, используемых для описания погрешностей измерений.

Содержание работы

1. Докажите, что справедливы следующие преобразования выражений:

– для среднего квадратического отклонения

$$S(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{\left(\sum_{i=1}^n x_i \right)^2}{n} \right]},$$

– для асимметрии

$$D_3 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^3 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^3 - \frac{3}{n} \left(\sum_{i=1}^n x_i^2 \right) \sum_{i=1}^n x_i + \frac{2}{n^2} \left(\sum_{i=1}^n x_i \right)^3 \right],$$

– для эксцесса

$$D_4 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^4 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^4 - \frac{4}{n} \left(\sum_{i=1}^n x_i^3 \right) \sum_{i=1}^n x_i + \frac{6}{n^2} \sum_{i=1}^n x_i^2 \sum_{i=1}^n x_i^2 - \frac{3}{n^3} \left(\sum_{i=1}^n x_i \right)^4 \right],$$

– для коэффициента парной корреляции

$$r_{x_1 x_2} = \frac{\frac{1}{n-1} \sum_{i=1}^n (x_{1i} - \bar{x}_1)(x_{2i} - \bar{x}_2)}{S(x_1)S(x_2)} = \frac{\sum_{i=1}^n x_{1i} x_{2i} - \frac{1}{n} \sum_{i=1}^n x_{1i} \cdot \sum_{i=1}^n x_{2i}}{\sqrt{\left[\sum_{i=1}^n x_{1i}^2 - \frac{1}{n} \left(\sum_{i=1}^n x_{1i} \right)^2 \right] \left[\sum_{i=1}^n x_{2i}^2 - \frac{1}{n} \left(\sum_{i=1}^n x_{2i} \right)^2 \right]}}.$$

Внимание! В одном из выражений намеренно допущена ошибка. Найдите ее.

2. Докажите, что для нормально распределенной случайной величины x с плотностью вероятности

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma(x)} \exp \left[-\frac{(x - \bar{x})^2}{2\sigma^2(x)} \right]$$

второй центральный D_2 момент в точности равен $\sigma^2(x)$.

Рекомендации: Для упрощения выкладок удобно перейти к стандартному нормальному распределению. В этом случае $D_2 = \sigma^2(z) = 1$.

3. Вычислите коэффициенты эксцесса для следующих распределений:

– нормального (плотность вероятности приведена в п. 2),

– равномерного

$$f_p(x) = \begin{cases} 1/(b-a), & x \in [a, b], \\ 0, & x \notin [a, b]; \end{cases}$$

– треугольного

$$f_T = \begin{cases} \frac{4(x-a)}{(b-a)^2}, & x \in [a, \bar{x}], \\ \frac{4(b-\bar{x})}{(b-a)^2}, & x \in [\bar{x}, b], \\ 0, & x \notin [a, b], \end{cases}$$

– распределения Лапласа

$$f_L(x) = \frac{1}{\sqrt{2} \cdot \sigma_L(x)} \exp\left[-\frac{\sqrt{2}}{\sigma_L(x)} |x - \bar{x}|\right].$$

Рекомендация: Как и в п. 2, для упрощения расчетов центрируйте все распределения. Например, для равномерного и треугольного распределений удобно задать $a = -\theta, b = \theta$.

Ответы: 0; -1,2; -0,6; 3,0.

4. Решите задачи:

1. Случайная величина A имеет нормальное распределение. $\bar{A} = 30, \sigma(A) = 10$. С какой вероятностью случайная величина заключена в интервале $[20, 40]$?

2. Случайная величина A имеет равномерное распределение на отрезке $[10, 50]$. С какой вероятностью она находится в интервале $[20, 40]$?

3. Случайная величина имеет треугольное распределение на интервале $[10, 50]$. С какой вероятностью она заключена в том же интервале $[20, 40]$?

Ответы: 0,68; 0,50; 0,75.

Литература: [2], раздел 2, [5], [6], [14].

ИССЛЕДОВАНИЕ ЗАВИСИМОСТИ ВЫБОРОЧНЫХ СРЕДНЕГО И СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ ОТ ОБЪЕМА ВЫБОРКИ

Известно, что результат измерения случайной величины, полученный по выборке

$$\bar{x} = \left(\sum_{i=1}^n x_i \right) / n, \quad (1)$$

и выборочное среднее квадратическое отклонение (СКО)

$$S(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2} \quad (2)$$

при увеличении объема выборки n стремятся, соответственно, к истинному значению x_0 и генеральному СКО $\sigma(x)$. (Величины \bar{x} и $S(x)$ являются оценками x_0 и $\sigma(x)$: $\bar{x} = \hat{x}_0$, $S(x) = \hat{\sigma}(x)$). Конкретные значения \bar{x} и $S(x)$, вычисленные по ограниченным выборкам, могут существенно отличаться от x_0 и $\sigma(x)$.

Целью настоящего задания является изучение изменчивости \bar{x} и $S(x)$ при изменении объема выборки от малых значений n ($n = 4 \dots 8$) до сравнительно больших ($n > 100$).

Ограничимся рассмотрением нормально распределенной случайной величины. Для решения задачи воспользуемся методами статистического численного моделирования.

Будем моделировать выборку нормально распределенной случайной величины. В силу центральной предельной теоремы нормальное распределение случайной величины можно получить, если формировать каждое значение, как сумму бесконечно большого числа других случайных величин. Практически удобно использовать с этой целью равномерно распределенные случайные числа. Образует значение x_i как

$$x_i = \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i, \quad \mu = \overline{1, m_1}, \quad i = \overline{1, n},$$

где a_{μ} – равномерно распределенные на некотором интервале случайные числа, m_1 – число суммируемых значений. Уже при $m_1 > 6 \dots 10$ распределение x_i оказывается соответствующим нормальному. Обычно a_{μ} задают на интервале $[0,1]$. Для получения a_{μ} либо используют таблицы случайных чисел, либо, что современнее, специальные программы на ПК. Эти программы называют генераторами случайных чисел или RND.

Для равномерно распределенной на интервале $[0,1]$ случайной величины ее среднее значение $\bar{a} = 0,5$, а дисперсия $\sigma^2(a) = 1/12$. Следовательно, при $n \rightarrow \infty$ значение $x_0 = m_1 \bar{a} = 0,5 m_1$, а $\sigma^2(x) = m_1 / 12$. Изучение отличий выборочных \bar{x} и $S^2(x)$ от этих значений и составляет предмет настоящего задания.

Принято считать, что выборочные значения $S(x)$ обычно меньше $\sigma(x)$. Учащемуся предоставляется возможность убедиться в ошибочности этого суждения.

Содержание работы

1. Сформируйте ряд нормально распределенной случайной величины x_i , $i = \overline{1, n}$. (Удобно задать n как $n = 2^p$, где, например, $p = 7, 8, \dots$).

Если в работе используется RND, то вычислительную процедуру удобно построить следующим образом. В темпе поступления x_i следует рассчитывать:

$$\sum_{i=1}^L x_i, \quad x_i^2, \quad \sum_{i=1}^L x_i^2, \quad \bar{x}_L = \left(\sum_{i=1}^L x_i \right) / L, \quad S_L^2(x) = \left(\frac{\sum_{i=1}^L x_i^2 - \left(\sum_{i=1}^L x_i \right)^2 / L}{L-1} \right), \quad S_{\bar{x},L} = S_L(x) / \sqrt{L},$$

$$t_{1-\alpha/2,f} = z_{1-\alpha/2} + \frac{b_{1-\alpha/2}}{L-2}, \quad \varepsilon_{\bar{x},L} = t_{1-\alpha/2,f} \cdot S_{\bar{x},L},$$

$z_{1-\alpha/2}$ – квантиль стандартного нормального распределения.

Для справки: $b_{1-0,05/2} = 2,40$; $b_{1-0,10/2} = 1,30$; $z_{1-0,05/2} = 1,96$; $z_{1-0,10/2} = 1,64$.

Задавая $L = 2^2, 2^3, \dots, 2^p$, выведите на печать значения $L, \bar{x}_L, S_{\bar{x},L}$ и $\varepsilon_{\bar{x},L}$. Таким образом вы получили информацию, необходимую для анализа.

2. Одновременно и параллельно с вычислением указанных характеристик организуйте группировку x_i по интервалам $[x_j, x_{j+1}]$, $j = \overline{0, k}$, где k – число интервалов. Рассчитайте n_j – число x_i , попадающих в интервал $[x_j, x_{j+1}]$. Формально, случайные величины x_i изменяются в пределах от 0 до m_1 , фактически, в значительно более узком интервале, вблизи $\bar{x} = m_1 / 2$. Число интервалов k рекомендуется согласовывать с объемом выборки ($k = 1 + 3,2 \lg L$). Однако при решении данной задачи удобно выбирать k постоянным, не зависящим от L . Задайте k "с запасом": $k = 2m_1$.

Выведите на печать значение $n_j(L)$, $j = \overline{1, k}$ снова для тех же L – ных объемов выборки, что и выше. Постройте гистограммы по этим данным. Проанализируйте, как изменяется форма гистограммы по мере увеличения выборки.

3. Операции по пункту 1 повторите 4...5 раз. Результаты всех серий представьте на графиках зависимостей \bar{x} , $S_{\bar{x}}$, $\varepsilon_{\bar{x}}$ от L . Нанесите на эти графики характеристики генеральной совокупности $x_0 = m_1 / 2 = \text{const}(L)$, $\sigma(x) = \sqrt{m_1 / 12} = \text{const}(L)$, $\sigma_{\bar{x}} = \sigma(x) / \sqrt{L}$ и $\varepsilon_{\bar{x}} = z_{1-\alpha/2} \cdot \sigma_{\bar{x}}$, соответственно.

4. Выполните анализ полученных результатов.

Примечание

1. Часто RND построены таким образом, что при повторном запуске программы последовательность вырабатываемых случайных чисел оказывается повторяющейся. Чтобы избежать этого, генератор следует погонять сначала "на холостом ходу", то есть начинать суммирование a_μ после произвольного неповторяющегося числа срабатываний RND.

2. При отсутствии умения или возможности работы на ПК задание можно выполнить, используя таблицы равномерно распределенных случайных чисел – см. Приложение V.

Варианты исходных данных

Вариант	m_1	p	α
1	12	7	0,05
2	12	8	0,05
3	12	9	0,10
4	24	7	0,10
5	24	8	0,05
6	24	9	0,05

Литература: [2], разделы 2, 3.

ДОВЕРИТЕЛЬНЫЕ ГРАНИЦЫ ПОГРЕШНОСТИ РЕЗУЛЬТАТА ИЗМЕРЕНИЙ И СРЕДНЕГО КВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ

Обработка рядов измерений (выборок) заключается в оценивании численных значений параметров статистических распределений (точечные оценки) и доверительных границ интервалов, внутри которых с заданной вероятностью содержатся истинные значения этих параметров (интервальное оценивание).

Доказывается, что для нормального распределения точечные оценки истинного значения \hat{x}_0 , среднего квадратического отклонения результата отдельного измерения $\hat{\sigma}(x)$, среднего квадратического отклонения результата измерения $\hat{\sigma}_{\bar{x}}$ и средней квадратической погрешности среднего квадратического отклонения $\hat{\sigma}_{\sigma}$ вычисляются, соответственно, по формулам:

$$\hat{x}_0 \equiv \bar{x} = \left(\sum_{i=1}^n x_i \right) / n, \quad (1)$$

$$\hat{\sigma}(x) \equiv S(x) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2}, \quad (2)$$

$$\hat{\sigma}_{\bar{x}} \equiv S_{\bar{x}} = \frac{S(x)}{\sqrt{n}}, \quad (3)$$

$$\hat{\sigma}_{\sigma} \equiv S_{\sigma} = \frac{S(x)}{\sqrt{2(n-1)}}. \quad (4)$$

Здесь n – объем выборки.

Доверительные границы погрешности результата измерения (ДГПРИ), при заранее известном СКО $\sigma(x)$ и заданном уровне значимости α , составляют:

– для результата единичного измерения

$$\varepsilon(x) = z_{1-\alpha/2} \cdot \sigma(x),$$

– для результата многократных измерений (среднего арифметического значения \bar{x})

$$\varepsilon_{\bar{x}} = z_{1-\alpha/2} \cdot \sigma_{\bar{x}} = z_{1-\alpha/2} \frac{\sigma(x)}{\sqrt{n}}, \quad (5)$$

где $z_{1-\alpha/2}$ – квантиль стандартного нормального распределения, соответствующий вероятности $p = 1 - \alpha/2$.

При неизвестном заранее СКО используют его выборочные значения $S(x)$, в этом случае ДГПРИ вычисляются по формуле

$$\varepsilon_{\bar{x}} = t_{1-\alpha/2, f} \cdot S(x) / \sqrt{n}. \quad (6)$$

Здесь $t_{1-\alpha/2, f}$ – коэффициент Стьюдента, f – число степеней свободы $S(x)$, $f = n - 1$.

Когда закон распределения измеряемой физической величины неизвестен, для расчета ДГПРИ можно использовать формулы, вытекающие из простого и усиленного неравенств Чебышева. Соответственно,

$$\varepsilon_{\bar{x}} = \sqrt{\frac{1}{cn}} \cdot \sigma(x), \quad (7a)$$

$$\varepsilon_{\bar{x}} = \frac{2}{3} \sqrt{\frac{1}{cn}} \cdot \sigma(x). \quad (7b)$$

Доверительные границы погрешности СКО $\sigma(x)$ (ДГПСКО), определяются соотношением

$$P \left\{ S(x) \sqrt{\frac{f}{\chi_{1-\alpha/2, f}^2}} \leq \sigma(x) \leq S(x) \sqrt{\frac{f}{\chi_{\alpha/2, f}^2}} \right\} = 1 - \alpha. \quad (8)$$

Здесь $\chi_{\alpha/2, f}^2$ и $\chi_{1-\alpha/2, f}^2$ – квантили χ^2 – распределения.

Целью работы является закрепление сведений из теории, приобретение навыков вычисления точечных и интервальных оценок и освоение правил округления результатов вычислений.

Имитационная модель. Будем формировать ряд нормально распределенной случайной величины. Элементы этого ряда обозначим $x_i, i = \overline{1, n}$, n – длина ряда (объем выборки). Подробно процедура формирования x_i описана в предыдущем задании. Приведем здесь только алгоритм вычисления.

$$x_i = \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i, \quad (9)$$

где a_{μ} – равномерно распределенные случайные числа, вырабатываемые RND ($0 \leq a_{\mu} \leq 1$).

Истинное значение случайной величины (при $n \rightarrow \infty$) $x_0 = m_1 / 2$, а генеральное СКО $\sigma(x) = \sqrt{m_1 / 12}$.

Содержание работы

Вариант А. Выполнение работы с использованием ПК.

1. Расчет ДГПРИ и ДГПСКО.

1.1. Используя имитационную модель (9), сформируйте короткий ряд (малую выборку) при $n_k = 4, \dots, 6$.

Найдите выборочные значения \bar{x} и $S(x)$. (Техника вычислений описана в предыдущем задании).

Вычислите ДГПРИ при $\alpha = 0,05$:

- по формуле (5), предполагая, что известен закон распределения случайной величины и СКО $\sigma(x)$;
- по формуле (6), предполагая, что закон распределения по-прежнему нормальный, а $\sigma(x)$ неизвестно;
- по формуле (7а), предполагая, что закон распределения неизвестен, но известно значение $\sigma(x)$;
- по формуле (7б), предполагая, что неизвестное распределение случайной величины является одномодальным, $\sigma(x)$ – известно.

Сравните между собой значения $\varepsilon_{\bar{x}}$, рассчитанные по разным формулам. Представьте результаты в стандартизированной форме, округленные в соответствии с рекомендациями ГОСТа.

Вычислите ДГПСКО, определяемые двойным неравенством в левой части формулы (8).

Квантильные значения распределений Стьюдента и χ^2 можно найти в Приложениях II и III.

1.2. Сформируйте ряд “средней” длины ($n_{\text{ср.}} = 15 \dots 20$) и выполните те же расчеты, что и в п. 1.1.

Оцените, как изменяются доверительные границы погрешности с увеличением длины ряда.

2. Анализ выборочных параметров распределения.

Выполните статистический численный эксперимент, многократно формируя однотипные ряды длиной n_k . Для каждого ряда вычислите значения $\bar{x}_v, S_v^2(x), S_v(x)$, где v – номер ряда, $v = \overline{1, N}$, N – число сформированных рядов. Задайте N достаточно большим: $N \geq 10^4 \dots 10^5$. Для дальнейшего анализа одновременно и последовательно рассчитайте величины:

- $z_v = (\bar{x}_v - x_0) / \sigma_{\bar{x}}$, где $\sigma_{\bar{x}} = \sigma(x) / \sqrt{n_k}$, z_v – центрированная безразмерная величина, аналогичная аргументу стандартного нормального распределения;

– $t_v = (\bar{x}_v - x_0) / S_{\bar{x}_v}$, $S_{\bar{x}_v} = S_v(x) / \sqrt{n_k}$, где t_v – эмпирический аналог аргумента распределения Стьюдента;

$$\chi_v^2 = (n_k - 1) S_v^2(x) / \sigma^2(x),$$

где χ_v^2 – аналог аргумента теоретического χ^2 -распределения.

Одновременно с получением текущих значений z_v, t_v, χ_v^2 группируйте их по интервалам. Формально z_v и t_v могут изменяться от $-\infty$ до $+\infty$. Практически достаточно рассмотреть их повторяемость в диапазоне от -3 до $+3$. Задайте при группировке интервалы $h_z, h_t \leq 0,25$. Значения χ_v^2 теоретически могут изменяться от 0 до $+\infty$. Практически можно ограничиться диапазоном от 0 до $3n_k$ при $h_{\chi^2} \leq 0,2$. Для каждого из перечисленных параметров рассчитайте N_{jA} – число случаев попадания величины $A (A = z_v, t_v, \chi^2)$ в j -тый интервал, $j = \overline{1, k}$. Далее постройте графики повторяемости различных значений $N_{jA} / (Nh_A)$, относя их к середине j -того интервала. Для сравнения нанесите на эти же графики соответственно кривые теоретических плотностей распределения:

– стандартного нормального

$$\varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

– Стьюдента

$$f(t) = \frac{1}{\sqrt{\pi(n_k - 1)}} \cdot \frac{\Gamma\left(\frac{n_k}{2}\right)}{\Gamma\left(\frac{n_k - 1}{2}\right)} \cdot \left(1 + \frac{t^2}{n_k - 1}\right)^{-\frac{n_k}{2}},$$

– Пирсона

$$f(\chi^2) = \frac{1}{2^{n_k/2} \Gamma\left(\frac{n_k}{2}\right)} \cdot (\chi^2)^{\frac{n_k - 2}{2}} \cdot \exp\left(-\frac{\chi^2}{2}\right).$$

Проанализируйте полученные результаты. Оцените степень соответствия эмпирических распределений теоретическим.

Вариант Б. Выполнение работы без использования ПК.

1. Сформируйте короткий ряд измерений, используя для определения a_μ таблицу равномерно распределенных случайных чисел (см. Приложение V).

Выполните все операции, указанные в п. 1. варианта А.

2. Постройте графики.

2.1. Графики плотности вероятности распределения Стьюдента $f(t)$ для одного из малых значений $n (n_k = 2 \dots 5)$ и одного из средних ($n_{cp} = 10 \dots 15$). Формально $-\infty < t < \infty$, практически достаточно построить эту зависимость в интервале $[-3, 3]$. Рассчитайте по графику (проинтегрируйте численно) или по таблицам вероятность, с которой величина t находится в интервале $[-2, 2]$ – для обоих значений n_k и n_{cp} .

2.2. Графики плотности вероятности χ^2 -распределения при тех же n_k и n_{cp} , что и в п.2.1. По графику (или таблицам) найдите верхние $\chi_{1-\alpha/2, f}^2$ и нижние $\chi_{\alpha/2, f}^2$ квантильные значения при $\alpha = 0,10$.

3. Решите задачи.

3.1. Точность измерения величины L характеризуется ее СКО $\sigma(L) = 0,15$. По результатам 9 измерений вычислено $\bar{L} = 6,06$. В каком доверительном интервале находится истинное значение L_0 при $p = 0,95$?

АНАЛИЗ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Часть I. Графические методы анализа

Выполнив многократные измерения физической величины, получим ряд значений x_i , $i = \overline{1, n}$, n — длина ряда (объем выборки). Для дальнейшего использования результаты измерений следует представить в компактной форме. С этой целью находят наилучшие оценки среднего значения (результата измерений), характеристику рассеяния — дисперсию, а также моменты более высокого порядка (асимметрия, эксцесс). Далее следует оценить доверительные интервалы, в которых с заданной вероятностью содержатся истинные значения этих параметров. Чтобы выполнить такие оценки, необходимо знать закон распределения случайной величины. Обычно этот закон заранее неизвестен. Его можно определить, только устремив к бесконечности объем выборки. Разумеется, практически это нереально. Вместо этого строят выборочную функцию распределения и подбирают для нее какой-либо теоретический закон. Чаще всего при обработке результатов измерений в качестве такого закона используют нормальное распределение. Однако в каждом конкретном случае следует убедиться в том, что нормальное распределение является хорошей аппроксимацией результатов эксперимента.

В настоящем задании рассматриваются простые графические методы анализа эмпирических распределений и проводится сравнение этих распределений с нормальным. Наиболее распространенными формами представления результатов измерений являются гистограмма и эмпирическая функция распределения. Для построения гистограммы результаты измерений группируют по k интервалам, где

$$k \approx 1 + 3,2 \lg n. \quad (1)$$

Находят число элементов ряда n_j , $j = \overline{1, k}$, попадающих в j -тый интервал. Вычисляют центры интервалов \bar{x}_j . Откладывают \bar{x}_j по оси абсцисс, а соответствующее значение n_j/n по оси ординат. Через каждую точку $(\bar{x}_j, n_j/n)$ параллельно оси абсцисс проводят отрезки, равные ширине интервала. Концы отрезков вертикальными линиями соединяют с осью абсцисс. Такой график называется гистограммой.

Величина

$$F_j = \sum_{j=1}^j (n_j/n) \quad (2)$$

называется эмпирической функцией распределения. Значения F_j при соответствующих \bar{x}_j наносят на график в виде ступенчатой функции, подобно гистограмме.

Качественный анализ гистограммы и эмпирической функции распределения позволяет выбрать подходящее теоретическое распределение. По выборке вычисляются параметры такого распределения. Далее рассчитываются p_j — теоретическая вероятность попадания результатов измерений в каждый из интервалов и $F(\bar{x}_j)$ — теоретическая функция распределения.

$$p_j = \int_{\bar{x}_j - h/2}^{\bar{x}_j + h/2} f(x) dx \approx f(\bar{x}_j)h, \quad (3)$$

$$F(x) = \int_{-\infty}^{\bar{x}_j} f(x) dx = \sum_{j=1}^j p_j, \quad (4)$$

где ширина интервала $h = \bar{x}_{j+1} - \bar{x}_j$, $f(x)$ — плотность вероятности теоретического распределения. Для нормального распределения при достаточно большом n ($n > 30$)

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot S(x)} \exp \left[-\frac{(x_j - \bar{x})^2}{2S^2(x)} \right]. \quad (5)$$

Рассчитанные значения p_j и $F(\bar{x}_j)$ наносят соответственно на гистограмму и график эмпирической функции распределения, по точкам проводят плавные кривые. Теперь согласованность или различие между эмпирическим и теоретическим распределениями становятся наглядными.

Часто интерес представляют различия между эмпирическим и теоретическим распределениями при больших отклонениях от среднего. В этих случаях более выразительными являются графики, построенные на вероятностной сетке.

Вероятностной сеткой называют такое преобразование оси ординат, при котором теоретическая функция распределения изображается прямой линией. Если эмпирическая функция распределения соответствует предполагаемому теоретическому закону, то точки $(\bar{x}_j + h/2, F(\bar{x}_j))$ лягут на прямую линию. Существенная нелинейность зависимости свидетельствует о несоответствии эмпирического распределения предполагаемому теоретическому.

Техника построения вероятностных сеток изложена в учебнике [2].

Имитационные модели эксперимента. Для того чтобы заранее знать действительный закон выборочного распределения и его параметры, откажемся здесь от использования каких-либо реальных результатов измерений. Вместо этого будем формировать ряды, являющиеся выборками из генеральных совокупностей, которые соответствуют тому или иному теоретическому закону.

– Нормальное распределение. Нормально распределенную случайную величину x_{1i} , $i = \overline{1, n}$ будем получать с помощью генератора равномерно распределенных случайных чисел (RND)

$$x_{1i} = \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i \quad (6)$$

Здесь a_{μ} – равномерно распределенные случайные числа, $a_{\mu} \in [0, 1]$, $\mu = \overline{1, m_1}$, m_1 – число суммируемых чисел. Истинное значение $x_{01} = m_1/2$, $\sigma(x_1) = \sqrt{m_1/12}$, $A_{01} = 0$, $\mathcal{E}_{01} = 0$. Естественно, что выборочные значения \bar{x}_1 , $S(x_1)$, A_1 и \mathcal{E}_1 будут отличаться от истинных.

– Равномерное распределение. Равномерное распределение легко получается с помощью RND. По определению $x \equiv a$. Чтобы истинное значение и СКО соответствовали нормальному распределению (6), сформируем ряд следующим образом:

$$x_{2i} = (a_i - 0,5) \cdot \sqrt{m_1} + m_1/2, \quad (7)$$

При этом $x_{01} = x_{02}$, $\sigma(x_{01}) = \sigma(x_{02})$, $A_{01} = A_{02}$, $\mathcal{E}_{01} \neq \mathcal{E}_{02}$, $\mathcal{E}_{02} = -1,2$.

– Распределение Лапласа. Чтобы получить распределение Лапласа, снова воспользуемся RND. Элементы ряда этого распределения можно рассчитать по формуле

$$x_{3i} = m_1/2 + \sqrt{m_1/24} \cdot \ln(a'/a'')_i, \quad (8)$$

где a' и a'' – случайные равномерно распределенные числа, вырабатываемые RND на i – том шаге.

Для такого распределения $x_{03} = x_{01}$, $\sigma(x_3) = \sigma(x_1)$, $A_{03} = 0$, $\mathcal{E}_{03} = 3$.

– Асимметричное распределение. Сгенерируем распределение, обладающее ненулевыми коэффициентами асимметрии и эксцесса. Получим сначала ряд вспомогательных значений

$$\tilde{x}_i = \sqrt{(x'_{1i} - m_1/2)^2 + (x''_{1i} - m_1/2)^2}, \quad (9)$$

где x'_{1i} и x''_{1i} – нормально распределенные случайные числа, рассчитанные по формуле (6) с теми же значениями x_{01} и $\sigma(x_1)$. Величина \tilde{x} подчиняется распределению Рэлея. Характеристики этого распределения: $\tilde{x}_0 = 1,25\sigma(x_1)$, $\sigma(\tilde{x}) = 0,66\sigma(x_1)$, $\tilde{A}_0 = 0,63$, $\tilde{\mathcal{E}}_0 = 0,25$.

Чтобы “привязать” это распределение к трем предыдущим, следует центрировать его на математическое ожидание, нормировать на СКО и сместить на x_{01}

$$x_{4i} = \frac{\tilde{x}_i - 1,25\sigma(x_1)}{0,66\sigma(x_1)} + \frac{m_1}{2}. \quad (10)$$

При выполнении работы задайте $m_1 = 12$, в этом случае $x_{01} = 6$, $\sigma(x_1) = 1$.

Содержание работы

1. Используя формулу (6), сформируйте ряд нормально распределенной случайной величины длиной $n_1 = 50$. Сгруппируйте его по интервалам. Вычислите выборочные среднее \bar{x} и СКО $S(x)$. Для сгруппированных рядов можно воспользоваться формулами

$$\bar{x} = \sum_{j=1}^k \bar{x}_j n_j / n, \quad (11)$$

$$S(x) = \sqrt{\frac{1}{n-1} \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 - \frac{h^2}{12}}, \quad (12)$$

где k – число интервалов, h – ширина интервала, $h = \bar{x}_{j+1} - \bar{x}_j$. В общем случае число интервалов k следует находить по формуле (1). Однако при выполнении этой работы задайте k “с запасом” $k = 2m_1$ (реально число интервалов, в которые будут попадать элементы ряда, окажется существенно меньше), при этом $h = 0,5$.

По полученным данным постройте гистограмму и эмпирическую функцию распределения.

Рассчитайте и нанесите на эти графики соответственно теоретическую вероятность попадания результатов “измерений” в j -тый интервал p_j и теоретическую функцию распределения $F(\bar{x}_j)$.

Оцените качественно степень соответствия эмпирического и теоретического распределений.

2. Сформируйте новый ряд значений, используя модель равномерного распределения – формула (7). Разбейте этот ряд по интервалам. При $m_1 = 12$ величина x_{2i} изменяется в диапазоне от $(6 - \sqrt{3})$ до $(6 + \sqrt{3})$. Удобно разбить этот диапазон на семь интервалов ($k = 7$). Тогда ширина интервала $h = 0,5$ (точнее $h = 0,494$). При этом \bar{x}_j будут отличаться от соответствующих значений в п. 1. Постройте гистограмму и эмпирическую функцию распределения, вычислите значения \bar{x} и $S(x)$ для этого ряда. Рассчитайте новые значения p_j и $F(\bar{x}_j)$ и нанесите их на гистограмму и эмпирическую функцию распределения соответственно. Оцените качественно расхождение между эмпирическим и теоретическим (нормальным) распределениями.

3,4. Проведите аналогичный анализ рядов, сформированных по алгоритмам (8) и (10).

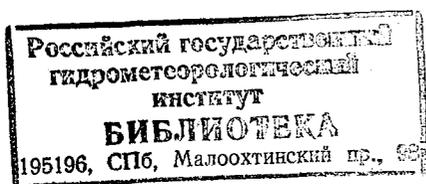
Сравните все результаты между собой. Обратите внимание на характерные особенности распределений, отличающихся от нормального по асимметрии и эксцессу.

5. Постройте вероятностную сетку нормального распределения. Нанесите на нее все четыре эмпирические функции распределения, предварительно умножив $F(\bar{x}_j)$ на коэффициент $n/(n+1)$, чтобы исключить возможность получения эмпирических функций распределения, равных единице.

Обратите внимание на то, как выглядят на вероятностной сетке эмпирические распределения, отличающиеся от нормального по асимметрии и эксцессу.

6. Повторите все операции для $n_2 = 2n_1$. В силу случайности формирования рядов с помощью RND, каждый исполнитель получит неповторяющиеся выборки, поэтому в работе не задаются варианты исходных данных.

Литература: [2], раздел 4.



АНАЛИЗ ЭМПИРИЧЕСКИХ РАСПРЕДЕЛЕНИЙ

Часть II. Проверка гипотезы о нормальности распределения результатов измерений

Существуют различные методы проверки гипотезы о соответствии эмпирического распределения случайной величины предполагаемому теоретическому закону. Ограничимся здесь только проверкой гипотезы о нормальном распределении. Рассмотрим методы, рекомендованные ГОСТом.

1. При длине ряда $13 \dots 15 < n < 50$ рекомендуется использовать составной критерий "I + II".

Критерий I. Вычисляется величина

$$\tilde{d} = \frac{\sum_{i=1}^n |x_i - \bar{x}|}{\sqrt{n(n-1)} \cdot S(x)}. \quad (1)$$

Выбор гипотезы осуществляется в соответствии с решающим правилом

$$d_{\alpha_I/2,n} \leq \tilde{d} \leq d_{1-\alpha_I/2,n} \Rightarrow H_0, \quad \tilde{d} < d_{\alpha_I/2,n}; \text{ или } \tilde{d} > d_{1-\alpha_I/2,n} \Rightarrow H_1, \quad (2)$$

где H_0 – нулевая гипотеза: распределение нормальное; H_1 – альтернативная гипотеза: распределение не соответствует нормальному; $d_{\alpha_I/2,n}$ и $d_{1-\alpha_I/2,n}$ – критические значения d , определяемые уровнем значимости α_I и длиной ряда n . Значения $d_{\alpha_I/2,n}$ и $d_{1-\alpha_I/2,n}$ приведены в табл. 1. Их также можно вычислить по формуле, аппроксимирующей эту таблицу

$$d_{\alpha_I/2,n}, d_{1-\alpha_I/2,n} = 0,8 \left(1 \mp \frac{1}{b_{\alpha_I} \sqrt{n}} \right). \quad (3)$$

При $\alpha_I = 0,10$ $b_{0,10} = 2,4$, при $\alpha_I = 0,02$ $b_{0,02} = 1,2$.

Критерий II. Находится число \tilde{m} разностей $y_i = |x_i - \bar{x}| - z_{(1+p)/2} \cdot S(x)$, больших нуля, т.е.

$$\tilde{m} = \sum_{i=1}^n \theta(y_i), \text{ где } \theta(y_i) = \begin{cases} 1, & y_i > 0, \\ 0, & y_i \leq 0, \quad i = \overline{1, n}. \end{cases} \quad (4)$$

Значение \tilde{m} сравнивается с критическим $m_{\alpha_{II}/2,n}$ (см. табл. 2). Решающее правило выбора гипотез

$$\tilde{m} \leq m_{\alpha_{II}/2,n} \Rightarrow H_0; \quad \tilde{m} > m_{\alpha_{II}/2,n} \Rightarrow H_1. \quad (5)$$

Эмпирическое распределение соответствует нормальному, если оба критерия дают нулевую гипотезу. Если хотя бы один критерий указывает на альтернативную гипотезу, то распределение не может считаться нормальным.

2. При длине ряда $n \geq 50$ используется критерий согласия χ^2 . Рассчитывается величина

$$\chi^2 = \sum_{j=1}^k \frac{(n_j - np_j)^2}{np_j}. \quad (6)$$

Для вычисления χ^2 ряд измерений группируется по интервалам. Находятся n_j – число элементов ряда, попавших в j -тый интервал, и теоретическая вероятность попадания результатов в этот интервал

$$p_j = \Phi\left(\bar{z}_j + \frac{h_z}{2}\right) - \Phi\left(\bar{z}_j - \frac{h_z}{2}\right) \quad (7)$$

или проще, но грубее

$$p_j = \varphi(\bar{z}_j) \cdot h_z, \quad (8)$$

где $\bar{z}_j = (\bar{x}_j - \bar{x})/S(x)$, $h_z = h/S(x)$, $h = \bar{x}_{j+1} - \bar{x}_j$, $\varphi(\bar{z}_j) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{\bar{z}_j^2}{2}\right)$, $\Phi(z)$ – функция Лапласа.

Значение χ^2 сравнивается с односторонним (правосторонним) критическим значением $\chi_{1-\alpha, f}^2$, где число степеней свободы $f = k - 3$, k – число интервалов. Если в каких-либо интервалах $n_j < 5$, то эти интервалы объединяют с соседними. Соответственно число k уменьшается.

Решающее правило:

$$\chi^2 \leq \chi_{1-\alpha, f}^2 \Rightarrow H_0; \chi^2 > \chi_{1-\alpha, f}^2 \Rightarrow H_1. \quad (9)$$

Для расчета $\chi_{1-\alpha, f}^2$ при $f \leq 30$ можно использовать формулы:

$$\chi_{1-0,05; f}^2 = 3,8 + 2,2(f-1)^{0,86}, \quad (10a)$$

$$\chi_{1-0,10; f}^2 = 2,7 + 1,9(f-1)^{0,89}. \quad (10b)$$

3. При $n > 50$ ГОСТ отдает предпочтение критерию согласия ω^2 .

Чтобы найти ω^2 , неупорядоченный ряд x_i , $i = \overline{1, n}$ дополняют значениями $x_0 = -\infty$ и $x_{n+1} = +\infty$ и перестраивают в возрастающий вариационный ряд x_l , $l = \overline{0, n+1}$, $x_l \leq x_{l+1}$. Вычисляется значение

$$\omega^2 n = \sum_{l=1}^n \left(\Phi(z_l) + \frac{n-2l+1}{2n} \right)^2, \quad (11)$$

где $z_l = (x_l - \bar{x})/S(x)$. Критические значения $(\omega^2 n)_{1-\alpha}$ зависят только от уровня значимости:

α	0,10	0,05	0,01
$(\omega^2 n)_{1-\alpha}$	0,35	0,46	0,74

Решающее правило:

$$\omega^2 n \leq (\omega^2 n)_{1-\alpha} \Rightarrow H_0; \omega^2 n > (\omega^2 n)_{1-\alpha} \Rightarrow H_1. \quad (12)$$

Целью настоящей работы является оценка согласованности критериев "I + II", χ^2 и ω^2 между собой. При объеме выборки около 50 критерии считаются одинаково пригодными. Воспользуемся методами статистического моделирования. Будем моделировать ряды, представляющие собой выборки из генеральных совокупностей, имеющих определенно нормальное распределение и распределение с очевидными асимметрией и эксцессом.

Воспользуемся теми же имитационными моделями эксперимента, которые приведены в задании 5. С помощью RND могут быть сформированы выборки из генеральных совокупностей, имеющих следующие распределения:

– нормальное

$$x_{li} = \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i, \quad (13)$$

– равномерное

$$x_{2i} = (a_i - 0,5) \cdot \sqrt{m_1} + m_1 / 2, \quad (14)$$

– Лапласа

$$x_{3i} = m_1 / 2 + \sqrt{m_1 / 24} \cdot \ln(a'_i / a''_i), \quad (15)$$

– Рэлея (смещенное)

$$x_{4i} = \frac{\sqrt{(x'_{1i} - m_1 / 2)^2 + (x''_{1i} - m_1 / 2)^2} - 1,25\sigma(x_1)}{0,66\sigma(x_1)} + \frac{m_1}{2}. \quad (16)$$

Здесь a_μ , $\mu = \overline{1, m_1}$, a_i , $i = \overline{1, n}$, a'_i , a''_i – вырабатываемые RND равномерно распределенные случайные числа, $0 \leq a \leq 1$, m_1 – число суммируемых a_μ .

Далее в работе рекомендуется использовать $m_1 = 12$, $n = 50$.

Содержание работы

1. Смоделируйте по формуле (13) ряд x_{1i} , $i = \overline{1, n}$. Такой ряд является выборкой из нормальной генеральной совокупности. Для последующих расчетов создайте файл, содержащий x_i . Рассчитайте значения \bar{x}_1 и $S(x_1)$.

1.1. Рассчитайте \tilde{d} и \tilde{m} для критерия "I + II" и сравните их с критическими $d_{\alpha_1/2, n}$ или $d_{1-\alpha_1/2, n}$ и $m_{\alpha_2/2, n}$ при $\alpha_I / 2 \approx \alpha_{II} / 2 \approx 0,05$. Распределение считается нормальным, если H_0 принимается по обоим критериям – I и II.

1.2. Рассчитайте χ^2 . Для этого сгруппируйте ряд по $2m_1$ интервалам ($h = 0,5$), найдите n_j , $j = \overline{1, 2m_1}$. Крайние интервалы объедините таким образом, чтобы в них было не менее 5 элементов. Определите окончательное число интервалов k . Рассчитайте p_j (см. формулы (7) или (8)). Найденное значение χ^2 сравните с критическим $\chi^2_{1-\alpha, f}$, здесь $f = k - 3$, $\alpha = 0,05$. Выберите соответствующую гипотезу.

1.3. Составьте вариационный ряд и вычислите $\omega^2 n$ по формуле (11). Сравните его с критическим значением $(\omega^2 n)_{1-\alpha}$, $\alpha = 0,05$. Поскольку исследуемый ряд является выборкой из нормально распределенной генеральной совокупности, то с вероятностью $(1 - \alpha)$ нулевая гипотеза должна быть подтверждена всеми тремя критериями. Если какой-либо из этих критериев отвергает нулевую гипотезу, то определите значение α , при котором нулевая гипотеза может быть принята.

2. Смоделируйте ряд, определенно отличающийся от нормального по асимметрии и эксцессу. С этой целью воспользуйтесь одной из имитационных формул (14), (15), (16) по вашему выбору.

Проверьте полученный ряд на соответствие (несоответствие) нормальному распределению. Найдите для этого ряда среднее значение и СКО и выполните снова все операции по п.п. 1.1, 1.2, 1.3.

Проанализируйте полученные результаты.

Приложение к заданию

Таблица 1

Критические значения $d_{\alpha_1/2, n}$ и $d_{1-\alpha_1/2, n}$ при различных n и α

$\alpha_1 / 2,$ $1 - \alpha_1 / 2$	n									
	11	16	21	26	31	36	41	46	51	61
0,01	0,67	0,68	0,69	0,70	0,71	0,72	0,72	0,72	0,73	0,73
0,05	0,72	0,72	0,73	0,74	0,74	0,74	0,75	0,75	0,75	0,76
0,95	0,91	0,89	0,88	0,87	0,86	0,86	0,85	0,85	0,85	0,84
0,99	0,94	0,91	0,90	0,89	0,88	0,88	0,88	0,87	0,86	0,86

Критические значения m и вероятность P при различных объемах
выборки n и уровнях значимости α_{II}

$\alpha_{II} / 2$	n								
	10	11-14	15-20	21-22	23	24-27	28-32	33-35	36-49
	$m_{\alpha_{II}, n}$								
	1	1	1	2	2	2	2	2	2
0,01	0,98	0,99	0,99	0,98	0,98	0,98	0,99	0,99	0,99
0,02	0,98	0,98	0,99	0,97	0,98	0,98	0,98	0,98	0,99
0,05	0,96	0,97	0,98	0,96	0,97	0,97	0,97	0,98	0,98

Для удобства пользования табл. 2 приведем здесь переход от P к $z_{(1+P)/2}$

P	0,96	0,97	0,98	0,99
$z_{(1+P)/2}$	2,06	2,17	2,33	2,58

Литература: [2], раздел 4, [13].

ПОГРЕШНОСТИ КОСВЕННЫХ ИЗМЕРЕНИЙ

Задача заключается в том, чтобы найти среднее значение и доверительные границы погрешности измерения случайной величины y , функционально связанной с факторами $x_j, j = \overline{1, k}$ (k – число факторов). При этом x_j являются результатами прямых измерений

$$y = y(x_j, j = \overline{1, k}). \quad (1)$$

На практике используются два пути решения этой задачи: метод линеаризации и метод приведения.

1. Классический метод линеаризации заключается в том, что исходная функция (1) заменяется ее линейным разложением в ряд Тейлора

$$y = \bar{y} + \sum_{j=1}^k \frac{\partial y}{\partial \bar{x}_j} (x_j - \bar{x}_j). \quad (2)$$

СКО результата косвенного измерения вычисляется по формуле

$$S_{\bar{y}}^{(n)} = \sqrt{\sum_{j=1}^k \sum_{u=1}^k r_{x_j x_u} \frac{\partial y}{\partial \bar{x}_j} \frac{\partial y}{\partial \bar{x}_u} S_{\bar{x}_j} S_{\bar{x}_u}}, \quad (3)$$

где $r_{x_j x_u}$ – парный коэффициент корреляции между x_j и x_u , $S_{\bar{x}_j} = S(x_j) / \sqrt{n_j}$, $S(x_j)$ – СКО, $j, u = \overline{1, k}$.

Значимость коэффициентов корреляции проверяется при сравнении их с критическими значениями

$$r_{1-\alpha/2, f} = \left(\frac{n-2}{t_{1-\alpha/2, f}^2} + 1 \right)^{-0,5}, \quad f = n-2. \quad (4)$$

Решающее правило

$$\left| r_{x_j x_u} \right| \leq r_{1-\alpha/2, f} \Rightarrow H_0, \quad \left| r_{x_j x_u} \right| > r_{1-\alpha/2, f} \Rightarrow H_1. \quad (5)$$

Если принимается нулевая гипотеза, то коэффициент корреляции задается равным нулю.

Для расчета доверительных границ погрешности результата косвенных измерений рекомендуется использовать формулу

$$\varepsilon_{\bar{y}}^{(n)} = \sqrt{\sum_{j=1}^k \sum_{u=1}^k r_{x_j x_u} \frac{\partial y}{\partial \bar{x}_j} \frac{\partial y}{\partial \bar{x}_u} \varepsilon_{\bar{x}_j} \cdot \varepsilon_{\bar{x}_u}}, \quad (6)$$

где

$$\varepsilon_{\bar{x}_j} = t_{1-\alpha/2, f_j} \cdot S_{\bar{x}_j}. \quad (7)$$

Здесь $f_j = n_j - 1$, n_j – длина j -го ряда.

2. Метод приведения заключается в том, что по результатам совместных измерений в каждом i -том испытании рассчитывается величина y_i

$$y_i = y(x_{ji}, j = \overline{1, k}, i = \overline{1, n}). \quad (8)$$

Таким образом, формируется ряд измерений y_i , обработку которого выполняют обычным методом:

$$\bar{y}^{(n)} = \left(\sum_{i=1}^n y_i \right) / n, \quad (9)$$

$$S^{(n)}(y) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y}^{(n)})^2}, \quad (10)$$

$$S_{\bar{y}}^{(n)} = \frac{S^{(n)}(y)}{\sqrt{n}}, \quad (11)$$

$$\varepsilon_{\bar{y}}^{(n)} = t_{1-\alpha/2, f} \cdot S^{(n)}(y) / \sqrt{n}. \quad (12)$$

Здесь, как обычно, $f = n - 1$.

Недостатком метода приведения является то, что при неравных длинах рядов ($n_j \neq n_u$) длина ряда y_i будет ограничена длиной самого короткого из рядов x_j . Остальная информация при обработке теряется.

Имитационные модели. Для получения "рядов измерений" нормально распределенных величин x_j используем, как и ранее, RND

$$x_{ji} = b_j \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i, \quad \mu = \overline{1, m_1}, \quad i = \overline{1, n_j}. \quad (13)$$

где a_{μ} – равномерно распределенные случайные числа, вырабатываемые RND ($0 \leq a_{\mu} \leq 1$), b_j – масштабный множитель. Полученные таким образом ряды измерений являются некоррелированными.

Чтобы смоделировать взаимно коррелированные ряды, можно воспользоваться алгоритмом

$$x_{ui} = b_u \left[x_{ji} \pm b_j \left(\sum_{\mu=m_1+1}^{m_1+m_2} a_{\mu} \right)_i \right], \quad j, u = \overline{1, k}. \quad (14)$$

Коэффициент корреляции между рядами x_j и x_u определяется соотношением между m_1 и m_2 . При малых m_2 получим ряды с высокой корреляцией. Увеличивая m_2 , можно получить ряды с незначимой корреляцией. Ряды x_{ji} и x_{ui} удобно формировать одновременно.

Ряды, распределения которых отличаются от нормального, можно получить, например, по формуле

$$x_{ji} = b_j \left[\frac{\sqrt{(x_{1i} - m_1/2)^2 + (x_{2i} - m_1/2)^2} - 1,25\sqrt{m_1/12}}{0,66 \cdot \sqrt{m_1/12}} + \frac{m_1}{2} \right], \quad (15)$$

где x_{1i} и x_{2i} находятся по формуле

$$x_{1(2)i} = \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i.$$

Содержание работы

1. Докажите, что для линейной зависимости $y = x_1 - x_2$ (либо $y = x_1 + x_2$), при $n_1 = n_2 = n$ СКО, найденные методом линеаризации (по формуле (3)) и методом приведения (по формуле (11)), равны друг другу.

2. Для функции, заданной в таблице вариантов исходных данных, получите выражение для расчета $\varepsilon_{\bar{y}}^{(n)}$ методом линеаризации.

3. Смоделируйте “ряды измерений” по формуле (13), используя значения m_1 , b_j и n_j , указанные в таблице вариантов исходных данных.

3.1. Вычислите $\bar{y}^{(n)}$ и ДГПРИ $\varepsilon_{\bar{y}}^{(n)}$.

3.2. Рассчитайте ряд y_i методом приведения. Найдите $\bar{y}^{(n)}$ и $\varepsilon_{\bar{y}}^{(n)}$. Сравните между собой результаты, полученные разными методами.

4. Используя формулы (13) и (14), смоделируйте коррелированные ряды измерений. Выполните операции, аналогичные п.п. 3.1, 3.2. Оцените, как изменяются ДГПРИ при наличии значимой корреляции между факторами.

5. Смоделируйте ряды измерений, распределения которых отличаются от нормального. Используйте для этого формулу (15). Выполните операции, аналогичные п.п. 3.1, 3.2. Обратите внимание на то, как изменяются значения $\bar{y}^{(n)}$, $\bar{y}^{(n)}$, $\varepsilon_{\bar{y}}^{(n)}$ и $\varepsilon_{\bar{y}}^{(n)}$ в зависимости от вида распределений рядов измерений.

Таблица 1

Варианты исходных данных

Вариант	Функция	$b_{j(u)}$	m_1	m_2	n_j
1	$y = x_1^2 \cdot x_2$	$b_1 = 1$ $b_2 = 0,5$	12	3	$n_1 = 50$ $n_2 = 60$
2	$y = x_1^2 + x_2^3$	$b_1 = 0,5$ $b_2 = 0,1$	12	4	$n_1 = 50$ $n_2 = 70$
3	$y = x_1 / x_2^2$	$b_1 = 1$ $b_2 = 0,5$	12	4	$n_1 = 60$ $n_2 = 50$
4	$y = x_1 \cdot x_2^3$	$b_1 = 1$ $b_2 = 0,5$	12	5	$n_1 = 70$ $n_2 = 50$
5	$y = x_1 \cdot x_2 \cdot x_3$	$b_1 = 1$ $b_2 = 1$ $b_3 = 0,5$	10	2	$n_1 = n_2 = n_3 = 50$
6	$y = x_1 + x_2^2 \cdot x_3^2$	$b_1 = 1$ $b_2 = 0,5$ $b_3 = 0,5$	24	6	$n_1 = n_2 = 50$ $n_3 = 70$

Литература: [2], раздел 6, [4], [6], [13].

СУММИРОВАНИЕ СИСТЕМАТИЧЕСКИХ И СЛУЧАЙНЫХ ПОГРЕШНОСТЕЙ

Суммарная систематическая погрешность θ вычисляется как композиция отдельных систематических погрешностей $\theta_\mu, \mu = \overline{1, m}$, m – число погрешностей, определяемых классом точности средств измерений, условиями квантования или кодирования и т.п., по формуле:

$$\theta = \bar{k}_\alpha \sqrt{\sum_{\mu=1}^m \theta_\mu^2}. \quad (1)$$

Коэффициент \bar{k}_α зависит от уровня значимости α :

α	0,10	0,05	0,01
\bar{k}_α	0,95	1,1	1,4

Если рассматривается только одна систематическая погрешность θ_1 ($m = 1$), то

$$\theta \approx (1 - \alpha/2)\theta_1. \quad (1a)$$

При малых значениях α практически $\theta \cong \theta_1$.

Доверительные границы суммарной (систематические + случайная) погрешности результата прямых равноточных измерений рекомендуется вычислять по формулам:

$$\Delta\alpha = \frac{\varepsilon_{\bar{x}} + \theta}{S_{\bar{x}} + \sqrt{\theta^2/3}} \sqrt{S_{\bar{x}}^2 + \theta^2/3} \quad (2)$$

либо

$$\Delta'\alpha = K'_\alpha(\varepsilon_{\bar{x}} + \theta). \quad (3)$$

Здесь $\varepsilon_{\bar{x}} = t_{1-\alpha/2, f} S_{\bar{x}}$ – доверительные границы случайной составляющей погрешности результата измерений, $f = n - 1$, n – объем выборки, K'_α – коэффициент, зависящий от уровня значимости α и соотношения между $S_{\bar{x}}$ и θ .

Значения K'_α при различных $\theta/S_{\bar{x}}$ и α

$\theta/S_{\bar{x}}$	0,5	0,75	0,8	1	2	3	4	5	6	7	8
$K'_{0,05}$	0,81	0,77	0,76	0,74	0,71	0,73	0,76	0,78	0,79	0,80	0,85
$K'_{0,01}$	0,87	0,85	0,84	0,82	0,80	0,81	0,82	0,83	0,83	0,84	0,85

Следуя общему принципу суммирования доверительных границ погрешности, можно использовать выражение

$$\tilde{\Delta}_\alpha = \sqrt{\varepsilon_{\bar{x}}^2 + \theta^2}. \quad (4)$$

При вычислении доверительных границ результата косвенных измерений величины y , определяемой функцией

$$y = y(x_j, j = \overline{1, k}), \quad (5)$$

сначала следует найти $\varepsilon_{\bar{y}}$ и $\theta(y)$

$$\varepsilon_{\bar{y}} = \left(\sum_{j=1}^k \sum_{u=1}^k r_{x_j x_u} \frac{\partial y}{\partial \bar{x}_j} \frac{\partial y}{\partial \bar{x}_u} \varepsilon_{\bar{x}_j} \varepsilon_{\bar{x}_u} \right)^{0,5} \quad (6)$$

Полагая, что корреляция между систематическими погрешностями отсутствует, общую систематическую погрешность при косвенных измерениях найдем как

$$\theta(y) = \sqrt{\sum_{j=1}^k \left(\frac{\partial y}{\partial \bar{x}_j} \right)^2} \theta_j, \quad (7)$$

где θ_j определяется формулой (1) для j -ого аргумента. Заменяя в формулах (2), (3), (4) $\varepsilon_{\bar{x}}$ на $\varepsilon_{\bar{y}}$ и θ на $\theta(y)$, легко можно найти Δ_{α} , Δ'_{α} или $\tilde{\Delta}_{\alpha}$.

Целью работы является изучение методики и приобретение навыков суммирования систематических и случайных погрешностей.

Содержание работы

1. Сравните между собой формулы (2), (3), (4). С этой целью для числа степеней $f = n - 1$ (значения n_1 заданы в таблице вариантов) рассчитайте Δ_{α} , Δ'_{α} и $\tilde{\Delta}_{\alpha}$ при различных отношениях $\theta/S_{\bar{x}}$, $\alpha = 0,05$. Для удобства вычислений задайте $S_{\bar{x}} = 1$. Результаты расчетов представьте в виде графиков.

ГОСТ рекомендует использовать формулу (2) при $0,8 < \theta/S_{\bar{x}} < 8$, предлагая считать, что при $\theta/S_{\bar{x}} < 0,8$ $\Delta_{\alpha} \cong \varepsilon_{\bar{x}}$, а при $\theta/S_{\bar{x}} > 8$ $\Delta_{\alpha} \cong \theta$. При каких значениях $\theta/S_{\bar{x}}$ эти приближения можно принять для формул (3), (4)?

2. Решите задачи.

2.1. Измерение температуры осуществляется резисторным термометром со шкалой от -40°C до $+60^{\circ}\text{C}$, класс точности прибора 1,0. Случайная составляющая погрешности характеризуется СКО $\sigma(t) = 2,5^{\circ}\text{C}$. Найдите то число измерений, при превышении которого точность измерений практически не повышается (доверительные границы погрешности перестают уменьшаться), $\alpha = 0,05$. Расчеты выполните по каждой из формул (2), (3), (4).

2.2. На некотором участке электрической цепи выполняются измерения силы тока I и разности потенциалов U . Используются миллиамперметр класса точности γ со шкалой 0...100 мА и вольтметр класса точности δ со шкалой 0...10,0 В (значения γ и δ заданы в таблице вариантов).

Получены следующие ряды совместных измерений

i	1	2	3	4	5	6	7	8	9	10
I , мА	58,5	61,0	59,8	60,2	61,2	58,8	59,1	60,4	59,2	62,0
U , В	7,12	7,22	7,18	7,09	7,24	7,10	7,15	7,14	-	-

Найдите как результаты косвенных измерений средние значения и доверительные границы погрешности а) сопротивления \bar{R} и б) выделяющейся на нем мощности \bar{W} при $\alpha = 0,05$.

Расчеты выполните по той из формул (2), (3), (4), которая вам показалась наиболее репрезентативной. Обоснуйте свой выбор.

Варианты исходных данных

варианты	n_1	γ	δ
1	6	1,0	2,0

2	6	0,5	2,0
3	10	1,0	1,0
4	10	0,5	1,0
5	15	1,0	1,5
6	15	0,5	1,5

Литература: [2], раздел 6, [13], [14].

ОЦЕНКА ОДНОРОДНОСТИ И РАВНОТОЧНОСТИ РЯДОВ ИЗМЕРЕНИЙ

Часть 1. Параметрические методы

Часто возникает задача сравнения двух или более рядов измерений одной и той же физической величины. Пусть, например, имеются k рядов измерений некоторой случайной величины $x_{ji}, j = \overline{1, k}, i = \overline{1, n_j}, n_j$ — длина j -того ряда. Вычисленные по ограниченным выборкам значения \bar{x}_j , так же как и $S^2(x_j)$, окажутся неравными друг другу. Являются ли эти различия случайными, незначимыми, либо отражают влияние каких-то факторов? Такие задачи решаются методами теории проверки статистических гипотез. Различают два вида методов: параметрические и непараметрические. В этом задании рассматриваются параметрические методы.

1. Сравнение дисперсий

Если анализируются выборочные дисперсии $S^2(x_j)$, далее, для краткости, просто $S_j^2, j = \overline{1, k}$, то проверяется нулевая гипотеза (H_0) об однородности генеральных дисперсий σ_j^2 против альтернативной (H_1) о их значимом различии (о неравноточности измерений):

$$H_0 : \sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2; \quad H_1 : \sigma_1^2 \neq \sigma_2^2 \neq \dots \neq \sigma_k^2. \quad (1)$$

1.1. Сравнение двух дисперсий ($k = 2$) проводят по критерию Фишера: рассчитывается отношение $F = S_1^2 / S_2^2$, которое затем сравнивается с критическим (табличным) значением $F_{1-\alpha, f_1, f_2}$, где $f_1 = f_{S_1^2} = n_1 - 1$ и $f_2 = f_{S_2^2} = n_2 - 1$, при этом индекс 1 присваивается большей из дисперсий. Решающее правило

$$F \leq F_{1-\alpha, f_1, f_2} \Rightarrow H_0; \quad F > F_{1-\alpha, f_1, f_2} \Rightarrow H_1. \quad (2)$$

Значения $F_{1-\alpha, f_1, f_2}$ можно найти в Приложении IV. При $\alpha = 0,05$ для $f_1, f_2 > 3$ табличное значение аппроксимируется приближенной формулой

$$F_{1-\alpha, f_1, f_2} = \frac{f_1 + 20,5}{f_1 + 6,0} + \frac{9,3}{f_2 - 1,7}. \quad (3)$$

1.2. Если число дисперсий больше двух ($k > 2$), а длины рядов разные ($n_j \neq const$), то используют критерий Бартлетта. Рассчитывается отношение

$$B/C = \left(f \ln S_0^2 - \sum_{j=1}^k f_j \ln S_j^2 \right) / \left[1 + \frac{1}{3(k-1)} \left(\sum_{j=1}^k \frac{1}{f_j} - \frac{1}{f} \right) \right], \quad (4)$$

где $f_j = n_j - 1, f = \sum_{j=1}^k f_j, S_0^2 = \frac{1}{f} \sum_{j=1}^k f_j S_j^2$. Поскольку отношение B/C имеет распределение близкое к χ^2 -распределению, то его сравнивают с $\chi_{1-\alpha, f}^2, f = k - 1$, выбирая гипотезу в соответствии с решающим правилом:

$$B/C \leq \chi_{1-\alpha, f}^2 \Rightarrow H_0; \quad B/C > \chi_{1-\alpha, f}^2 \Rightarrow H_1. \quad (5)$$

Значения $\chi^2_{1-\alpha, f}$ приведены в Приложении III. Их также можно рассчитать по формулам:

$$\chi^2_{1-0,05; f} = 3,8 + 2,2(f-1)^{0,86}, \quad \chi^2_{1-0,10; f} = 2,7 + 1,9(f-1)^{0,89}. \quad (6)$$

1.3. Если $k > 2$ и $f_j = const$, то рекомендуется использовать критерий Кочрена

$$G = S^2_{\max} / \left(\sum_{j=1}^k S_j^2 \right). \quad (7)$$

Критические значения $G_{1-\alpha, f, k}$ можно найти в справочниках.

2. Сравнение средних

Гипотезы формулируются для истинных значений x_{0j}

$$H_0 : x_{01} = x_{02} = \dots = x_{0k}; \quad H_1 : x_{01} \neq x_{02} \neq \dots \neq x_{0k}. \quad (8)$$

2.1. Сравнение двух средних ($k = 2$). Пусть имеются два ряда измерений, для которых определены $\bar{x}_1, \bar{x}_2, S_1^2, S_2^2, n_1$ и n_2 .

2.1.1. Наиболее распространенным является t -критерий (Стьюдента). Предварительно следует проверить гипотезу об однородности дисперсий (по критерию Фишера – см. п. 1). Если дисперсии однородны, то для односторонней критической области при

$$|\bar{x}_1 - \bar{x}_2| \leq t_{1-\alpha, f} \sqrt{\frac{S_1^2(n_1-1) + S_2^2(n_2-1)}{n_1+n_2-2} \cdot \frac{n_1+n_2}{n_1n_2}} \quad (9)$$

принимается нулевая гипотеза. В противном случае – альтернативная. Здесь $f = f_1 + f_2 = n_1 + n_2 - 2$.

Если дисперсии неоднородны, решающее правило имеет вид:

$$|\bar{x}_1 - \bar{x}_2| \leq T \Rightarrow H_0; \quad |\bar{x}_1 - \bar{x}_2| > T \Rightarrow H_1, \quad (10)$$

где

$$T = \frac{t_{1-\alpha, f_1} S_{\bar{x}_1}^2 + t_{1-\alpha, f_2} S_{\bar{x}_2}^2}{\sqrt{S_{\bar{x}_1}^2 + S_{\bar{x}_2}^2}}, \quad (11)$$

$$S_{\bar{x}_1}^2 = S_1^2 / \sqrt{n_1} \quad \text{и} \quad S_{\bar{x}_2}^2 = S_2^2 / \sqrt{n_2}.$$

2.1.2. Более общим является универсальный критерий UT , в соответствии с которым рассчитывается величина

$$\tilde{\varepsilon}_{\bar{y}} = \sqrt{\left(t_{1-\alpha, f_1} S_{\bar{x}_1} \right)^2 - 2r_{x_1 x_2} t_{1-\alpha, f_1} t_{1-\alpha, f_2} S_{\bar{x}_1} S_{\bar{x}_2} + \left(t_{1-\alpha, f_2} S_{\bar{x}_2} \right)^2}, \quad (12)$$

где $\bar{y} = |\bar{x}_1 - \bar{x}_2|$. Парный коэффициент корреляции рассчитывается по формуле

$$r_{x_j x_u} = \left[\frac{1}{n_{\min} - 1} \sum_{i=1}^{n_{\min}} (x_{ji} - \bar{x}_j)(x_{ui} - \bar{x}_u) \right] / \left(\overset{\circ}{S}_j \overset{\circ}{S}_u \right), \quad (13)$$

где $n_{\min} = \min\{n_j, n_u\}$ – длина более короткого ряда, $\overset{\circ}{S}_j, \overset{\circ}{S}_u$ – СКО, рассчитанные для рядов длиной n_{\min} . Вычисленные значения $r_{x_j x_u}$ следует проверить на значимость.

Решающее правило:

$$|\bar{x}_1 - \bar{x}_2| \leq \tilde{\epsilon}_{\bar{y}} \Rightarrow H_0; |\bar{x}_1 - \bar{x}_2| > \tilde{\epsilon}_{\bar{y}} \Rightarrow H_1. \quad (14)$$

2.2. Сравнение нескольких ($k > 2$) средних.

Если дисперсии однородны, то используют метод однофакторного дисперсионного анализа. Рассчитываются $S_0^2 = \frac{1}{f} \sum_{j=1}^k f_j S_j^2$ – дисперсия воспроизводимости (как и в п. 1.2) и межгрупповая дисперсия

$$S_M^2 = \frac{1}{k-1} \sum_{j=1}^k n_j (\bar{x}_j - \bar{\bar{x}})^2, \quad (15)$$

как мера рассеяния средних по группам относительно общего среднего $\bar{\bar{x}} = \frac{1}{n} \sum_{j=1}^k n_j \bar{x}_j$.

Теперь процедура проверки гипотезы о равенстве средних сводится к проверке гипотезы об однородности дисперсий S_M^2 и S_0^2 . С этой целью находят отношение Фишера

$F = S_M^2 / S_0^2$, которое сравнивают с критическим $F_{1-\alpha, f_{S_M^2}, f_{S_0^2}}$, числа степеней свободы $f_{S_M^2} = k - 1$,

$f_{S_0^2} = f$. Если $S_M^2 < S_0^2$, то однозначно принимается нулевая гипотеза.

Имитационные модели. Приведенные выше критерии предназначены для сравнения рядов измерений нормально распределенных случайных величин. Такие ряды легко могут быть смоделированы с помощью RND.

Модельные некоррелированные ряды измерений x_{ji} или x_{ui} , $j, i = \overline{1, k}$, $j \neq u$, $i = \overline{1, n_{j(u)}}$, $n_{j(u)}$ – длина j – того (u – того) ряда, могут быть получены с помощью алгоритма

$$x_{j(u)i} = \left(\sum_{\mu=1}^{m_{j(u)}} a_{\mu} \right)_i, \quad (16)$$

где m_j – число суммируемых случайных равномерно распределенных чисел a_{μ} ($0 \leq a_{\mu} \leq 1$). Генеральные средние x_{0j} и дисперсии σ_j^2 определяются значениями m_j : $x_{0j} = m_j / 2$, $\sigma_j^2 = m_j / 12$.

Последовательно полученные ряды x_{ji} при одинаковых m_j характеризуются одинаковыми x_{0j} и σ_j^2 , но, естественно, разными выборочными средними \bar{x}_j и дисперсиями S_j^2 .

Для того чтобы получить два ряда x_{ji} и x_{ui} , имеющие одинаковые генеральные средние $x_{0j} = x_{0u}$ и разные дисперсии $\sigma_j^2 \neq \sigma_u^2$, достаточно воспользоваться алгоритмом

$$x_{ui} = \left(\sum_{\mu=1}^{m_j} a_{\mu} + \sum_{\mu=m_j+1}^{(m_j+m_u)/2} a_{\mu} - \sum_{\mu=(m_j+m_u)/2+1}^{m_u} a_{\mu} \right)_i, \quad (17a)$$

при $m_u > m_j$, ($\sigma_u^2 > \sigma_j^2$) или

$$x_{ji} = \left(\sum_{\mu=1}^{m_u} a_{\mu} + \sum_{\mu=m_u+1}^{(m_j+m_u)/2} a_{\mu} - \sum_{\mu=(m_j+m_u)/2+1}^{m_j} a_{\mu} \right)_i, \quad (17b)$$

при $m_j > m_u$, ($\sigma_j^2 > \sigma_u^2$). В обоих случаях $m_u + m_j$ – четное число.

Рассмотрим теперь формирование коррелированных рядов. Для того чтобы ряды x_{ji} и x_{ui} оказались взаимно коррелированными, парные элементы этих рядов должны включать в себя вклады общего фактора и случайности.

$$x_{ji} = \left\{ \sum_{\mu=1}^{m^*} a_{\mu} + \sum_{\mu=m^*+1}^{m_j} a_{\mu} \right\}_i, \quad (18a1)$$

$$x_{ui} = \left\{ \sum_{\mu=1}^{m^*} a_{\mu} + \sum_{\mu=m_j+1}^{m_u+(m_j-m^*)} a_{\mu} \right\}_i. \quad (18a2)$$

Здесь первое слагаемое представляет собой вклад общих факторов и является одинаковым для обоих рядов (одним и тем же). Величина m^* определяет число суммируемых случайных чисел, входящих в оба ряда. Вторые слагаемые отражают влияние случайности. Коэффициент корреляции ρ_{ju} тем выше, чем ближе значение m^* к m_j и m_u . Ряды, полученные по алгоритмам (18a1) и (18a2), коррелированы положительно.

Для получения отрицательной корреляции можно преобразовать эти алгоритмы к виду

$$x'_{ji} = \left\{ \sum_{\mu=1}^{m^*} a_{\mu} + \sum_{\mu=m^*+1}^{m_j} a_{\mu} \right\}_i, \quad (18b1)$$

$$x'_{ui} = \left\{ - \sum_{\mu=1}^{m^*} a_{\mu} - \sum_{\mu=m_j+1}^{m_u+(m_j-m^*)} a_{\mu} + m_u \right\}_i. \quad (18b2)$$

Выражения (18) применимы при $m_u > m_j$. Генеральные характеристики рядов x_{ji} , x'_{ji} так же как x_{ui} и x'_{ui} , определяемые выражениями (18a) и (18b), являются одинаковыми.

Содержание работы

1. Смоделируйте четыре ряда измерений, отличающихся друг от друга генеральными средними, дисперсиями и парными коэффициентами корреляции:

1.1. Ряд x_{1i} по алгоритму (16) при m_1 , заданном в таблице вариантов исходных данных;

1.2. Некоррелированный с x_{1i} ряд x_{2i} , отличающийся от первого дисперсией, по алгоритму (17a) или (17b) в зависимости от соотношения между m_1 и m_2 ;

1.3. Взаимно коррелированные ряды x_{3i} и x_{4i} по алгоритму (18) при указанном в таблице вариантов знаке коэффициента корреляции ((18a) при $\rho_{34} > 0$ и (18b) при $\rho_{34} < 0$).

Таким образом $x_{01} = x_{02} = x_{03}$, $\sigma_1^2 = \sigma_3^2$, $\sigma_2^2 = \sigma_4^2$, $\rho_{12} = \rho_{13} = \rho_{14} = \rho_{23} = \rho_{24} = 0$, $\rho_{34} \neq 0$.

1.4. В этом задании при сравнении рядов потребуются значения выборочных средних \bar{x}_j , дисперсий S_j^2 и коэффициентов корреляции r_{ju} , $j, u = \overline{1,4}$. Необходимые для расчета этих величин

промежуточные суммы $\sum_{i=1}^{n'} x_{ji}$, $\sum_{i=1}^{n'} x_{ji}^2$ и $\sum_{i=1}^{n'} x_{ji}x_{ui}$ можно вычислять непосредственно в темпе обра-

ботки значений x_{ji} (без занесения x_{ji} в файлы памяти). В этом случае величины \bar{x}'_j , $S_j'^2$ и r'_{ju} легко можно сосчитать при любом текущем значении n' ($1 \leq n' \leq n_j$), используя формулы

$$\bar{x}'_j = \left(\sum_{i=1}^{n'} x_{ji} \right) / n',$$

$$S_j'^2 = \left(\sum_{i=1}^{n'} x_{ji}^2 - \left(\sum_{i=1}^{n'} x_{ji} \right)^2 / n' \right) / (n' - 1),$$

$$r'_{ju} = \left(\sum_{i=1}^{n'} x_{ji} x_{ui} - \left(\sum_{i=1}^{n'} x_{ji} \right) \left(\sum_{i=1}^{n'} x_{ui} \right) / n' \right) / (n' - 1) S_j' S_u'.$$

Для дальнейшего анализа выведите на печать эти величины при указанных в таблице вариантов значениях n' (короткие ряды) и n (длинные ряды). Для каждого варианта задается $n_j = \text{const} = n$.

1.5. Проверьте гипотезу о значимости коэффициентов корреляции.

2. По критерию Фишера выполните попарное сравнение всех четырех дисперсий (S_1^2 и S_2^2 , S_1^2 и S_3^2 , S_1^2 и S_4^2 , S_2^2 и S_3^2 , S_2^2 и S_4^2 , S_3^2 и S_4^2) для короткого и длинного рядов. Уровень значимости здесь и далее $\alpha = 0,05$.

3. Проверьте гипотезу об однородности всех четырех дисперсий по критерию Бартлетта (для коротких и длинных рядов).

4. Выполните попарное сравнение выборочных средних по критерию Стьюдента (t или T – в зависимости от однородности или неоднородности дисперсий) и критерию UT (для коротких и длинных рядов).

5. Проверьте гипотезу о равенстве всех четырех средних методом однофакторного дисперсионного анализа.

Проанализируйте полученные результаты. Обратите внимание на то, как согласуются между собой различные критерии, и на то, как зависит выбор гипотезы от объема выборки.

Таблица

Варианты исходных данных

Вариант	$m_1,$ m_3	$m_2,$ m_4	m^*	n'	n	ρ_{34}
1	12	14	10	10	50	>0
2	10	12	8	8	50	>0
3	14	16	12	6	40	>0
4	12	15	10	10	60	≤ 0
5	24	28	20	12	50	≤ 0
6	24	26	18	10	50	≤ 0

Литература: [2], раздел 7, [5], [6].

ОЦЕНКА ОДНОРОДНОСТИ И РАВНОТОЧНОСТИ РЯДОВ ИЗМЕРЕНИЙ

Часть II. Непараметрические методы

Непараметрические методы построены на сравнении некоторых обобщающих характеристик, рассчитываемых по всем элементам анализируемых рядов. При этом в конечном итоге, как и в задании 9, проверяются гипотезы о равенстве средних или однородности дисперсий. Достоинство непараметрических методов состоит в том, что они могут быть использованы при произвольных (не обязательно нормальных) распределениях случайных величин.

1. Ранговый критерий Вилкоксона. Критерий предназначен для проверки гипотезы о равенстве средних двух эмпирических распределений

$$H_0 : x_{01} = x_{02}; \quad H_1 : x_{01} \neq x_{02}. \quad (1)$$

Пусть имеются два ряда измерений: $x_{1i}, i = \overline{1, n_1}$, $x_{2l}, l = \overline{1, n_2}$. Присвоим индекс 1 более короткому ряду, так что $n_1 \leq n_2$. Для наглядности представления введем обозначения $x_{1i} = a_i$, $x_{2l} = b_l$. Из двух рядов измерений (a_i и b_l) построим один возрастающий вариационный ряд длиной $n_1 + n_2$ и проранжируем его, присваивая элементам ряда порядковые номера $r^{(B)}$ от $r^{(B)} = 1$ для наименьшего (крайнего слева) до $r^{(B)} = n_1 + n_2$ для наибольшего (крайнего справа). Например,

Ряд	a_1	a_2	a_3	b_1	a_4	b_2	...	a_n	...	b_{n_2-2}	b_{n_2-1}	b_{n_2}
Ранги	1	2	3	4	5	6	$n_1 + n_2 - 2$	$n_1 + n_2 - 1$	$n_1 + n_2$

(2)

Если соседние элементы вариационного ряда имеют одинаковые численные значения, то всем им присваиваются одинаковые средние ранги $\bar{r}^{(B)}$. Далее просуммируем все $r_i^{(B)}$ – ранги элементов короткого ряда. Получим сумму рангов Вилкоксона $R^{(B)}$

$$R^{(B)} = \sum_{i=1}^{n_1} r_i^{(B)}. \quad (3)$$

Случайная величина $R^{(B)}$ имеет нормальное распределение с математическим ожиданием

$$R_0 = (n_1(n_2 + n_1 + 1) - 1) / 2 \quad (4)$$

и СКО

$$\sigma(R) = \sqrt{\frac{n_1 n_2}{12} (n_2 + n_1 + 1)}. \quad (5)$$

Решающее правило выбора гипотез при задании одностороннего критерия

$$R_0 - z_{1-\alpha} \sigma(R) \leq R^{(B)} \leq R_0 + z_{1-\alpha} \sigma(R) \Rightarrow H_0 : \bar{a} = \bar{b},$$

$$R^{(B)} < R_0 - z_{1-\alpha} \sigma(R) \Rightarrow H_1 : \bar{a} < \bar{b},$$

$$R^{(B)} > R_0 + z_{1-\alpha} \sigma(R) \Rightarrow H_1 : \bar{a} > \bar{b}.$$

Напомним, что $\bar{a} \equiv \bar{x}_1$ и $\bar{b} \equiv \bar{x}_2$. Критерий Вилкоксона можно использовать при $n_1 \geq 5$.

2. Ранговый критерий Сиджела-Тьюки. Этот критерий используется для проверки гипотезы об однородности дисперсий двух рядов

$$H_0 : \sigma_1^2 = \sigma_2^2; \quad H_1 : \sigma_1^2 \neq \sigma_2^2.$$

Как и в предыдущем случае из двух рядов измерений строится вариационный ряд. Однако ранжирование осуществляется иначе. Малые ранги присваиваются крайним (слева и справа) элементам

общего ряда, а большие – центральным. Покажем процедуру ранжирования на примере того же ряда (2)

Ряд	a_1	a_2	a_3	b_1	...	a_{n_1}	...	b_{n_2-3}	b_{n_2-2}	b_{n_2-1}	b_{n_2}
Ранги	1	4	5	8	7	6	3	2

Для вычисления критерия $R^{(C-T)}$ – критерия Сиджела-Тьюки – снова суммируется число рангов для короткого ряда

$$R^{(C-T)} = \sum_{i=1}^{n_1} r_i^{(C-T)}. \quad (6)$$

Значения R_0 и $\sigma(R)$ для обоих критериев одинаковы, они рассчитываются, соответственно, по формулам (4) и (5). Похожим остается и решающее правило

$$\begin{aligned} R_0 - z_{1-\alpha}\sigma(R) \leq R^{(C-T)} < R_0 + z_{1-\alpha}\sigma(R) &\Rightarrow H_0: \sigma_1^2 = \sigma_2^2 \\ R^{(C-T)} < R_0 - z_{1-\alpha}\sigma(R) &\Rightarrow H_1: \sigma_1^2 > \sigma_2^2, \\ R^{(C-T)} > R_0 + z_{1-\alpha}\sigma(R) &\Rightarrow H_1: \sigma_1^2 < \sigma_2^2. \end{aligned} \quad (7)$$

Условия применения те же, что и для критерия Вилкоксона.

Вычисление сумм рядов Вилкоксона $R^{(B)}$ и Сиджела-Тьюки $R^{(C-T)}$ при работе на ПК

Описанные здесь процедуры вычисления $R^{(B)}$ и $R^{(C-T)}$ являются простыми и наглядными. Однако, если длина вариационного ряда $N = n_1 + n_2$ сравнительно велика, то нахождение рангов “вручную” становится громоздким.

При работе на ПК для вычисления рангов $r_i^{(B)}$ достаточно сравнить значение элемента a_i со всеми остальными элементами этого ряда a_j , $j = \overline{1, n_1}$ и со всеми элементами ряда b_l , $l = \overline{1, n_2}$

$$r_i^{(B)} = \sum_{j=1}^{n_1} \theta\{a_i - a_j\} + \sum_{l=1}^{n_2} \theta\{a_i - b_l\}, \quad (8)$$

где $\theta\{y\}$ – функция Хэвисайда: $\theta\{y\} = \begin{cases} 1, & y \geq 0, \\ 0, & y < 0. \end{cases}$ Сумма рангов $R^{(B)}$ рассчитывается по формуле (3).

Анализ показывает, что между $r_i^{(B)}$ и $r_i^{(C-T)}$ существует определенная связь. Ее можно использовать для непосредственного перехода от $r_i^{(B)}$ к $r_i^{(C-T)}$.

Формулы перехода имеют вид:

1. При $1 \leq r_i^{(B)} < N/2$

$$r_i^{(C-T)} = \begin{cases} 2r_i^{(B)} - 1, & \text{если число } r_i^{(B)} \text{ – нечетное} \\ 2r_i^{(B)}, & \text{если число } r_i^{(B)} \text{ – четное.} \end{cases} \quad (9)$$

2. При $N/2 < r_i^{(B)} \leq N$ выбор формулы зависит от четности или нечетности N :
для N четного

$$r_i^{(C-T)} = \begin{cases} 2(N+1-r_i^{(B)}) - 1, & \text{если } r_i^{(B)} \text{ – нечетное,} \\ 2(N+1-r_i^{(B)}), & \text{если } r_i^{(B)} \text{ – четное.} \end{cases} \quad (10)$$

Для N нечетного при $r_i^{(B)} \neq \frac{N+1}{2}$

$$r_i^{(C-T)} = \begin{cases} 2(N+1-r_i^{(B)}) & \text{если } r_i^{(B)} - \text{нечетное,} \\ 2(N+1-r_i^{(B)})-1 & \text{если } r_i^{(B)} - \text{четное.} \end{cases} \quad (11)$$

И наконец, для N нечетного и $r_i^{(B)} = \frac{N+1}{2}$

$$r_i^{(C-T)} = 2(N+1-r_i^{(B)})-1. \quad (12)$$

Сумма рангов $R^{(C-T)}$ рассчитывается по формуле (6).

Чтобы "научить" ПК различать четные и нечетные числа, можно использовать следующий простой алгоритм. Пусть D – анализируемое число. Разность $[D/2 - \text{int}(D/2)]$ при четном D равна нулю, а при нечетном больше нуля.

3. χ^2 – критерий используется для сравнения нескольких рядов измерений, отличающихся средними значениями или формами распределения. Чтобы исключить громоздкие вычисления, ограничимся здесь только сравнением двух рядов. Пусть имеются ряды: $x_{1i}, i = \overline{1, n_1}$ и $x_{2i}, i = \overline{1, n_2}$. Разобьем оба ряда на одинаковые интервалы $\bar{x}_\omega \pm h_\omega/2$, $\omega = \overline{1, M}$. Найдем для каждого интервала $n_{j\omega}$ – число результатов, попадающих в ω – вый интервал для j – ого ряда, $j = \overline{1, 2}$. Рассчитаем

$$\chi^2 = n_1 n_2 \sum_{\omega=1}^M \frac{1}{n_{1\omega} + n_{2\omega}} \left(\frac{n_{1\omega}}{n_1} - \frac{n_{2\omega}}{n_2} \right)^2. \quad (13)$$

Полученное значение χ^2 сравнивается с критическим $\chi_{1-\alpha, f}^2$, $f = M - 1$. Выбор гипотезы осуществляется в соответствии с решающим правилом

$$\chi^2 \leq \chi_{1-\alpha, f}^2 \Rightarrow H_0, \quad \chi^2 > \chi_{1-\alpha, f}^2 \Rightarrow H_1. \quad (14)$$

Принятие альтернативной гипотезы свидетельствует о значимом различии средних \bar{x}_1 и \bar{x}_2 , либо о неоднородности дисперсий.

Применение критерия обуславливается объемами выборок $n_1 + n_2 > 50$.

Имитационные модели. Как уже отмечалось, использование непараметрических методов не ограничивается сравнением нормально распределенных случайных величин. Поэтому в настоящем задании предлагается сравнивать модельные ряды измерений, имеющие различные распределения.

Нормально распределенные взаимно коррелированные ряды x_{ji} и x_{ui} можно получить, как и в задании 9, по алгоритмам: при положительной корреляции ($\rho_{ju} > 0$)

$$x_{ji} = \left\{ \sum_{\mu=1}^{m^*} a_\mu + \sum_{\mu=m^*+1}^{m_j} a_\mu \right\}_i, \quad (15a1)$$

$$x_{ui} = \left\{ \sum_{\mu=1}^{m^*} a_\mu + \sum_{\mu=m_j+1}^{m_u+(m_j-m^*)} a_\mu \right\}_i, \quad (15a2)$$

при отрицательной корреляции ($\rho_{ju} < 0$)

$$x'_{ji} = \left\{ \sum_{\mu=1}^{m^*} a_\mu + \sum_{\mu=m^*+1}^{m_j} a_\mu \right\}_i, \quad (15b1)$$

$$x'_{ui} = \left\{ - \sum_{\mu=1}^{m^*} a_{\mu} - \sum_{\mu=m_j+1}^{m_j+m^*} a_{\mu} + m_u \right\}_i \quad (1562)$$

Ряд x_{ji} можно сформировать, используя алгоритм генерирования распределения Лапласа с ярко выраженным эксцессом ($\Theta = 3$)

$$x_{ji} = \frac{m_j}{2} + \sqrt{\frac{m_j}{24}} \ln(a'/a'')_i, \quad (16)$$

где a' и a'' – случайные равномерно распределенные числа, последовательно вырабатываемые RND на i – том шаге.

Ряд измерений x_{ji} , соответствующий асимметричному распределению (см. задание 5), можно смоделировать по алгоритму

$$x_{ji} = \frac{\sqrt{(x'_i - 6)^2 + (x''_i - 6)^2} - 1,25}{0,66} + \frac{m_j}{2}, \quad (17)$$

где $x'_i = \left(\sum_{\mu=1}^{12} a_{\mu} \right)_i$, $x''_i = \left(\sum_{\mu=1}^{12} a_{\mu} \right)_i$.

Обратите внимание на то, что в этих суммах последовательно вырабатываемые RND a_{μ} являются разными.

В выражениях (15)...(17) $j, u = \overline{1, k}$, k – число моделируемых рядов, $i = \overline{1, n_j}$, n_j – длина соответствующего ряда, m_j , m_u , m^* – число суммируемых равномерно распределенных случайных чисел.

Содержание работы

1. Используя имитационные модели, сформируйте четыре ряда измерений: x_{1i} и x_{2i} , следуя алгоритмам (15), x_{3i} и x_{4i} соответственно по алгоритмам (16), (17).

Значения m_1 , m_2 , m^* , $n_1 = n_2 = n$ и знак коэффициента корреляции ρ_{12} указаны в таблице вариантов исходных данных.

Занесите эти ряды в память ПК.

2. Рассчитайте для каждого ряда выборочные средние \bar{x}_j и дисперсии S_j^2 .

$$\bar{x}_j = \left(\sum_{i=1}^n x_{ji} \right) / n,$$

$$S_j^2 = \left(\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 \right) / (n - 1).$$

3. Проверьте гипотезы о равенстве средних в этих рядах для всех шести пар по ранговому критерию Вилкоксона. Уровень значимости здесь и далее $\alpha = 0,05$.

4. Проверьте гипотезы об однородности всех шести пар дисперсий по критерию Сиджела-Тьюки.

5. Проверьте гипотезы об однородности всех шести пар рядов по критерию χ^2 . При принятии альтернативной гипотезы для какой-либо пары объясните, чем определяется различие распределений — неравенством средних или неоднородностью дисперсий.

Проанализируйте полученные результаты.

Таблица

Варианты исходных данных

Вариант	m_1, m_3	m_2, m_4	m^*	n	ρ_{12}
1	12	14	10	50	> 0
2	10	12	8	50	< 0
3	14	16	12	40	> 0
4	12	15	10	60	< 0
5	24	28	20	50	> 0
6	24	26	18	50	< 0

Литература: [2], раздел 7, [5], [6], [13].

ДИСПЕРСИОННЫЙ АНАЛИЗ

Дисперсионный анализ — это раздел теории эксперимента, в котором рассматриваются методы проверки гипотез зависимости исследуемой величины от контролируемых факторов. Суть методов заключается в том, что неслучайное влияние того или иного фактора чисто формально выражают в терминах случайности через дисперсию фактора σ_{Φ}^2 . Далее используют аппарат сравнения дисперсий. Если дисперсия фактора значимо больше нуля, то принимают гипотезу о влиянии фактора

$$H_0 : \sigma_{\Phi}^2 = 0, \quad H_1 : \sigma_{\Phi}^2 > 0. \quad (1)$$

Достоинства дисперсионного анализа в полной мере проявляются при реализации планов многофакторного эксперимента.

Распространенной формой таких планов служат ортогональные квадраты. Латинскими называют квадраты, в которых каждый уровень дополнительного фактора встречается в любых строке и столбце только один раз. Два ортогональных латинских квадрата составляют греко-латинский квадрат, три — гипер-греко-латинский и т. д. Приведем здесь в качестве примера гипер-греко-латинский квадрат 4×4 — табл. 1.

Таблица 1

	A_1	A_2	A_3	A_4
B_1	$a\alpha 1$	$b\beta 4$	$c\gamma 3$	$d\delta 2$
B_2	$b\gamma 2$	$a\delta 3$	$d\alpha 4$	$c\beta 1$
B_3	$c\delta 4$	$d\gamma 1$	$a\beta 2$	$b\alpha 3$
B_4	$d\beta 3$	$c\alpha 2$	$b\delta 1$	$a\gamma 4$

Эта матрица представляет собой план пятифакторного эксперимента, в котором факторы A, B, C, D, E задаются на четырех уровнях каждый:

$$A_j, \quad B_i, \quad j, i = \overline{1,4}; \quad C_v = a, b, c, d; \quad D_{\mu} = \alpha, \beta, \gamma, \delta, \quad E_{\pi} = 1, 2, 3, 4.$$

Практическая реализация этого плана состоит в том, что измерения исследуемой величины выполняются при тех сочетаниях уровней факторов, которые приведены в табл. 1. После проведения экспериментов будем иметь m^2 значений $x_{i,j,v,\mu,\pi}$.

Процедура обработки заключается в следующем. Найдем суммы результатов измерений: по столбцам табл. 1 как $X_j = \sum_{i=1}^m x_{ji}$ (индексы при x , не участвующие в операции сложения опускаем);

по строкам — $X_i = \sum_{j=1}^m x_{ji}$, далее по каждому из уровней остальных факторов $X_v = \sum_{j=1}^m x_{jv}$,

$X_{\mu} = \sum_{j=1}^m x_{j\mu}$, $X_{\pi} = \sum_{j=1}^m x_{j\pi}$ (суммирование можно выполнить и по i).

Затем найдем общие суммы

$$Q_0 = \sum_{j=1}^m \sum_{i=1}^m x_{ji}^2, \quad Q_1 = \frac{1}{m^2} \left(\sum_{j=1}^m X_j \right)^2 = \frac{1}{m^2} \left(\sum_{i=1}^m X_i \right)^2$$

и суммы квадратов по факторам

$$Q_A = \frac{1}{m} \sum_{j=1}^m X_j^2; \quad Q_B = \frac{1}{m} \sum_{i=1}^m X_i^2; \quad Q_C = \frac{1}{m} \sum_{v=1}^m X_v^2; \quad Q_D = \frac{1}{m} \sum_{\mu=1}^m X_{\mu}^2; \quad Q_E = \frac{1}{m} \sum_{\pi=1}^m X_{\pi}^2.$$

Общая формула для расчета дисперсии воспроизводимости имеет вид

$$S_0^2 = \frac{Q_0 + (k-1)Q_1 - \sum_{\Phi=1}^k Q_{\Phi}}{(m-1)(m+1-k)}, \quad (2)$$

где k – число исследуемых факторов, $\Phi = A, B, C, \dots$

Дисперсии, определяющие влияние k – того фактора, смешанного со случайностью, рассчитывают по формулам

$$S_{\Phi}^2 = \frac{Q_{\Phi} - Q_1}{m-1}. \quad (3)$$

Для проверки гипотез находят отношения Фишера $F_{\Phi} = S_{\Phi}^2 / S_0^2$. Решающее правило

$$F_{\Phi} \leq F_{1-\alpha, f_{S_{\Phi}^2}, f_{S_0^2}} \Rightarrow H_0, \quad F_{\Phi} > F_{1-\alpha, f_{S_{\Phi}^2}, f_{S_0^2}} \Rightarrow H_1. \quad (4)$$

В настоящем задании учащемуся предлагается освоить процедуру расчетов и оценить эффективность метода при различном числе исследуемых факторов.

Имитационная модель. Зададим ее простой линейной зависимостью

$$x = x_0 + g_0 Z + g_1 A + g_2 B + g_3 C + g_4 D + g_5 E. \quad (5)$$

Здесь g_0, g_1, \dots, g_5 – коэффициенты, определяющие вклад каждого из факторов: x_0 – характерное (исходное или среднее) значение x , Z – случайная составляющая, определяемая влиянием всех неконтролируемых факторов. В экспериментах могут быть контролируемые либо факторы, либо часть их. Контролируемые факторы в "испытаниях" задаются на уровнях, определяемых планом эксперимента. Неконтролируемые факторы принимают, как и Z , случайные значения, увеличивая тем самым рассеяние x . Для вычисления Z и значений неконтролируемых факторов воспользуемся генератором случайных чисел (RND), формируя нормально распределенные случайные величины

$$Z = \sum_{\mu=1}^{m_1} a_{\mu}, \quad \Phi = \sum_{\mu=1}^{m_1} a_{\mu}. \quad (6)$$

Здесь, как и в других заданиях, индекс μ означает номер последовательно вырабатываемых RND равномерно распределенных чисел $a_{\mu}, \mu = \overline{1, m_1}$ (не путать с такими же обозначениями уровней фактора D).

Зададим далее $m_1 = 12$, а уровни контролируемых факторов одинаковыми рядами чисел: 1, 2, ... m (то есть $i, j, \mu, v, \pi = \overline{1, m}$).

Содержание работы

1. Составьте план эксперимента для двух факторов — A и B ($k = 2$), задавая каждый из них на пяти уровнях ($m = 5$). Заполните матрицу плана 5×5 результатами эксперимента, вычисленными по имитационной модели (5). Значения факторов Z, C, D, E рассчитайте по формулам (6).

Найдите X_j и X_i ; Q_0, Q_1, Q_A и Q_B ; S_0^2, S_A^2 и S_B^2 по формулам (2) и (3). По критерию Фишера $F = S_{\Phi}^2 / S_0^2$ проверьте гипотезу зависимости x от факторов A и B .

Скрытый смысл этого задания состоит в том, чтобы на этапе проведения двухфакторного эксперимента получить незначимые зависимости x от обоих факторов A и B , либо, по крайней мере, от одного фактора B . Если при заданных в таблице вариантов значениях коэффициентов g_1 и g_2 эта цель оказалась не достигнутой, измените их набор таким образом, чтобы влияние A и B , либо только B оказалось незначимым.

2. Убедитесь теперь в эффективности многофакторного эксперимента. Составьте греко-латинский квадрат 5×5 , включив в него факторы C и D . Рассчитайте новые значения $x_{i,j,v,\mu}$. Вычислите последовательно $X_i, X_j, X_v, X_\mu, Q_0, Q_1, Q_A, Q_B, Q_C, Q_D$, а затем $S_0^2, S_{\Phi}^2, \Phi = A, B, C, D$. Проверьте снова гипотезы о значимом влиянии факторов. Если имитационная модель была настроена хорошо, то влияние факторов с большими коэффициентами должно оказаться значимым.

3. Продлите эти исследования. Составьте гипер-греко-латинский квадрат 5×5 . Реализуйте новый план эксперимента, рассчитав значения $x_{i,j,v,\mu,\pi}$. Вычислите, как и ранее, S_0^2 и все $S_{\Phi}^2, \Phi = A, B, C, D, E$ и снова по критерию Фишера проверьте гипотезы о влиянии всех пяти факторов.

Проанализируйте полученные результаты. Обратите внимание на то, как изменяется S_0^2 при увеличении числа исследуемых факторов. Сформулируйте ваши предложения по изменению коэффициентов.

Таблица 2

Варианты исходных данных

Вариант	Коэффициенты					
	g_0	g_1	g_2	g_3	g_4	g_5
1	1	1	-0,8	0,6	0,4	-0,2
2	1	1	-0,6	0,5	-0,4	0,3
3	1	0,9	-0,8	0,7	0,6	-0,5
4	0,75	1	-0,8	0,6	-0,5	0,4
5	0,75	0,8	-0,7	0,6	0,4	-0,2
6	1,25	1	-1	0,6	-0,6	0,3

Во всех вариантах $x_0 = 10$.

Литература: [2], раздел 10, [1], [12].

КОРРЕЛЯЦИЯ МЕЖДУ СЛУЧАЙНЫМИ ВЕЛИЧИНАМИ

Если случайные величины y и x связаны между собой таким образом, что с изменением x изменяется функция распределения y , то такая связь называется стохастической (вероятностной). Если закон распределения сохраняется, а изменяется только среднее значение, то такую стохастическую связь называют корреляционной.

Мерой тесноты линейной связи между случайными величинами y и x_j , $j = \overline{1, k}$ является парный коэффициент корреляции ρ_{yx_j} . Выборочный парный коэффициент корреляции $r_{yx_j} \equiv r_{yj}$ вычисляется по формуле

$$r_{yj} = \frac{\frac{1}{n-1} \left(\sum_{i=1}^n (y_i - \bar{y})(x_{ji} - \bar{x}_j) \right)}{S(y)S(x_j)} = \frac{\sum_{i=1}^n y_i x_{ji} - \left(\sum_{i=1}^n y_i \right) \left(\sum_{i=1}^n x_{ji} \right) / n}{\sqrt{\left[\sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 / n \right] \left[\sum_{i=1}^n x_{ji}^2 - \left(\sum_{i=1}^n x_{ji} \right)^2 / n \right]}}, \quad (1)$$

где $i = \overline{1, n}$, n – объем выборки.

Коэффициент корреляции изменяется в пределах от -1 до 1.

Значимость парного коэффициента корреляции проверяется сравнением его с критическим значением $r_{кр} = r_{1-\alpha/2, f}$, зависящим от уровня значимости α и числа степеней свободы $f = n - 2$

$$r_{1-\alpha/2, f} = \left(\frac{n-2}{t_{1-\alpha/2, f}^2} + 1 \right)^{-0,5}. \quad (2)$$

Здесь $t_{1-\alpha/2, f}$ – коэффициент Стьюдента. Решающее правило

$$\begin{aligned} |r_{yj}| \leq r_{1-\alpha/2, f} &\Rightarrow H_0 : \rho_{yj} = 0, \\ |r_{yj}| > r_{1-\alpha/2, f} &\Rightarrow H_1 : \rho_{yj} \neq 0. \end{aligned} \quad (3)$$

Если величина y изменяется под влиянием нескольких факторов, то парный коэффициент корреляции r_{yj} не всегда отражает истинную силу связи между y и x_j . Такая связь может быть оценена, если влияние остальных факторов исключается. Эта цель достигается вычислением частных коэффициентов корреляции $r_{yj(1,2,\dots,j-1,j+1,\dots,k)}$. В индексе в скобках перечисляются номера исключаемых факторов. Для нахождения частных коэффициентов корреляции следует рассчитать все парные коэффициенты r_{yj} , $j = \overline{1, k}$, а также r_{ju} , $u = \overline{1, k}$. Если ограничиться зависимостью y только от двух факторов x_1 и x_2 , то частные коэффициенты $r_{y1(2)}$ и $r_{y2(1)}$ вычисляются по формулам

$$r_{y1(2)} = \frac{r_{y1} - r_{y2}r_{12}}{\sqrt{(1-r_{y2}^2)(1-r_{12}^2)}}, \quad (4)$$

$$r_{y2(1)} = \frac{r_{y2} - r_{y1}r_{12}}{\sqrt{(1-r_{y1}^2)(1-r_{12}^2)}}. \quad (5)$$

Легко видеть, что $r_{y1(2)} = r_{y1}$ только в том случае, когда r_{y2} и r_{12} равны нулю, то есть когда отсутствует корреляция между y и x_2 , а факторы x_1 и x_2 ортогональны. Значимость частных коэффициентов корреляции проверяется, как и парных, по формулам (2) и (3). Обобщающей характеристикой связи случайной величины y со всеми факторами x_j , $j = 1, k$ является множественный коэффициент корреляции R . В случае двух факторов множественный коэффициент корреляции вычисляется по формуле

$$R = \sqrt{\frac{r_{y1}^2 + r_{y2}^2 - 2r_{y1}r_{y2}r_{12}}{1 - r_{12}^2}}. \quad (6)$$

Коэффициент множественной корреляции изменяется в пределах от 0 до 1.

Перечисленные коэффициенты корреляции отражают только линейную связь между переменными. Если связь нелинейная, то мерой тесноты связи между двумя переменными может служить корреляционное отношение

$$\eta_{yx} = \frac{S_M}{S(y)}, \quad (7)$$

где

$$S_M = \sqrt{\frac{1}{n-1} \sum_{\omega=1}^M n_{\omega} (\bar{y}_{\omega} - \bar{y})^2}, \quad (8)$$

$$S(y) = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2}. \quad (9)$$

Здесь S_M есть корень квадратный из межгрупповой дисперсии, которая находится следующим образом. Поле корреляции делится на M интервалов по x . Подсчитывается число точек n_{ω} , попавших в каждый интервал, а также среднее значение величины y внутри интервалов \bar{y}_{ω} , $\omega = \overline{1, M}$. Полученные значения подставляются в формулу (8).

Корреляционное отношение, как и R , изменяется от 0 до 1. При $\eta_{yx} = 0$ связь между y и x отсутствует, при $\eta_{yx} = 1$ связь функциональная. Если связь между y и x линейная, то $\eta_{yx} = r_{yx}$. Когда связь между y и x нелинейная, $\eta_{yx} > r_{yx}$.

Имитационные модели. Простейшую линейную модель зависимости y от x_1 и x_2 представим в виде

$$y = a_0 + a_1 x_1 + a_2 x_2 + Z. \quad (10a)$$

Более сложную нелинейную модель зададим как

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + Z. \quad (10b)$$

Здесь $a_0, a_1, a_2, b_0, b_1, b_2, b_{12}, b_{11}, b_{22}$ – коэффициенты, а Z – слагаемое, имитирующее случайный шум. Нормальное распределение случайной величины Z может быть задано с помощью генератора случайных чисел как

$$Z_i = b_3 \left(\sum_{\mu=1}^{m_1} a_{\mu} \right)_i, \quad (11)$$

где m_1 – количество суммируемых равномерно распределенных чисел a_μ , $\mu = \overline{1, m_1}$. Значение исследуемой величины в i -том испытании, то есть y_i , получается подстановкой в выражения (10), значений факторов x_{1i} и x_{2i} , $i = \overline{1, n}$.

Уровни факторов x_1 и x_2 будем задавать безразмерными и центрированными относительно нуля: ..., -2, -1, 0, 1, 2, ...

Планы эксперимента, то есть сочетания уровней факторов x_1 и x_2 представим либо в узлах прямоугольной сетки (квадрат), либо косоугольной (ромб) (рис. 1).

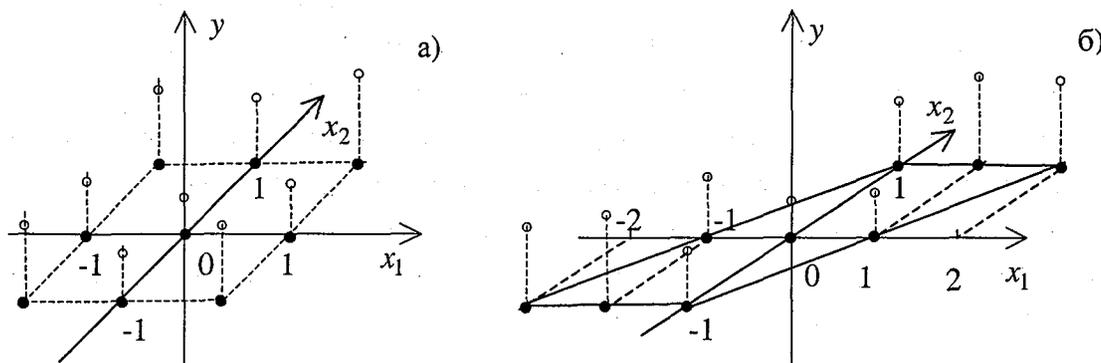


Рис. 1. План эксперимента: а) – квадрат, б) – ромб.

• – координаты точек в плоскости x_1x_2 ; ◦ – значения y (откладываются по аппликате).

Содержание работы

1. Выполните расчеты и анализ на базе линейной имитационной модели (10а).

1.1. Составьте план эксперимента в виде квадрата 3×3 с уровнями факторов $x_1, x_2 = -1; 0; +1$.

1.2. Рассчитайте значения y_i во всех $n = 3^2$ точках плана, принимая $Z = 0$. Занесите результаты в таблицу вида

i	x_1	x_2	y
1	x_{11}	x_{21}	y_1
2	x_{12}	x_{22}	y_2
...
n	x_{1n}	x_{2n}	y_n

1.3. Постройте график вида 1а.

1.4. Рассчитайте СКО $S(x_1)$, $S(x_2)$, $S(y)$.

1.5. Вычислите парные коэффициенты корреляции r_{12} , r_{y1} , r_{y2} .

1.6. Вычислите частные коэффициенты корреляции $r_{y1(2)}$ и $r_{y2(1)}$.

1.7. Вычислите множественный коэффициент корреляции R .

1.8. Постройте план эксперимента в виде ромба, задавая $x_1 = -2; -1; 0$ при $x_2 = -1$; $x_1 = -1; 0; 1$ при $x_2 = 0$ и $x_1 = 0; 1; 2$ при $x_2 = 1$.

1.9. Повторите операции по п.п. 1.2, ..., 1.7.

1.10. Проанализируйте все результаты. Сравните парные и частные коэффициенты корреляции величины y с факторами x_1 и x_2 . Оцените значимость этих коэффициентов. Какие из полученных результатов можно было предсказать заранее?

1.11. Реализуйте план в виде ромба с учетом случайной составляющей Z , задавая $b_3 = 2/m_1$.

1.12. Выполните все операции по п.п. 1.8, 1.9, 1.10 при $Z \neq 0$.

2. Проведите имитационные эксперименты по нелинейной модели.

2.1. Постройте план эксперимента в виде ромба, задавая

$$x_1 = -4; -3; -2; -1; 0 \text{ при } x_2 = -2,$$

$$x_1 = -3; -2; -1; 0; 1 \text{ при } x_2 = -1,$$

$$x_1 = -2; -1; 0; 1; 2 \text{ при } x_2 = 0,$$

$$x_1 = -1; 0; 1; 2; 3 \text{ при } x_2 = 1,$$

$$x_1 = 0; 1; 2; 3; 4 \text{ при } x_2 = 2.$$

Представьте результаты расчетов в виде таблицы и графика, подобного 1б. Расчеты выполняйте сразу с учетом случайной величины Z , задавая ее, как и в п. 1.11.

2.2. Как и в п.1, рассчитайте последовательно $S(x_1)$, $S(x_2)$, $S(y)$, r_{y1} , r_{y2} , $r_{y1(2)}$, $r_{y2(1)}$, R .

2.3. Проанализируйте полученные результаты. Сравните между собой r_{y1} и $r_{y1(2)}$, r_{y2} и $r_{y2(1)}$.

3. Для плана, реализованного в п. 2, найдите корреляционные отношения η_{y1} и η_{y2} .

3.1. Постройте двумерные поля корреляции $y = y(x_1)$ и $y = y(x_2)$. Разбейте их на интервалы, соответственно, по x_1 и x_2 . Найдите средние значения \bar{y}_ω , $\omega = \overline{1, M}$, где M – число интервалов.

3.2. Рассчитайте S_{M1} и S_{M2} . Значение $S(y)$ найдено ранее (см. п. 2).

3.3. Рассчитайте η_{y1} и η_{y2} , сравните их с r_{y1} и r_{y2} . Как меняется оценка тесноты связи при переходе от линейной характеристики к нелинейной?

Варианты исходных данных

Вариант	Коэффициенты								
	a_0	a_1	a_2	b_0	b_1	b_2	b_{12}	b_{11}	b_{22}
1	0	1	1	10	-1	4	0	-1	0
2	0	4	1	10	+1	4	0	-1	0
3	0	1	4	20	-1	-4	0	+1	0
4	5	-4	1	20	+1	-4	0	+1	0
5	5	-1	4	10	+3	+3	-1	0	0
6	5	-2	2	10	+4	+4	-1	-1	-1

Литература: [2], раздел 11, [6].

АВТОКОРРЕЛЯЦИОННАЯ И СПЕКТРАЛЬНАЯ ФУНКЦИИ

Целью настоящей работы является знакомство с понятиями корреляционной и спектральной функций, приобретением навыков расчета и анализа этих характеристик.

Предположим, что анализируются результаты измерений некоторой физической величины x . При этом известно, что истинное значение изменяется со временем под действием ряда факторов. Измерения выполняются в моменты времени t_i с шагом $\Delta = t_{i+1} - t_i$, $\Delta = const$. Таким образом формируется последовательный ряд значений x_i , $i = \overline{1, n}$, n – длина ряда. Такой временной ряд рассматривается как реализация случайного процесса (точнее, одна из множества его возможных реализаций). Если математическое ожидание, значение x_0 и дисперсия $\sigma^2(x)$ сохраняются постоянными во времени, то процесс считается *стационарным*.

Представим результаты измерений в виде графика зависимости x_i от времени или порядкового номера в ряду. Если ряд достаточно длинный, то визуальный анализ временного хода оказывается довольно трудным. Как правило, можно убедиться, что изменения величины x носят нерегулярный характер, но в то же время, они не являются строго случайными. Между элементами ряда, сдвинутыми на время $\tau = \Delta \cdot s$ (или на порядковый номер s), обнаруживается некоторая связь. В качестве меры такой связи можно использовать коэффициент парной корреляции. Образум из одного и того же ряда x_i , $i = \overline{1, n}$ две выборки: x_i и x_{i+s} , $i = \overline{1, n-s}$ – и найдем $r_{x_i x_{i+s}} \equiv r_s$. Для стационарного случайного процесса получим

$$r_s = \left[\frac{1}{n-s-1} \sum_{i=1}^{n-s} (x_i - \bar{x})(x_{i+s} - \bar{x}) \right] / S^2(x), \quad (1)$$

где

$$\bar{x} = \left(\sum_{i=1}^n x_i \right) / n \quad \text{и} \quad S^2(x) = \left[\sum_{i=1}^n (x_i - \bar{x})^2 \right] / (n-1).$$

Величину r_s называют *коэффициентом автокорреляции*. Последовательность значений r_s , рассчитанных для различных сдвигов s , называют *корреляционной функцией*. Обозначим ее $R(s)$. Формально выражение (1) можно использовать до $s = n-1$. Однако погрешность вычисления $R(s)$ пропорциональна множителю $1/(n-s)^2$. Поэтому на практике рекомендуется использовать формулу (1) до значений $s < n/3$. При выборе предельного значения s в каждом конкретном случае следовало бы исходить из условия значимости корреляционной функции. Четких рекомендаций по оценке значимости $R(s)$ не существует. Рассчитанные для реальных временных рядов функции $R(s)$ весьма разнообразны. Если исходная функция $x(t)$ периодическая, то периодической окажется и $R(s)$. Для белого шума, то есть отсутствия какой-либо связи между элементами ряда, $R(s) = 0$ при $s > 0$.

Для метеорологических процессов типичная зависимость $R(s)$ на графике имеет вид затухающей знакопеременной кривой, стремящейся при больших s к нулю. Чтобы оценить значимость r_s , достаточно сравнить его с критическим значением для парных коэффициентов корреляции

$$r_{1-\alpha/2, f} = \left(\frac{f}{t_{1-\alpha/2, f}^2} + 1 \right)^{-0,5}, \quad f = n - s - 2, \quad (2)$$

следуя решающему правилу

$$\begin{aligned} |r_s| \leq r_{1-\alpha/2, f} &\Rightarrow H_0 \text{ (коэффициент незначим)}, \\ |r_s| > r_{1-\alpha/2, f} &\Rightarrow H_1 \text{ (коэффициент значим)}. \end{aligned}$$

Однако для $R(s)$, многократно пересекающей ось абсцисс, участки функции, близкие к точкам пересечения, очевидным образом окажутся незначимыми. Для них приведенное решающее правило непригодно. В качестве компромисса предлагается считать функцию $R(s)$ значимой до тех пор, пока

$R(s)$ значима в своих экстремумах: $|R(s)_{\text{макс},j}| > r_{1-\alpha/2,f}$, $j = \overline{1, h}$, h – число экстремумов. Наличие значимых максимумов на кривой $R(s)$ может свидетельствовать о появлении квазипериодических колебаний (колебаний с «плавающей» частотой) в анализируемом случайном процессе.

Частоты колебаний, вносящих основной вклад в изменчивость случайного процесса, выявляются при построении спектральной функции. Спектральная функция может быть получена в результате прямого преобразования Фурье случайного процесса (исходного временного ряда).

Более эффективным является преобразование Фурье корреляционной функции. В этом случае сразу рассчитывается спектр мощности случайного процесса. Сглаженная спектральная плотность мощности $\bar{P}(v_k)$ рассчитывается по формуле

$$\bar{P}(v_k) = 2\Delta \sum_{s=0}^L w(s)R(s)\cos(2\pi v_k \Delta \cdot s). \quad (3)$$

Здесь v_k – частота, $v_k = k/(\Delta \cdot L) = kv_1$, $k = \overline{0, L/2}$, где v_1 – основная частота и шаг по частоте $v_1 = 1/(L\Delta)$, максимальная частота (частота Найквиста) $v_H = 1/(2\Delta)$. Величина $\bar{P}(v_k)$ отражает долю дисперсии случайного процесса, приходящуюся на единичный интервал частоты. Размерность $[\bar{P}(v_k)] = \Gamma \text{ц}^{-1}$. Корреляционное окно $w(s)$ часто задают в форме окна Тьюки

$$w(s) = \frac{1}{2}(1 + \cos(\pi s/L)) \text{ при } s \leq L. \quad (4)$$

В формулах (3), (4) параметр L – точка отсечения корреляционной функции, то есть значение s , которым ограничиваются при расчетах спектра. Параметр L определяет разрешающую способность спектральной функции, ее детализацию и погрешность вычисления. Точка отсечения задается с учетом значимости $R(s)$. Удобно округлять L до четного числа.

Имитационная модель. Сформируем дискретный временной ряд x_i , как сумму двух гармонических колебаний с амплитудами A_1 , A_2 и частотами v_{01} , v_{02} . «Испортим» этот ряд наложением на него случайной нормально распределенной величины Z , $Z_i = \left(\frac{2}{m_1} \sum_{\mu=1}^{m_1} a_{\mu} - 1 \right)_i$. Здесь a_{μ} – вырабатываемые RND равномерно распределенные числа ($0 \leq a_{\mu} \leq 1$).

$$x_i = A_1 \sin(2\pi v_{01} \Delta \cdot i) + A_2 \sin(2\pi v_{02} \Delta \cdot i) + BZ_i \quad (5)$$

Коэффициент B определяет соотношение между амплитудами гармонических колебаний и случайных возмущений.

Если исключить случайность ($B = 0$) и рассматривать бесконечно длинный ряд, то задача получит строгое теоретическое решение: спектральная функция будет представлена двумя δ -функциями на частотах v_{01} , v_{02} . Для ограниченного дискретного ряда с наложенной случайностью спектральная функция окажется размытой по частотам.

Содержание работы

1. Подставьте в формулу (5) значения параметров, указанные в таблице вариантов исходных данных. Сформируйте ряд значений x_i , $i = \overline{1, n}$. Представьте этот ряд или часть его в виде графика $x_i = x(t_i)$, где $t_i = i \cdot \Delta$. Вычислите \bar{x} и $S^2(x)$.

2. По формуле (1) рассчитайте значения коэффициента автокорреляции r_s , $s < n/3$. Постройте график корреляционной функции $R(s)$. Проанализируйте его. Выделите значимую часть $R(s)$. С этой целью рассчитайте по формуле (2) несколько значений $r_{1-\alpha/2,f}$ при различных s . Нанесите каждое значение с положительным и отрицательным знаками на график $R(s)$. По этим точкам постройте «коридор незначимости» $R(s)$, проведите кривые по значениям $+r_{1-\alpha/2,f}$ и $-r_{1-\alpha/2,f}$. Вы-

делите на кривой $R(s)$ максимумы и минимумы функций, отвечающие условию значимости. Являются ли значимыми максимумы, соответствующие частотам ν_{01} , ν_{02} ? Выберите точку отсечения корреляционной функции L . Например, найдите последний значимый максимум (п.з.м.) на кривой $R(s)$. Снимите с графика соответствующее ему значение сдвига $s_{\text{п.з.м.}}$. Отсекая $R(s)$ с некоторым запасом, найдите $L \cong (1,5...2)s_{\text{п.з.м.}}$ (L – четные).

3. Рассчитайте спектральную функцию $\bar{P}(\nu_k)$ для всех частот $\nu_k = k/(\Delta \cdot L)$, $k = 0, L/2$. Постройте график $\bar{P}(\nu_k)$. Проанализируйте его. Если в соответствии с имитационной моделью на частотах ν_{01} , ν_{02} (или близких к ним, когда ν_k не совпадает в точности с этими частотами) четко проявляются максимумы спектра мощности, вам повезло – работа закончена. Впрочем, для проверки качества $\bar{P}(\nu_k)$ вы можете повторить расчеты с другими значениями L (больше и меньше использованного).

Если максимумы спектра на частотах ν_{01} , ν_{02} не выделяются или выделяются нечетко, то следует увеличить в несколько раз длину ряда. Рассчитав новую функцию $R(s)$, найдите новое значение L и построьте новую спектральную функцию. Если достанет сил, то повторите этот расчет для различных L .

Таблица 1

Варианты исходных данных

Вариант	Δ, c	A_1	A_2	$\nu_{01}, \text{Гц}$	$\nu_{02}, \text{Гц}$	B
1	1,0	1,0	1,0	0,15	0,35	2,0
2	1,0	1,0	0,5	0,20	0,35	1,5
3	1,0	0,5	0,3	0,15	0,33	1,0
4	10^{-3}	1,0	0,5	150	330	1,5
5	10^{-3}	0,7	0,5	180	320	2,0

Для всех вариантов $n = 200$, $m = 12$.

Литература: [2], раздел 11, [3].

ИНТЕРПОЛЯЦИЯ И СГЛАЖИВАНИЕ

Результаты экспериментов, определяющие зависимость исследуемой величины от контролируемых факторов удобно представить в виде некоторой аналитической функции. Если измерения выполняются с высокой точностью, так что дисперсия воспроизводимости исследуемой величины пренебрежимо мала по сравнению с ее общей дисперсией, то для описания зависимости достаточно воспользоваться формулами интерполяции. Когда точность измерений сравнительно невелика, построение аппроксимирующей функции, проходящей непосредственно через точки измерений, становится бессмысленным, а часто и невозможным. В этих случаях уместнее использовать сглаживание. Сглаживающие аналитические зависимости называют уравнениями регрессии или функциями отклика.

Целью настоящей работы является изучение методов интерполяции и сглаживания, приобретение практических навыков реализации этих методов.

1. Интерполяция. Пусть y – измеряемая величина, x – контролируемый фактор. На каждом уровне фактора x_i , $i = \overline{1, n}$ (в каждом узле сетки) имеется результат измерения y_i . При этом реализуется план эксперимента, в котором $x_i < x_{i+1}$. Задача заключается в том, чтобы построить функцию $y(x)$, позволяющую найти y при любом значении x . Интерполяционная функция $y(x)$ должна удовлетворять условию $y(x_i) = y_i$ во всех узлах сетки на отрезке $[x_1, x_n]$. Обычно функцию $y(x)$ ищут в классе полиномов.

1.1. Полином Лагранжа. Интерполяционный полином Лагранжа задается в виде

$$y(x) = y_1 P_1(x) + y_2 P_2(x) + \dots + y_n P_n(x) = \sum_{j=1}^n y_j P_j(x), \quad (1)$$

где, в свою очередь, $P_j(x)$ также являются полиномами степени $n-1$

$$P_j(x) = \prod_{\substack{i=1 \\ i \neq j}}^n \frac{(x - x_i)}{(x_j - x_i)}. \quad (2)$$

Здесь для сокращения записи индекс i при y и $P(x)$ мы заменили на j , $j = \overline{1, n}$. Это сделано с той целью, чтобы при $i = j$ из произведения (2) легко исключался сомножитель $(x - x_i)/(x_j - x_i)$. Сначала по формуле (2) вычисляются все $P_j(x)$, $j = \overline{1, n}$. В формулу (2) для каждого фиксированного x_j подставляются значения x_i , $i = \overline{1, n}$, за исключением $i = j$. Далее $P_j(x)$ подставляют в формулу (1). Формула (1) пригодна для вычисления $y(x)$ на всем отрезке $[x_1, x_n]$.

1.2. Глобальный кубический сплайн. Сплайном называют функцию, составленную из набора полиномов, каждый из которых используется только на интервале между двумя соседними узлами сетки. Качество сплайна определяется тем, насколько хорошо стыкуются (склеиваются) между собой последовательные участки функции. Глобальный кубический сплайн удовлетворяет требованиям непрерывности собственно функции, а также ее первой и второй производных. Общая формула этого сплайна имеет вид

$$y_{i-1}(x) = m_{i-1} \frac{(x_i - x)^3}{6h_{i-1}} + m_i \frac{(x - x_{i-1})^3}{6h_{i-1}} + \left(y_{i-1} - m_{i-1} \frac{h_{i-1}^2}{6} \right) \frac{(x_i - x)}{h_{i-1}} + \left(y_i - m_i \frac{h_{i-1}^2}{6} \right) \frac{(x - x_{i-1})}{h_{i-1}}, \quad x_{i-1} \leq x \leq x_i, \quad i = \overline{2, n}. \quad (3)$$

Здесь $h_{i-1} = x_i - x_{i-1}$ – шаг по x между узлами сетки.

Однако прежде чем проводить расчеты по этой формуле, необходимо вычислить значения вторых производных $y''(x_i) \equiv m_i$, входящих в нее. С этой целью придется решить следующую систему уравнений, неизвестными в которой являются вторые производные

$$\left\{ \begin{array}{l} m_1 \frac{h_1}{6} + m_2 \frac{h_1 + h_2}{3} + m_3 \frac{h_2}{6} = \frac{y_3 - y_2}{h_2} - \frac{y_2 - y_1}{h_1}, \\ m_2 \frac{h_2}{6} + m_3 \frac{h_2 + h_3}{3} + m_4 \frac{h_3}{6} = \frac{y_4 - y_3}{h_3} - \frac{y_3 - y_2}{h_2}, \\ \dots, \\ m_{i-1} \frac{h_{i-1}}{6} + m_i \frac{h_{i-1} + h_i}{3} + m_{i+1} \frac{h_i}{6} = \frac{y_{i+1} - y_i}{h_i} - \frac{y_i - y_{i-1}}{h_{i-1}}, \\ \dots, \\ m_{n-2} \frac{h_{n-2}}{6} + m_{n-1} \frac{h_{n-2} + h_{n-1}}{3} + m_n \frac{h_{n-1}}{6} = \frac{y_n - y_{n-1}}{h_{n-2}} - \frac{y_{n-1} - y_{n-2}}{h_{n-2}}, \end{array} \right. \quad (4)$$

$i = \overline{2, n-1}$.

Система (4) насчитывает $n-2$ уравнений, но она незамкнута, поскольку в ней n неизвестных. Значения двух из них задают заранее: $m_1 = m_n = 0$.

1.3. Локальный кубический сплайн. В этом сплайне выполняются условия непрерывности только для функции и ее первой производной

$$y_n(x) = \frac{(x_i - x)^2 [2(x - x_{i-1}) + h_{i-1}]}{h_{i-1}^3} y_{i-1} + \frac{(x - x_{i-1})^2 [2(x_i - x) + h_{i-1}]}{h_{i-1}^3} y_i +$$

$$+ \frac{(x_i - x)^2 (x - x_{i-1})}{h_{i-1}^2} \mu_{i-1} - \frac{(x - x_{i-1})^2 (x - x_i)}{h_{i-1}^2} \mu_i, \quad x_{i-1} \leq x \leq x_i, \quad i = \overline{2, n}. \quad (5)$$

Значения первых производных $y'(x_i)$ в узлах сетки обозначены здесь через μ_i , $i = \overline{1, n}$. Величины μ_i рассчитывают методами численного дифференцирования. В частности, при $h_{i-1} = h_i = h$ для $i = \overline{2, n-1}$ по формулам $\mu_i = (y_{i+1} - y_{i-1})/2h$. На границах отрезка используются выражения $\mu_1 = (4y_2 - 3y_1 - y_3)/2h$ и $\mu_n = (3y_n + y_{n-2} - 4y_{n-1})/2h$.

2. Уравнение регрессии. Уравнение регрессии рассматривают как оценку реальной зависимости y от x . Используем для него традиционное обозначение \hat{y} . Остановимся здесь на представлении уравнения регрессии полиномами вида

$$\hat{y} = \sum_{j=0}^k a_j x^j. \quad (6)$$

Максимальная степень полинома $k \leq n-1$. Коэффициенты a_j находят методом наименьших квадратов в соответствии с условиями

$$\frac{\partial}{\partial a_j} \sum_{i=1}^n (y_i - \hat{y})^2 = 0, \quad j = \overline{0, k}. \quad (7)$$

Задавая в (6) степень k и осуществляя $k+1$ раз операцию (7), получим систему нормальных уравнений

$$\begin{array}{l} a_0 n + a_1 \sum_{i=1}^n x_i + a_2 \sum_{i=1}^n x_i^2 + \dots + a_k \sum_{i=1}^n x_i^k = \sum_{i=1}^n y_i, \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 + a_2 \sum_{i=1}^n x_i^3 + \dots + a_k \sum_{i=1}^n x_i^{k+1} = \sum_{i=1}^n x_i y_i, \\ a_0 \sum_{i=1}^n x_i^2 + a_1 \sum_{i=1}^n x_i^3 + a_2 \sum_{i=1}^n x_i^4 + \dots + a_k \sum_{i=1}^n x_i^{k+2} = \sum_{i=1}^n x_i^2 y_i, \\ \dots, \\ a_0 \sum_{i=1}^n x_i^k + a_1 \sum_{i=1}^n x_i^{k+1} + a_2 \sum_{i=1}^n x_i^{k+2} + \dots + a_k \sum_{i=1}^n x_i^{k+k} = \sum_{i=1}^n x_i^k y_i. \end{array} \quad (8)$$

Эта система содержит $k + 1$ уравнений. Например, для линейной зависимости при $k = 1$ в системе (8) остаются только два уравнения

$$\begin{aligned} a_0 n + a_1 \sum_{i=1}^n x_i &= \sum_{i=1}^n y_i, \\ a_0 \sum_{i=1}^n x_i + a_1 \sum_{i=1}^n x_i^2 &= \sum_{i=1}^n x_i y_i. \end{aligned} \quad (9)$$

Аналогично система (8) конкретизируется для любого числа k . Чтобы найти неизвестные a_j , предварительно следует рассчитать все суммы $\sum_{i=1}^n x_i^{j+u}$, $\sum_{i=1}^n x_i^j y_i$, $j, u = \overline{0, k}$. Обычно систему уравнений (8) решают методом Гаусса (последовательным исключением неизвестных).

Практически выбор оптимальной степени полинома осуществляют путем последовательного усложнения уравнения регрессии. Качество уравнения оценивается дисперсией адекватности

$$S_{\text{ад}}^2 = \frac{1}{n-l} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad l = k + 1. \quad (10)$$

Дисперсию адекватности сравнивают с дисперсией воспроизводимости $S_0^2(y)$ по критерию Фишера. Если $F = \frac{S_{\text{ад}}^2}{S_0^2(y)} \leq F_{1-\alpha, f, S_{\text{ад}}^2, f, S_0^2(y)}$, то уравнение адекватно. В противном случае степень полинома увеличивают на единицу (в уравнении регрессии добавляется еще одно слагаемое). Операции повторяются до тех пор, пока не будет получено адекватное уравнение регрессии. Доверительные границы линейного уравнения регрессии вычисляются по формуле

$$\varepsilon_{\hat{y}} = t_{1-\alpha/2, f} \cdot S_{\text{ад}} \left[\frac{1}{n} + \frac{(x - \bar{x})^2}{(n-1)S^2(x)} \right]^{1/2}, \quad (11)$$

где $t_{1-\alpha/2, f}$ – коэффициент Стьюдента, $S^2(x) = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$.

Имитационные модели. Ниже предлагается использовать два вида зависимости y от x . Значения y задаются в узлах сетки $x_i = i$, $i = \overline{1, n}$; n – число узлов.

1. *Импульсная модель:* в каждом узле сетки y принимает одно из двух значений 0 или 1.

2. *Модель со случайностью:* на функциональную зависимость $\overset{\circ}{y} = \overset{\circ}{y}(x)$ накладывается эффект случайности. Значения y_i в узлах сетки рассчитываются по формуле

$$y_i = \overset{\circ}{y}(x_i) + bZ_i. \quad (12)$$

Здесь $Z_i = \left(\frac{2}{m_1} \sum_{\mu=1}^{m_1} a_{\mu} - 1 \right)_i$ – нормально распределенная случайная величина, получаемая путем суммирования вырабатываемых RND равномерно распределенных случайных чисел a_{μ} , $a_{\mu} \in [0, 1]$.

Вид функции $\overset{\circ}{y} = \overset{\circ}{y}(x)$ задается в таблице вариантов исходных данных.

Содержание работы

1. Интерполяция и сглаживание импульсной функции.

1.1. Значения импульсной функции, задаваемые в таблице вариантов, нанесите на график зависимости y_i от x_i .

1.2. Подставьте значения y_i и x_i в формулы (1), (2) и получите выражение для интерполяционного полинома Лагранжа. Убедитесь для контроля, что функция проходит через все узлы (x_i, y_i) сетки. Рассчитайте $y_{i+0,5}$ для значений x , находящихся между узлами в точках $x_{i+0,5}$, $i = 1, n-1$. Проведите через все точки графика плавную кривую.

1.3. Подставляя в уравнения (3),(4) все известные значения y_i и x_i , а также задавая $m_1 = m_n = 0$, составьте систему уравнений и найдите m_i , $i = 1, n-1$. Убедитесь в правильности вычислений, проверяя согласованность $y(x_i)$ и y_i в узлах сетки. Вычислите значения $y_{i+0,5}$ в точках $x_{i+0,5}$, нанесите эти значения на график и проведите плавную кривую. Сравните ее с предыдущей интерполяцией по формуле Лагранжа.

1.4. Прodelайте те же операции по формуле локального сплайна (5).

1.5. По тем же исходным данным постройте уравнения регрессии от линейного ($k = 1$) до степени $k = n - 2$. (При $k = n - 1$ уравнение регрессии совпадает с интерполяционным полиномом Лагранжа). Рассчитайте значения \hat{y}_{k_i} , $k = 1, n - 2$, при различных x_i , $i = 1, n$ и $x_{i+0,5}$, $i = 1, n - 1$. Результаты нанесите на график. (Чтобы не загромождать рисунок, расчеты по уравнениям регрессии можно представить на отдельном графике).

Проанализируйте все результаты. Сделайте заключение о степени приближения интерполяционных и сглаживающих функций к «результатам эксперимента», особенно между узлами сетки.

2. По формуле (12) сгенерируйте пары значений (x_i, y_i) , $x_i = i$, $i = 1, n$, для аналитической функции, суммируемой со случайной величиной. Результаты расчета представьте на графике.

2.1. По «результатам эксперимента» постройте интерполяционный полином Лагранжа, рассчитайте как и в п.1 значения $y_{i+0,5}$ и постройте на графике интерполирующую кривую.

2.2, 2.3. Выполните то же самое для глобального и локального сплайнов.

2.4. Постройте линейное уравнение регрессии.

2.4.1. По формуле (11) рассчитайте доверительные границы линейного уравнения регрессии. «Результаты эксперимента», линию регрессии и доверительные границы нанесите на один график. По формуле (10) рассчитайте дисперсию адекватности линейного уравнения регрессии. Вычислите также дисперсию воспроизводимости. Полагая, что рассеяние определяется только изменчивостью Z_i ,

найдем ее аналитически: $\sigma_0^2 = \frac{4b^2}{m_1} \sigma^2(a) = \frac{b^2}{3m_1}$, $f_{\sigma_0^2} = \infty$. Оценку адекватности выполните по критерию Фишера.

2.4.2. Если линейное уравнение окажется неадекватным, постройте квадратическое уравнение регрессии. Нанесите кривую, рассчитанную по этому уравнению, на график. Оцените адекватность квадратического уравнения. Если и оно окажется неадекватным, перейдите к кубическому и т. д.

Проанализируйте полученные результаты.

Варианты исходных данных

Вариант	Импульсная модель. Значения y_i при x_i						Функциональная модель со случайностью			
	1	2	3	4	5	6	$y = y(x)$	n	b	m_1
1	0	1	0	0	0	-	$\sin\left(\frac{\pi x}{6}\right)$	6	1	12
2	0	1	1	0	0	0	$\cos\left(\frac{\pi x}{6}\right)$	5	1	12
3	0	1	-1	0	0	0	$1 - \exp\left(-\frac{x}{2}\right)$	5	5	12

4	0	1	0	1	0	0	$1 - \exp\left(-\frac{x^2}{8}\right)$	5	0,5	12
5	0	1	0	0	1	0	$10 - \frac{x^2}{4}$	5	1,5	12

Литература: [2], раздел 12, [1], [6].

**ПЛАНИРОВАНИЕ АКТИВНОГО ЭКСПЕРИМЕНТА ПРИ
ИЗУЧЕНИИ ПОВЕРХНОСТИ ОТКЛИКА**

**Часть I. Планирование эксперимента при построении
функции отклика**

Построим план активного эксперимента для нахождения функции отклика $\hat{y} = \hat{y}(x_j, j = \overline{1, k^*})$, где k^* – число контролируемых факторов. Проведем n опытов, задавая в каждом из них значения факторов x_{ji} на разных уровнях, $i = \overline{1, n}$. Перейдем от x_{ji} , представляющих собой величины разных размерностей, к безразмерным кодированным переменным по формуле

$$\tilde{x}_{ji} = \frac{2x_{ji} - (x_{j\max} + x_{j\min})}{x_{j\max} - x_{j\min}}. \quad (1)$$

Здесь $x_{j\max}$ и $x_{j\min}$ – верхняя и нижняя границы интервалов варьирования каждого фактора. При $x_{ji} = x_{j\max}$ кодированное значение $\tilde{x}_{ji} = +1$, при $x_{ji} = x_{j\min}$ значение $\tilde{x}_{ji} = -1$. Если ограничиться построением линейной зависимости, то логично уровни факторов задать равными этим крайним значениям: $\tilde{x}_{ji} = +1, -1$. Для построения нелинейной зависимости число уровней должно быть увеличено.

Будем искать функцию отклика в виде полинома, отражающего линейную зависимость от факторов и учитывающего взаимодействия между ними (парные, тройные и более высокого порядка)

$$\begin{aligned} \hat{y} = & b_0 + b_1\tilde{x}_1 + b_2\tilde{x}_2 + \dots + b_{k^*}\tilde{x}_{k^*} + b_{12}\tilde{x}_1\tilde{x}_2 + b_{13}\tilde{x}_1\tilde{x}_3 + \dots + b_{1k^*}\tilde{x}_1\tilde{x}_{k^*} + \\ & + b_{23}\tilde{x}_2\tilde{x}_3 + \dots + b_{2k^*}\tilde{x}_2\tilde{x}_{k^*} + b_{123}\tilde{x}_1\tilde{x}_2\tilde{x}_3 + \dots \end{aligned} \quad (2)$$

Для унификации записи слагаемые с взаимодействиями будем рассматривать как дополнительные модельные факторы, обозначая их следующим образом: $b_{12}\tilde{x}_1\tilde{x}_2 = b_{k^*+1}\tilde{x}_{k^*+1}$,

$b_{13}\tilde{x}_1\tilde{x}_3 = b_{k^*+2}\tilde{x}_{k^*+2}$ и т. д. до $b_k\tilde{x}_k$. Введем также виртуальный фактор

$\tilde{x}_0 \equiv 1$. Таким образом, полином содержит $k+1$ слагаемых. Теперь функцию отклика можно записать в виде

$$\hat{y} = \sum_{j=0}^k b_j \tilde{x}_j. \quad (3)$$

Составим матрицу плана полного факторного эксперимента $m = 2^{k^*}$. В этой матрице столбцами являются \tilde{x}_j , а строками – уровни факторов в i -том испытании. Потребуем, чтобы такая матрица удовлетворяла условиям центральности и ортогональности

$$\sum_{i=1}^n \tilde{x}_{ji} = 0, \quad (4)$$

$$\sum_{i=1}^n \tilde{x}_{ji}\tilde{x}_{ui} = 0, \quad j, u = \overline{1, k}, \quad j \neq u. \quad (5)$$

по всем столбцам (кроме \tilde{x}_0). В этом случае коэффициенты полинома вычисляются независимо друг от друга по формуле

$$b_j = \frac{\sum_{i=1}^n \tilde{x}_{ji} y_i}{\sum_{i=1}^n \tilde{x}_{ji}^2}. \quad (6)$$

Пример простейшего плана ПФЭ $n = 2^2 = 4$ представлен в табл. 1.

Таблица 1

i	\tilde{x}_0	\tilde{x}_1	\tilde{x}_2	$\tilde{x}_3 = \tilde{x}_1 \tilde{x}_2$	y
1	+	+	+	+	y_1
2	+	+	-	-	y_2
3	+	-	+	-	y_3
4	+	-	-	+	y_4

Знаки +, - означают +1, -1, соответственно.

Функция отклика, которая может быть получена в результате реализации этого плана

$$\hat{y} = b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 + b_{12} \tilde{x}_1 \tilde{x}_2. \quad (7)$$

В случае трех факторов формула плана $n = 2^3 = 8$. При реализации такого плана может быть получена функция отклика вида

$$\hat{y} = b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 + b_3 \tilde{x}_3 + b_{12} \tilde{x}_1 \tilde{x}_2 + b_{13} \tilde{x}_1 \tilde{x}_3 + b_{23} \tilde{x}_2 \tilde{x}_3 + b_{123} \tilde{x}_1 \tilde{x}_2 \tilde{x}_3. \quad (8)$$

Функции отклика, подобные (7), (8), в узлах сетки в точности совпадают со значениями y_i , включая погрешности измерений. Чтобы получить сглаженную функцию отклика, следует исключить из нее (приравнять нулю) слагаемые с незначимыми коэффициентами b_j . Проверка коэффициентов на значимость осуществляется по решающему правилу

$$|b_j| \leq \varepsilon_{b_j} \rightarrow H_0 : \text{коэффициент незначим}, |b_j| > \varepsilon_{b_j} \rightarrow H_1 : \text{коэффициент значим}, \quad (9)$$

где $\varepsilon_{b_j} = t_{1-\alpha/2, f_0} \cdot S_0(y) / \sum_{i=1}^n \tilde{x}_{ji}^2$.

После исключения слагаемых с незначимыми коэффициентами функцию отклика следует проверить на адекватность. С этой целью рассчитывается дисперсия адекватности

$$S_{\text{ад}}^2 = \frac{1}{n-l} \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (10)$$

где l - число оставшихся коэффициентов в функции отклика. Число степеней свободы дисперсии адекватности $f_{S_{\text{ад}}}^2 = n-l$. Далее дисперсия адекватности сравнивается с дисперсией воспроизводимости $S_0^2(y)$. Уравнение считается адекватным, если отношение Фишера

$$F = S_{\text{ад}}^2 / S_0^2(y) < F_{1-\alpha, f_{S_{\text{ад}}}^2, f_{S_0^2}^2}. \quad (11)$$

Функция отклика может быть проверена на адекватность также путем постановки дополнительных опытов, например, в центре плана при всех $\tilde{x}_j = 0$. Если разность между измеренным и рассчитанным значениями в центре плана окажется недопустимо большой, то следует перейти к построению нелинейной (параболической) функции отклика. При этом, чтобы сохранить свойства центральности и ортогональности, доработка плана до нелинейного осуществляется проведением опытов в центре факторного пространства и в звездных точках - точках, расположенных на осях факторов. Длина звездного плеча l^* зависит от числа факторов. Приведем несколько значений

k^*	2	3	4	5
l^*	1,000	1,215	1,414	1,547

Таким образом, формула плана для построения квадратической функции отклика имеет вид $m = 2^{k^*} + 2k^* + 1$.

Нелинейный план для двух факторов, ядром которого является линейный план $n = 2^2$, приведен в табл. 2.

Таблица 2

i	\tilde{x}_0	\tilde{x}_1	\tilde{x}_2	$\tilde{x}_3 = \frac{\tilde{x}_1 \tilde{x}_2}{\tilde{x}_1 \tilde{x}_2}$	$\tilde{x}_4 = \frac{\tilde{x}_1^2 - \tilde{x}_2^2}{\tilde{x}_1^2 - \tilde{x}_2^2}$	$\tilde{x}_5 = \frac{\tilde{x}_2^2 - \tilde{x}_1^2}{\tilde{x}_2^2 - \tilde{x}_1^2}$	$\tilde{x}_6 = \frac{\tilde{x}_4 \tilde{x}_5}{\tilde{x}_4 \tilde{x}_5}$	$\tilde{x}_7 = \frac{\tilde{x}_4 \tilde{x}_2}{\tilde{x}_4 \tilde{x}_2}$	$\tilde{x}_8 = \frac{\tilde{x}_5 \tilde{x}_1}{\tilde{x}_5 \tilde{x}_1}$	y
1	+	+	+	+	1/3	1/3	1/9	1/3	1/3	y_1
2	+	+	-	-	1/3	1/3	1/9	-1/3	1/3	y_2
3	+	-	+	-	1/3	1/3	1/9	1/3	-1/3	y_3
4	+	-	-	+	1/3	1/3	1/9	-1/3	-1/3	y_4
5	+	0	0		-2/3	-2/3	4/9	0	0	y_5
6	+	+	0		1/3	-2/3	-2/9	0	-2/3	y_6
7	+	-	0		1/3	-2/3	-2/9	0	2/3	y_7
8	+	0	+		-2/3	1/3	-2/9	-2/3	0	y_8
9	+	0	-		-2/3	1/3	-2/9	2/3	0	y_9

Ортогональность плана, представленного в этой таблице, достигается введением квадратических слагаемых $\tilde{x}_4 = \tilde{x}_1^2 - \tilde{x}_2^2$ и $\tilde{x}_5 = \tilde{x}_2^2 - \tilde{x}_1^2$. Как легко убедиться, $\overline{\tilde{x}_1^2} = \overline{\tilde{x}_2^2} = 2/3$. В столбцах $\tilde{x}_6, \tilde{x}_7, \tilde{x}_8$ представлены взаимодействия линейных и нелинейных переменных. Полная нелинейная функция отклика, которая может быть получена после реализации этого плана, имеет вид

$$\hat{y} = b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 + b_3 \tilde{x}_1 \tilde{x}_2 + b_4 \left(\tilde{x}_1^2 - \tilde{x}_2^2 \right) + b_5 \left(\tilde{x}_2^2 - \tilde{x}_1^2 \right) + b_6 \left(\tilde{x}_1^2 - \tilde{x}_2^2 \right) \left(\tilde{x}_2^2 - \tilde{x}_1^2 \right) + b_7 \left(\tilde{x}_1^2 - \tilde{x}_2^2 \right) \tilde{x}_2 + b_8 \left(\tilde{x}_2^2 - \tilde{x}_1^2 \right) \tilde{x}_1. \quad (12)$$

Коэффициенты b_j рассчитываются, как и ранее, по формуле (6). Проверка коэффициентов на значимость и функции отклика на адекватность выполняется описанными выше способами.

Имитационная модель. Зададим неизвестную экспериментатору зависимость y от x_j выражением

$$y_i = y_0 + x_{1i} \sin^2 \left(\frac{\pi x_{1i}}{30} \right) \sin \left(\frac{\pi x_{2i}}{30} \right) \sin \left[\left(\frac{\pi x_{3i}}{20} \right) (1 - 0,01(x_{1i} + x_{2i})) \right] + Z_i. \quad (13)$$

Здесь Z – случайная нормально распределенная величина, отражающая влияние неучтенных факторов, в частности, погрешности измерения,

$$Z_i = \frac{1}{2} \left(\sum_{\mu=1}^{m_1} a_{\mu} - \frac{m_1}{2} \right)_i, \quad (14)$$

где $a_{\mu}, \mu = \overline{1, m_1}$ – равномерно распределенные случайные числа ($0 \leq a_{\mu} \leq 1$), вырабатываемые RND.

Дисперсия воспроизводимости для отклика может быть оценена без проведения повторных опытов при фиксированных уровнях факторов, поскольку в этом конкретном случае рассеяние y определяется только случайной величиной Z . Поэтому $\sigma_0^2(y) = \sigma_0^2(Z)$. Напомним, что $\sigma^2(a) = 1/12$, откуда

$$\sigma^2(Z) = \frac{1}{4} m_1 \sigma^2(a) = \frac{m_1}{48}. \quad (15)$$

Содержание работы

1. Найдите линейную функцию отклика для двух факторов x_1 и x_2 , принимая x_3 постоянным, например, $x_3 = 10$. С этой целью постройте план ПФЭ $n = 2^2$ самостоятельно или воспользуйтесь табл. 1. Диапазоны изменений факторов в эксперименте заданы в таблице вариантов исходных данных. "Реализуйте" этот план, подставляя в имитационную модель (13) значения $x_{j\text{макс}}$ и $x_{j\text{мин}}$ в соответствии с табл. 1. Вычислите коэффициенты функции отклика (7) и проверьте их на значимость по формуле (9), принимая $S_0^2(y) = \sigma^2(Z)$, а $f_0 = \infty$. Если какие-то из коэффициентов оказались незначимы, то проверьте адекватность функции отклика после их исключения.

"Поставьте" пятый опыт в центре плана и найдите разность $(y_5 - \hat{y}_5)$. Если эта разность окажется больше доверительных границ погрешности измерения $\varepsilon_{\bar{y}} = z_{1-\alpha/2} \cdot \sigma(Z) / \sqrt{n}$, то перейдите к построению нелинейного плана.

2. Достройте линейный план до нелинейного (квадратического) $n = 2^2 + 2 \cdot 2 + 1$. Реализуйте этот план, выполнив опыты в звездных точках. Вычислите коэффициенты функции отклика (12). Проверьте их на значимость (см. п. 1) и оцените адекватность нелинейной функции отклика. Изобразите поверхность отклика графически. С этой целью проведите координатные оси x_1 и x_2 под острым углом ($\approx 45^\circ$) друг к другу. Точку пересечения осей примите за центр плана ($\tilde{x}_1 = \tilde{x}_2 = 0$). Отложите на осях значения $+1, -1$. Проведите через центр плана вертикальную ось y . Отложите значения y_i и \hat{y}_i в каждом из девяти узлов построенной горизонтальной сетки. Соедините каждые три точки \hat{y}_i , лежащие в вертикальных плоскостях параллельных x_1 и x_2 , плавными кривыми.

3. Составьте линейный план ПФЭ $n = 2^3 = 8$, в котором предусматривается одновременное изменение уровней всех трех факторов. Реализуйте его на основе имитационной модели (13). Найдите коэффициенты функции отклика (8). Далее, как и выше, проверьте значимость коэффициентов функции отклика и ее адекватность.

4. Совершите подвиг Геракла. Составьте нелинейный план $n = 2^3 + 2 \cdot 3 + 1$, определите вид функции отклика, которая может быть получена на его основе. Реализуйте план и выполните все операции, связанные с оценками значимости коэффициентов и адекватности полученной функции отклика.

Проанализируйте все полученные результаты.

Таблица 3

Варианты исходных данных

Вариант	$x_{1\text{мин}}$	$x_{1\text{макс}}$	$x_{2\text{мин}}$	$x_{2\text{макс}}$	$x_{3\text{мин}}$	$x_{3\text{макс}}$
1	5	15	5	15	5	15
2	5	15	10	20	5	15
3	10	20	10	20	10	20
4	10	20	5	15	10	20
5	15	25	10	20	5	15
6	15	25	15	25	10	20

Во всех вариантах $y_0 = 5$, $m_1 = 12$.

Литература: [2], раздел 13, [1].

ПЛАНИРОВАНИЕ АКТИВНОГО ЭКСПЕРИМЕНТА ПРИ ИЗУЧЕНИИ ПОВЕРХНОСТИ ОТКЛИКА

Часть II. Планирование эксперимента при поиске экстремума

Часто задачей эксперимента, особенно на начальном этапе, является не столько описание поверхности отклика во всем факторном пространстве, сколько нахождение такого сочетания уровней факторов, при котором исследуемый отклик принимает максимальное (или минимальное) значение. Например, нахождение наибольшей глубины водоема, зоны максимальной отражаемости в градовом облаке, параметров измерительного преобразователя, обеспечивающих его устойчивую работу, и т. д.

Простейший, но и самый неэффективный путь решения такой задачи состоит в использовании приемов однофакторного эксперимента. При этом для достижения цели требуется провести большое число испытаний. Более эффективными являются планы многофакторного эксперимента, при осуществлении которых одновременно и согласованно изменяются уровни всех факторов. Рассмотрим два возможных варианта такого подхода.

1. Крутое восхождение или движение по градиенту.

Представим себе бесхитростного, но сильного гнома, которому приказано как можно быстрее вскарабкаться на вершину горы. Какой путь выберет гном? По-видимому, самый короткий, то есть по самому крутому склону. Этот пример объясняет смысл образного термина «крутое восхождение». При проведении многофакторного эксперимента визуальный выбор «крутого склона» оказывается невозможным. Формализация решения этой задачи состоит в том, что сначала в окрестности исходной точки находят направление максимального изменения отклика, то есть направление градиента. Затем, перемещаясь в факторном пространстве по этому направлению, ставят новые опыты.

Процедура нахождения градиента заключается в следующем. Пусть объект исследования в некоторый момент характеризуется сочетанием уровней факторов (исходных уровней) $\overset{\circ}{x}_j$, $j = \overline{0, k^*}$, k^*

— число варьируемых факторов. Сочетание уровней $\overset{\circ}{x}_j$ определяет точку в факторном пространстве. Обозначим ее $\overset{\circ}{0}$. Зададим вокруг точки $\overset{\circ}{0}$ небольшое (локальное) факторное пространство. Границы этого пространства определим как $x_{j(l)} = \overset{\circ}{x}_j \pm \Delta_j$, где Δ_j — интервал варьирования, $\Delta_j = (x_{j\max(l)} - x_{j\min(l)})/2$. Индекс (л) означает локальность пространства.

Если выполнить кодирование факторов по формуле

$$\tilde{x}_{ji(l)} = \left(x_{ji(l)} - \overset{\circ}{x}_j \right) / \Delta_j, \quad (1)$$

то $\tilde{x}_{j\max(l)} = +1$, а $\tilde{x}_{j\min(l)} = -1$. Далее индекс (л) опустим.

Составим линейный план полного факторного эксперимента $\overset{\circ}{n} = 2^{k^*}$, отвечающего условиям центральности и ортогональности, как в предыдущем задании. Реализуем этот план — выполним измерения исследуемой величины y_i , $i = \overline{1, \overset{\circ}{n}}$ во всех точках плана. Результаты измерения используются для построения линейной функции отклика вида

$$\hat{y} = b_0 + b_1 \tilde{x}_1 + b_2 \tilde{x}_2 + \dots + b_{k^*} \tilde{x}_{k^*}. \quad (2)$$

Коэффициенты функции отклика вычисляются по формуле

$$b_j = \frac{\sum_{i=1}^{\overset{\circ}{n}} \tilde{x}_{ji} y_i}{\sum_{i=1}^{\overset{\circ}{n}} \tilde{x}_{ji}^2}. \quad (3)$$

Градиент функции отклика

$$\text{grad}\hat{y} = \sum_{j=1}^{k^*} \frac{\partial \hat{y}}{\partial \tilde{x}_j} \frac{\tilde{x}_j}{|\tilde{x}_j|}, \quad (4)$$

где $\frac{\tilde{x}_j}{|\tilde{x}_j|}$ – единичный вектор.

Поскольку $\frac{\partial \hat{y}}{\partial \tilde{x}_j} = b_j$, модуль градиента

$$|\text{grad}\hat{y}| = \sqrt{b_1^2 + b_2^2 + \dots + b_{k^*}^2}, \quad (5)$$

направление градиента определяется знаками коэффициентов b_j .

Начнем движение по градиенту. Выполним $\binom{\circ}{n+1}$ -ый опыт в точке с координатами

$$\tilde{x}_{j\binom{\circ}{n+1}} = \tilde{d}_{j\binom{\circ}{n+1}}, \quad j = \overline{1, k^*}. \quad (6)$$

или для размерных переменных

$$x_{j\binom{\circ}{n+1}} = x_j + \tilde{d}_{j\binom{\circ}{n+1}} \Delta_j, \quad (6a)$$

где $\tilde{d}_{j\binom{\circ}{n+1}}$ – расстояние по оси j от центра плана до новой точки. Зададим $\tilde{d}_{j\binom{\circ}{n+1}}$ таким образом,

чтобы выйти за пределы локального факторного пространства. Очевидно, что все \tilde{d}_j связаны между собой. Выберем в качестве опорного фактор, для которого $|b_j|$ является достаточно большим (не обязательно максимальным). Пусть, например, x_1 – опорный фактор. Тогда

$$\tilde{d}_{ji} = \frac{b_j}{b_1} \tilde{d}_{1i}. \quad (7)$$

Формула (7) справедлива для движения по градиенту к максимуму. Если целью эксперимента является поиск минимума отклика, то знаки коэффициентов b_j следует изменить на обратные. Ниже для определенности будем говорить о поиске максимума отклика.

Итак, рассчитаем все $\tilde{d}_{j\binom{\circ}{n+1}}$ и $x_{j\binom{\circ}{n+1}}$. Выполним эксперимент в этой точке и сравним полу-

ченное значение $y_{\binom{\circ}{n+1}}$ со всеми значениями y_i . Если $y_{\binom{\circ}{n+1}} > y_i$, $i = \overline{1, n}$, то зададим координаты

следующей точки в направлении градиента и выполним в ней измерение $y_{\binom{\circ}{n+2}}$. Если

$y_{\binom{\circ}{n+2}} > y_{\binom{\circ}{n+1}}$, то продолжим движение по градиенту. Когда $y_{\binom{\circ}{n+l+1}}$ окажется меньше $y_{\binom{\circ}{n+l}}$,

следует вокруг точки $\binom{\circ}{n+l}$ задать новое локальное факторное пространство, получить новую функцию отклика (уже с другими коэффициентами b_j), найти новое направление градиента и про-

UNIT 1: THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1.1 THE HISTORY OF THE WORLD

1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

1.1 THE HISTORY OF THE WORLD

UNIT 2: THE HISTORY OF THE WORLD

UNIT 2

Year	Event	Location	Significance
1492	Columbus discovered America	North America	Start of European colonialism
1776	Declaration of Independence	Philadelphia, USA	Birth of the United States
1848	Revolution in France	Paris, France	End of the Bourbon monarchy
1914	Start of World War I	Europe	Global conflict
1945	End of World War II	Germany	Victory of the Allies
1949	Founding of the People's Republic of China	China	End of the Republic of China
1979	Revolutions in Iran and China	Iran and China	End of the Cold War
1989	Fall of the Berlin Wall	Berlin, Germany	End of the Cold War
1991	Dissolution of the Soviet Union	Soviet Union	End of the Soviet Union
2001	9/11 attacks	New York, USA	Start of the War on Terror
2008	Global Financial Crisis	Global	Global economic downturn
2011	Arab Spring	Middle East	End of authoritarianism
2015	Paris Agreement	Paris, France	Global climate agreement
2020	COVID-19 pandemic	Global	Global health crisis

Далее легко рассчитать по формуле (3) коэффициенты линейной функции отклика вида (2). Заметим, что $\sum_{i=1}^{\overset{\circ}{n}} \tilde{x}_{0i}^2 = \overset{\circ}{n}$, а $\sum_{i=1}^{\overset{\circ}{n}} \tilde{x}_{ji}^2 = 0,5$, $j = \overline{1, k^*}$.

Легко видеть, что для получения линейной функции отклика при симплексном планировании число опытов оказывается существенно меньше, чем в планах ПФЭ. Платой за это является меньшая точность функции отклика. Для симплексного плана $S_{b_j}^2 = 2S_0^2(y)$.

Теперь перейдем к использованию симплексных планов для нахождения экстремума. Воспользуемся следующим свойством симплекса. Если любую из вершин симплекса зеркально отразить относительно противоположащей грани, то новая точка вместе с оставшимися образует новый симплекс той же размерности. Это означает, что для перемещения симплекса в пространстве достаточно перенести одну из его вершин. Будем перемещать симплекс в сторону увеличения отклика. Сравнивая значения отклика в вершинах исходного симплекса, выделим «худшую» точку, в которой y_i имеет наименьшее значение. Найдем ее зеркальное отражение, чтобы построить новый симплекс. Координаты этой точки вычисляются по формуле.

$$x_{j \left(\overset{\circ}{n}+1 \right)} = \frac{2}{k^*} \sum_{\substack{i=1 \\ i \neq u}}^{\overset{\circ}{n}} x_{ji} - x_{ju}, \quad (9)$$

где i – номер вершины симплекса, $\overset{\circ}{n} = k^* + 1$ – число вершин, u – номер худшей точки (вершины), x_{ju} – j -тая координата худшей точки, $j = \overline{1, k^*}$. Обратите внимание на то, что здесь x_j – некодированные уровни факторов.

Далее процедура заключается в том, что в новом симплексе снова находят худшую точку и ее отражение. Операции повторяются до тех пор, пока в очередной отраженной точке значение y_i не окажется меньше, чем в отражаемой. Чтобы убедиться в том, что максимум отклика достигнут, можно отразить следующую худшую точку. Если значение y_i снова окажется меньше, то и по этой координате максимум достигнут. Процедура завершается нахождением центра последнего симплекса. Координаты этого центра ближе всего отражают положение максимума отклика.

$$x_{j\text{ц}} = \sum_{v=1}^{\overset{\circ}{n}} x_{jv}, \quad (10)$$

где v – номера вершин последнего симплекса, $v = \overline{1, \overset{\circ}{n}}$. Остается найти $y_{\text{ц}}$, поставив опыт в центре этого симплекса.

Имитационная модель. Используем здесь ту же модель, что и в предыдущем задании. Чтобы не создавать дополнительные трудности, исключим из нее слагаемое, отражающее влияние случайных факторов. Таким образом,

$$y_i = y_0 + x_{1i} \sin^2 \left(\frac{\pi x_{1i}}{30} \right) \sin \left(\frac{\pi x_{2i}}{30} \right) \sin \left[\left(\frac{\pi x_{3i}}{20} \right) (1 - 0,01(x_{1i} + x_{2i})) \right]. \quad (11)$$

Содержание работы

1. Найдите максимальное значение исследуемой величины $y_{\text{макс}}$ и его координаты x_j , $j = \overline{1, k^*}$ методом крутого восхождения. «Результаты эксперимента» получите по имитационной модели (11). Координаты точки x_j и интервалы варьирования Δ_j заданы в таблице вариантов исходных данных (табл. 2).

1.1. Выполните крутое восхождение, изменяя только факторы x_1 и x_2 . Уровень фактора x_3 задайте постоянным и равным x_3^0 . Сначала выполните кодирование уровней факторов и составьте локальный линейный план ПФЭ $n = 2^2$. Найдите значения $y_i, i = \overline{1,4}$ в точках плана.

Вычислите коэффициенты функции отклика вида (2) при $k^* = 2$. Знаки коэффициентов b_1 и b_2 определяют направление градиента. Сделайте первый шаг по градиенту. Для этого определите координаты пятой ($4+1=5$) точки, лежащей за пределами локального плана. Задайте $\tilde{d}_{15} \geq 1$, а \tilde{d}_{25} рассчитайте по формуле (7). По формуле (6а) перейдите к некодированным координатам и, подставляя их в выражение (11), найдите значение y_5 . Аналогичным образом получите y_6 . Если $y_6 > y_5$, поставьте опыт в точке 7 и т. д. Когда получите $y_{(4+l+1)} < y_{(4+l)}$, вернитесь к точке с координатами $x_{1(4+l)}, x_{2(4+l)}$ и постройте вокруг нее новый локальный план. Реализуйте его, вычислите коэффициенты новой функции отклика, снова определите направление градиента, продолжайте опыты в этом направлении и т. д. до тех пор, пока не будет достигнут максимум отклика (коэффициенты b_1 и b_2 новой локальной функции отклика окажутся близкими к нулю).

1.2. Выполните движение по градиенту в трехмерном пространстве. С этой целью постройте план ПФЭ $n = 2^3$. Реализуйте его и найдите коэффициенты функции отклика (2) при $k^* = 3$. Далее осуществите движение по градиенту, как это описано выше.

2. Из той же исходной точки $x_j, j = \overline{1, k^*}$ осуществите поиск экстремума (максимума) методом симплекс-планирования.

2.1. Постройте двухфакторный симплекс (равносторонний треугольник) с центром в исходной точке. Задайте кодированные координаты вершин этого треугольника в соответствии с данными табл. 1. Перейдите от кодированных уровней факторов к некодированным x_{ji} . Реализуйте план и найдите худшую точку. Перенесите эту точку в факторном пространстве, отражая ее зеркально относительно противоположной стороны треугольника. Вычислите координаты отраженной точки, используя формулу (9). Найдите значение $y_{\left(\begin{smallmatrix} \circ \\ n+1 \end{smallmatrix} \right)}$. Найдите худшую точку в новом симплексе и отразите ее в фак-

торном пространстве и т.д. Процедуру повторяйте до тех пор, пока не будет достигнут максимум (пока в отраженной точке значение отклика не окажется меньше, чем в отражаемой). Найдите центр последнего симплекса (см. формулу (10)) и значение $y_{\text{макс}}$ в этом центре.

2.2. Постройте симплекс для трехфакторного пространства (тетраэдр). По табл. 1 определите кодированные координаты вершин симплекса, перейдите к некодированным координатам и реализуйте план эксперимента. Определите наихудшую точку и, отражая ее, постройте новый симплекс. Найдите значение отклика в этой новой точке. Далее повторяйте операции как в п. 2.1.

Проанализируйте все полученные результаты.

Таблица 2

Варианты исходных данных

Варианты	x_1	x_2	x_3
1	5	5	5
2	5	10	5
3	5	15	5
4	10	5	10
5	10	15	10
6	20	5	20

Во всех вариантах $\Delta_1 = \Delta_2 = 2, \Delta_3 = 1, y_0 = 5$.

Литература: [2] раздел 13, [1].

ПОСТРОЕНИЕ ДИСКРИМИНАНТНОЙ ФУНКЦИИ

Назовем *дискриминантной функцией* уравнение поверхности, разделяющей многофакторное пространство на две области, в одной из которых вероятность осуществления случайного события больше, чем вероятность его отсутствия, а в другой наоборот. В случае двухфакторного пространства такая поверхность редуцируется в линию — прямую или кривую, а однофакторного — в точку. В метеорологической практике дискриминантная функция используется для выделения и прогнозирования условий, при которых может произойти то или иное опасное явление погоды (гроза, град, гололед и т. п.).

Рассмотрим принципы построения дискриминантной функции. Предположим, что ведутся наблюдения за некоторым явлением погоды G . Будем обозначать факт наблюдения этого явления 1, а отсутствия — 0. Возникновение явления G зависит от большого числа факторов, контролируемых и неконтролируемых, а может быть и неизвестных. Пусть k — число контролируемых факторов.

Результаты совместных измерений этих факторов представлены рядами x_{ji} , $j = \overline{1, k}$, $i = \overline{1, n}$, n — длина рядов. Влияние неконтролируемых факторов приводит к тому, что зависимость G от x_j становится неоднозначной, появляется случайный шум. Задача заключается в том, чтобы построить функцию $L = L(x_j, j = \overline{1, k})$, которая наилучшим образом отделяла бы условия возникновения явления от условий его отсутствия.

Существуют различные методы построения таких функций. Остановимся здесь только на одном простом, но достаточно эффективном методе. Суть его заключается в том, что дискриминантная функция L находится как разность двух функций отклика, одна из которых отражает повторяемость появления события, а другая — его отсутствие.

Обратимся к результатам пассивного эксперимента, когда значения x_{ji} изменяются произвольно от $x_{j\max}$ до $x_{j\min}$. Найдем центр факторного пространства как точку $(x_{j1/2}, j = \overline{1, k})$, где $x_{j1/2} = (x_{j\max} + x_{j\min})/2$. Далее процедура зависит от того, какой вид функций отклика, а следовательно, и дискриминантной функции, мы собираемся получить. Обычно ограничиваются линейной моделью. В этом случае все значения x_{ji} достаточно сгруппировать относительно двух уровней факторов $x_{j1/4}$ и $x_{j3/4}$, которые можно получить как $x_{j1/4} = (x_{j\min} + x_{j1/2})/2$ и $x_{j3/4} = (x_{j1/2} + x_{j\max})/2$.

Перейдем к кодированным — центрированным и безразмерным — переменным

$$\tilde{x}_{ji} = \frac{2x_{ji} - (x_{j3/4} + x_{j1/4})}{x_{j3/4} - x_{j1/4}} \quad (1)$$

В соответствии с формулой (1) $\tilde{x}_{j1/4} = -1$, а $\tilde{x}_{j3/4} = +1$. Таким образом после группировки и кодирования все $x_{ji} > x_{j1/2}$ приобретут значения $\tilde{x}_{ji} = \tilde{x}_{j3/4} = +1$, а все $x_{ji} < x_{j1/2}$ превратятся в $\tilde{x}_{ji} = \tilde{x}_{j1/4} = -1$. Если имеется несколько значений $x_{ji} = x_{j1/2}$, то одной половине из них присваиваются значения +1, а другой —1.

Далее можно перейти к построению функции отклика на основе линейных планов полного факторного эксперимента (ПФЭ) 2^k . Для этого следует рассчитать число случаев возникновения явления и число случаев его отсутствия при различных сочетаниях уровней факторов. Технику вычислений проиллюстрируем на примере трехфакторного эксперимента. В табл. 1 приведены результаты измерений x_{ji} с отметкой в каждой строке факта наличия явления $y_i = 1$ или его отсутствия $y_i = 0$. В соседних столбцах таблицы приведены кодированные значения \tilde{x}_{ji} .

Таблица 1

Результаты измерений и их кодирование

i	x_{1i}	\tilde{x}_{1i}	x_{2i}	\tilde{x}_{2i}	x_{3i}	\tilde{x}_{3i}	y_i
1	x_{11}	+1	x_{21}	-1	x_{31}	-1	0
2	x_{12}	-1	x_{22}	+1	x_{32}	-1	1
3	x_{13}	+1	x_{23}	+1	x_{33}	+1	1
...
n	x_{1n}	+1	x_{2n}	+1	x_{3n}	-1	0

Сосчитаем число нулей и единиц в столбце y_i при каждом из сочетаний кодированных уровней факторов \tilde{x}_{ji} , т.е. при сочетании уровней факторов +++, ++-, и т. д. (для краткости записи единицы опущены). Общее число таких сочетаний для трехфакторного эксперимента $N = 2^3 = 8$. Результаты сведем в табл. 2.

Таблица 2

Представление результатов эксперимента матрицей плана ПФЭ 2^3

ν	\tilde{x}_0	\tilde{x}_1	\tilde{x}_2	\tilde{x}_3	$\tilde{x}_4 = \tilde{x}_1 \cdot \tilde{x}_2$	$\tilde{x}_5 = \tilde{x}_1 \cdot \tilde{x}_3$	$\tilde{x}_6 = \tilde{x}_2 \cdot \tilde{x}_3$	$\tilde{x}_7 = \tilde{x}_1 \cdot \tilde{x}_2 \cdot \tilde{x}_3$	$y^{(1)}$	$y^{(0)}$
1	+	+	+	+	+	+	+	+	$y_1^{(1)}$	$y_1^{(0)}$
2	+	+	+	-	+	-	-	-	$y_2^{(1)}$	$y_2^{(0)}$
3	+	+	-	+	-	+	-	-	$y_3^{(1)}$	$y_3^{(0)}$
4	+	+	-	-	-	-	+	+	$y_4^{(1)}$	$y_4^{(0)}$
5	+	-	+	+	-	-	+	-	$y_5^{(1)}$	$y_5^{(0)}$
6	+	-	+	-	-	+	-	+	$y_6^{(1)}$	$y_6^{(0)}$
7	+	-	-	+	+	-	-	+	$y_7^{(1)}$	$y_7^{(0)}$
8	+	-	-	-	+	+	+	-	$y_8^{(1)}$	$y_8^{(0)}$

Присвоим значениям в строках таблицы индекс $\nu = \overline{1,8}$. Столбцы $\tilde{x}_1, \tilde{x}_2, \tilde{x}_3$ отражают физические уровни кодированных факторов. Каждая строка отражает сочетание уровней этих факторов. Столбцы $\tilde{x}_4, \dots, \tilde{x}_7$ учитывают взаимодействие факторов, \tilde{x}_0 – формальный фактор, которому присваивается постоянное значение +1.

Табл. 2 используется для построения функций отклика

$$\hat{y}^{(1)} = b_0^{(1)} + b_1^{(1)} \tilde{x}_1 + b_2^{(1)} \tilde{x}_2 + b_3^{(1)} \tilde{x}_3 + b_{12}^{(1)} \tilde{x}_1 \tilde{x}_2 + b_{13}^{(1)} \tilde{x}_1 \tilde{x}_3 + b_{23}^{(1)} \tilde{x}_2 \tilde{x}_3 + b_{123}^{(1)} \tilde{x}_1 \tilde{x}_2 \tilde{x}_3, \quad (2)$$

$$\hat{y}^{(0)} = b_0^{(0)} + b_1^{(0)} \tilde{x}_1 + b_2^{(0)} \tilde{x}_2 + b_3^{(0)} \tilde{x}_3 + b_{12}^{(0)} \tilde{x}_1 \tilde{x}_2 + b_{13}^{(0)} \tilde{x}_1 \tilde{x}_3 + b_{23}^{(0)} \tilde{x}_2 \tilde{x}_3 + b_{123}^{(0)} \tilde{x}_1 \tilde{x}_2 \tilde{x}_3. \quad (3)$$

В формулах (2), (3) взаимодействия выписаны в явном виде. Далее будем использовать сокращенную запись $b_{12} = b_4$, а $\tilde{x}_1 \tilde{x}_2 = \tilde{x}_4$, $b_{13} = b_5$, $\tilde{x}_1 \tilde{x}_3 = \tilde{x}_5$ и т. д. в соответствии с обозначениями табл. 2. Коэффициенты b_j , $j = \overline{0,7}$ функций отклика (2), (3) вычисляются по формуле

$$b_j = \frac{\sum_{\nu=1}^8 \tilde{x}_{j\nu} \cdot y_\nu}{\sum_{\nu=1}^8 \tilde{x}_{j\nu}^2}. \quad (4)$$

Теперь легко находится дискриминантная функция

$$L = \hat{y}^{(1)} - \hat{y}^{(0)} = c_0 + c_1 \tilde{x}_1 + c_2 \tilde{x}_2 + c_3 \tilde{x}_3 + c_{12} \tilde{x}_1 \tilde{x}_2 + c_{13} \tilde{x}_1 \tilde{x}_3 + c_{23} \tilde{x}_2 \tilde{x}_3 + c_{123} \tilde{x}_1 \tilde{x}_2 \tilde{x}_3, \quad (5)$$

где $c_j = b_j^{(1)} - b_j^{(0)}$.

Подставляя в выражение (5) $L = 0$, получим условия равенства повторяемости наличия и отсутствия явления G . При $L > 0$ вероятность развития явления больше, чем его отсутствия, при $L < 0$ – наоборот.

Формальный анализ требует проверки всех коэффициентов b_j в выражениях (2), (3) на значимость. Поскольку данные, необходимые для вычисления дисперсий воспроизводимости, отсутствуют, оценку значимости можно осуществить, сравнивая b_j (кроме $j = 0$) между собой и считая незначимыми те из них, которые по абсолютному значению существенно меньше остальных.

Имитационная модель. Исследуемое явление будем характеризовать величиной G . Пусть эта величина определяется тремя контролируруемыми факторами x_1 , x_2 , x_3 и множеством неконтролируемых факторов, влияние которых представим одной нормально распределенной случайной величиной Z . В каждом i -том испытании значение G_i рассчитаем по формуле

$$G_i = G_0 + K_1(\sin x_{1i} + \sin x_{2i} + \sin x_{3i})/3 + K_2(|\sin x_{1i} - \sin x_{2i}|) + K_3(|\sin x_{1i} - \sin x_{3i}|) + K_4(|\sin x_{2i} - \sin x_{3i}|) + Z_i, \quad i = \overline{1, n}, \quad (6)$$

где K_1, \dots, K_4 – «ключи», то есть коэффициенты, принимающие значения 0 и 1 в соответствии с таблицей вариантов исходных данных, n – длина ряда. Случайная величина Z_i формируется с помощью генератора случайных чисел

$$Z_i = \frac{1}{2} \left(\sum_{\mu=1}^{m_1} a_{\mu} - \frac{m_1}{2} \right)_i, \quad \mu = \overline{1, m_1}. \quad (7)$$

Здесь a_μ – равномерно распределенные случайные числа ($0 \leq a_\mu \leq 1$). Моделируя пассивный эксперимент, будем считать, что значения x_{ji} равномерно распределены в интервале от 0 до $\pi/2$.

$$x_{ji} = a_{ji} \pi / 2, \quad (8)$$

где a_{ji} – число, вырабатываемое в i -том опыте для j -той переменной.

Удобно формировать Z_i и x_{ji} в одном цикле, удливив его до $m_1 + 3$. Тогда $a_{ji} = a_{m_1+j}$.

Пусть явление G отмечается наблюдателем, если G_i превышает пороговое значение $G_{\text{пор}}$. Тогда

$$y_i = \begin{cases} 1, & \text{если } (G_i - G_{\text{пор}}) > 0, \\ 0, & \text{если } (G_i - G_{\text{пор}}) < 0. \end{cases} \quad (9)$$

(Если G_i вычисляется с точностью до трех-четырех знаков после запятой, то ситуации, когда $G_i = G_{\text{пор}}$, при ограниченных выборках становятся маловероятными).

При выполнении работы рекомендуется задавать $m_1 = 12$, $G_0 = 5$, $n = 100$.

Содержание работы

1. Используя имитационную модель, сформируйте ряды x_{1i} , x_{2i} , x_{3i} , рассчитайте G_i . Результаты занесите в таблицу, подобную табл. 1.

2. Разбейте ряды на две части: обучающую $n_{об.}$ и контрольную $n_{к.}$ выборки $n = n_{об.} + n_{к.}$. Длину контрольной выборки задайте равной $n_{к.} = (1/4 \dots 1/3)n$.

3. Проанализируйте значения G_i обучающей выборки. Выберите такое пороговое значение $G_{пор.}$, чтобы примерно половина G_i , $i = \overline{1, n_{об.}}$ оказалась больше (меньше) $G_{пор.}$. Дополните табл. 1 столбцом y_i . Для всех значений $G_i > G_{пор.}$, $i = \overline{1, n}$ присвойте $y_i = 1$, для $G_i < G_{пор.}$ $y_i = 0$.

4. Проанализируйте ряды x_{ji} обучающей выборки. Найдите по выборкам значения $x_{j\min}$ и $x_{j\max}$. Сгруппируйте и закодируйте все x_{ji} обучающей выборки: присвойте значение +1 всем \tilde{x}_{ji} , если $x_{ji} > x_{j1/2}$, и -1 всем \tilde{x}_{ji} , если $x_{ji} < x_{j1/2}$.

5. Скопируйте табл. 2 и заполните в ней столбцы $y_v^{(1)}$ и $y_v^{(0)}$, рассчитав эти значения для соответствующих сочетаний уровней факторов.

6. По формуле (4) рассчитайте коэффициенты $b_j^{(1)}$ и $b_j^{(0)}$ для функций отклика (2), (3).

Оцените значимость всех коэффициентов. Слагаемые с незначимыми коэффициентами исключите из функций отклика.

7. Постройте дискриминантную функцию по формуле (5).

8. Проверьте качество дискриминантной функции. С этой целью перейдите в дискриминантной функции от кодированных переменных к физическим (формула (1)).

Дополните табл. 1 контрольной выборки столбцом прогнозируемых значений $y_{iпр.}$. Подставляя значения x_{ji} контрольной выборки в дискриминантную функцию, рассчитайте значения L_i , $i = \overline{1, n_{к.}}$. Если $L_i > 0$, то $y_{iпр.} = 1$, если $L_i < 0$, то $y_{iпр.} = 0$.

Найдите процент оправдываемости прогнозов.

Таблица 3

Варианты исходных данных

Вариант	K_1	K_2	K_3	K_4
1	0	1	1	0
2	0	1	0	1
3	0	0	1	1
4	1	1	1	0
5	1	1	0	1
6	1	0	1	1

Литература: [2], раздел 14, [11].

Значение функции Лапласа. Квантильные значения стандартного распределения $z_{1-\alpha}$, удовлетворяющие условию $F(z_{1-\alpha}) = 0,5 + \Phi(z_{1-\alpha}) = 1 - \alpha$

z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$	z	$\Phi(z)$
0,00	0,0000	0,64	0,2389	1,29	0,4015	1,94	0,4738
0,01	0,0040	0,65	0,2422	1,30	0,4032	1,95	0,4744
0,02	0,0080	0,66	0,2454	1,31	0,4049	1,96	0,4750
0,03	0,0120	0,67	0,2486	1,32	0,4066	1,97	0,4756
0,04	0,0160	0,68	0,2517	1,33	0,4082	1,98	0,4761
0,05	0,0199	0,69	0,2549	1,34	0,4099	1,99	0,4767
0,06	0,0239	0,70	0,2580	1,35	0,4115	2,00	0,4772
0,07	0,0279	0,71	0,2611	1,36	0,4131	2,02	0,4783
0,08	0,0319	0,72	0,2642	1,37	0,4147	2,04	0,4793
0,09	0,0359	0,73	0,2673	1,38	0,4162	2,06	0,4803
0,10	0,0398	0,74	0,2703	1,39	0,4177	2,08	0,4812
0,11	0,0438	0,75	0,2734	1,40	0,4192	2,10	0,4821
0,12	0,0478	0,76	0,2764	1,41	0,4207	2,12	0,4830
0,13	0,0517	0,77	0,2794	1,42	0,4222	2,14	0,4838
0,14	0,0557	0,78	0,2823	1,43	0,4236	2,16	0,4846
0,15	0,0596	0,79	0,2852	1,44	0,4251	2,18	0,4854
0,16	0,0636	0,80	0,2881	1,45	0,4265	2,20	0,4861
0,17	0,0675	0,81	0,2910	1,46	0,4279	2,22	0,4868
0,18	0,0714	0,82	0,2939	1,47	0,4292	2,24	0,4875
0,19	0,0753	0,83	0,2967	1,48	0,4306	2,26	0,4881
0,20	0,0793	0,84	0,2995	1,49	0,4319	2,28	0,4887
0,21	0,0832	0,85	0,3023	1,50	0,4332	2,30	0,4893
0,22	0,0871	0,86	0,3051	1,51	0,4345	2,32	0,4898
0,23	0,0910	0,87	0,3078	1,52	0,4357	2,34	0,4904
0,24	0,0948	0,88	0,3106	1,53	0,4370	2,36	0,4909
0,25	0,0987	0,89	0,3133	1,54	0,4382	2,38	0,4913
0,26	0,1026	0,90	0,3159	1,55	0,4394	2,40	0,4918
0,27	0,1064	0,91	0,3186	1,56	0,4406	2,42	0,4922
0,28	0,1103	0,92	0,3212	1,57	0,4418	2,44	0,4927
0,29	0,1141	0,93	0,3238	1,58	0,4429	2,46	0,4931
0,30	0,1179	0,94	0,3264	1,59	0,4441	2,48	0,4934
0,31	0,1217	0,95	0,3289	1,60	0,4452	2,50	0,4938
0,32	0,1255	0,96	0,3315	1,61	0,4463	2,52	0,4941
0,33	0,1293	0,97	0,3340	1,62	0,4474	2,54	0,4945
0,34	0,1331	0,98	0,3365	1,63	0,4484	2,56	0,4948
0,35	0,1368	0,99	0,3389	1,64	0,4495	2,58	0,4951
0,36	0,1406	1,00	0,3413	1,65	0,4505	2,60	0,4953
0,36	0,1443	1,01	0,3438	1,66	0,4515	2,62	0,4956
0,37	0,1480	1,02	0,3461	1,67	0,4525	2,64	0,4959
0,38	0,1517	1,03	0,3485	1,68	0,4535	2,66	0,4961
0,39	0,1554	1,04	0,3508	1,69	0,4545	2,68	0,4963
0,40	0,1591	1,05	0,3531	1,70	0,4554	2,70	0,4965
0,41	0,1628	1,06	0,3554	1,71	0,4564	2,72	0,4967
0,42	0,1628	1,07	0,3577	1,72	0,4573	2,74	0,4969
0,43	0,1664	1,08	0,3599	1,73	0,4582	2,76	0,4971
0,44	0,1700	1,09	0,3621	1,74	0,4591	2,78	0,4973
0,45	0,1736	1,10	0,3643	1,75	0,4599	2,80	0,4974
0,46	0,1772	1,11	0,3665	1,76	0,4608	2,82	0,4976
0,47	0,1808	1,12	0,3686	1,77	0,4616	2,84	0,4977
0,48	0,1844	1,13	0,3708	1,78	0,4625	2,86	0,4979
0,49	0,1879	1,14	0,3729	1,79	0,4633	2,88	0,4980
0,50	0,1915	1,15	0,3749	1,80	0,4641	2,90	0,4981
0,51	0,1950	1,16	0,3770	1,81	0,4649	2,92	0,4982
0,52	0,1985	1,17	0,3790	1,82	0,4656	2,94	0,4984
0,53	0,2019	1,18	0,3810	1,83	0,4664	2,96	0,49846
0,54	0,2054	1,19	0,3830	1,84	0,4671	2,98	0,49856
0,55	0,2088	1,20	0,3849	1,85	0,4678	3,00	0,49865
0,56	0,2123	1,21	0,3869	1,86	0,4686	3,20	0,49931
0,57	0,2157	1,22	0,3883	1,87	0,4693	3,40	0,49966
0,58	0,2190	1,23	0,3907	1,88	0,4699	3,60	0,49984
0,59	0,2224	1,24	0,3925	1,89	0,4706	3,80	0,499928
0,60	0,2257	1,25	0,3944	1,90	0,4713	4,00	0,499968
0,61	0,2291	1,26	0,3962	1,91	0,4719	5,00	0,499997
0,62	0,2324	1,27	0,3980	1,92	0,4726		
0,63	0,2357	1,28	0,3997	1,93	0,4732		

Коэффициенты Стьюдента. Критические значения

 $t_{1-\alpha/2, f}$ и $t_{1-\alpha, f}$ распределения Стьюдента

Число степеней свободы f	Уровень значимости α (двусторонняя критическая область)					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.31	12.7	31.82	63.7	318.3	637.0
2	2.92	4.30	6.97	9.92	22.33	31.6
3	2.38	3.18	4.54	5.84	10.22	12.9
4	2.13	2.78	3.75	4.60	7.17	8.61
5	2.01	2.67	3.37	4.03	5.89	6.86
6	1.94	2.45	3.14	3.71	5.21	5.96
7	1.89	2.36	3.00	3.50	4.79	5.40
8	1.86	2.31	2.90	3.36	4.80	5.04
9	1.83	2.26	2.82	3.25	4.30	4.78
10	1.81	2.23	2.76	3.17	4.14	4.59
11	1.80	2.20	2.72	3.11	4.03	4.44
12	1.78	2.18	2.68	3.05	3.93	4.32
13	1.77	2.16	2.65	3.01	3.85	4.22
14	1.76	2.14	2.62	2.98	3.79	4.14
15	1.75	2.13	2.60	2.95	3.73	4.07
16	1.75	2.12	2.58	2.92	3.69	4.01
17	1.74	2.11	2.57	2.90	3.65	3.96
18	1.73	2.10	2.55	2.88	3.61	3.92
19	1.73	2.09	2.54	2.86	3.58	3.88
20	1.73	2.09	2.53	2.85	3.55	3.85
21	1.72	2.08	2.52	2.83	3.53	3.82
22	1.72	2.07	2.51	2.82	3.51	3.79
23	1.71	2.07	2.50	2.81	3.49	3.77
24	1.71	2.06	2.49	2.80	3.47	3.74
25	1.71	2.06	2.49	2.79	3.45	3.72
26	1.71	2.06	2.48	2.78	3.44	3.71
27	1.71	2.05	2.47	2.77	3.42	3.69
28	1.70	2.05	2.46	2.76	3.40	3.66
29	1.70	2.05	2.46	2.76	3.40	3.66
30	1.70	2.04	2.46	2.75	3.39	3.65
40	1.68	2.02	2.42	2.70	3.31	3.55
60	1.67	2.00	2.39	2.66	3.23	3.46
120	1.66	1.98	2.36	2.62	3.17	3.37
∞	1.64	1.96	2.33	2.58	3.09	3.29
	0.05	0.025	0.01	0.005	0.001	0.0005
	Уровень значимости α (односторонняя критическая об-					

Квантильные значения $\chi^2_{1-\alpha, f}$ распределения Пирсона.

Число степеней свободы f	Уровень значимости α							
	0,01	0,025	0,05	0,10	0,90	0,95	0,975	0,99
1	6.6	5.0	3.8	2.71	0.0158	0.0039	0.00098	0.00016
2	9.2	7.4	6.0	4.60	0.211	0.103	0.051	0.020
3	11.3	9.4	7.8	6.25	0.584	0.352	0.216	0.115
4	13.3	11.1	9.5	7.78	1.06	0.711	0.484	0.297
5	15.1	12.8	11.1	9.24	1.61	1.15	0.831	0.554
6	16.8	14.4	12.6	10.6	2.20	1.64	1.24	0.872
7	18.5	16.0	14.1	12.0	2.83	2.17	1.69	1.24
8	20.1	17.5	15.5	13.4	3.49	2.73	2.18	1.65
9	21.7	19.0	16.9	14.7	4.17	3.33	2.70	2.09
10	23.2	20.5	18.3	16.0	4.86	3.94	3.25	2.58
11	24.7	21.9	19.7	17.3	5.58	4.57	3.82	3.05
12	26.2	23.3	21.0	18.6	6.30	5.23	4.40	3.57
13	27.7	24.7	22.4	19.8	7.04	5.89	5.01	4.11
14	29.1	26.1	23.7	21.1	7.79	6.57	5.63	4.66
15	30.6	27.5	25.0	22.3	8.55	7.26	6.26	5.23
16	32.0	28.8	26.3	23.6	9.31	7.96	6.91	5.81
17	33.4	30.2	27.6	24.8	10.1	8.67	7.56	6.41
18	34.8	31.5	28.9	26.0	10.9	9.39	8.23	7.01
19	36.2	32.9	30.1	27.2	11.7	10.1	8.91	7.63
20	37.6	34.2	31.4	28.4	12.4	10.9	9.59	8.26
21	38.9	35.5	32.7	29.6	13.2	11.6	10.3	8.90
22	40.3	36.8	33.9	30.8	14.0	12.3	11.0	9.54
23	41.6	38.1	35.2	32.0	14.8	13.1	11.7	10.2
24	43.0	39.4	36.4	33.2	15.7	13.8	12.4	10.9
25	44.3	40.6	37.7	34.4	16.5	14.6	13.1	11.5
26	45.6	41.9	38.9	35.6	17.3	15.4	13.8	12.2
27	47.0	43.2	40.1	36.7	18.1	16.2	14.6	12.9
28	48.3	44.5	41.3	38.0	19.0	16.9	15.3	13.6
29	49.6	45.7	42.6	39.1	19.8	17.7	16.0	14.3
30	50.9	47.0	43.8	40.3	20.6	18.5	16.8	15.0

Приложение IV

Квантили распределения Фишера $F_{1-\alpha, f_1, f_2}$ для $\alpha = 0,05$

(f_1, f_2 – число степеней свободы большей и меньшей дисперсий, соответственно)

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
11	4,8	4,0	3,6	3,4	3,2	3,1	2,8	2,6	2,4
12	4,8	3,9	3,5	3,3	3,1	3,0	2,7	2,5	2,3
13	4,7	3,8	3,4	3,2	3,0	2,9	2,6	2,4	2,2
14	4,6	3,7	3,3	3,1	3,0	2,9	2,5	2,3	2,1
15	4,5	3,7	3,3	3,1	2,9	2,8	2,5	2,3	2,1
16	4,5	3,6	3,2	3,0	2,9	2,7	2,4	2,2	2,0
17	4,5	3,6	3,2	3,0	2,8	2,7	2,4	2,2	2,0
18	4,4	3,6	3,2	2,9	2,8	2,7	2,3	2,1	1,9
19	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
20	4,4	3,5	3,1	2,9	2,7	2,6	2,3	2,1	1,8
22	4,3	3,4	3,1	2,8	2,7	2,6	2,2	2,0	1,8
24	4,3	3,4	3,0	2,8	2,6	2,5	2,2	2,0	1,7
26	4,2	3,4	3,0	2,7	2,6	2,4	2,1	1,9	1,7
28	4,2	3,3	2,9	2,7	2,6	2,4	2,1	1,9	1,6
30	4,2	3,3	2,9	2,7	2,5	2,4	2,1	1,9	1,6
40	4,1	3,2	2,9	2,6	2,5	2,3	2,0	1,8	1,5
60	4,0	3,2	2,8	2,5	2,4	2,3	1,9	1,7	1,4
120	3,9	3,1	2,7	2,5	2,3	2,2	1,8	1,6	1,3
∞	3,8	3,0	2,6	2,4	2,2	2,1	1,8	1,5	1,0

Приложение V

Равномерно распределенные случайные числа

10	09	73	25	33	76	52	01	35	86	34	67	35	48	76	80	95	90	91	17
37	54	20	48	05	64	89	47	42	96	24	80	52	40	37	20	63	61	04	02
08	42	26	89	53	19	64	50	93	03	23	20	90	25	60	15	95	33	47	64
99	01	90	25	29	09	37	67	07	15	38	31	13	11	65	88	67	67	43	97
12	80	79	99	70	80	15	73	61	47	64	03	23	66	53	98	95	11	68	77
66	06	57	47	17	34	07	27	68	50	36	69	73	61	70	65	81	33	98	85
31	06	01	08	05	45	57	18	24	06	35	30	34	26	14	86	79	90	74	39
85	26	97	76	02	02	05	16	56	92	68	66	57	48	18	73	05	38	52	47
63	57	33	21	35	05	32	54	70	48	90	55	35	75	48	28	46	82	87	09
73	79	64	57	53	03	52	96	47	78	35	80	83	42	82	60	93	52	03	44
98	52	01	77	67	14	90	56	86	07	22	10	94	05	58	60	97	09	34	33
11	80	50	54	31	39	80	82	77	32	50	72	56	82	48	29	40	52	42	01
83	45	29	96	34	06	28	89	80	83	13	74	67	00	78	18	47	54	06	10
88	68	54	02	00	86	50	75	84	01	36	76	66	79	51	90	36	47	64	93

Продолжение Приложения V

99	59	46	73	48		87	51	76	49	69		91	82	60	89	28		93	78	56	13	68
65	48	11	76	74		17	46	85	09	50		58	04	77	69	74		73	03	95	71	86
80	12	43	56	35		17	72	70	80	15		45	31	82	23	74		21	11	57	82	53
74	35	09	98	17		77	40	27	72	14		43	23	60	02	10		45	52	16	42	37
69	91	62	68	03		66	25	22	91	48		36	93	68	72	03		76	62	11	39	90
09	89	32	05	05		14	22	56	85	14		46	42	75	67	88		96	29	77	88	22
91	49	91	45	23		68	47	92	76	86		46	16	28	35	54		94	75	08	99	23
80	33	69	45	98		26	94	03	68	58		70	29	73	41	35		53	14	03	33	40
44	10	48	19	49		85	15	74	79	54		32	97	92	65	75		57	60	04	08	81
12	55	07	37	42		11	10	00	20	40		12	86	07	46	97		96	64	48	94	39
63	60	64	93	29		16	50	53	44	84		40	21	95	25	63		43	65	17	70	82
61	19	69	04	46		26	45	74	77	74		51	92	43	37	29		65	39	45	95	93
15	47	44	52	66		95	27	07	99	53		59	36	78	38	48		82	39	61	01	18
94	55	72	85	73		67	89	75	43	87		54	62	24	44	31		91	19	04	25	92
42	48	11	62	13		97	34	40	87	21		16	86	84	87	67		03	07	11	20	59
23	52	37	83	17		73	20	88	98	37		68	93	59	14	16		26	25	22	96	63
04	49	35	24	94		75	24	63	38	24		45	86	25	10	25		61	96	27	93	35
00	54	99	76	54		64	05	18	81	59		96	11	96	38	96		54	69	28	23	91
35	96	31	53	07		26	89	80	93	54		33	35	13	54	62		77	97	45	00	24
59	80	80	83	91		45	42	72	68	42		83	60	94	97	00		13	02	12	48	92
46	05	88	52	36		01	39	09	22	86		77	28	14	40	77		93	91	08	36	47
32	17	90	05	97		87	37	92	52	41		05	56	70	70	07		86	74	31	71	57
69	23	46	14	06		20	11	74	52	04		15	95	66	00	00		18	74	39	24	23
19	56	54	14	30		01	75	87	53	79		40	41	92	15	85		66	67	43	68	06
45	15	51	49	38		19	47	60	72	46		43	66	79	45	43		59	04	79	00	33
94	86	43	19	94		36	16	81	08	51		34	88	88	15	53		01	54	03	54	56
98	08	62	48	26		45	24	02	84	04		44	99	90	88	96		39	09	47	34	07
33	18	51	62	32		41	94	15	09	49		89	43	54	85	81		88	69	54	19	94
80	95	10	04	06		96	38	27	07	74		20	15	12	33	87		25	01	62	52	98
79	75	24	91	40		71	96	12	82	96		69	86	10	25	91		74	85	22	05	39
18	63	33	25	37		98	14	50	65	71		31	01	02	46	74		05	45	56	14	27
74	02	94	39	02		77	55	73	22	70		97	79	01	71	19		52	52	75	80	21
54	17	84	56	11		80	99	33	71	43		05	33	51	29	69		56	12	71	92	55
11	66	44	98	83		52	07	98	48	27		59	38	17	15	39		09	97	33	34	40
48	32	47	79	28		31	24	96	47	10		02	29	53	68	70		32	30	75	75	46
69	07	49	41	38		87	63	79	19	76		35	58	40	44	01		10	51	82	16	15

СПИСОК ЛИТЕРАТУРЫ

1. Ахназарова С. Л., Кафаров В. В. Методы оптимизации эксперимента в химической технологии. — М.: Высшая школа, 1985. — 327 с.
2. Бекряев В. И. Основы теории эксперимента. Учебное пособие. — СПб.: Изд. РГГМУ, 2002. — 265 с.
3. Белоцерковский А. В. Спектральный анализ в гидрометеорологии. — СПб.: Изд. РГГМИ, 1993. — 64 с.
4. Вадзинский Р. Н. Справочник по вероятностным распределениям. — СПб.: Наука, 2001. — 295 с.
5. Вентцель Е. С., Овчаров Л. А. Теория вероятностей и ее инженерные приложения. — М.: Высшая школа, 2000. — 480 с.
6. Гмурман В. Е. Теория вероятностей и математическая статистика. — М.: Высшая школа, 1999. — 479 с.
7. ГОСТ 16263–70 ГСИ. Метрология. Термины и определения. — 54 с.
8. ГОСТ 8.417–81 ГСИ. Единицы физических величин. — 40 с.
9. ГОСТ 8.401–80 ГСИ. Классы точности средств измерения. Общие понятия. — 13 с.
10. Основные термины в области метрологии. Словарь–справочник. Юдин М. Ф., Селиванов М. Н., Тищенко О. Ф., Скороходов А. И.; под ред. Тарбеева Ю. В. — М.: Изд. стандартов, 1989. — 113 с.
11. Пановский Г. А., Брайер Г. В. Статистические методы в метеорологии / Пер. с англ. — Л.: Гидрометеоздат, 1972. — 209 с.
12. Пустыльник Е. И. Статистические методы анализа и обработки наблюдений. — М.: Наука, 1968. — 288 с.
13. Селиванов М. Н., Фридман А. Э., Кудряшова Ж. Ф. Качество измерений. Метрологическая справочная книга. — Л.: Лениздат, 1987. — 295 с.
14. Сергеев А. Г., Крохин В. В. Метрология: Учебное пособие для вузов. — М.: Логос, 2000. — 408 с.

Учебное издание

ПРАКТИКУМ ПО ОСНОВАМ ТЕОРИИ ЭКСПЕРИМЕНТА

Составитель: Виктор Иванович Бекряев,

Редактор: И. Г. Максимова.

ЛР № 020309 от 30.12.96

Подписано в печать ...10.04.2003..... Формат 60 × 90¹/₁₆ Бумага офсетная. Печать офсетная.
Печ. л. 4,5 Уч.-изд. л. 4,25 Тираж 200 экз. Зак. 38.

195196, СПб, Малоохтинский пр. 98. РГГМУ.
Отпечатано ООО «Дарин»
